Human Interaction and Emerging Technologies (IHIET 2025), Vol. 197, 2025, 204–212
https://doi.org/10.54941/ahfe1006713

EchoXR: A Collaborative VR Framework for Spatial Acoustics in Architectural Design

Fabio Scotto, Chao Chia-Hsuan, Giacomo Montiani, Achilleas Xydis, Fabio Gramazio, and Matthias Kohler

Swiss Federal Institute of Technology (ETH), Zürich, Switzerland

ABSTRACT

This paper introduces EchoXR, a multiplayer Virtual Reality (VR) framework enabling real-time, collaborative exploration of architectural acoustics. By leveraging advanced tracking and spatial audio technologies, participants can co-experience how design changes—such as adding absorptive panels, altering partitions, or varying materials—impact the acoustics of a virtual environment. Although acoustic simulations often require extensive processing time, EchoXR integrates optimized algorithms with VR's immersive capabilities to deliver real-time auralization. Notably, users can hear each other's voices according to the simulated acoustic conditions, providing an immediate, immersive sense of how level in the space. The system supports multiple concurrent users, allowing designers, clients, and stakeholders to engage in synchronous, spatially coherent discussions. Through intuitive 3D user interfaces, participants can collaboratively adjust design elements and instantly perceive the resulting acoustic effects. This shared acoustic experience fosters more informed decision-making, minimizing the need for costly late-stage interventions or hastily added acoustic solutions that can disrupt the overall design concept and function of the space. The preliminary implementation demonstrates the feasibility and potential of collaborative VR auralization for architectural design workflows. By facilitating a deeper understanding of acoustics in the early design phase, EchoXR underlines the transformative role immersive technologies can play in shaping the future of built environments.

Keywords: Virtual reality, Collaborative architectural design, Immersive acoustics, Real-time auralization, Spatial audio

INTRODUCTION

The integration of Virtual Reality (VR) in architectural design has increasingly transformed the way designers and stakeholders conceptualize, iterate, and evaluate spatial experiences. Immersive technologies offer a powerful platform for visualizing architectural models at scale, enabling users to assess form, function, and human flow within a spatially coherent framework before construction begins (Portman et al., 2015). By embedding design decisions in a sensory-rich context, VR enhances not only visualization but also decision-making among interdisciplinary stakeholders. However, one critical aspect of built environments remains underrepresented in early design workflows: acoustics (Vorländer et al., 2014). Acoustic performance is traditionally addressed in the late stages of a project or even post-construction, often leading to ad hoc solutions (Badino et al., 2020) such as sound-absorbing panels, baffles, or dropped ceilings. These interventions, while functionally effective, are frequently treated as afterthoughts, ultimately compromising both the visual coherence and conceptual integrity of the architectural vision. The lack of intuitive, early-stage acoustic assessment tools severely limits the designer's ability to account for auditory qualities in parallel with visual aesthetics (Peters et al., 2021). While high-fidelity simulation tools like ODEON (DTU Science Park, 2025), EASE (AFMG Technologies GmbH, 2025) or Pachyderm (McNeel Europe, 2025) offer precise acoustic modeling, they often require substantial computational resources and specialized knowledge, rendering them impractical for iterative, collaborative design phases (Aspök et al., 2014). Moreover, these tools operate in isolation from immersive environments, restricting how users perceive and understand acoustic properties as part of their embodied experience within a space (Vorländer et al., 2014).

EchoXR addresses this critical gap by introducing a VR-based, collaborative framework for spatial acoustic exploration, with an emphasis on interaction and real-time feedback. Designed explicitly to integrate acoustic simulation into early design phases, EchoXR enables multiple users to simultaneously experience and manipulate architectural acoustics in real-time within an immersive environment. Through integrated motion tracking, advanced spatial audio rendering techniques, and intuitive gesture-based interactions, participants can modify surfaces, materials, and spatial layouts—and immediately perceive the acoustic consequences of their choices. This immediate feedback loop enhances not only decision-making but also dialogue across disciplines, offering a shared ground for evaluating and negotiating design strategies. By embedding acoustic analysis within collaborative VR sessions, EchoXR promotes a holistic, human-centered approach to architecture that anticipates both how spaces look and how they sound. As immersive technologies evolve, EchoXR exemplifies how multisensory design feedback can shift professional practice, empowering architects to make acoustically informed decisions from the very earliest stages of design.

BACKGROUND & RELATED WORK

The integration of Virtual Reality into architectural design workflows has increasingly enabled designers, engineers, and clients to collaboratively explore and assess spatial qualities at a human scale (Berg & Vance, 2016). Unlike traditional design tools, VR provides immersive, real-time visualization, offering deeper insight into spatial proportions, circulation, light dynamics, and experiential qualities well before ground is broken (Whyte & Nikolić, 2018). Despite these strengths, many immersive platforms remain predominantly visual, neglecting essential sensory dimensions such as acoustics, which play a fundamental role in shaping how spaces are perceived and used (Doğan & Sorguç, 2023). The inability to simulate auditory performance in real time often delays acoustic considerations until after schematic or detailed design phases have concluded. High-fidelity acoustic modeling tools like ODEON, EASE, and Pachyderm remain the gold standard for aural simulation but are typically decoupled from visual environments, require expert knowledge, and rely on non-interactive workflows. These tools are thus poorly suited to dynamic design sessions or participatory evaluations involving non-specialists (Aspök et al., 2014).

At the same time, a growing body of literature explores how natural interaction methods can enhance design cognition and creativity in immersive contexts. Multi-user VR platforms have been developed to support collaborative reviews, yet many still rely on menu-driven, controller-based interfaces that constrain embodied interaction (Milovanovic et al., 2017). Recent advances in gesture tracking and spatial hand interfaces suggest promising alternatives. Studies have shown that gesture-based interfaces improve spatial memory, task performance, and collaborative communication in virtual design environments (Shin et al., 2024); (Bowman & McMahan, 2007). Likewise, technologies like OptiTrack and passthrough-enabled headsets have advanced spatial awareness and safety, supporting increasingly complex multi-user sessions with reduced risk of physical collisions (Scavarelli & Teather, 2017). Few existing systems, however, integrate real-time audio feedback with intuitive gesture interfaces and motion tracking into a cohesive architectural design framework. EchoXR builds on these insights, merging real-time auralization with embodied, multi-user VR interaction to create an intuitive, responsive environment in which participants can experience—and reshape—acoustics in tandem with spatial form.

THE ECHOXR FRAMEWORK

EchoXR is a novel Virtual Reality (VR) system designed to facilitate collaborative, acoustically informed design decisions in the early stages of architectural workflows. The framework integrates binaural auralization of virtual sound sources, hands-free interaction, and multiplayer networking, allowing designers and stakeholders to co-experience and evaluate auditory and spatial qualities of designed environments in real-time. Developed primarily in Unity and utilizing Steam Audio for realistic spatial audio rendering and the Meta Quest SDK for immersive VR interactions, EchoXR bridges the gap between static visualization and embodied multisensory engagement. The system further incorporates OptiTrack motion capture technology to ensure precise tracking of user movements, enhancing spatial accuracy and immersion for all participants (Figure 1).

At the core of EchoXR lies its capacity for real-time auralization. Users hear each other's voices and environmental sounds dynamically processed according to the spatial geometry and material composition of the designed environment. This simulation is tightly synchronized with the system's visual and interaction components, ensuring that any design change—such as adjusting materials, introducing partitions, or adding sound absorptive elements—results in an immediate and perceptible acoustic update. Crucially, these interactions occur without disrupting the immersive session, creating a seamless experience where feedback is multisensory, shared, and iterative. This capability is particularly valuable during collaborative reviews, where real-time negotiation between stakeholders can directly shape the evolving spatial and acoustic configuration of a project.

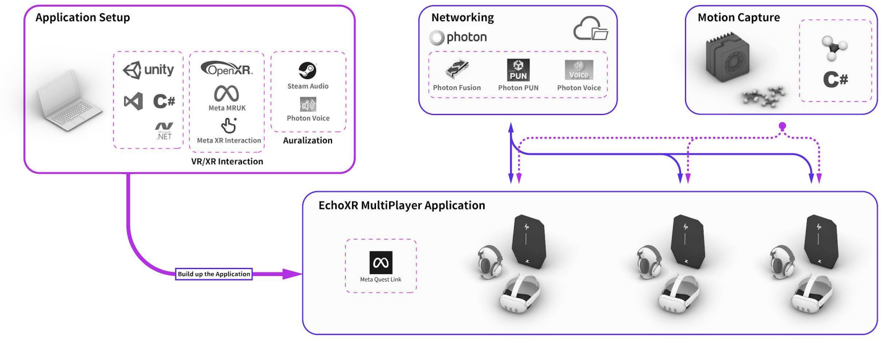

Figure 1: The software architecture of EchoXR.

Multiplayer/Networking

Multiplayer networking is core to EchoXR's collaborative design paradigm. Users join shared virtual sessions where spatial audio is synchronized and positional tracking ensures consistent auditory experiences across participants. Roles can be distributed—such as lead designer, client, or acoustics consultant—each capable of interacting with the environment and contributing to the design process. Voice communication is not only preserved but enhanced: sound is spatially located according to the simulated acoustics, making discussions more natural, immersive, and contextually grounded (Deacon & Barthet, 2023).

The software architecture for networking functionalities is built upon Photon Fusion, a networking solution that extends the Unity API. These functionalities are categorized into two types: Networked Events and Networked Tracking. Networked Events manage discrete interactions initiated by local players that affect shared environmental states. These are synchronized through a shared authority model in which clients request and receive State Authority before modifying networked content, thereby maintaining consistency across distributed sessions (Exit Games, n.d.)

Networked Tracking ensures the continuous synchronization of avatar elements—headsets, hands, and body posture—across users. These elements are instantiated and managed through Photon's event systems, and their ownership is dynamically assigned to maintain real-time coherence. EchoXR further extends Photon to incorporate Meta's hand tracking system, enabling fine-grained gesture interaction across users with minimal latency (Figure 2). The combined system supports expressive, multi-user interaction without the need for physical controllers, which reinforces natural communication and co-design in shared virtual environments (Meta, 2025).

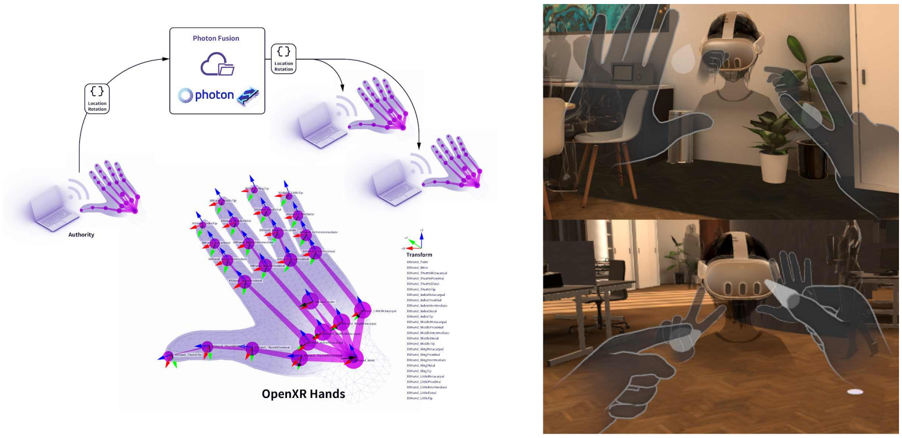

Figure 2: The workflow of hands tracking and networked synchronising.

Interaction

EchoXR emphasizes hands-free interaction through gesture-based navigation and manipulation. Instead of relying on traditional VR controllers, users engage with the environment using tracked hand gestures to walk, point, select, and alter architectural elements. This interaction paradigm is designed to reduce physical fatigue, improve user accessibility, and foster more fluid and intuitive spatial engagement. Hands-free control also aligns better with the collaborative nature of design reviews, where continuous communication and shared focus are critical.

The gesture system is specifically optimized for core architectural operations, including selecting materials from a UI menu and applying them to surfaces with natural gestures. From a user experience perspective, EchoXR introduces an ambient, gesture-activated interface that appears within the virtual space. Anchored to the user's left palm, this UI emerges when the user turns their hand upward (Figure 3), revealing context-sensitive tools for material selection, geometry adjustment, and acoustic configuration. The UI is designed to minimize occlusion, support both standing and seated postures, and maintain high visibility within a range of lighting and spatial conditions. By lowering cognitive and physical barriers, this interaction model promotes active experimentation and real-time co-creation among participants.

To support dynamic spatial manipulation in virtual environments, the interface integrates two primary modes of control: On/Off Toggles and a Drop-Down List (Figure 4 left). The On/Off Toggles allow users to selectively add or remove environmental acoustics elements such as desk panels, wall panels, ceiling panels, and curtains, enabling on-the-fly customization of the workspace (Figure 4 right). The Drop-Down List further enhances flexibility by offering predefined spatial configurations (Figure 4 middle), including options such as "Walls + Glazing," "Acoustic Curtains," and "No Partitions." These interaction mechanisms are accessed through a floating control panel that responds to intuitive hand gestures, including a scaling animation to call up or dismiss the interface. As shown in the accompanying figure, this UI setup enables seamless, low-effort interaction within immersive VR scenarios, supporting personalized design exploration and collaborative decision-making. Crucially, changes made through the interface—whether toggling acoustic treatments or selecting spatial configurations—not only affect the visual layout but also alter the perceived acoustics of the environment in real time. This reinforces EchoXR's central aim: to link architectural design decisions with their acoustic consequences through embodied, interactive, and collaborative experiences.

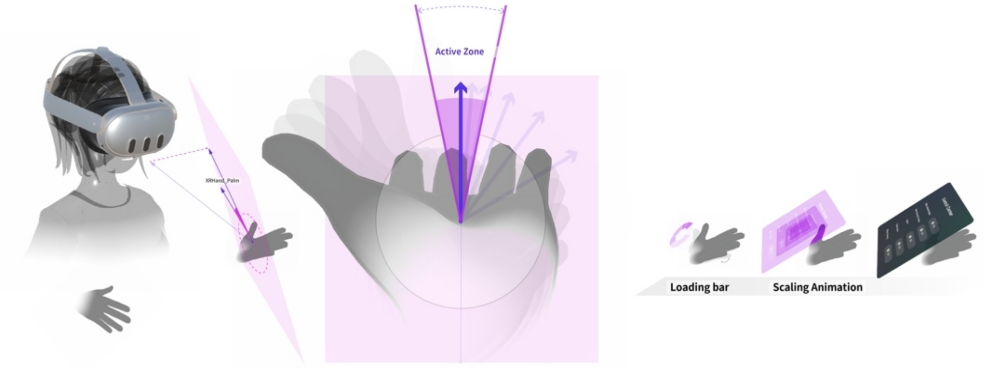

Figure 3: Palm-based UI activation in EchoXR.

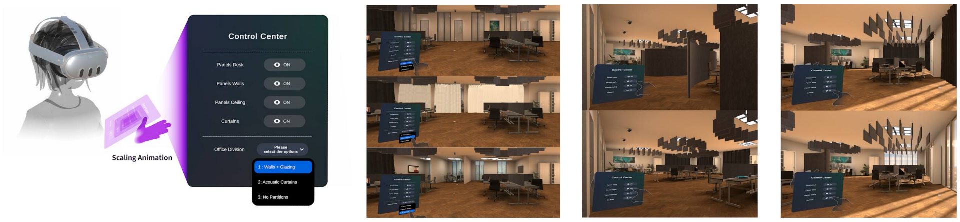

Figure 4: Hand-driven user interface for real-time acoustic environment control in EchoXR.

Tracking

To ensure physical safety and enhance spatial awareness during co-located multiplayer sessions, EchoXR integrates robust tracking solutions using both external motion capture and onboard headset capabilities. Each user's physical position is tracked via OptiTrack and broadcast to all other users, creating a shared spatial coordinate system that aligns physical and virtual spaces. Locally, the Meta Mixed Reality Utility Kit (MRUK) manages user position and orientation through a prior scan of the user's room-scale environment, improving reliability and reducing drift over time.

Crucially, EchoXR incorporates a proximity-aware passthrough feature, designed to mitigate collision risks and heighten user trust. When users approach each other beyond a predefined threshold, passthrough is automatically activated, revealing a live feed of the physical surroundings. This mode is smoothly integrated with the immersive experience, allowing users to maintain awareness without breaking presence. Such hybrid spatial safety systems are essential for multi-user VR applications and lay the groundwork for safe, scalable use of immersive tools in shared architectural workflows.

CHALLENGES AND LIMITATIONS

While EchoXR presents a compelling vision for interactive, acoustically informed design, several technical and practical challenges remain. One key limitation involves balancing the computational demands of immersive, real-time environments with the need for smooth interactivity. High-fidelity acoustic feedback, dynamic scene updates, and simultaneous user interactions all place pressure on available resources, especially as project complexity increases. Achieving scalable performance without compromising responsiveness remains a core challenge for deployment in large-scale architectural contexts.

Another concern lies in the learning curve associated with hands-free interaction. Although gesture-based control enhances embodiment and collaboration, users unfamiliar with non-controller paradigms may initially struggle to perform design tasks fluidly. Effective onboarding and customization of gesture vocabularies will be crucial for broader adoption. Similarly, while multi-user synchronization is generally robust, occasional latency or network inconsistencies can affect positional accuracy and coordination, particularly in low-bandwidth settings.

Finally, the integration of EchoXR into existing architectural workflows demands careful alignment with established practices and tools. Architects may hesitate to adopt new systems that do not readily interface with their modeling environments or documentation standards. As such, further development must include export functionalities, BIM interoperability, and workflow integration strategies to fully embed EchoXR in professional practice. Despite these challenges, the platform's capacity to reduce reliance on costly post-construction interventions and foster early-stage, multisensory collaboration positions it as a powerful tool for the future of architectural design.

OUTLOOK AND CONCLUSION

EchoXR significantly enhances collaborative architectural workflows by integrating VR interactivity with real-time acoustic feedback, addressing critical issues related to delayed acoustic considerations in traditional design processes. As the technology matures, there are numerous opportunities to refine and expand its capabilities further within VR environments. Future research could focus on improving scalability and computational efficiency to support larger, more complex architectural projects and enhanced interactivity among multiple simultaneous users. The integration of advanced

AI-driven predictive simulations could allow for even more accurate and responsive acoustic feedback, dynamically informing design decisions based on iterative and collaborative adjustments. Additionally, improvements in gesture-based interactions could further reduce user learning curves, increasing the accessibility and efficiency of collaborative sessions.

While VR remains the primary focus, a potential direction could involve extending EchoXR into Augmented Reality (AR). An AR application would enable users to visualize sound behavior directly overlaid onto real-world environments, creating a complementary tool to VR-based exploration. Visualizing acoustic fields and reverberation paths in situ could enhance the understanding of sound propagation in existing buildings, opening new pathways for on-site acoustic diagnostics, retrofit evaluation, and accessible design for the hearing impaired. Such an extension could further bridge the gap between virtual acoustic simulations and physical spaces, reinforcing EchoXR's goal of delivering comprehensive, intuitive, and multisensory design experiences.

ACKNOWLEDGMENT

The authors gratefully acknowledge the ETH Foundation for funding the research project. They also wish to thank the Center for Augmented Computational Design in Architecture, Engineering and Construction (Design++) at ETH Zurich for providing access to the infrastructure of the Immersive Design Lab during the development and testing of EchoXR. In addition, the authors acknowledge the use of AI-based tools, including ChatGPT and Grammarly, for text review and editorial suggestions.

CONTRIBUTIONS

Fabio Scotto (FS) and Achilleas Xydis (AX) jointly defined the original idea for the project. FS led the XR development, managed the project, and designed the virtual reality environment. AX was responsible for the acoustic dimension and served as the lead in that domain. FS and AX jointly supervised the overall research. Chao Chia-Hsuan (CC), Giacomo Montiani (GM) were responsible for the implementation and integration of the system components. They explored, tested, and evaluated different technical approaches, applying the most effective methods to achieve a cohesive user experience. GM focused on integrating and optimizing the acoustic features within the XR environment, while CC led the development of interactive elements and ensured seamless integration within the VR system. All authors reviewed and approved the final manuscript.

REFERENCES

AFMG Technologies (GmbH), 2025. *EASE-Evac*. [Online] Available at: https://www.afmg.eu/en/ease-evac?_gl=1*py4jtk*_up*MQ..*_ga*MTg2OTM1Njk0NC4xNzQ0NTQ0OTUz*_ga_BGT6XW7ZST*MTc0NDU0NDk1Mi4xLjEuMTc0NDU0NDk1OC4wLjAuMA. [Accessed 13 April 2025].

Aspök, L., Pelzer, S., Wefer, F. & Vorländer, M., 2014. *A Real-Time Auralization Plugin for Architectural Design and Education*. Berlin, Universitätsverlag der TU Berlin.

Badino, E., Shtrepi, L. & Astolfi, A., 2020. Acoustic Performance-Based Design: A Brief Overview of the Opportunities and Limits in Current Practice. *Acoustics,* 2(2), pp. 246–278.

Berg, L. P. & Vance, J. M., 2016. Industry use of virtual reality in product design and manufacturing: a survey. *Virtual Reality,* Volume 21, pp. 1–17.

Bowman, D. A. & McMahan, R. P., 2007. Virtual Reality: How Much Immersion Is Enough?. *Computer,* 40(7), pp. 36–43.

Deacon, T. & Barthet, M., 2023. Spatial Design Considerations for Interactive Audio in Virtual Reality. In: M. Geronazzo & S. Serafin, eds. *Sonic Interactions in Virtual Environments*. s.l.: Springer, pp. 181–217.

Doğan, A. Z. & Sorguç, A. G., 2023. *Sound perception in virtual environments*. Chiba, Institute of Noise Control Engineering.

DTU Science Park, 2025. *ODEON Room Acoustics Software*. [Online] Available at: https://odeon.dk[Accessed 13 April 2025].

Exit Games, n.d.. *Photon Engine*. [Online] Available at: https://www.photonengine.com/pun [Accessed 13 April 2025].

McNeel Europe, 2025. *Pachydrem Acoustical Simulation (by Arthur)*. [Online] Available at: https://www.food4rhino.com/en/app/pachyderm-acoustical-simulation [Accessed 13 April 2025].

Meta, 2025. *Meta Developers Hands Interactions*. [Online] Available at: https://developers.meta.com/horizon/design/hands [Accessed 13 April 2025].

Milovanovic, J., Moreau, G., Siret, D. & Miguet, F., 2017. *Virtual and Augmented Reality in Architectural Design and Education*. Instanbul, Turkey, Springer.

Peters, B., Nguyen, J. & Omar, R., 2021. *Parametric Acoustics: Design techniques that integrate modelling and simulation*. Madeira, Portugal, s.n.

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

Scavarelli, A. & Teather, R. J., 2017. *VR Collide! Comparing Collision-Avoidance Methods Between Co-located Virtual Reality Users*. Denver, Colorado, Association for Computing Machinery.

Shin, J. et al., 2024. A Methodological and Structural Review of Hand Gesture Recognition Across Diverse Data Modalities. *IEEE Access,* Volume 12, pp. 142606–142639.

Vorländer, M., Schröder, D., Pelzer, S. & Wefers, F., 2014. Virtual Reality for Architectural Acoustics. *Journal of Building Performance Simulation,* 8(1), pp. 15–25.

Whyte, J. & Nikolić, D., 2018. *Virtual Reality and the Built Environment*. 2nd Edition ed. London: Routledge.