Plug the Guitar In Before You Put the Headset On: Co-designing a Mixed Reality Collaborative Application with Professional Musicians

Alberto Boem
Department of Information Engineering and Computer Science
University of Trento
Trento, Italy
alberto.boem@unitn.it

Victor Casado
SAE Barcelona
Barcelona, Spain
v.casado@sae.edu

Luca Bucciarelli
SAE Barcelona
Barcelona, Spain
l.bucciarelli@sae.edu

Luca Turchet
Department of Information Engineering and Computer Science
University of Trento
Trento, Italy
luca.turchet@unitn.it

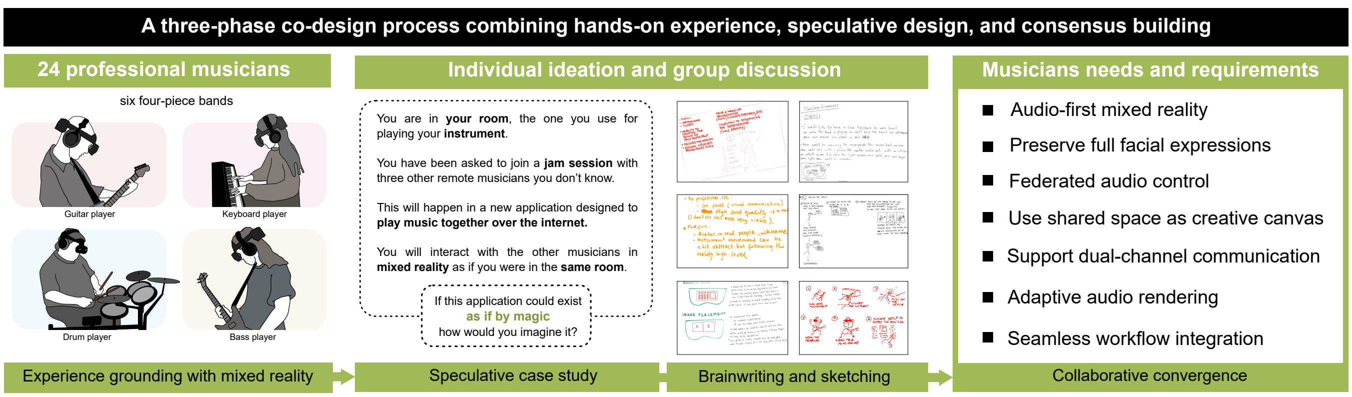

Figure 1: How professional musicians envision a mixed reality musical collaboration application. A three-phase co-design process combining: hands-on experience, speculative design, and consensus building with 24 professional musicians organized into six four-piece bands.

Abstract

Mixed reality (MR) technologies hold promise for distributed musical collaboration, yet the needs of professional musicians remain underexplored. We conducted a set of co-design workshops with 24 professional musicians, organized into six four-piece bands, to envision future MR systems for networked music performance. Using a three-phase process (hands-on MR experience, individual brainwriting, and collaborative consensus-building), we identified core requirements that challenge current MR paradigms. Professional musicians operate from an "Audio-First" framework where sonic experience constitutes the primary reality, contradicting visual-first MR approaches. Five unanimous requirements emerged: audio-first workflow, support for non-verbal communication, personalized audio control, spatial layout customization, and plug-and-play simplicity. Other high-consensus themes included real-time voice communication and realistic avatar representation. Taken together, our findings reveal that professional creative collaboration demands fundamentally different design priorities than traditional office-centered collaborative MR applications, requiring systems that support established workflows and nonverbal communication channels while prioritizing workflow preservation over visual immersion.

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3793871

CCS Concepts

• **Human-centered computing → Mixed / augmented reality**; **Collaborative interaction**; • **Applied computing →** *Sound and music computing*.

Keywords

mixed reality, telepresence, musical metaverse, networked music performance, musical xr, head-mounted display, social xr, co-design

ACM Reference Format:
Alberto Boem, Victor Casado, Luca Bucciarelli, and Luca Turchet. 2026. Plug the Guitar In Before You Put the Headset On: Co-designing a Mixed Reality Collaborative Application with Professional Musicians. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3772318.3793871

1 Introduction

Musical collaboration fundamentally relies on complex real-time interactions that integrate temporal synchronization, spatial awareness, and sophisticated nonverbal communication channels. Professional musicians have developed embodied expertise through years of practice in shared physical spaces, where subtle visual cues, breathing patterns, and micro-gestures enable collective musical creativity–a process common across diverse traditions [97].

The advent of internet technologies opened new possibilities for distributed musical collaboration, establishing networked music performance (NMP) [26, 93] as a research field dedicated to enabling musicians to perform together across geographic distances. In recent years, systems like JackTrip [21] or LoLa [41] demonstrated that ultra-low latency audio transmission could support basic musical synchronization, establishing technical foundations for remote collaboration. The COVID-19 pandemic accelerated adoption from niche research to mainstream necessity, revealing both the potential and fundamental limitations of audio-only remote collaboration technologies. However, current NMP systems, relying on either sole audio streams or audio with 2D videos, fail to address the embodied, spatial, and social dimensions that define collaborative musical experience.

In this context, mixed reality (MR) technologies [100] appear particularly promising for musical collaboration, as they enable musicians to continue using their physical instruments while augmenting their environment with immersive telepresence capabilities that could address key limitations of current NMP systems such as spatial awareness and reduced social presence (e.g., [16, 22, 96, 109]). This convergence of immersive technologies and distributed musical practice points toward emerging visions such as the "Musical Metaverse," which refers to persistent, interconnected virtual and physical environments designed specifically for collaborative musical creation and live performance [104]. These developments were possible because of recent technological advances, which have made MR collaboration increasingly viable: video passthrough head-mounted displays (HMDs) have reached consumer accessibility (e.g., Meta Quest 3, Apple Vision Pro), tracking technologies have become more sophisticated, and research in social XR and computer-supported cooperative work (CSCW) has advanced our understanding of distributed collaboration systems [43].

However, a critical gap exists. Current research on collaborative MR focuses primarily on office-based scenarios that differ fundamentally from professional musical collaboration requirements. Musical performance presents unique demands that extend beyond traditional CSCW and telepresence applications: musicians must share precise temporal experiences, maintain high-fidelity audio transmission for real-time creative expression, and preserve the social and emotional dimensions that drive collaborative creativity. Current research on MR musical collaboration remains constrained to technical feasibility studies, coupled with limited user experience evaluations involving small pools of participants (e.g., [4, 16, 22, 25, 58, 96, 107, 110]). These limitations stem from complex integration challenges between collaborative MR and NMP systems, where stringent musical requirements create technical contradictions that constrain research scope and practical implementation [17]. More critically, in contrast with the increased interest in MR systems for music [15], there is no systematic investigation of professional musicians' requirements or design preferences for MR musical collaboration systems. Without understanding how expert practitioners envision these systems, MR musical collaboration risks developing along trajectories that are not able to properly support actual professional creative work, and potentially excluding the very users whose adoption could validate these technologies for broader musical communities. These gaps motivate our central research questions:

- **RQ1:** What design requirements do professional musicians identify as essential for MR musical collaboration systems?
- **RQ2:** Are the identified requirements challenging or extending current collaborative MR approaches?

To address these questions, we conducted a co-design study with 24 professional musicians, organized into six four-piece band configurations. Our methodology employs a three-phase approach that bridges current technological constraints with future design possibilities by combining three complementary components:

- Experience Grounding: Direct embodied experience with existing MR and audio technology during a collaborative performance;
- Individual Ideation: Structured speculative design activities to envision ideal future systems as if they already existed "*by magic*" [5, 15, 72];
- Collaborative Convergence: Group negotiation to identify consensus-based requirements that transcend personal preferences.

This approach enabled participants to ground their design requirements after a direct experience with MR technology, while at the same time engaging in speculative activities that encouraged them to imagine systems unconstrained by current technical limitations (see Figure 1).

Our findings reveal that for professional musicians the sonic dimension of the experience constitutes the primary reality, with visual elements remaining secondary, directly contradicting prevailing collaborative MR design assumptions that prioritize visual immersion. The remarkable consensus achieved across diverse professional musicians provides concrete guidance for MR platform developers while contributing to the broader understanding of how immersive technologies can genuinely support, rather than constrain, professional creative practice.

This work makes three primary contributions to the HCI and music technology communities:

(1) An empirically-grounded, consensus-based set of design requirements for professional MR musical collaboration derived through systematic co-design processes with expert practitioners.

(2) A methodology combining participatory and speculative design approaches to bridge current technological constraints and future possibilities.
(3) Evidence that professional creative collaboration demands fundamentally different design priorities than entertainment or productivity applications, challenging current MR development paradigms.

As MR technologies mature, these requirements provide foundational principles for developing systems that enable geographically distributed musicians to collaborate as if they were sharing the same physical space. By considering collaborative MR through the demanding requirements of professional musical practice, this work provides insights for researchers and practitioners into how immersive and collaborative musical systems can support authentic real-time activity and human expression in domains that extend well beyond the musical one.

2 Background

This work is positioned at the intersection of research on collaborative MR, NMP, and participatory design methodologies for music technologies. By reviewing existing research across these domains, we show critical gaps at their intersections—gaps that inform our research questions and methodological approach.

2.1 The Promise and Limitations of Current Collaborative MR

Research in collaborative MR has established foundational understanding of multi-user interaction in remote and co-located immersive spaces [43], yet it focuses predominantly on office-based scenarios that differ fundamentally from musical collaboration requirements (e.g., [49, 50, 61, 66]). Recent studies have especially examined whiteboard collaboration [50] and object manipulation in hybrid workspaces [61], as well as facing formations for large group meetings [117]. While advancing our understanding of traditional CSCW tasks –such as group meetings and document sharing– they reveal an implicit assumption that collaboration follows patterns of turn-taking, explicit task division, and artifact-centered interaction. Musical collaboration, by contrast, demands continuous real-time synchronization and shared temporal experience that extends beyond traditional productivity paradigms.

Previous work on collaborative MR has recognized the importance of social presence and co-presence as foundational components for interacting with remote users in immersive spaces [83, 113], identifying eye contact, facial expressions, and body language as critical factors for effective collaboration and affective communication [11, 67, 84, 92]. Studies have also shown that realistic avatar representations can increase social presence [75, 84], even with low-fidelity implementations [91], while gaze and facial expression capture improve communication effectiveness when applied to 3D embodied avatars [11, 92]. However, these insights have not been systematically translated to the design of creative collaborative contexts for music, where nonverbal communication serves more than mere coordination: it enables professional musicians to negotiate musical interpretations in real-time through gestures, breathing patterns, and postural shifts, exceeding what current collaborative MR systems can support. This gap between office-centric design and creative collaboration motivates our first research focus: understanding what professional musicians actually require from MR collaborative systems.

2.2 Networked Musical Performance: Technical Achievements and User Experience Gaps

NMP research has established technical foundations for distributed musical collaboration [93]. Previous work has demonstrated that ultra-low latency and high-quality audio transmission can effectively support musical synchronization among displaced musicians [26], with research consistently identifying sub-30 millisecond end-to-end latency as necessary for rhythmic coordination in ensemble performance [27]. This stringent requirement has shaped all subsequent NMP technical research and continues to define the success or failure of such systems. Widely adopted systems (such as JackTrip [21] and LoLa [41]) can provide low-latency uncompressed audio transmission, though they might require specific network configurations and dedicated bandwidth allocation. However, the vast majority of current NMP systems operate primarily in audio-only contexts. When visual communication is included (e.g., [39, 41]), it typically consists of basic video conferencing capabilities rather than more immersive configurations with embodied representations of remote users.

This situation reflects both technical constraints and implicit assumptions about the needs of musicians. While integration of sophisticated visual feedback remains limited by bandwidth and latency constraints, it also reflects a research trajectory prioritizing technical feasibility over user-centered research and systematic understandings of musicians' needs. MR research has primarily explored office-based collaboration with limited investigation of musical performance, while NMP research has focused on musical requirements but with less systematic exploration of how immersive technologies might reshape collaborative practices.

2.3 Emerging MR Musical Systems: From Technical Feasibility to User Requirements

Recent research has begun exploring the intersection between collaborative MR and NMP, though work remains constrained by several challenges, mostly technical. While the majority of the research has explored the use of simplified avatars with real-time audio [22, 23, 57, 109], technical investigations have identified latency mismatches between audio and visual aspects of the experience, particularly when point cloud capture and streaming are combined with networked audio [107], which represents the most data-intensive but also the most realistic scenario [16, 96]. While musicians require sub 30 ms of latency in audio signals, volumetric capture and streaming system work at much higher latencies and users can tolerate up to 900 ms of end-to-end delay [36, 112].

As a result, technical integration challenges have narrowed the scope and methodology of research in this area. Studies remain confined to controlled laboratory environments with simplified scenarios, focusing necessarily on technical feasibility (e.g., [16, 23, 96, 107]). More critically, existing work typically involves small participant pools mostly playing in co-located settings [22, 23, 110], or where data streaming only includes the visual but not the audio component of the experience [95, 96], preventing systematic identification of requirements that transcend individual preferences and specific technical issues. Even if some studies have explored aspects of the user experience such as immersion and social presence [16, 23, 96, 109] they have examined mostly dyadic improvisations [58, 96, 107, 109], single-musician interactions with virtual agents [25, 57, 110], and in few occasions small ensemble performance [22, 23].

While technical constraints have limited user research to small-scale feasibility studies, building systems without understanding professional needs risks producing tools unsuited to actual musical practice.

2.4 Participatory Design in Music Technology: Lessons and Opportunities

In order to move beyond purely technology-driven research, the field of music has embraced participatory design methods to place users at the center of the development of new systems. Studies have engaged musicians in designing digital musical instruments [44, 45, 106, 108], accessible interfaces [101, 105], music information retrieval systems [68], as well as musical VR systems for creative practices [8, 89].

Some research has combined participatory with speculative design for future-facing issues. Lepri and McPherson [72] developed the "making up instruments" workshops, while others integrated design fiction strategies [15, 63, 71], demonstrating that musicians can productively engage with future possibilities when balancing current experience with speculative imagination.

However, most of these works focus on artifacts—whether physical or virtual instruments—rather than communication platforms like NMP systems. This means that such methodologies may not automatically translate to the design of collaborative infrastructure.

While participatory workshops have successfully elicited MR collaboration requirements from diverse groups—collaboration designers [118], UX designers [119], ethics experts [69], and children [47]—even without complete prototypes, professional musicians have never been systematically engaged as co-designers for MR collaboration platforms. Previous work has only explored their broad needs in the context of the Musical Metaverse [15].

Musicians possess embodied expertise about collaborative performance that cannot be inferred from office-based patterns or technical capabilities alone. Since technical challenges currently constrain prototype-based research, participatory approaches offer a way to gather requirements unconstrained by implementation limitations.

3 Methodology

We conducted a study using a qualitative methodology to understand how expert musicians envision the characteristics of an MR application for networked musical collaboration, focusing on design requirements and interaction concepts. Our approach positioned musicians as expert co-designers along with the authors, enabling deep engagement with both current collaborative practices and future possibilities.

3.1 Participants

We recruited 24 professional musicians (9 participants identified as women, 15 as men; age range 18-54, M = 27.6, SD = 12.4) through purposive sampling targeting active performers with extensive experience in synchronous group performance. Recruitment occurred through SAE Barcelona, local music conservatories, and professional music networks in Barcelona and the region of Catalonia.

Participants reported an average 12.7 years of musical experience (SD = 11.1), with regular activities including live performances, studio recording, and collaborative rehearsals across genres including rock, pop, funk, jazz, Latin, and metal. This experience base ensured familiarity with the embodied, temporal, and social dynamics characteristic of professional musical collaboration. Twenty participants were from Spain, while four were from South America. Only five participants had previous experience with HMDs, even if sporadic and in the context of gaming in VR.

Participants were organized into six four-piece bands, each representing a standard ensemble formation: drums, bass, guitar, and keyboard. This configuration was selected because it represents a common professional arrangement across multiple genres while maintaining analytical tractability for consensus-building activities. The study received approval from the University of Trento ethical committee, with all participants providing informed consent for participation, audio recording, and use of created materials.

3.2 Procedure

Our study was presented to participants as a fictional scenario describing a situation they were asked to design. The scenario was grounded in previous work on MR networked music systems [17, 107] and exploratory studies that have converged on a common setup: a color video passthrough HMD, audio delivered through headphones [16, 23, 96, 110], and an embedded system for network audio transmission [23, 58, 107]. The scenario was communicated to the participants through the reading of the following text:

> You are in your room, the one you use for playing your instrument, practicing, and recording. You have been asked to join a jam session with three other remote musicians you don't know. This will happen through a new application designed to play music together over the internet but also capable of making you feel as if you were in the same room with them, both visually and auditorily. You will see the other musicians through a mixed reality head-mounted display and you will hear them through a pair of stereo headphones.

Consequently, we formalized the request to participants using this prompt:

> If this application could exist as *if by magic*, how would you imagine it?

The use of a *magic* prompt after a direct hands-on experience with current technology was intentionally designed to create productive tension. Rather than producing naive wishful thinking, this approach enabled participants to elaborate ideas and requirements

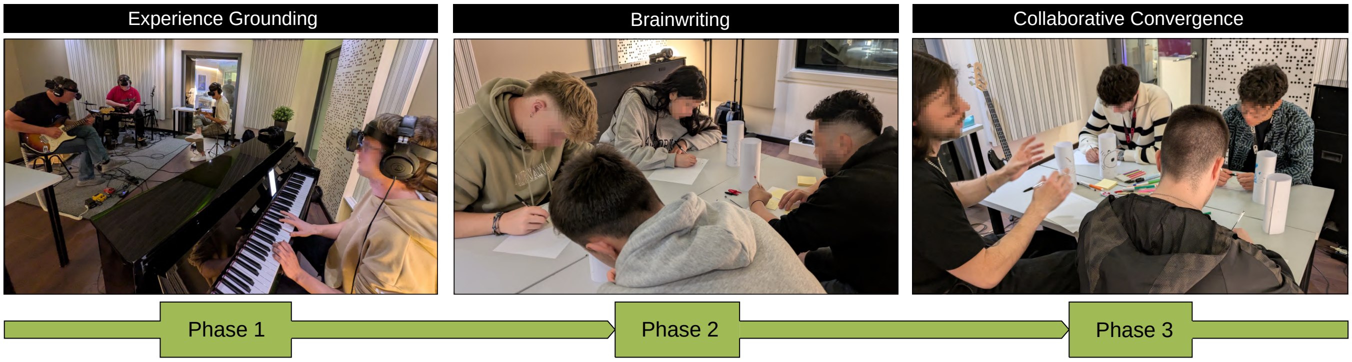

Figure 2: The procedure used in the study. At first, musicians were introduced to an experience grounding activity (Phase 1), then to a *brainwriting* activity (Phase 2), and then to a collaborative convergence phase (Phase 3).

while being aware of current limitations. Their expertise enabled them to distinguish fundamental collaborative needs from technological flights of fancy.

All group sessions began with a structured introduction (of approximately 10 minutes) made by two members of the research team describing the goal of the research project and the role that MR and NMP technologies play in our project. Subsequently, the research activity started. It was divided into three phases, as illustrated in Figure 2:

- **Phase 1 – Experience grounding** (~45 min): Musicians played together in the same room, wearing an MR headset with video passthrough and a pair of headphones for the mixed audio. Audio routing ensured that the sound of a musician was correctly heard by all the others at the same time. The purpose was to give participants a shared, first-hand reference experience with the devices that musicians would need to use apart from their instrument in a collaborative MR application.
- **Phase 2 – Individual ideation and sketching** (~90 min): Participants were seated individually with A4 sketch papers, colored markers, and post-it notes. By referring to the fictional scenario and the prompt provided by the facilitators, musicians were asked to engage in a *brainwriting* activity [111], to design from their perspective the different components of a future application for networked music using MR technologies.
- **Phase 3 - Collaborative convergence** (~90 min): Each musician presented the sketches produced in the previous phase to the other members of the group. Then, the group discussed together overlaps and differences of each idea, negotiating a shared set of features they would prioritize in an ideal system. The final output of each workshop was a list of agreed features and principles.

The activity occurred over the period of two weeks, with a total of four days. Each group session lasted approximately four hours, including breaks, and the time dedicated to preparation of the room and setup of the instruments.

All activities were conducted at SAE Barcelona. Two of the authors, researchers from the University of Trento, facilitated the workshop. English was the primary language used; however, particularly during Phases 2 and 3, some parts were conducted in Spanish (Castellano), which both researchers spoke.

Figure 3: The two devices used in Phase 1: a Lynx R1 HMD, and a pair of studio professional headphones.

3.3 Phase 1: Experience Grounding

In the first activity, participants were asked to familiarize themselves with the two primary technology components that would serve as the foundation for the application they were tasked with designing: an MR HMD and a pair of stereo headphones for audio, as shown in Figure 3. This gave musicians direct experience of what it means to play with an HMD, as described by Bailenson and colleagues as *"seeing the world through digital prisms"* [10]. Headphones, already familiar to professional musicians, can enable high-quality rendering and spatial audio delivery.

Since the technical integration between NMP and MR systems has limited the majority of previous work, we deliberately excluded networking aspects since this remains technically challenging for MR with physical musical instruments. We assume this will be solved in the future as research develops. While this approach cannot fully simulate next distributed collaboration systems, it enables the systematic exploration of how MR hardware interfaces can integrate with established musical practices. Moreover, by focusing on local MR experiences, we isolated the effects of wearing HMDs and headphones during musical performance without confounding network-induced artifacts.

Musicians conducted Phase 1 activities in a dedicated room (referred here as Performance Room) while audio mixing was managed

Figure 4: On the left, the Performance Room where musicians played in Phase 1. On the right, the Control Room used by sound engineers to route and mix the audio streams coming from the Performance Room.

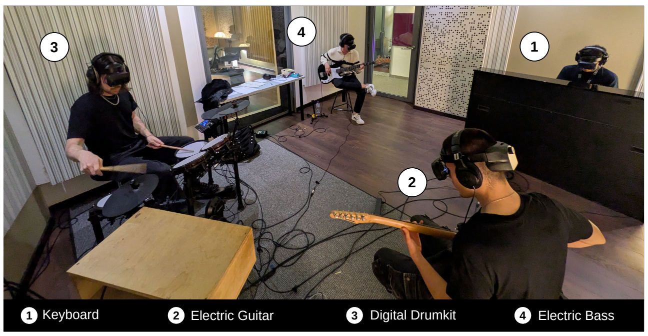

Figure 5: Members of Group 4 jamming during Phase 1. Numbered positions show: 1) Keyboard player at digital piano, 2) Electric guitarist, 3) Digital drum kit player, and 4) Electric bass player. All musicians are wearing MR headsets and headphones.

by a sound engineer of SAE in an adjacent Control Room, simulating the behavior of an NMP system (see Figure 4). Before the beginning of each music session, five minutes were dedicated to a single and group soundcheck, since basic setup and signal routing were done in advance by SAE technicians.

For the MR experience participants were provided with Lynx R1 HMDs. These devices can provide video passthrough that can be experienced through specially designed 4-fold catadioptric freeform prism optics. This provides a ~90° circular field of view. The Lynx R1 features 1600x1600 LCD displays (90Hz), Qualcomm Snapdragon XR2 Gen processor, and Wi-Fi/Bluetooth support. We developed a Unity application used only to display video passthrough feed without additional virtual overlay to be used during this phase. Audio was provided via Audio-Technica ATH-M20x high-quality stereo headphones, chosen for their ease of integration with the HMD. Through this setting, each musician was able to hear their sound mixed with the one of the others in real-time.

Since our focus was on the experience of musicians playing their own physical instruments, we used only electric and digital instruments in order to minimize disturbance, since audio was delivered through headphones. While SAE provided several instruments (including guitar, bass, keyboard and a digital drum kit), participants were free to bring their own personal instruments and sound effects (e.g., guitar pedals).

Participants received 10-minute guided explanation of the MR HMD, including hand tracking, object manipulation, and basic spatial mapping features. Following this tutorial, musicians were asked to engage in 15 minutes of collaborative musical improvisation while wearing headsets and headphones (see Figure 5). Participants were then asked to choose a song or a musical structure, rather than forcing specific styles. For example, the second group improvised on *"Oye como va"* by Santana following the guitar player's proposal, while the third group agreed on a jazz standard proposed by the keyboard player after a brief discussion. Upon completion, musicians removed HMDs, headphones, and instruments. After a 15-minute break they were asked to elaborate on their experience during an unstructured interview session conducted by two of the authors.

3.4 Phase 2: Individual Ideation and Sketching

First, musicians were asked to sit together around a table. The two researchers illustrated the goals and how the activity was going to unfold. Second, musicians were presented with four sections of the imaginary application. We selected such sections to help musicians maintain focus and to better encourage the exploration of different design dimensions. The four sections were the following:

(1) **Preparation**: Setup requirements before starting the application, including instrument, HMD, and headphone preparation sequences;
(2) **Join a Session**: Requirements for starting the application and joining a playing session;
(3) **Play in a Session**: Collaboration needs including spatial arrangement, player appearance, audio preferences, and real-time collaborative features;
(4) **Exit a Session**: Steps needed to leave a session and quit the application.

Each section lasted for a variable time between 10-20 minutes depending on how much time musicians took to complete the activity, and if they needed more clarifications from the researchers. Participants were asked to think about the features they would like to see in the future MR application through words and sketches. To develop their ideas, participants were provided with white A4 paper sheets, colored markers, ball pens, sticky notes, and basic prototyping materials including cardboard, and tape. Participants were encouraged to think beyond current technical limitations and previous knowledge, and to focus on ideal design goals. Figure 6 presents a series of sketches made by the participants.

3.5 Phase 3: Collaborative convergence

In the final phase, participants were asked to share their sketches and ideas with the group. Following each presentation, a group discussion took place. Such a discussion was aimed at building on and combining individual ideas. Researchers facilitated with additional questions, and helped to identify key requirements and define consolidated design concepts representing group consensus.

3.6 Data Collection

Data were collected as follows: In Phase 1, we recorded audio of the final interview session, and one researcher took detailed notes. In Phase 2, we collected only the sketches made by participants,

Figure 6: Examples of sketches produced by the participants for the four sections explored during Phase 2.

which were later digitized through high-resolution scanning for archiving. In Phase 3, we recorded audio of the entire discussion, and one researcher took detailed notes. Participants provided explicit consent for audio recording and sketch usage, understanding materials would be used solely for research analysis.

3.7 Data Analysis

We analyzed data from Phase 1 and Phase 3 using thematic analysis [19] to identify patterns in participants' experiences and consensus-based design requirements.

3.7.1 Analysis of Experience Grounding (Phase 1). We conducted a thematic analysis of the recordings and notes from Phase 1 interviews. Two researchers from the University of Trento independently performed open coding on the audio transcripts and discussion notes using an inductive approach. Initial codes were descriptive labels capturing discrete ideas and concepts mentioned by participants. The analysis proceeded through three stages: (1) independent coding by both researchers, (2) two collaborative meetings to discuss findings, convergences, and divergences, and iteratively refine categories, and (3) a final meeting to resolve disagreements through joint re-examination of source material. Final themes were selected based on prevalence across participants, conceptual coherence, and distinctness from other themes. Both researchers examined cross-group consistency and instrument-specific patterns.

3.7.2 Analysis of Collaborative Convergence (Phase 3). The two researchers who performed the first analysis independently identified design requirements mentioned during group convergence sessions, coding both individual presentations and final group agreements. Consensus levels were calculated based on individual participants: unanimous (24/24), near-unanimous (23/24), and high consensus (20-22/24). When groups showed competing requirements, we documented both majority positions and dissenting perspectives, noting instrument-specific variations. Disagreements were resolved through discussion and eventually by revising the source materials. Finally, we identified commonalities and –if any– differences across groups.

Following the identification of 12 initial themes, researchers performed axial coding to establish relationships and connections between these themes. This process grouped them into broader conceptual categories based on their underlying patterns, revealing higher-level concepts.

4 Results

Our analysis showed distinct patterns in how professional musicians experience and envision MR musical collaboration systems.

4.1 Experience Grounding Findings

The thematic analysis of Phase 1 revealed five major themes regarding how participants responded to playing instruments using HMDs and headphones, highlighting both challenges and opportunities in MR-mediated musical collaboration. The themes are summarized in Figure 7 and presented as follows:

(1) **Visuo-Motor Conflict**: The most prevalent theme involved a fundamental conflict between the visual attention required for instrument control and the need for social coordination with other musicians, as reported by 22 participants. This conflict manifested differently across instruments, and represents the primary barrier to seamless MR musical performance. Keyboard players experienced the most severe challenges. They exhibited exaggerated head movements when alternating between monitoring hand positions and maintaining visual contact with the other musicians. One keyboard player explained: "*I had to make extreme movements—down to watch my fingers, up to see the others—which made me dizzy.*" Guitar players reported difficulty seeing their instruments, especially the guitar neck, while maintaining awareness of other musicians, describing the transition between direct and mediated vision as "*weird and a bit disturbing.*" Drum players faced unique spatial challenges, often describing the constant trade-off between "*seeing the drums or seeing the others*" as a persistent difficulty.

(2) **Social Disruption**: Fundamental disruptions to social connection characterized the experiences of 19 participants. Social isolation emerged as the most significant aspect, with multiple participants describing feelings of being "*distant and alone*" despite the collaborative nature of the activity. According to the musicians, this was caused mostly by the HMD and the need to concentrate more on the visual aspect of the activity. One guitarist's observation—"*it's a little bit sad for me to think that we play alone*"—captured the emotional dimension of this isolation. Physical discomfort also contributed significantly to such disruption, with participants reporting headset weight and ergonomic problems. Drummers were particularly affected by equipment weight and movement constraints, with one noting, *"since I move a lot, I don't want extra weight/cables, I got a bit dizzy.*" The introduction of the HMDs elicited ambivalent responses to immersion, with some embracing the cinematic quality of the visual experience through HMDs, while others experienced it as creating barriers to authentic musical interaction.

(3) **Audio Quality Matters**: In contrast to the visual and social challenges, the quality of the audio transmission emerged as a notable success, with 19 participants reporting positive responses. Participants consistently reported that audio quality was "*better than typical live performances*," suggesting that MR networked systems should aim to provide the best audio experience possible. The quality of the audio delivery provided a stable foundation that allowed musicians to maintain basic coordination, even when visual communication was compromised. The reliability of audio channels allowed participants to focus on adapting to other challenges without simultaneously struggling with audio deficits, fostering optimism about overcoming remaining limitations through technological development.

(4) **Loss of Musical Coordination**: Communication breakdown between musicians emerged as a critical theme, as reported by 16 participants. This highlighted the fundamental importance of non-verbal communication channels, which were disrupted in current MR implementations due to the use of HMDs. The loss of eye contact was the most unanimously reported deficit across all instruments. One bass player emphasized, "*I need to see the face of the people, especially the eyes*," while a guitarist noted, "*There is no visual contact...I can't see the eyes.*" These challenges were particularly pronounced in improvisational contexts, where spontaneous coordination relies heavily on subtle visual cues. One guitarist observed that visual communication is "*especially important during jams/improvisations*." The breakdown extended beyond immediate performance coordination to also affect the emotional dimensions of musical collaboration.

(5) **Professional vs. Recreational Use**: Eight participants explicitly differentiated between professional and recreational applications, revealing context-dependent technology acceptance. Concerns about professional suitability emerged as a significant barrier, with participants describing current systems as appropriate for "*fun*" rather than "*professional stuff*." One pianist evaluated the system as being "*a level as fun, but not really professional.*" Educational applications emerged as a promising middle ground, with participants identifying teaching contexts as potentially viable. This professional-recreational divide varied across instruments, with piano players showing the strongest skepticism and guitar players demonstrating greater willingness to consider professional applications once limitations are addressed.

Additionally, we observed that adaptation to HMDs varied significantly across instruments, with 12 participants mentioning possibilities for adjustment. Bass players demonstrated the most rapid adaptation, while piano players faced the greatest challenges, requiring specialized accommodations rather than simple practice-based improvement. Drummers showed variable but generally positive adaptation patterns, with participants expressing confidence in eventual adjustment through practice. Guitar players demonstrated mixed outcomes, depending more on their musical roles than instrument specific factors. Lead guitarists, who required visual communication for coordination, reported greater difficulty than rhythmic guitarists, who could rely more on audio-based coordination. This variation suggests that MR systems should be designed to accommodate different musical roles within the same instrument category.

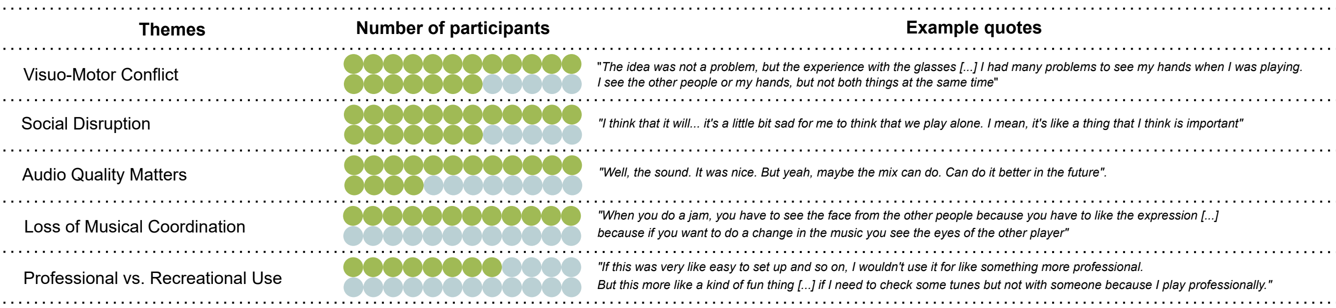

Figure 7: The themes derived from the analysis of participants' experience during Phase 1.

Figure 8: Consensus-based design requirements derived from Phase 3 collaborative sessions.

5 Consensus Patterns in MR Musical Collaboration

The analysis of the results of Phase 3 revealed strong consensus among musicians regarding core requirements for MR musical collaboration systems (see Fig. 8). These requirements form an interdependent hierarchical structure organized around four core findings that span the entire interaction lifecycle of the experience and transcend individual instrument characteristics and musical genres.

5.1 Workflow Integration as a Non-Negotiable Foundation

The first and most critical finding is that musicians are not interested in engaging with MR collaboration until their established professional audio workflows are fully preserved and operational. This represents a fundamental design constraint: MR systems must adapt to musicians' proven practices rather than imposing novel paradigms.

This workflow preservation manifests across the entire setup sequence through what we term **Audio-First Workflow** (24/24 participants). Sound verification should precede any visual or social engagement. As one pianist explained: "*I would plug the instrument first and the last step will be open the app. It's more relaxed knowing that all the setup is working*". A guitarist reinforced: "*After properly setting up the instrument, headset and headphones, check the sound and adjust it. Finally, start the session*". Musicians treat their entire signal chain as a unified system where audio quality cannot be compromised, viewing the HMD as just another piece of professional equipment.

Yet alongside this identified workflow, musicians demand radical simplicity in execution through **Plug-and-Play Setup** (24/24). As one guitarist stated: "*Realistically, I probably don't want to connect the guitar every time* [...] *so I would prefer to be connected already and push a button and the system goes right away*". Therefore, MR systems must be streamlined while respecting established workflows, and not eliminating them.

The resolution emerges through intelligent defaults and context-aware **Presets and Default Configurations** (20/24): musicians want systems that can remember configurations for different musical scenarios (e.g., genre or activity), and that can be instantly recalled but at the same time remaining fully customizable. This adaptiveness should enable the same musician to seamlessly transition between different collaborative contexts without extensive reconfiguration, as noted by a drummer "*I think personal configuration is really important* [...] *because you don't want to play with the same people every time*". Musicians also emphasized physical integration through **Integrate Musical Equipment Workflow** (22/24), such as dedicated charging solutions and protective storage that fit within existing studio layouts. The universal concern is that MR equipment should enhance rather than disrupt their established creative processes, from initial setup through final teardown.

5.2 Communication and Social Presence

Given a proper audio foundation, the second finding reveals that musical collaboration requires simultaneous visual and audio representational fidelity that musicians consider non-negotiable.

Support for **Visual Communication** achieved complete consensus (24/24). A guitarist emphasized: "*The precision of the movement I see is more important than the appearance* [...] *It's more important to have the consistency between the gesture*". Musicians require this aspect not for social pleasantness but for functional coordination. Visual communication proved equally critical as musicians requested **Realistic Avatar Representation** (22/24), though context-dependent. A drummer stated: "*For me, realism is a bit more important when playing professionally* [...] *but just to play for fun, monkeys and dragons or like whatever avatar you want to have is fine*". This reveals a cross-cutting dimension: professional contexts demand maximum realism for transmitting subtle coordination cues, while casual sessions allow stylization, suggesting that collaborative systems should adapt fidelity levels to session type rather than applying uniform settings.

Real-Time Voice Support (23/24) complements rather than replacing visual channels. Musicians need both explicit coordination through voice (particularly valuable in improvisation where verbal cues can be used to initiate new sections) and implicit coordination through visual observation. Voice also serves pre- and post-session functions: discussing musical ideas, organizing roles, and building relationships when collaborators meet for the first time. As a guitar player noted "*this is very important* [...] *maybe a gate or something that when you talk or when you shout while you're playing, they hear you.*"

Together, these multimodal requirements show that musicians conceptualize remote collaboration systems to be used for supporting in-person dynamics as completely as possible, not for creating diminished "remote" experiences.

5.3 Customizable Music Spaces

The third finding highlights that musicians demand complete personal control over their audio and visual environment while maintaining a shared collaborative space.

Personalized Audio Control showed perfect consensus (24/24). A guitarist noted: "*I would like to see like a mixer in my glasses and I can just adjust what I need when I need* [...] *some of elements I can go and adjust it and I have like fader knobs*". Musicians require independent control not only of their own instrument's input and output, but also of the audio channels from the connected musicians. This mirrors traditional mixing board paradigms where each performer crafts their personal monitor mix without affecting others' experiences.

Similarly to what was observed with audio, **Spatial Layout Customization** achieved equal consensus (24/24), with musicians demanding agency over spatial arrangements. As one pianist explained: "*What I want is to change positions, to drop them differently and to have the option to change also to*". Since collaborators appear as embodied avatars or volumetric representations superimposed on musicians' physical space, positioning must remain flexible rather than imposed—even conventional arrangements like semicircles should be supported and customizable.

Despite spatial audio capabilities enabled by MR, **Stereo Mix Priority** (23/24) showed that musicians prefer familiar stereo mixing as their default. A drummer clarified: "*Stereo mix because it's as you hear when you are playing live*". Most musicians prefer established stereo mixing techniques as their default, with spatial audio retained as an optional alternative. This hierarchical preference—familiarity as default, innovation as choice—allows each musician to select their preferred paradigm individually without forcing consensus within the group. These control requirements reflect a core principle of professional music-making: collaborative performance depends on individual autonomy. Each musician must craft their optimal listening and viewing conditions to contribute effectively to the collective output.

5.4 Session Management

The final finding reveals that musicians require proper closure in both musical and social dimensions, treating session endings as opportunities to preserve collaborative outcomes and maintain relationships for future work.

Recording and Playback capabilities proved essential (20/24), with requests spanning from being able to save simple stereo mixes to multi-track sessions with individual instrument isolation, and video recordings as well. A guitarist noted: "*At the end of the session, the video is uploaded to the cloud, allowing access to complete recordings*". Musicians also want to replay recordings within the application for rehearsal and play-along activities, extending the system's utility beyond live collaboration. This emphasis on capture reflects musicians' broader expectation that MR systems preserve rather than replace their established creative workflows—recording is not an add-on feature but a fundamental professional requirement.

Departure Coordination (22/24) addressed the social dimension of endings. Musicians need clear notification when colleagues exit, but critically, they prefer gradual transitions over abrupt disappearances. A drummer explained: "*In the case of musicians present in the scene, their exit is animated with a fade-out*". This preference reflects how sudden visual changes can disrupt concentration and musical flow. Graceful departures should maintain social continuity—preserving the quality of relationships established through the multimodal presence described earlier—while clearly communicating the session's conclusion.

These requirements can work interdependently: departure coordination often includes deciding about recordings before leaving, while recording capabilities provide tangible artifacts that validate the collaborative effort and enable continued musical development after the session ends.

6 Discussion

Our findings reveal fundamental differences between current MR design paradigms and the specialized requirements of professional musical collaboration. The consensus achieved across 24 professional musicians provides empirical evidence that complements and extends previous work on collaborative MR, CSCW, and telepresence, while identifying pathways toward systems that genuinely support expert creative practice. We discuss the Phase 1 results first, followed by the themes that emerged from our Phase 3 analysis.

6.1 The Limits of Passthrough MR for Social Musicianship

Why did experienced musicians struggled with video passthrough MR during collaborative performance? The visual-motor conflicts reported by 22 participants in Phase 1, reveal a conflict between technological mediation and the embodied expertise that characterizes professional creative practice.

This finding aligns with previous observations of professional musicians using video passthrough HMDs (e.g., [16, 96]) as well as see-through glasses [115]. Research showed that guitar players reduce head movement when wearing passthrough HMDs to avoid issues such as latency or motion blur, which can negatively impact performance [16]. The introduction of HMDs as mediating artifacts created fundamental contradictions within established activity systems of musical collaboration.

Embodied music cognition [70] illuminates why these disruptions proved so consequential. Musical performance represents a paradigmatic type of embodied interaction, where meaning emerges through physical engagement with both instruments and collaborators [64]. The tight coupling between physical gesture and sonic production –what Godøy calls *motor-mimetic processes* [48]– creates meaning-making systems particularly vulnerable to technological disruption. Since these processes rely on seamless coordination of vision, movement, and sound [20, 76], even minor occlusions, visual distortions, or latencies in MR systems can break the experience the performers depend on. As one player noted, the transition between direct and mediated vision was "*weird and a bit disturbing*," revealing how new mediating artifacts disrupt established musical collaboration activities.

Our results suggest that for musicians, an HMD represents a novel device that requires integration into their body schema. Additionally, our findings indicate that such expertise might create conservative adaptation pressures where technological intervention must try to preserve existing embodied coordination systems. The successful audio experience that 19 participants reported validates this principle, since audio delivery through headphones preserved rather than disrupted established musical coordination patterns. However, these vulnerabilities to technological mediation found in skilled users are not yet well-documented in the HCI literature. Participants also reported social disruption caused by HMDs impeding clear perception of others, noting that they felt alone even when present in the same room. These observations align with findings by Santoso and Bailenson [94] indicating that video passthrough contributes to feelings of social absence.

While these findings regarding Phase 1 may be dependent on the particular HMD used in the study, future work in MR and music technologies should better examine the causes behind the issues we observed and develop alleviating designs. For instance, it remains unclear what impact video passthrough latency (including camera capture, image processing, and display rendering pipeline) has on visuomotor coordination. Although this concern may be particularly acute for music-making, studying skilled users such as musicians can inform broader development of MR hardware, as well as immersive telepresence and collaborative systems.

6.2 Design for the Ear: Towards Audio-First MR

The consensus across all groups converged on a setup phase where audio should be configured first, with the HMD worn as the final step (see Figure 9). As one participant explained, "*I need to know my sound is working before I put anything on my head*." This reflects established professional protocols where sonic infrastructure must be verified before proceeding further. The consensus across diverse musical backgrounds demonstrates that this represents a shared professional knowledge rather than an individual preference. Moreover, this reflects an epistemological stance where sonic experience constitutes the primary reality [59], with visual elements serving supporting rather than foundational roles.

This prioritization of audio over video, challenges the *ocularcentrism* [87] of current design of MR collaborative MR, which emphasizes visual fidelity as foundational for presence and immersion [9, 49, 75, 83, 98, 99]. Our findings suggest that MR systems for professional musical work must instead adopt an "*Audio-First*" design philosophy [24], prioritizing technological transparency [18].

However, current commercially available MR headsets typically allocate computational resources primarily to visual rendering, treating audio as secondary [17]. Our findings suggest that professional musical applications require inverted resource allocation, where audio processing receives primary optimization. While productivity applications may succeed through visual-first approaches, professional musical collaboration requires audio-first architectures that preserve established tool-activity relationships. The notable absence of audio-related issues in our study validates this principle: high-quality audio provided a stable foundation that enabled coordination even when visual communication was compromised.

These findings invert Chion's concept of *added value* in audiovisual perception [29]: rather than sound adding value to the image, our participants positioned visual elements as potentially adding value only to an already-established sonic experience.

However, these findings must be contextualized within our participant demographics. The "*Audio-First*" design principle may not be universal across all musicians—D/deaf and hard-of-hearing musicians [46] would likely prioritize other sensory channels [85]. Future work should include D/deaf musicians in co-design processes to understand how MR systems can support multiple sensory hierarchies simultaneously.

6.3 When Faces Disappear: Nonverbal Communication in MR

Participants unanimously identified non-verbal communication support as essential for musical collaboration, with particular emphasis on eye contact and facial expression visibility. Research on music ensembles consistently shows that interactions between musicians rely heavily on visual cues, including body movements and hand gestures [13, 14, 38, 116]. The emphasis given by participants on facial expressions aligns with foundational work on nonverbal communication [42], while extending it into creative collaboration where micro-expressions serve not only emotional communication, but also aesthetic negotiation and creative decision-making [97].

Research has consistently highlighted the importance of gaze and facial expressions during musical collaboration [30, 79, 90, 114], and in immersive systems in general, showing how facial expressions

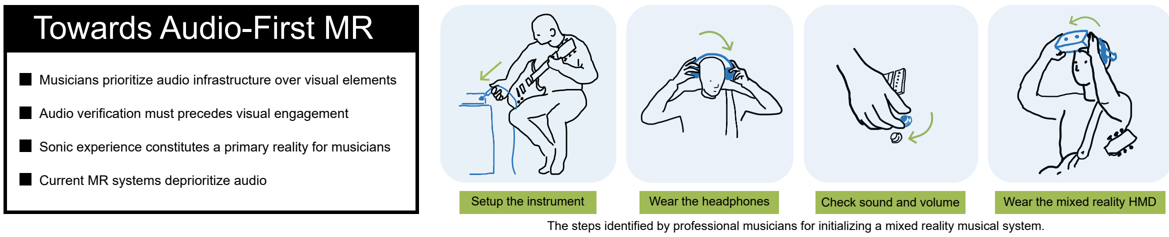

Figure 9: Towards Audio-First MR: Musicians prioritize audio infrastructure over visual elements in MR setup workflow.

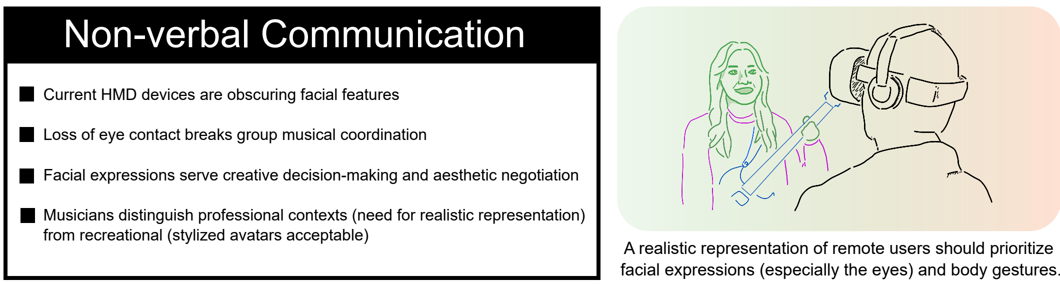

Figure 10: Non-verbal Communication: Current MR devices challenge essential non-verbal communication channels in musical collaboration.

can increase social presence [84]. As shown in Phase 1, HMDs mask portions of faces such as the eyes, thus limiting facial expressions. The loss of these channels represents a breakdown of the subtle social choreography that enables collaborative creation. Notably, several musicians expressed a preference for "*fake eyes than no eyes*," revealing that gaze is essential for maintaining basic human social interaction, not just technical coordination (see Figure 10). This finding underscores that even imperfect visual representation may be preferable when it comes to preserving social connection. Previous work showed that features such as user gaze [11, 92] and head nodding [1], as well as facial expression rendering [33, 75], can be simulated using motion capture systems. However, translating these elements into collaborative MR remains an open challenge. Beyond technical feasibility, musicians' responses revealed nuanced preferences about the nature of visual representation in collaborative contexts.

The preference expressed for *realistic* representations indicates an understanding of how visual authenticity serves different functions across contexts. Participants distinguished between *professional* and *recreational* contexts: professional interactions require authentic representation to enable transmission of non-verbal cues essential for coordination, while recreational contexts might accommodate stylized or fantasy avatars. This requirement concerns coherence of body movements rather than purely visual photorealism. This emphasis reveals that professional creative collaboration requires heightened social awareness encompassing cognitive, affective, aesthetic, and expressive communication dimensions beyond traditional social presence frameworks [12, 83].

For this, volumetric point-cloud representations have been explored in MR collaborative music [16, 96, 107], reproducing some of the effects observed in non-musical contexts [62, 91, 102]. However, the level of visual fidelity necessary for musical collaboration remains unclear. While studies of office meetings and learning activities suggest high visual fidelity as a non-essential requirement [62, 91], these findings may not transfer to musical contexts where subtle cues are critical. Furthermore, current real-time volumetric systems introduce latencies that exceed acceptable thresholds for live musical performance [36] and require technical configurations that are prohibitively complex for musicians to deploy independently [31, 86, 112].

Apple Vision Pro's spatial "Personas" point towards an alternative approach based on semantic communication [74], which is capable of achieving remarkable efficiency while maintaining realistic representation of remote users [28]. However, musical collaboration presents unique set of challenges for this approach. Unlike facial expressions, which can be efficiently encoded as discrete semantic tokens, musical instruments require capturing complex, continuous body movements with high spatial and temporal fidelity. A guitarist's fretting technique or a drummer's stick dynamics demand substantially richer data streams than standard human communication gestures, potentially eliminating the bandwidth efficiencies that make semantic communication viable.

Beyond capturing human movements, musical collaboration requires accurately representing the instruments themselves—an aspect largely unaddressed in previous MR collaboration research. Musicians' instruments are not merely held objects but extensions of their bodies, central to both their performance technique and communicative gestures. Current volumetric capture and streaming systems struggle to simultaneously represent fine hand movements, facial expressions, and the physical instruments with sufficient fidelity.

Addressing the loss of facial expressions and micro-gestures in MR environments thus represents not merely a technical problem requiring better sensing technologies, but a systemic challenge demanding coordinated advances in capture fidelity, and transmission efficiency, specifically designed for musical collaboration.

6.4 Balancing Individual Control and Collective Awareness in Collaborative Music Systems

Participants require control over both their own audio signal and personal monitor mix while maintaining an awareness of the collective sound. In particular, they unanimously demanded granular control over the audio stream coming from each peer (including volume, equalization, and panning), similar to capabilities found in Digital Audio Workstations (DAW) and professional mixing consoles. This need for individual autonomy within collaborative contexts aligns with established CSCW principles on personal and shared workspace design [51–53]. However, these requirements cannot be satisfied through predetermined system configurations or presets alone, since musicians need dynamic, real-time control adapted to each performance context. As several participants noted, musicians typically do not control the final audio mix in concerts or performances, as this is usually managed by specialized audio engineers. However, in this context this is considered an important feature.

While participants demanded access to basic parameters, current approaches may not scale well beyond three or four participants. The challenge lies in providing sophisticated functionality that remains accessible and intuitive for expert users in a multi-user environment accommodating an undefined number of audio sources and instrument types. For personal control, participants envisioned graphical overlays controlled by hand gestures or custom physical controllers like knobs or faders. This parallels current professional practice where musicians increasingly use personal digital mixers to control individual In-Ear Monitor mixes, though extending this model to collaborative MR environments presents unique spatial and interaction design challenges. For controlling other musicians' audio, participants proposed directly selecting the avatar representation of remote users to access specialized controls (see Figure 11).

This dual requirement –individual audio control and collective musical coherence– challenges traditional client-server architectures in NMP where a central mixer distributes identical audio to all participants. Instead, our findings suggest the need for what we term a "*federated audio architecture*" where each musician maintains independent control over their local audio mix while simultaneously contributing to the shared musical performance. Crucially, these individual mix adjustments affect only the musician's personal monitoring experience and do not alter what other participants hear, enabling each performer to optimize their listening environment without disrupting others' mixes or the collective sound.

Implementing federated audio architecture in collaborative MR contexts represents a major technical challenge: the system must support simultaneous, independent audio processing pipelines for each participant while maintaining synchronization across the network for coherent ensemble performance. This extends beyond user interface design to fundamental system architecture, requiring new approaches that balance personal optimization with collective musical coherence.

6.5 Spatial Authoring: From Mediated Spaces to Musical Places

The majority of participants expressed the need to control how remote musicians appear in their room and where they are positioned both visually and auditorily. While most participants were not strict about specific placements (e.g., drums in front, keyboard on side), they would like to be capable of rearranging them, requiring spatial authoring capabilities (see Figure 12). This aspect is poorly considered in collaborative MR research, which typically studies dyadic interactions or small groups where users can move freely in space. Musicians rarely change position during sessions, particularly when playing drum kits or keyboards, making initial placement vital for defining both visual perspective and auditory experience.

The ability to arrange virtual collaborators reflects a proxemic approach to both space and musical relationships [54]. Some participants expressed the desire to be able to save specific spatial arrangements for future sessions, though each musician may organize remote peers differently. This extends the classical CSCW distinction between "space" and "place" [55]: musicians require control over collaborative environments to transform spatial arrangements into meaningful contexts to support creative activities.

Recent studies on collaborative MR have addressed similar issues through techniques for aligning users in incongruent spaces [49], avatar placement re-targeting [121], and face-formations for large groups of participants [118]. However, these techniques derive from office-centric use cases [49, 61, 66] where spatial arrangements are often standardized and functional: musical practice requires culturally and practice-informed *space* that transforms into meaningful *place* through use. Musicians expressed the need to shape their environment based on genre conventions, interpersonal relationships between players, and the professional workflows of specific ensembles.

Professional musicians bring embodied spatial knowledge developed over decades of performance experience. MR systems must therefore accommodate, rather than override, this expertise by providing flexible spatial affordances. Successful platforms should act as *spatial instruments* that musicians can tune and configure like acoustic instruments [15], adapting their established practices to MR environments rather than forcing adaptation to rigid technological constraints.

6.6 Real-Time Voice Support as a Fundamental Communication Infrastructure

Almost all participants identified voice communication as a fundamental infrastructure that enables organization and coordination activities in collaborative music-making (see Figure 13). Recent studies have substantiated the importance of verbal communication in collaborative immersive environments for social presence [77] and task perception [78]. Although music is inherently non-verbal, in collaborative musical settings verbal communication serves crucial

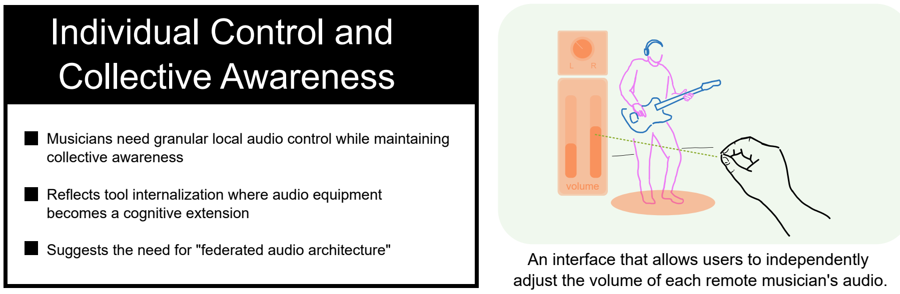

Figure 11: Individual Control and Collective Awareness: Musicians require granular local audio control while maintaining collective awareness in collaborative systems.

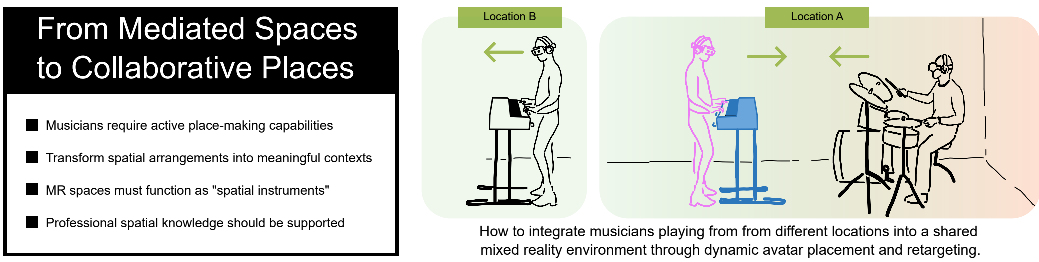

Figure 12: From Mediated Spaces to Collaborative Places: Musicians require dynamic place-making capabilities to transform spatial arrangements into meaningful creative contexts.

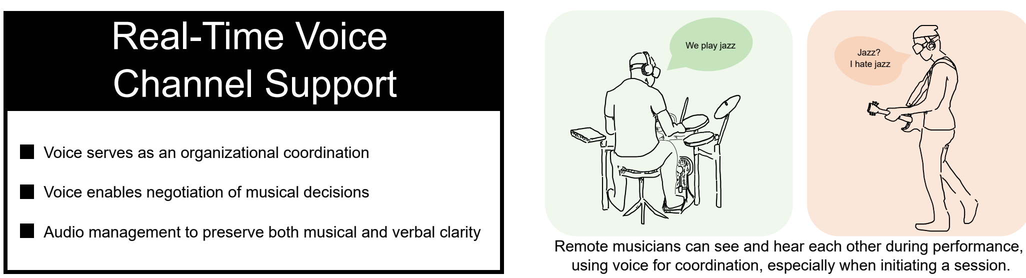

Figure 13: Real-time Voice Channel Support: Voice communication supports organizational coordination and enables negotiation of musical decisions during collaborative activities.

functions. Research showed that conductors find verbal communication more effective than gestures during rehearsals [80], while singers appear to respond better to verbal instruction than gestural cues [120]. In improvisational contexts, voice calls can initiate new sections, signal tempo changes, or coordinate complex transitions requiring precise timing beyond visual cues alone—common practices in jazz and flamenco traditions.

However, integrate voice communication effectively into MR musical collaboration is a distinct area of development. Unlike musical audio, speech is less sensitive to jitter and latency disruption, and dedicated codecs can employ robust error concealment techniques. [2, 88]. Moreover, humans generally tolerate higher latencies in speech due to multisensory integration mechanisms [37, 73]. While current standalone MR HMDs include microphone arrays for capturing speech, and speech communication is standard in social immersive applications [3]. These technical advantages do not eliminate design challenges in musical settings.

Musical collaboration systems must enable clear voice communication without compromising instrument audio quality. Poorly managed voice audio can mask musical detail, disrupt performers' flow, or create distracting acoustic artifacts during performance. How musicians inject speech during playing sessions remains an open design question. Some participants suggested using dynamic gates on voice streams –such as automatically activating voice transmission only when speaking– an approach underexplored in collaborative MR research and design. Talk-back systems, standard in recording studios for selective communication between control rooms and performance spaces, are rarely found in music collaborative systems –to our knowledge, the only system that supports this was the currently discontinued Elk Live– representing new interaction design challenges for MR applications beyond music.

6.7 Balancing Stereo and Spatial Audio in MR Collaboration Design

The high preference for audio rendering given to stereo audio mix presents a series of interesting design considerations. Extensive literature demonstrates that spatial audio enhances immersion, presence, source localization [35, 65], and represents a fundamental component of multimodal telepresence systems [34]. However, according to those findings, spatial audio does not appear equally important in collaborative and multi-user immersive systems for music.

While in 2D video meetings, spatial audio can increase the perception of interactivity and shared space [82], in social VR and AR environments participants reported no significant differences between stereo and spatial audio [6, 60]. Similarly, while musicians prefer spatial audio when playing in audio-only situations [103], comparisons between stereo and spatial audio in guitar players using MR found no significant differences [16]. These findings suggest that in immersive environments, visuospatial features might override auditory spatial cues, potentially explaining why our participants gravitated toward stereo.

This preference might also reflects *path dependence* in professional practice [81], where decades of expertise development around stereo paradigms create resistance to change. Professional musicians have invested years developing listening skills, mixing techniques, and aesthetic preferences based on stereo reproduction, making spatial audio transitions potentially disruptive to established creative processes.

However, participants expressed the desire for the possibility of switching between audio modes (stereo and spatial) while maintaining stereo as the default (see Figure 14). This suggests that MR musical collaboration systems should provide spatial audio as an optional enhancement rather than a complete replacement of traditional stereo workflows.

6.8 Preserving the Flow: Design for Unencumbered Musical Interactions

Professional musicians require a critical balance: sophisticated tools capable of handling complex tasks must coexist with plug-and-play simplicity. This requirement appeared particularly consistent in the Preparation phase where it assumed critical importance (see Figure 15). Martinez-Avila et al. [7] observed that professional guitarists consistently prioritize their instrument as primary focus, with all technological interactions remaining secondary and supportive to music-making.

Musicians maintain instruments "at hand" even when engaging with digital resources –such as holding a guitar while adjusting the controls of a DAW– creating what they have termed *encumbered interactions*, situations where the physical instrument constrains how musicians can interact with technology [7]. In our study, participants anticipated similar encumbrances in MR contexts, identifying potential friction points in setup procedures, HMD operation while controlling instrument sound, and managing incoming audio from connected musicians.

These encumbrances demand that MR musical collaboration systems operate as *transparent tools* that fade into the background. Even brief technological interruptions can disrupt the embodied practices central to professional expertise and interfere with what Collins [32] describes as "*interaction ritual chains*", the repeated collaborative encounters that generate emotional energy, shared focus, and collective effervescence within creative musical communities.

The emphasis placed by participants on defaults and presets like "jazz quartet" or "rehearsal for rock trio" demonstrates how this transparency can be achieved. Such domain-specific configurations should enable users to leverage their professional expertise for rapidly setting up appropriate contexts without extensive technical decision-making, aligning with cognitive load management research [56]. This preference for quick and modifiable defaults reflects necessary conditions for maintaining creative flow through minimal technical friction.

However, this apparent tension between simplicity and sophistication reflects a common pattern in professional tools rather than a contradiction. Participants' request for presets alongside granular audio control suggests a progressive disclosure model: immediate functionality through intelligent defaults, with advanced controls accessible but not mandatory—mirroring successful professional audio tools like DAWs. For musicians, *plug-and-play* does not imply simplicity; rather, it signifies that complexity should be managed appropriately rather than eliminated.

6.9 Implications for Design

The seven identified requirements form an interdependent system: "Audio-First" architectures where computational resources prioritize audio processing over visual rendering; preservation of nonverbal communication through facial tracking and expression rendering to overcome HMD occlusion; federated audio control enabling individual optimization within collaborative contexts; spatial customization capabilities allowing musicians to transform mediated spaces into meaningful creative places; support for speech alongside musical streams to maintain verbal communication channels; selectable audio rendering systems adapting to musicians' perceptual needs; and transparent workflow integration through intelligent defaults and plug-and-play operation.

These requirements extend CSCW research by revealing how musical collaboration intensifies traditional workspace awareness [40, 52], and demands interaction techniques beyond current collaborative and social MR paradigms [43]. The real-time, expressive nature of musical collaboration requires heightened social presence indicators and individual agency within group coordination, distinguishing professional creative work from conventional task-oriented collaboration.

6.10 Limitations

The adopted participatory co-design approach, while essential for capturing expert perspectives, introduced both opportunities and constraints requiring careful examination.

The focus on four-piece band configurations, while providing analytical clarity and consensus validity, limits generalization to other

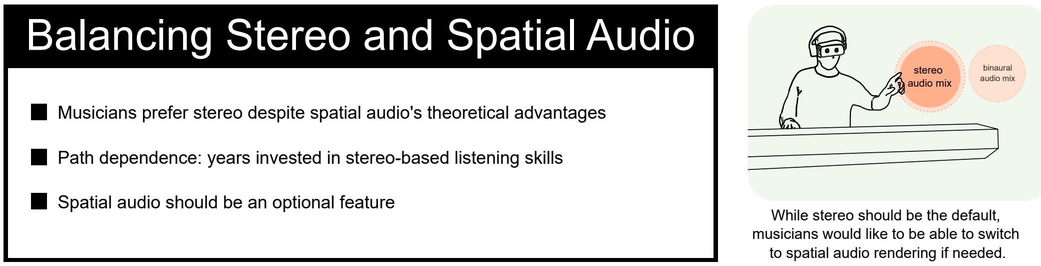

Figure 14: Balancing Stereo and Spatial Audio: Musicians prefer stereo mix despite spatial audio's theoretical advantages due to established professional practices.

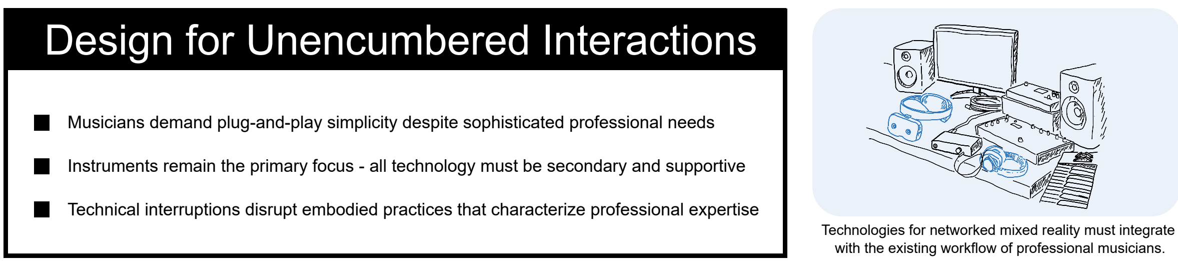

Figure 15: Design for Unencumbered Interactions: Musicians demand plug-and-play simplicity despite sophisticated professional needs, keeping instruments as primary focus.

musical contexts. These include orchestral performance or larger ensembles with hierarchical structures and conductor-centered communication, since they present fundamentally different collaboration patterns from the ones examined here. Non-Western musical traditions with distinct group structures, improvisation practices, and communication conventions may reveal culturally specific design needs beyond our focus on Western popular music.

Moreover, recruiting professional musicians with established practices may have introduced conservatism toward technological change. Hobbyist or new generations of musicians that have developed their skills primarily or exclusively through digital collaboration, might express different priorities compared to professionals, especially on elements such as visual representation, spatial interaction, and workflow integration.

The technology and setting used in Phase 1 proved valuable but also limiting. While the familiarization phase with video passthrough HMDs provided a shared experiential grounding for participants, it could not fully simulate seamless networked MR collaboration. We deliberately used a co-located setup to isolate HMD-specific challenges from well-documented network and audio issues in distributed MR-NMP systems, establishing a baseline before exploring future possibilities in Phase 2 and 3. Additionally, characteristics specific to the Lynx R1 HMD –such as its optics and weight, may have influenced the experience of the participants. This unique optical design provides a distinct visual experience compared to similar HMDs on the market that make use of use more conventional pancake lenses. While devices such as the Meta Quest 3 might offer improved passthrough quality, they share similar weight ranges (450-650 g). We think that the core issues participants identified (visuo-motor conflicts and facial occlusion) are inherent to the HMD form factor and the use of video passthrough technology rather than device-specific.

Moreover, the co-located scenario might have amplified feelings of isolation, as musicians experienced the paradox of physical proximity mediated by headsets. This limitation would likely be less pronounced with the availability of functional distributed MR systems for music collaboration.

Our participant pool, while diverse in genre and instruments, did not include musicians with visual or hearing impairments or other disabilities. This have limited our claims about universal requirements: the audio-first approach might need to coexist with visual-first or haptic-first modes to ensure inclusive professional music-making.

In this study, we did not provide implementation or testing of the designs generated during the workshop. This limitation stems mostly from the current lack of gold standards or systems capable of supporting functional prototyping of MR collaborative music systems. The required implementations should involve not only novel interaction techniques, but also integration of NMP systems with MR platforms, advancement of audio and network architectures, and technologies for acquisition and streaming compatible with music's hard real-time demands. Such developments must ultimately be tested in coherent settings with multiple musicians performing from distributed locations.

The rapidly evolving nature of both MR technologies and professional music practices creates fundamental challenges for requirements research in this domain. These findings represent professional needs at a particular technological moment, requiring ongoing validation and refinement as both technical capabilities and professional practices continue evolving together.

7 Conclusion

This study provides a first exploratory, but empirically grounded, set of design requirements for professional MR musical collaboration systems through formal consensus-building with professional musicians. The strong consensus achieved highlights these requirements' importance and provides a foundation for future development.

Our findings reveal that professional musical collaboration demands fundamentally different technological priorities than office-based and productivity-oriented MR applications. Musicians operate from an audio-first paradigm where sonic experience constitutes the primary reality, contradicting commercial MR platforms that typically prioritize visual aspects of the experience. The unanimous consensus on supporting non-verbal communication –including facial expressions– demonstrates that human connection remains essential for creative collaboration alongside technical capabilities.

Musicians require individual audio control within collaborative contexts (balancing personal optimization with group coordination) while preserving the full spectrum of communication necessary for musical expression. Voice communication should serve both organizational coordination and artistic expression, extending beyond purely sonic interaction. The demand for spatial layout customization indicates that musicians need agency over their environment to transform mediated spaces into meaningful creative places, though they accept pre-configured arrangements aligned with their musical context. Professional adoption also requires plug-and-play simplicity that enables immediate creative flow over technical mastery, despite musicians' technical sophistication.

The seven shared design requirements that emerged from Phase 3 represent some non-negotiable foundations for professional adoption, while instrument-specific patterns indicate the need for customizable interfaces accommodating different musical roles. These requirements can inform novel MR systems that enable geographically distributed musicians to collaborate as if sharing physical space—the vision motivating this research—while preserving the essential human elements that make musical collaboration meaningful, creative, and professionally viable.

Acknowledgments

We acknowledge the support of the MUSMET project funded by the EIC Pathfinder Open scheme of the European Union (grant agreement n. 101184379). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Innovation Council. Neither the European Union nor the European Innovation Council can be held responsible for them.

References

[1] Nadine Aburumman, Marco Gillies, Jamie A Ward, and Antonia F de C Hamilton. 2022. Nonverbal communication in virtual reality: Nodding as a social signal in virtual interactions. *International Journal of Human-Computer Studies* 164 (2022), 102819. doi:10.1016/j.ijhcs.2022.102819

[2] Carlo Aironi, Leonardo Gabrielli, Samuele Cornell, and Stefano Squartini. 2025. Complex-bin2bin: A Latency-Flexible Generative Neural Model for Audio Packet Loss Concealment. *IEEE Transactions on Audio, Speech and Language Processing* 33 (2025), 199–210. doi:10.1109/TASLP.2024.3515794

[3] Abbas M. Al-Ghaili, Hairoladenan Kasim, Naif Mohammed Al-Hada, Zainuddin Bin Hassan, Marini Othman, Jakir Hussain Tharik, Rafiziana Md. Kasmani, and Ibraheem Shayea. 2022. A Review of Metaverse's Definitions, Architecture, Applications, Challenges, Issues, Solutions, and Future Trends. *IEEE Access* 10 (2022), 125835–125866. doi:10.1109/ACCESS.2022.3225638

[4] Nicusor Amarie, Gianluca Fadda, Vlad Popescu, Jean-Marius Ghenta, Maurizio Murroni, and Anderson Augusto Simiscuka. 2025. Live Feedback for Immersive Music Performances-A Case Study. In *ACM International Conference on Interactive Media Experiences Workshops (IMXw)*. SBC, 26–29. doi:10.5753/imxw.2025.8520

[5] Kristina Andersen and Ron Wakkary. 2019. The Magic Machine Workshops: Making Personal Design Knowledge. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300342

[6] Stephanie Arévalo Arboleda, Christian Kunert, Jakob Hartbrich, Christian Schneiderwind, Chenyao Diao, Christoph Gerhardt, Tatiana Surdu, Florian Weidner, Wolfgang Broll, Stephan Werner, and Alexander Raake. 2024. Beyond Looks: A Study on Agent Movement and Audiovisual Spatial Coherence in Augmented Reality. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 502–512. doi:10.1109/VR58804.2024.00071

[7] Juan Pablo Martinez Avila, Chris Greenhalgh, Adrian Hazzard, Steve Benford, and Alan Chamberlain. 2019. Encumbered Interaction: a Study of Musicians Preparing to Perform. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1—-13. doi:10.1145/3290605.3300706

[8] Hanif Baharin, Noris Norowi, Khatriza Ahmad Saffian, and Yap Eng Sim. 2024. Participatory Design Workshop to Create a Virtual Reality Musical Instrument Based on Tumbuk Kalang. In *Advances in Visual Informatics*, Halimah Badioze Zaman, Peter Robinson, Alan F. Smeaton, Renato Lima De Oliveira, Bo Nørregaard Jørgensen, Timothy K. Shih, Rabiah Abdul Kadir, Ummul Hanan Mohamad, and Mohammad Nazir Ahmad (Eds.). Springer Nature Singapore, Singapore, 142–151.

[9] Jeremy Bailenson. 2018. *Experience on Demand: What Virtual Reality Is, How It Works, and What It Can Do.* WW Norton & Company.

[10] Jeremy N. Bailenson, Benjamin Beams, Jonathan Brown, Christine DeVeaux, Eunice Han, Ana Carolina M. Queiroz, Rabindra Ratan, Michael Santoso, Tharini Srirangarajan, Yuqing Tao, and Peng Wang. 2024. Seeing the world through digital prisms: Psychological implications of passthrough video usage in mixed reality. *Technology, Mind, and Behavior* 5, 2 (2024), 1–16. doi:10.1037/tmb0000129

[11] Gary Bente, Felix Eschenburg, and Lisa Aelker. 2007. Effects of simulated gaze on social presence, person perception and personality attribution in avatar-mediated communication. In *Presence 2007: Proceedings of the 10th Annual International Workshop on Presence, October 25-27, 2007, Barcelona, Spain.* 207–14.

[12] Frank Biocca, Chad Harms, and Judee K. Burgoon. 2003. Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence: Teleoperators and Virtual Environments* 12, 5 (10 2003), 456–480. doi:10.1162/105474603322761270

[13] Laura Bishop. 2024. Togetherness in musical interaction [version 2; peer review: 5 approved, 4 approved with reservations]. *Routledge Open Research* 3, 16 (2024), 16. doi:10.12688/routledgeopenres.18202.2

[14] Laura Bishop, Victor González Sánchez, Bruno Laeng, Alexander Refsum Jensenius, and Simon Høffding. 2021. Move like everyone is watching: Social context affects head motion and gaze in string quartet performance. *Journal of New Music Research* 50, 4 (2021), 392–412. doi:10.1080/09298215.2021.1977338

[15] Alberto Boem, Matteo Tomasetti, Alessio Gabriele, Agostino di Scipio, Alessandra Micalizzi, and Luca Turchet. 2025. Imagining the Musical Metaverse: a study on the needs and creative practices of composers and musicians. *Digital Creativity* (2025), 1–27. doi:10.1080/14626268.2025.2542146

[16] Alberto Boem, Matteo Tomasetti, and Luca Turchet. 2025. Between immersion and usability: A comparative study of 2D and mixed reality interfaces for remote music making. *International Journal of Human-Computer Studies* 203 (2025), 103586. doi:10.1016/j.ijhcs.2025.103586

[17] Alberto Boem, Matteo Tomasetti, and Luca Turchet. 2025. Issues and Challenges of Audio Technologies for the Musical Metaverse. *J. Audio Eng. Soc* 73, 3 (2025), 94–114. doi:10.17743/jaes.2022.0193

[18] Jay David Bolter and Richard Grusin. 1999. *Remediation. Understanding New Media.* MIT Press.

[19] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (2019), 589–597. doi:10.1080/2159676X.2019.1628806

[20] Iballa Burunat, Elvira Brattico, Tuomas Puoliväli, Tapani Ristaniemi, Mikko Sams, and Petri Toiviainen. 2015. Action in Perception: Prominent Visuo-Motor Functional Symmetry in Musicians during Music Listening. *PLOS ONE* 10, 9 (09 2015), 1–18. doi:10.1371/journal.pone.0138238

[21] Juan-Pablo Cáceres and Chris Chafe. 2010. JackTrip: Under the Hood of an Engine for Network Audio. *Journal of New Music Research* 39, 3 (2010), 183–187. doi:10.1080/09298215.2010.481361

[22] Patrick Cairns, Anthony Hunt, Jacob Cooper, Daniel Johnston, Ben Lee, Helena Daffern, and Gavin Kearney. 2022. Recording music in the metaverse: a case study of XR BBC Maida Vale Recording Studios. In *Audio Engineering Society Conference: 2022 AES International Conference on Audio for Virtual and Augmented Reality*. Audio Engineering Society.

[23] Patrick Cairns, Anthony Hunt, Daniel Johnston, Jacob Cooper, Ben Lee, Helena Daffern, and Gavin Kearney. 2023. Evaluation of Metaverse Music Performance With BBC Maida Vale Recording Studios. *Journal of the Audio Engineering Society* 71, 6 (2023), 313–325. doi:10.17743/jaes.2022.0086

[24] Anıl Çamcı and Rob Hamilton. 2020. Audio-first VR: New perspectives on musical experiences in virtual environments. *Journal of New Music Research* 49, 1 (2020), 1–7. doi:10.1080/09298215.2019.1707234

[25] Adriaan Campo, Aleksandra Michałko, Bavo Van Kerrebroeck, Boris Stajic, Maja Pokric, and Marc Leman. 2023. The assessment of presence and performance in an AR environment for motor imitation learning: A case-study on violinists. *Computers in Human Behavior* 146 (2023), 107810. doi:10.1016/j.chb.2023.107810

[26] Alexander Carôt and Christian Werner. 2007. Network music performance-problems, approaches and perspectives. In *Proceedings of the "Music in the Global Village"-Conference, Budapest, Hungary*, Vol. 162. 23–10.

[27] Chris Chafe, Michael Gurevich, Grace Leslie, and Sean Tyan. 2004. Effect of Time Delay on Ensemble Accuracy. In *Proceedings of the International Symposium on Musical Acoustics*, Vol. 31. 46.

[28] Ruizhi Cheng, Nan Wu, Matteo Varvello, Eugene Chai, Songqing Chen, and Bo Han. 2024. A First Look at Immersive Telepresence on Apple Vision Pro. In *Proceedings of the 2024 ACM on Internet Measurement Conference* (Madrid, Spain) *(IMC '24)*. Association for Computing Machinery, New York, NY, USA, 555––562. doi:10.1145/3646547.3689006

[29] Michel Chion. 1994. *Audio-Vision. Sound on Screen.* Columbia University Press.

[30] Martin Clayton, Kelly Jakubowski, Tuomas Eerola, Peter E Keller, Antonio Camurri, Gualtiero Volpe, and Paolo Alborno. 2020. Interpersonal Entrainment in Music Performance: Theory, Method, and Model. *Music Perception* 38, 2 (11 2020), 136–194. doi:10.1525/mp.2020.38.2.136

[31] Mario Montagud Climent, Marc Martos, Álvaro Egea, and Sergi Fernández Langa. 2025. Social VR With Holographic Comms: Enablers for New Engaging Experiences Within the TV/Video Consumption Landscape. *IEEE Transactions on Broadcasting* 71, 3 (2025), 793–807. doi:10.1109/TBC.2025.3570869

[32] Randall Collins. 2005. *Interaction Ritual Chains.* Princeton University Press, Princeton. doi:10.1515/9781400851744

[33] Theo Combe, Rebecca Fribourg, Lucas Detto, and Jean-Marie Normand. 2024. Exploring the Influence of Virtual Avatar Heads in Mixed Reality on Social Presence, Performance and User Experience in Collaborative Tasks. *IEEE Transactions on Visualization and Computer Graphics* 30, 5 (2024), 2206–2216. doi:10.1109/TVCG.2024.3372051

[34] Jeremy R. Cooperstock. 2011. Multimodal Telepresence Systems. *IEEE Signal Processing Magazine* 28, 1 (2011), 77–86. doi:10.1109/MSP.2010.939040

[35] Gustavo Corrêa De Almeida, Vinicius Costa de Souza, Luiz Gonzaga Da Silveira Júnior, and Maurício Roberto Veronez. 2024. Spatial Audio in Virtual Reality: A systematic review. In *Proceedings of the 25th Symposium on Virtual and Augmented Reality* (Rio Grande, Brazil) *(SVR '23)*. Association for Computing Machinery, New York, NY, USA, 264––268. doi:10.1145/3625008.3625042

[36] Carlos Cortés, Irene Viola, Jesús Gutiérrez, Jack Jansen, Shishir Subramanyam, Evangelos Alexiou, Pablo Pérez, Narciso García, and Pablo César. 2024. Delay Threshold for Social Interaction in Volumetric eXtended Reality Communication. *ACM Trans. Multimedia Comput. Commun. Appl.* 20, 7, Article 206 (April 2024), 22 pages. doi:10.1145/3651164

[37] Michael J Crosse, Giovanni M Di Liberto, and Edmund C Lalor. 2016. Eye Can Hear Clearly Now: Inverse Effectiveness in Natural Audiovisual Speech Processing Relies on Long-Term Crossmodal Temporal Integration. *Journal of Neuroscience* 36, 38 (2016), 9888–9895. doi:10.1523/JNEUROSCI.1396-16.2016

[38] Sofia Dahl and Anders Friberg. 2007. Visual Perception of Expressiveness in Musicians' Body Movements. *Music Perception* 24, 5 (06 2007), 433–454. doi:10.1525/mp.2007.24.5.433

[39] Stefano Delle Monache, Luca Comanducci, Michele Buccoli, Massimiliano Zanoni, Augusto Sarti, Enrico Pietrocola, Filippo Berbenni, and Giovanni Cospito. 2019. A Presence-and Performance-Driven Framework to Investigate Interactive Networked Music Learning Scenarios. *Wireless Communications and Mobile Computing* 2019, 1 (2019), 4593853. doi:10.1155/2019/4593853

[40] Paul Dourish and Victoria Bellotti. 1992. Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work* (Toronto, Ontario, Canada) *(CSCW '92)*. Association for Computing Machinery, New York, NY, USA, 107––114. doi:10.1145/143457.143468

[41] Carlo Drioli, Claudio Allocchio, and Nicola Buso. 2013. Networked Performances and Natural Interaction via LOLA: Low Latency High Quality A/V Streaming System. In *Information Technologies for Performing Arts, Media Access, and Entertainment*, Paolo Nesi and Raffaella Santucci (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 240–250. doi:10.1007/978-3-642-40050-6_21

[42] Paul Ekman and Wallace V. Friesen. 1969. The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding. *Semiotica* 1, 1 (1969), 49–98. doi:10.1515/semi.1969.1.1.49

[43] Barrett Ens, Joel Lanir, Anthony Tang, Scott Bateman, Gun Lee, Thammathip Piumsomboon, and Mark Billinghurst. 2019. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies* 131 (2019), 81–98. doi:10.1016/j.ijhcs.2019.05.011

[44] Yuan-Ling Feng, Zhaoguo Wang, Yuan Yao, Hanxuan Li, Yuting Diao, Yu Peng, and Haipeng Mi. 2024. Co-designing the Collaborative Digital Musical Instruments for Group Music Therapy. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 698, 18 pages. doi:10.1145/3613904.3642649

[45] Andreas Förster and Norbert Schnell. 2025. Accessible guitar playing: Exploring participatory design of digital musical instruments in a special educational needs school. *International Journal of Human-Computer Studies* 196 (2025), 103435. doi:10.1016/j.ijhcs.2024.103435

[46] Robert Fulford, Jane Ginsborg, and Juliet Goldbart. 2011. Learning not to listen: the experiences of musicians with hearing impairments. *Music Education Research* 13, 4 (2011), 447–464. doi:10.1080/14613808.2011.632086

[47] Olga Gali, Gözde Tüter, Sena Beste Ercan, Óscar Delgado, Gray Atherton, Liam Cross, and Narcis Pares. 2024. Participatory Design of an Embodied Mixed Reality Experience Aimed to Assessing Individual and Collaborative Behaviors of Children. *Proc. ACM Hum.-Comput. Interact.* 8, CHI PLAY, Article 328 (Oct. 2024), 27 pages. doi:10.1145/3677093

[48] Rolf Inge Godøy. 2003. Motor-Mimetic Music Cognition. *Leonardo* 36, 4 (08 2003), 317–319. doi:10.1162/002409403322258781

[49] Jens Emil Sloth Grønbæk, Ken Pfeuffer, Eduardo Velloso, Morten Astrup, Melanie Isabel Sønderkær Pedersen, Martin Kjær, Germán Leiva, and Hans Gellersen. 2023. Partially Blended Realities: Aligning Dissimilar Spaces for Distributed Mixed Reality Meetings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 456, 16 pages. doi:10.1145/3544548.3581515

[50] Jens Emil Sloth Grønbæk, Juan Sánchez Esquivel, Germán Leiva, Eduardo Velloso, Hans Gellersen, and Ken Pfeuffer. 2024. Blended Whiteboard: Physicality and Reconfigurability in Remote Mixed Reality Collaboration. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 798, 16 pages. doi:10.1145/3613904.3642293

[51] Jonathan Grudin. 1988. Why CSCW applications fail: problems in the design and evaluationof organizational interfaces. In *Proceedings of the 1988 ACM Conference on Computer-Supported Cooperative Work* (Portland, Oregon, USA) *(CSCW '88)*. Association for Computing Machinery, New York, NY, USA, 85–93. doi:10.1145/62266.62273

[52] Carl Gutwin and Saul Greenberg. 2002. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW)* 11, 3 (2002), 411–446. doi:10.1023/A:1021271517844

[53] Carl Gutwin, Saul Greenberg, and Mark Roseman. 1996. Workspace Awareness in Real-Time Distributed Groupware: Framework, Widgets, and Evaluation. In *People and Computers XI*. Springer London, London, 281–298. doi:10.1007/978-1-4471-3588-3_18

[54] Edward T. Hall. 1973. *The Silent Language.* Knopf Doubleday Publishing Group.

[55] Steve Harrison and Paul Dourish. 1996. Re-place-ing space: the roles of place and space in collaborative systems. In *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work* (Boston, Massachusetts, USA) *(CSCW '96)*. Association for Computing Machinery, New York, NY, USA, 67–76. doi:10.1145/240080.240193

[56] Nina Hollender, Cristian Hofmann, Michael Deneke, and Bernhard Schmitz. 2010. Integrating cognitive load theory and concepts of human–computer interaction. *Computers in Human Behavior* 26, 6 (2010), 1278–1288. doi:10.1016/j.chb.2010.05.031 Online Interactivity: Role of Technology in Behavior Change.

[57] Torin Hopkins, Suibi Che-Chuan Weng, Rishi Vanukuru, Emma Wenzel, Amy Banic, Mark D Gross, and Ellen Yi-Luen Do. 2022. Studying the Effects of Network Latency on Audio-Visual Perception During an AR Musical Task. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 26–34. doi:10.1109/ISMAR55827.2022.00016

[58] Torin Hopkins, Suibi Che Chuan Weng, Rishi Vanukuru, Emma A Wenzel, Amy Banic, Mark D Gross, and Ellen Yi-Luen Do. 2022. AR Drum Circle: Real-Time Collaborative Drumming in AR. *Frontiers in Virtual Reality* 3 (2022), 847284. doi:10.3389/frvir.2022.847284

[59] Don Ihde. 2007. *Listening and Voice. Phenomenologies of Sound.* SUNY Press.

[60] Felix Immohr, Gareth Rendle, Christian Kehling, Anton Lammert, Steve Göring, Bernd Froehlich, and Alexander Raake. 2024. Subjective Evaluation of the Impact of Spatial Audio on Triadic Communication in Virtual Reality. In *2024 16th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 262–265. doi:10.1109/QoMEX61742.2024.10598292

[61] Andrew Irlitti, Mesut Latifoglu, Thuong Hoang, Brandon Victor Syiem, and Frank Vetere. 2024. Volumetric Hybrid Workspaces: Interactions with Objects in Remote and Co-located Telepresence. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 802, 16 pages. doi:10.1145/3613904.3642814

[62] Andrew Irlitti, Mesut Latifoglu, Qiushi Zhou, Martin N Reinoso, Thuong Hoang, Eduardo Velloso, and Frank Vetere. 2023. Volumetric Mixed Reality Telepresence for Real-time Cross Modality Collaboration. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 101, 14 pages. doi:10.1145/3544548.3581277

[63] Balazs Andras Ivanyi, Truls Bendik Tjemsland, Lloyd May, Matt Robidoux, and Stefania Serafin. 2024. Participatory Design of a Collaborative Accessible Digital Musical Interface with Children with Autism Spectrum Disorder. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, S M Astrid Bin and Courtney N. Reed (Eds.). Utrecht, Netherlands, Article 8, 9 pages. doi:10.5281/zenodo.13904778

[64] Alexander Refsum Jensenius and Marcelo M. Wanderley. 2010. Musical Gestures: Concepts and Methods in Research. In *Musical Gestures*. Routledge, 24–47.

[65] Jason Jerald. 2015. *The VR book: Human-centered design for virtual reality*. Association for Computing Machinery and Morgan & Claypool. 635 pages. doi:10.1145/2792790

[66] Janet G Johnson, Danilo Gasques, Tommy Sharkey, Evan Schmitz, and Nadir Weibel. 2021. Do You Really Need to Know Where "That" Is? Enhancing Support for Referencing in Collaborative Mixed Reality Environments. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 514, 14 pages. doi:10.1145/3411764.3445246

[67] Chiara Jongerius, Roy S Hessels, Johannes A Romijn, Ellen MA Smets, and Marij A Hillen. 2020. The Measurement of Eye Contact in Human Interactions: A Scoping Review. *Journal of Nonverbal Behavior* 44, 3 (2020), 363–389. doi:10.1007/s10919-020-00333-3

[68] Peter Knees and Kristina Andersen. 2016. Searching for Audio by Sketching Mental Images of Sound: A Brave New Idea for Audio Retrieval in Creative Music Production. In *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval* (New York, New York, USA) *(ICMR '16)*. Association for Computing Machinery, New York, NY, USA, 95–102. doi:10.1145/2911996.2912021

[69] Veronika Krauß, Pejman Saeghe, Alexander Boden, Mohamed Khamis, Mark McGill, Jan Gugenheimer, and Michael Nebeling. 2024. What Makes XR Dark? Examining Emerging Dark Patterns in Augmented and Virtual Reality through Expert Co-Design. *ACM Trans. Comput.-Hum. Interact.* 31, 3, Article 32 (Aug. 2024), 39 pages. doi:10.1145/3660340

[70] Marc Leman. 2007. *Embodied Music Cognition and Mediation Technology*. MIT press.

[71] Giacomo Lepri, John Bowers, Samantha Topley, Paul Stapleton, Peter Bennett, Kristina Andersen, and Andrew McPherson. 2022. The 10,000 Instruments Workshop - (Im)practical Research for Critical Speculation. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Andrew McPherson and Emma Frid (Eds.). Auckland, New Zealand, Article 50. doi:10.21428/92fbeb44.9e7c9ba3

[72] Giacomo Lepri and Andrew McPherson. 2019. Making Up Instruments: Design Fiction for Value Discovery in Communities of Musical Practice. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 113––126. doi:10.1145/3322276.3322353

[73] Jiacheng Liu, Vit Drga, and Ifat Yasin. 2021. Optimal Time Window for the Integration of Spatial Audio-Visual Information in Virtual Environments. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. 723–728. doi:10.1109/VR50410.2021.00099

[74] Xuewen Luo, Hsiao-Hwa Chen, and Qing Guo. 2022. Semantic Communications: Overview, Open Issues, and Future Research Directions. *IEEE Wireless Communications* 29, 1 (2022), 210–219. doi:10.1109/MWC.101.2100269

[75] Fang Ma and Xueni Pan. 2022. Visual Fidelity Effects on Expressive Self-avatar in Virtual Reality: First Impressions Matter. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 57–65. doi:10.1109/VR51125.2022.00023

[76] Alice Mado Proverbio, Marta Calbi, Mirella Manfredi, and Alberto Zani. 2014. Audio-visuomotor processing in the Musician's brain: an ERP study on professional violinists and clarinetists. *Scientific Reports* 4, 1 (2014), 5866. doi:10.1038/srep05866

[77] Christian Merz, Carolin Wienrich, and Marc Erich Latoschik. 2024. Does Voice Matter? The Effect of Verbal Communication and Asymmetry on the Experience of Collaborative Social XR. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 1127–1136. doi:10.1109/ISMAR62088.2024.00129

[78] Christian Merz, Carolin Wienrich, and Marc Erich Latoschik. 2025. Does Task Matter? Task-Dependent Effects of Cross-Device Collaboration on Social Presence. In *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 536–543. doi:10.1109/VRW66409.2025.00116

[79] Evan Morgan, Hatice Gunes, and Nick Bryan-Kinns. 2015. Using affective and behavioural sensors to explore aspects of collaborative music making. *International Journal of Human-Computer Studies* 82 (2015), 31–47. doi:10.1016/j.ijhcs.2015.05.002

[80] Jessica Napoles. 2014. Verbal Instructions and Conducting Gestures: Examining Two Modes of Communication. *Journal of Music Teacher Education* 23, 2 (2014), 9–20. doi:10.1177/1057083712474936

[81] Donald Norman. 2013. *The Design of Everyday Things*. MIT Press.

[82] Kate Nowak, Lev Tankelevitch, John Tang, and Sean Rintel. 2023. Hear We Are: Spatial Audio Benefits Perceptions of Turn-Taking and Social Presence in Video Meetings. In *Proceedings of the 2nd Annual Meeting of the Symposium on Human-Computer Interaction for Work* (Oldenburg, Germany) *(CHIWORK '23)*. Association for Computing Machinery, New York, NY, USA, Article 2, 10 pages. doi:10.1145/3596671.3598578

[83] Catherine S Oh, Jeremy N Bailenson, and Gregory F Welch. 2018. A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Frontiers in Robotics and AI* 5 (2018), 114. doi:10.3389/frobt.2018.00114

[84] Soo Youn Oh, Jeremy Bailenson, Nicole Krämer, and Benjamin Li. 2016. Let the Avatar Brighten Your Smile: Effects of Enhancing Facial Expressions in Virtual Environments. *PLOS ONE* 11, 9 (09 2016), 1–18. doi:10.1371/journal.pone.0161794

[85] Keita Ohshiro and Mark Cartwright. 2022. How people who are deaf, Deaf, and hard of hearing use technology in creative sound activities. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility* (Athens, Greece) *(ASSETS '22)*. Association for Computing Machinery, New York, NY, USA, Article 66, 4 pages. doi:10.1145/3517428.3550396

[86] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. 2016. Holoportation: Virtual 3D Teleportation in Real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 741–754. doi:10.1145/2984511.2984517

[87] Juhani Pallasmaa. 2012. *The Eyes of the Skin Architecture and the Senses*. John Wiley & Sons.

[88] C. Perkins, O. Hodson, and V. Hardman. 1998. A survey of packet loss recovery techniques for streaming audio. *IEEE Network* 12, 5 (1998), 40–48. doi:10.1109/65.730750

[89] Sophia Ppali, Vali Lalioti, Boyd Branch, Chee Siang Ang, Andrew J. Thomas, Bea S. Wohl, and Alexandra Covaci. 2022. Keep the VRhythm going: A musician-centred study investigating how Virtual Reality can support creative musical practice. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 220, 19 pages. doi:10.1145/3491102.3501922

[90] Erik Prytz, Susanna Nilsson, and Arne Jönsson. 2010. The importance of eye-contact for collaboration in AR systems. In *2010 IEEE International Symposium on Mixed and Augmented Reality*. 119–126. doi:10.1109/ISMAR.2010.5643559

[91] Holger Regenbrecht, Katrin Meng, Arne Reepen, Stephan Beck, and Tobias Langlotz. 2017. Mixed Voxel Reality: Presence and Embodiment in Low Fidelity, Visually Coherent, Mixed Reality Environments. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 90–99. doi:10.1109/ISMAR.2017.26

[92] Shane L Rogers, Rebecca Broadbent, Jemma Brown, Alan Fraser, and Craig P Speelman. 2022. Realistic Motion Avatars are the Future for Social Interaction in Virtual Reality. *Frontiers in Virtual Reality* 2 (2022), 750729. doi:10.3389/frvir.2021.750729

[93] Cristina Rottondi, Chris Chafe, Claudio Allocchio, and Augusto Sarti. 2016. An Overview on Networked Music Performance Technologies. *IEEE Access* 4 (2016), 8823–8843. doi:10.1109/ACCESS.2016.2628440

[94] Monique Santoso and Jeremy Bailenson. 2024. How Video Passthrough Headsets Influence Perception of Self and Others. *Cyberpsychology, Behavior, and Social Networking* 27, 11 (2024), 798–806. doi:10.1089/cyber.2024.0398

[95] Ruben Schlagowski, Kunal Gupta, Silvan Mertes, Mark Billinghurst, Susanne Metzner, and Elisabeth André. 2022. Jamming in MR: Towards Real-Time Music Collaboration in Mixed Reality. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 854–855. doi:10.1109/VRW55335.2022.00278

[96] Ruben Schlagowski, Dariia Nazarenko, Yekta Can, Kunal Gupta, Silvan Mertes, Mark Billinghurst, and Elisabeth André. 2023. Wish You Were Here: Mental and Physiological Effects of Remote Music Collaboration in Mixed Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 102, 16 pages. doi:10.1145/3544548.3581162

[97] Frederick A. Seddon. 2005. Modes of communication during jazz improvisation. *British Journal of Music Education* 22, 1 (2005), 47––61. doi:10.1017/S0265051704005984

[98] Mel Slater. 2018. Immersion and the illusion of presence in virtual reality. *British journal of psychology* 109, 3 (2018), 431–433. doi:10.1111/bjop.12305
[99] Mel Slater, Pankaj Khanna, Jesper Mortensen, and Insu Yu. 2009. Visual Realism Enhances Realistic Response in an Immersive Virtual Environment. *IEEE Computer Graphics and Applications* 29, 3 (2009), 76–84. doi:10.1109/MCG.2009.55
[100] Maximilian Speicher, Brian D. Hall, and Michael Nebeling. 2019. What is Mixed Reality?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1—-15. doi:10.1145/3290605.3300767
[101] Atau Tanaka and Adam Parkinson. 2016. Haptic Wave: A Cross-Modal Interface for Visually Impaired Audio Producers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2150–2161. doi:10.1145/2858036.2858304
[102] Balasaravanan Thoravi Kumaravel, Fraser Anderson, George Fitzmaurice, Bjoern Hartmann, and Tovi Grossman. 2019. Loki: Facilitating Remote Instruction of Physical Tasks Using Bi-Directional Mixed-Reality Telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 161––174. doi:10.1145/3332165.3347872
[103] Matteo Tomasetti and Luca Turchet. 2023. Playing With Others Using Headphones: Musicians Prefer Binaural Audio With Head Tracking Over Stereo. *IEEE Transactions on Human-Machine Systems* 53, 3 (2023), 501–511. doi:10.1109/THMS.2023.3270703
[104] Luca Turchet. 2023. Musical Metaverse: vision, opportunities, and challenges. *Personal and Ubiquitous Computing* 27, 5 (2023), 1811–1827. doi:10.1007/s00779-023-01708-1
[105] Luca Turchet, David Baker, and Tony Stockman. 2021. Musical Haptic Wearables for Synchronisation of Visually-impaired Performers: a Co-design Approach. In *Proceedings of the 2021 ACM International Conference on Interactive Media Experiences* (Virtual Event, USA) *(IMX '21)*. Association for Computing Machinery, New York, NY, USA, 20––27. doi:10.1145/3452918.3458803
[106] Luca Turchet and Mathieu Barthet. 2019. Co-Design of Musical Haptic Wearables for Electronic Music Performer's Communication. *IEEE Transactions on Human-Machine Systems* 49, 2 (2019), 183–193. doi:10.1109/THMS.2018.2885408
[107] Luca Turchet, Nicola Garau, and Nicola Conci. 2022. Networked Musical XR: where's the limit? A preliminary investigation on the joint use of point clouds and low-latency audio communication. In *Proceedings of the 17th International Audio Mostly Conference* (St. Pölten, Austria) *(AM '22)*. Association for Computing Machinery, New York, NY, USA, 226––230. doi:10.1145/3561212.3561237
[108] Luca Turchet, Andrew McPherson, and Mathieu Barthet. 2018. Co-design of a Smart Cajón. In *Journal of the Audio Engineering Society*, Vol. 66. 220–230. doi:10.17743/jaes.2018.0007
[109] Bavo Van Kerrebroeck, Giusy Caruso, and Pieter-Jan Maes. 2021. A Methodological Framework for Assessing Social Presence in Music Interactions in Virtual Reality. *Frontiers in Psychology* 12 (2021), 663725. doi:10.3389/fpsyg.2021.663725
[110] Bavo Van Kerrebroeck, Kristel Crombé, Stéphane M de Leymarie, Marc Leman, and Pieter-Jan Maes. 2023. The virtual drum circle: polyrhythmic music interactions in mixed reality. *Journal of New Music Research* 52, 4 (2023), 316–336. doi:10.1080/09298215.2024.2339244
[111] Arthur B VanGundy. 1984. Brain writing for new product ideas: an alternative to brainstorming. *Journal of Consumer Marketing* 1, 2 (1984), 67–74. doi:10.1108/eb008097
[112] Irene Viola, Jack Jansen, Shishir Subramanyam, Ignacio Reimat, and Pablo Cesar. 2023. VR2Gather: A Collaborative, Social Virtual Reality System for Adaptive, Multiparty Real-Time Communication. *IEEE MultiMedia* 30, 2 (2023), 48–59. doi:10.1109/MMUL.2023.3263943
[113] Ina Wagner, Wolfgang Broll, Giulio Jacucci, Kari Kuutii, Rod McCall, Ann Morrison, Dieter Schmalstieg, and Jean-Jacques Terrin. 2009. On the Role of Presence in Mixed Reality. *Presence* 18, 4 (2009), 249–276. doi:10.1162/pres.18.4.249
[114] Ashley E Walton, Michael J Richardson, Peter Langland-Hassan, and Anthony Chemero. 2015. Improvisation and the self-organization of multiple musical bodies. *Frontiers in Psychology* 6 (2015), 313. doi:10.3389/fpsyg.2015.00313
[115] Suibi Che-Chuan Weng, Torin Hopkins, Shih-Yu Ma, Chad Tobin, Amy Banic, and Ellen Yi-Luen Do. 2023. Investigating the Effects of Limited Field of View on Jamming Experience in Extended Reality. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 493–495. doi:10.1109/ISMAR-Adjunct60411.2023.00107
[116] Aaron Williamon and Jane W Davidson. 2002. Exploring Co-Performer Communication. *Musicae Scientiae* 6, 1 (2002), 53–72. doi:10.1177/102986490200600103
[117] Emily Wong, Adélaïde Genay, Jens Emil Sloth Grønbæk, and Eduardo Velloso. 2025. Spatial Heterogeneity in Distributed Mixed Reality Collaboration. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 980, 19 pages. doi:10.1145/3706598.3714033
[118] Emily Wong, Juan Sánchez Esquivel, Germán Leiva, Jens Emil Sloth Grønbæk, and Eduardo Velloso. 2024. Practice-informed Patterns for Organising Large Groups in Distributed Mixed Reality Collaboration. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1030, 18 pages. doi:10.1145/3613904.3642502
[119] Lei Xia, Xiaomei Li, Siye Zeng, Danyang Chen, and Ling Fan. 2025. Navigating Stakeholder Tensions in Spatial Computing: A Participatory Design Approach for Inclusive Mixed Reality Environments *(CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 415, 6 pages. doi:10.1145/3706599.3720268
[120] Cornelia Yarbrough and Katia Madsen. 1998. The Evaluation of Teaching in Choral Rehearsals. *Journal of Research in Music Education* 46, 4 (1998), 469–481. doi:10.2307/3345344
[121] Leonard Yoon, Dongseok Yang, Jaehyun Kim, ChoongHo Chung, and Sung-Hee Lee. 2022. Placement Retargeting of Virtual Avatars to Dissimilar Indoor Environments. *IEEE Transactions on Visualization and Computer Graphics* 28, 3 (2022), 1619–1633. doi:10.1109/TVCG.2020.3018458