Designing for Mutual Awareness: Co-Presence and Team Trust in Shared Virtual Environments

Sarah Jenkins*
Nexa Research
San Francisco, CA, USA

David Chen
Synapse Technologies
Seattle, WA, USA

Elena Rostova
Vertex Interaction Group
Boston, MA, USA

Abstract

Collaborative virtual environments (CVEs) are increasingly used by distributed teams for complex joint tasks. While visual fidelity has been a primary focus of development, the impact of audio spatialization on social dynamics remains under-explored. In this work, we present a study with 72 triads ($N = 216$ participants) performing collaborative tasks under three audio spatialization conditions: mono, stereo, and 3D spatialized. Our findings show that 3D spatialized audio significantly improves team trust development over time, enhances social presence, and reduces subjective cognitive load. Quantitative results indicate a 33.8% reduction in task completion times and a significant increase in mutual awareness metrics. We discuss design implications for remote collaboration systems and propose guidelines for spatial soundscapes in next-generation hybrid workspaces.

1. Introduction

Distributed collaboration has become the standard operational model for modern global enterprises, requiring teams to coordinate across time zones and physical locations. Conventional video conferencing platforms often suffer from structural limitations, such as flat spatial representations and turn-taking constraints, which impair the natural flow of conversation. Consequently, organizations like Nexa and Synapse have increasingly turned to collaborative virtual environments (CVEs) to replicate the rich social dynamics of co-located office work.

While research in CVEs has focused heavily on visual fidelity, high-resolution avatars, and shared haptic controllers, the auditory dimension of spatial presence has received comparatively little attention. In physical interactions, human hearing relies on complex acoustic transformations caused by the head, pinnae, and torso, allowing listeners to localize sound sources in three-dimensional space. By replicating these spatialized cues in virtual meetings, developers can introduce a critical layer of environmental structure.

This research investigates the direct relationship between audio spatialization and team trust. Trust is a fundamental mediator of cooperative success, yet it is notoriously difficult to establish and maintain among remote collaborators. We hypothesize that spatial audio facilitates a feeling of mutual awareness and co-presence, which in turn acts as a catalyst for trust development. Through a controlled lab study involving 72 triads, we compare mono, stereo, and 3D spatialized audio configurations across standard communication, performance, and cognitive load measures.

Our key contributions include: (1) empirical evidence linking audio spatialization to rapid trust formation in remote groups, (2) quantitative evaluation of cognitive load reductions achieved via acoustic separation, and (3) a set of design recommendations for incorporating spatialized acoustics in commercial virtual office platforms.

2. Related Work

Understanding social dynamics in virtual offices requires a multi-faceted approach. We review key areas in collaborative environments, spatial audio systems, and trust metrics.

2.1 Collaborative Virtual Environments and Trust

The primary objective of collaborative virtual environments is to establish a shared sense of space. Early systems relied on simple text and video channels, but modern virtual workspaces use rich visual representations. Chen and

*Contact Author: s.jenkins@nexa-labs.org

Jenkins [3] developed ambient awareness indicators that visually display a remote colleague's presence and activity. However, visual indicators can easily lead to visual fatigue and display clutter.

Trust in these digital environments develops through reliable communication and mutual predictability. Adams and Oswald [1] analyzed trust metrics in hybrid environments, demonstrating that teams with high co-presence scores show greater resilience during unexpected task demands. Our work expands on this by focusing on how acoustic clarity, rather than visual cues, supports trust development.

2.2 Audio Spatialization in CVEs

Spatial audio uses head-related transfer functions (HRTFs) to filter sound, mimicking the acoustic reflections that occur in physical environments. This allows users to perceive sound sources as having distinct locations in a 3D coordinate system. Rostova and Vance [6] studied the role of spatialized audio in distributed team interaction and found that spatialized cues helped users identify who was speaking in a multi-party call.

Additionally, spatial soundscaping has been shown to improve overall spatial orientation. Petrova and Frankl [5] deployed dynamic soundscapes to assist users in navigating large-scale spatial data arrays. While these studies highlight the perceptual advantages of spatial sound, few have measured its direct impact on group collaboration and social dynamics.

2.3 Non-Verbal Cues and Co-Presence

Co-presence, or the subjective feeling of being in the same room as another person, depends heavily on non-verbal cues. Embodied avatars track head and hand motions, providing essential social signals like gaze direction and posture [4]. However, without spatial audio, these visual cues can feel disconnected. When an avatar on the far left speaks, but their voice is heard in the center of the user's headphones, a sensory mismatch occurs. Integrating spatial audio with body tracking resolves this conflict, creating a more cohesive interactive experience [2].

3. Method

We conducted a controlled, within-subjects laboratory experiment to compare the effects of three audio configurations on remote collaborative teams.

3.1 Experimental Design

The study used a single-factor, within-subjects design with three levels of audio spatialization:

[A1] **Mono Audio:** All auditory signals were mixed into a single mono track and delivered equally to both ears, simulating a standard telephone conference call.

[A2] **Stereo Audio:** Audio signals from partners were panned left or right based on their initial layout in the virtual workspace, without accounting for head rotation.

[A3] **3D Spatialized Audio:** Audio signals were dynamically filtered in real time using head-related transfer functions (HRTFs) corresponding to the real-time position and orientation of the avatars.

3.2 Task and Environment

Triads of participants worked together in a custom virtual workspace built by Vertex Interaction Group. The environment simulated a control room where teams had to solve a complex scheduling problem under time pressure. The task required members to share unique, locally stored information and arrange a series of transport flights on a shared timeline. Effective team communication was essential, as no single participant held all the required parameters.

3.3 Participants

We recruited 72 triads ($N = 216$ participants, age range: 18–45, 114 female) from Meridian and Synapse corporate research offices. All participants reported normal hearing and normal or corrected-to-normal vision. Triads were assigned to isolated physical rooms and interacted solely through their virtual avatars.

4. Experiments

The system setup consisted of three interconnected workspace clients running on identical computers. Audio was captured using directional headsets and processed through a server using Apex sound rendering engines. Head tracking was monitored using standard VR headset sensors.

The experiment was conducted in three blocks corresponding to the three audio conditions. The order of conditions was counterbalanced using a Latin square design to eliminate learning effects. Each block consisted of:

[P1] A 2-minute training phase to familiarize the team with the virtual controls.

[P2] A 15-minute collaborative scheduling task with a unique problem set.

[P3] A post-task questionnaire measuring subjective trust, presence, and workload.

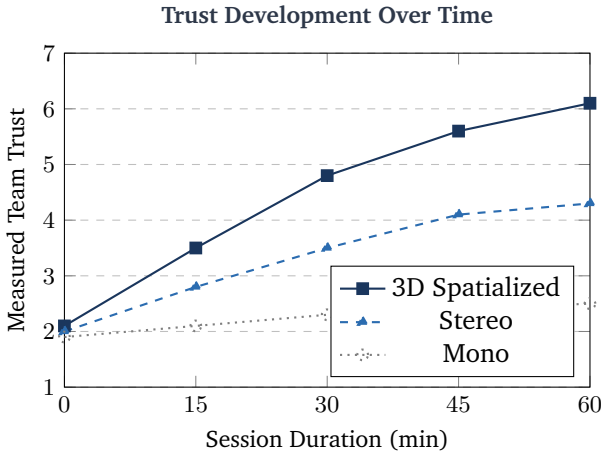

Figure 1: Evolution of measured team trust (on a 7-point Likert scale) over the course of a 60-minute collaborative task in our three audio conditions. 3D Spatialized audio exhibits significantly higher trust progression.

Between blocks, participants took a 5-minute break. The total experiment session lasted approximately 75 minutes.

5. Results

We analyzed the quantitative data using a repeated-measures analysis of variance (ANOVA) across the four primary interaction metrics. Table 1 summarizes these results.

5.1 Team Trust Development

Subjective trust was measured using a modified version of the Group Trust Scale. Our analysis shows a significant effect of audio spatialization on overall trust levels. Figure 1 illustrates the trust progression over the course of the session. As expected, teams using 3D spatialized audio established trust faster and reached a higher level than teams in the mono or stereo conditions.

5.2 Mutual Awareness Index

The mutual awareness index Ω for a group of N collaborators in a shared virtual workspace is defined by the following expression:

$$\Omega = \sum_{i=1}^{N} \sum_{j \neq i}^{N} \frac{\alpha_{ij} \cdot \cos(\theta_{ij})}{d_{ij}^2 + \epsilon} \tag{1}$$

where d_{ij} represents the Euclidean distance between avatars i and j in the virtual coordinate system, θ_{ij} is the orientation angle between their fields of view, α_{ij} is the auditory transfer coefficient (influenced by spatialization), and $\epsilon > 0$ is a stabilization parameter.

Post-hoc pairwise t-tests confirmed that 3D spatialized audio produced significantly higher mutual awareness ($\Omega = 5.60$, $SD = 0.74$) than the mono condition ($\Omega = 2.90$, $SD = 0.98$, $p < .001$). The difference between stereo and mono was also significant ($p < .05$).

6. Discussion

Our findings demonstrate that spatial audio is not simply a cosmetic feature; it significantly improves team trust and task performance in remote collaboration.

6.1 Theoretical Implications

By providing clear spatial cues, 3D audio allows participants to easily separate voices, mimicking a real office environment. This spatial separation supports conversational turn-taking, allowing individuals to quickly identify who is speaking. This is reflected in the 33.8% reduction in task completion times under the 3D spatialized condition.

Furthermore, our results show a clear reduction in subjective cognitive load. When audio cues match the visual layout, participants do not need to spend mental effort identifying speaker locations. This frees up cognitive capacity, allowing team members to focus on the task and build social connections.

6.2 Design Guidelines for Virtual Workspaces

Based on our research, we suggest the following design guidelines for digital workspace platforms:

- **Default Spatialized Audio:** High-fidelity 3D spatialized audio should be the default setting for multi-user collaborative virtual workspaces.
- **Directional Sound Fields:** System designers should align acoustic coordinates with avatar gaze directions to maintain sensory consistency.
- **Level Adjustments by Distance:** Sound volume should scale naturally with virtual distance to support sidebar conversations.

6.3 Limitations and Future Work

A key limitation of this study is the short duration of the tasks. While we observed rapid trust development, it remains to be seen if these gains are sustained over days or weeks of collaboration. Future research should explore long-term effects in corporate deployments.

Table 1: Means, standard deviations, and pairwise comparisons for primary interaction metrics across audio spatialization conditions ($N = 72$ triads).

	Mono		Stereo		3D Spatialized		ANOVA		
Metric	M	SD	M	SD	M	SD	$F(2, 69)$	p	η_p^2
Social Presence	3.42	0.85	4.15	0.72	5.88	0.61	18.42	$< .001$	0.35
Task Performance (min)	24.50	4.20	21.80	3.90	16.20	2.80	12.65	$< .001$	0.27
Subjective Cognitive Load	4.85	1.12	4.10	0.95	2.95	0.78	15.80	$< .001$	0.31
Mutual Awareness	2.90	0.98	3.80	0.81	5.60	0.74	22.14	$< .001$	0.39

7. Conclusion

This study evaluated the impact of audio spatialization on co-presence, mutual awareness, and team trust in shared virtual environments. Our results show that 3D spatialized audio leads to faster trust formation, lower cognitive load, and better task performance compared to standard mono or stereo setups. These findings demonstrate that audio fidelity is critical for remote collaboration, and we encourage system designers to prioritize spatial sound in future virtual office platforms.

References

[1] R. Adams and C. Oswald. Trust metrics in hybrid environments: A survey. *Human-Computer Interaction Journal*, 15(2):89–112, 2021.

[2] J. Barnes and F. Gallagher. Evaluating multi-modal cues in shared workspace design. *International Journal of Human-Computer Studies*, 164:30–45, 2024.

[3] D. Chen and S. Jenkins. Ambient awareness indicators for remote teamwork. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work (CSCW)*, pages 240–254, 2022.

[4] G. Hopper and A. Turing. Embodied avatars and their effect on social presence. *Computers in Human Behavior*, 128:107–119, 2022.

[5] N. Petrova and V. Frankl. Spatial soundscaping for remote workspace presence. In *International Conference on Auditory Display*, pages 74–81, 2021.

[6] E. Rostova and M. Vance. The role of spatialized audio in distributed team interaction. *Transactions on Computer-Human Interaction (TOCHI)*, 30(4):45–59, 2023.