Co-Designing Virtual Reality Musical Instruments and Spatial Layouts for Collaborative Music-Making

Alberto Boem
University of Trento
Trento, Italy
alberto.boem@unitn.it

Stavros Skouras
PatchXR AG
Volketswil, Switzerland
stav@patchxr.com

Gad Baruch Hinkis
PatchXR AG
Volketswil, Switzerland
gad@patchxr.com

Mélodie Mousset
PatchXR AG
Volketswil, Switzerland
melo@patchxr.com

Luca Turchet
University of Trento
Trento, Italy
luca.turchet@unitn.it

Figure 1: Three moments of the co-design activities done in PatchWorld. On the left, a screenshot of the instrument design workshop. On the center, the facilitator showing the resulting *Universal Rhythm Box*. On the right, a participant sketching the layout of a performative space during the second workshop.

Abstract

Collaborative music-making in Virtual Reality (VR) presents unique challenges that intertwine instrument design, spatial arrangement, and social dynamics. This paper presents findings from a participatory co-design study investigating rhythmic instruments and spatial layouts for collaborative VR music, conducted entirely within PatchWorld with nine participants from its user community. Through iterative prototyping across two workshops, participants in our study converged on design principles favoring VR-native interactions over approximations of physical instrument. The co-design process yielded two contributions: the *Universal Rhythm Box*, a collaborative rhythmic instrument whose design parameters were wholly derived from structured participant feedback; and a spatial environment arranged for mixed-skill ensembles incorporating tiered access zones and semi-circular arrangements. Our findings suggest that spatial layout and instrument design are inseparable in collaborative VR. We also identify attribution ambiguity as a key challenge in networked ensembles, and provide empirically-grounded guidelines prioritizing visual feedback as a primary expressive dimension for group playing in VR.

Keywords

musical xr, musical metaverse, social vr, co-design

NIME '26, June 23–26, 2026, London, UK

1 Introduction

The Musical Metaverse [66] promises to redefine musical performance and collaboration through immersive, technologically mediated social interactions [19]. Virtual Reality Musical Instruments (VRMIs) [4, 62, 68] expand the design space of musical instruments and performance environments beyond physical laws, where gesture, sound, and visual feedback can be coupled in ways impossible with acoustic instruments. Yet collaborative music-making in VR presents unique challenges that remain underexplored [20]. Unlike single-user scenarios, collaborative systems must simultaneously address instrument design, spatial arrangement, and social dynamics. These elements become intertwined in ways they rarely are in physical environments.

This paper presents findings from a participatory co-design study investigating two interconnected dimensions of collaborative VR music: (1) rhythmic instruments suitable for group performance, and (2) spatial layouts supporting mixed-skill group interactions. We conducted workshops entirely within PatchWorld by PatchXR[1], a VR application dedicated to musical creation, involving nine participants from its user community across two independent groups (see Figure 1). Unlike previous studies that moved between physical and virtual design spaces [2], our methodology kept the entire co-design process inside VR, engaging participants as expert users, in an ecologically valid scenario.

Our research addresses a gap at the intersection of three bodies of work developed largely in isolation. First, previous research on non-VR collaborative musical instruments showed that such instruments should facilitate *performer-performer* communication and coordination, not only *performer-instrument* interaction dynamics [13, 37]. Second, frameworks for virtual performance

[1]https://patchxr.com/ (accessed: 12/04/2026)

spaces [11, 72] and visualization in digital ensembles [6] primarily address projection-based systems where performers remain physically co-located, a condition fundamentally different from networked VR systems based on head-mounted displays (HMDs) where musicians experience an immersive environment egocentrically. Third, while participatory design has become central to NIME practice, prior work on VR music spaces [45–47] employed fixed designs evaluated by users rather than involving users as expert co-designers.

Through iterative prototyping, participants converged on principles challenging conventional VR instrument design. Rather than replicating physical instruments, successful prototypes embraced VR-native interactions, like 3D User Interfaces (UIs) that leverage spatial positioning and visual feedback as primary expressive dimensions. The co-design process yielded two concrete contributions: the *Universal Rhythm Box*, a collaborative rhythmic instrument featuring multi-axis gestural control and real-time visual feedback; and a validated spatial configuration for mixed-skill ensembles, incorporating tiered access zones and semicircular arrangements balancing visibility, collaboration, and skill differentiation.

This work contributes to NIME discourse on VRMIs and the Musical Metaverse in three ways. First, we provide empirically-grounded design guidelines for collaborative VRMIs that prioritize VR-native interaction over physical simulation. Second, we show that spatial layout and instrument design are inseparable in collaborative VR, as the semicircular arrangement emerged specifically because participants' preferred instruments required mutual visibility of real-time waveform displays. Third, we identify *attribution ambiguity* as a key challenge: in current networked VR ensembles, musicians cannot determine whether timing problems originate from their own performance, system delays, or collaborators' errors, disrupting the feedback loop essential to ensemble coordination.

2 Related Work

2.1 Collaborative Rhythm-making in VR

Shared Virtual Environments [70] and Social XR platforms [12] have extended collaborative music-making into immersive 3D spaces, building on single-user VRMI foundations [49, 62, 68]. Recent surveys [20, 41] reveal persistent challenges in group synchronization and networked instrument design, while the Musical Metaverse [66] has emerged as a focus area with work on enabling technologies [21, 26], interaction metaphors [18], especially in the context of web-based platforms [27, 52]. Early research on virtual percussion established that tactile feedback and system latency critically affect playability, with users preferring tangible objects for inertial feedback [42]. Subsequent work explored vibrotactile feedback in 3D sequencers [73] and gesture separation frameworks for preserving both expressivity and interface flexibility [8]. While foundational, these projection-based systems differ substantially from current HMD-based VR, where tracked controllers can be effectively used to support timing consistency [22].

Notably, these works addressed only single-user scenarios. Collaborative VR rhythm-making remains underexplored: LeMo [45] pioneered shared step-sequencers for co-located users, while recent studies have highlighted how ensemble timing demands compound the feedback limitations of current HMD systems [16]. Alternative interaction metaphors, such as using avatars as percussive surfaces [18], have also emerged. Our study addresses remaining gaps through participatory co-design of both instruments and spatial layouts.

2.2 VR Performance Spaces

Physical venues rely on centuries-old conventions to balance acoustic projection, visual communication, and social dynamics, whereas immersive environments offer malleable configurations that introduce new challenges [51].

Several frameworks developed in the area of NIME tried to address some of these. Berthaut and Hachet [10] proposed a *three circles* model distinguishing spatial needs of individual musicians, ensembles, and audiences. Zappi et al. [72] extended this with a *scenography* framework emphasizing cooperation, communication, and organization, while Mazzanti et al. [43] contributed the *Augmented Stage* concept. These frameworks address only projection-based setups where performers remain physically co-located.

Regarding networked VR based on HMDs, Men and Bryan-Kinns [46] identified emergent personal and group territories during collaborative VR composition, finding that *side-by-side* and *face-to-face* configurations arise naturally. Subsequent work has also explored the role of privacy in shared virtual environments for music making. Despite this work, critical questions remain: since these works tested only dyads, how can we manage the inclusion of a larger pool of users? Which strategies can be used for balancing accessibility, for example of novices in experts' workspace?

2.3 Co-Design and Participatory Approaches

Co-design methodologies have proven valuable in NIME for balancing technical innovation with performer's needs [56, 61]. Participatory approaches have yielded insights across instrument design, input/output techniques, and accessibility [32, 33, 36, 40, 65, 67, 74], though focusing on instruments rather than performance environments. Recent Musical Metaverse studies extended participatory methods but emphasized assessment over iterative design [17, 19].

In HCI, co-design has been applied to VR systems [39, 44, 53, 63], including immersive music environments [54] and traditional instrument recreation [2]. While prior work moved between physical and virtual design spaces, our methodology keeps the entire co-design process inside VR, engaging participants as experienced platform users in the simultaneous design of both instruments and spatial configurations.

2.4 PatchWorld

PatchWorld by PatchXR is an application available on the Meta Quest Store, dedicated to musical creation within virtual environments. Beyond single use, it enables multiple users to build virtual musical instruments and play them in shared immersive environments over networked sessions, with real-time audio and avatar presence synchronization across remote participants. PatchWorld features a library of devices and tools to be combined and interconnected using a visual patching paradigm, similar to those used in software such as Pure Data and Max/MSP. It has an active community, with regular meetups, performances, as well as a dedicated record label for music production done inside and with the platform. PatchWorld has been already used in several NIME-related research [3, 25, 34].

Figure 2: The different prototypes used in the first workshop. On the left, the vertical layout used to distribute instruments to the participants. On the right, the six instruments: (P1) Expressive Drum Pad, (P2) Jangler Grid, (P3) Pattern Grid Instrument, (P4) Bouncy Drums, (P5) Rhythm Variation Grid, (P6) Euclidean Rhythm Cube.

3 Preliminary Work

To inform the focus of the co-design workshops, two preliminary studies were conducted to investigate user needs for collaborative music-making in PatchWorld.

Survey: Fourteen musicians recruited from communities built around different VR musical creation applications (e.g., PatchXR, TribeXR[2], CsoundMeta [69]) completed a survey exploring their motivations, preferred session structures, and challenges when making music in VR. Results showed that participants valued VR for its creative freedom, global accessibility, and sense of co-presence. However, they consistently identified latency and synchronization as technical issues, and limited emotional expressivity of avatars as primary barriers to effective collaboration. Preferences for session structure varied, though most favored approximately one-hour sessions.

Jam Sessions. Twelve collaborative VR music sessions were conducted over four months using the PatchWorld platform, involving 4–11 participants per session. Instrument configurations and spatial layouts were iteratively modified across sessions to evaluate their effects on collaboration and creative flow. Three questionnaires administered after selected sessions gathered feedback on engagement, usability, and emotional connection. Findings indicated that circular stage layouts might improve mutual visibility and timing coordination, while hybrid instrument setups combining instruments like sequencers and gesture-based instruments could be helpful in sustaining engagement across different expertise levels.

Together, these studies identified three priority areas for design intervention: (1) rhythm and tempo tools to support temporal alignment, (2) spatial layout configurations that enhance coordination and presence, and (3) mechanisms to better support players with different skill levels. These priorities directly structured the subsequent workshops: rhythm and tempo tools motivated the instrument design focus of Workshop 1, spatial layout configurations motivated Workshop 2, and affective communication mechanisms informed the evaluation criteria applied across both.

4 Methodology

Our research employed a participatory co-design approach [48, 59], in which end users actively inform and shape design outcomes through structured engagement rather than post-hoc evaluation, and it was organized around two main topics:

[2]https://www.tribexr.com/ (accessed: 12/04/2026)

(1) **Workshop 1 - VR Musical Instrument Design**: develop an intuitive, accessible interface for creating foundational rhythms that mitigates inherent network and hardware latency through thoughtful interaction design and feedback mechanisms. The rationale for prioritizing rhythmic instrument design stemmed from observations obtained during the 12 collaborative jam sessions in PatchWorld conducted prior to the co-design workshops.
(2) **Workshop 2 - Staging Layout Configuration**: design an optimal spatial layout and staging configuration for distributed VR music-making sessions, to support more experienced players and beginners as well.

The two workshops employed distinct design approaches. Workshop 1 used evaluation-mediated co-design, where participants assessed pre-built prototypes and their structured feedback directly determined the design parameters of the resulting instrument, which was subsequently developed by the R&D team of PatchWorld. This first workshop was used as a starting point, since participants agreed it was more crucial to establish robust instrument design before addressing spatial arrangement. Workshop 2 employed a more direct co-design approach, with participants constructing and iterating spatial configurations in situ within the VR environment.

4.1 Participants

The co-design workshop sessions involved two independent groups, referred to as Group A and Group B, which comprised four and five participants respectively (total n = 9; 3 identified as non-binary, 2 as females, 4 as males; mean age = 40.1 years; SD = 13.3). Participants were volunteers recruited from PatchWorld user community, mainly through the official Discord server, including both experienced and amateur musicians. Efforts were made to ensure a range of musical backgrounds and VR familiarity, enabling the workshops to probe usability and expressivity requirements across diverse skill levels. Participants were connected from different countries across the European Union and the United States of America.

They were split into the two groups to obtain a size compatible with a focus group study while preserving participant diversity, but also to accommodate the different requests in terms of availability and time zone. All participants provided signed informed consent.

4.2 Workshop Design and Technical Setup

Workshops were held remotely over VR sessions using the PatchWorld application on Meta Quest 3 devices, chosen for its active musician community, built-in instrument prototyping capabilities, and native support for networked multi-user sessions. Each session lasted approximately one hour. One of the authors acted as a facilitator and workshop leader that managed all sessions, delivering standardized orientation to the VR environment and guiding participants through the prepared protocol. Another author played the role of the observer, taking notes and performing audio and video recordings of the entire workshops. Each participant (including facilitators and observer) connected from their home network.

4.3 Data Collection

Each workshop was video-recorded from two perspectives within the VR space (facilitator and silent observer). Audio was extracted and transcribed verbatim using OpenAI Whisper [3]. Transcripts were complemented by field notes maintained primarily by the silent observer, documenting key points and observations during sessions. This triangulated approach (video, transcripts, and field notes) captured both verbal discussions and musical interactions across the co-design process.

4.4 Data Analysis

Transcripts and field notes were analyzed using reflexive thematic analysis [23, 24]. Workshop recordings were transcribed and segmented into 5,207 discrete utterances. After removing utterances unrelated to the evaluation of collaborative music-making (e.g., greetings, technical login issues), 1,956 utterances remained for analysis. Each utterance was assigned to one or more thematic codes derived from the analysis of the pre-workshop surveys, and addressing themes such as usability, latency, collaboration, instrument design, and musical coordination. Through iterative review, codes were refined and grouped into higher-level categories based on conceptual similarity, resulting in 363 coded utterances. Code frequency and co-occurrence analyses were performed to identify patterns across the dataset.

5 Workshop 1: Instrument Design

Of all the challenges surfaced during preliminary work, creating a foundational rhythm using percussive gestures proved particularly problematic. Participants noticed that even slight timing deviations or system latencies would quickly disrupt the perceived rhythm, causing the group to lose cohesion. This was particularly problematic given that most participants anchored their playing around an initial beat serving as a shared metronome. These observations align with broader findings in VR musical instrument research on the challenges of latency and temporal precision [49, 62, 68], and directly motivated the selection of rhythmic instrument prototypes for Workshop 1. Six prototype instruments were selected from the existing library of PatchWorld. They were chosen by one of the authors (one of the developers of PatchWorld) to represent diverse interaction modes (see Figure 2). These included:

- **P1** : a velocity-sensitive drum pad with optional quantization and single-sample output;
- **P2** : a 3D grid of semi-transparent rectangles mapping hand depth to dynamics and horizontal position to pitch, supporting continuous gesture-based triggering;
- **P3** : a pattern grid generating repeating rhythmic patterns when held within a cell, with four control axes—pitch via lateral movement, subdivision rate via depth, dynamic amplitude via vertical position, and sample length via wrist rotation—supporting two-hand interaction;
- **P4** : a hybrid instrument combining movable nodes on an XY surface with pre-programmed rhythmic patterns;
- **P5** : an evolution of P3 remapping axes to rhythmic subdivision and conditional silence, augmented with low-pass filtering for dynamic shaping;
- **P6** : and a semi-transparent Euclidean rhythm cube distributing hits algorithmically across steps, with lateral movement controlling phase offset and two-sample layering per hand.

[3]Whisper: Robust Speech Recognition via Large-Scale Weak Supervision, https://github.com/openai/whisper (accessed: 12/04/2026).

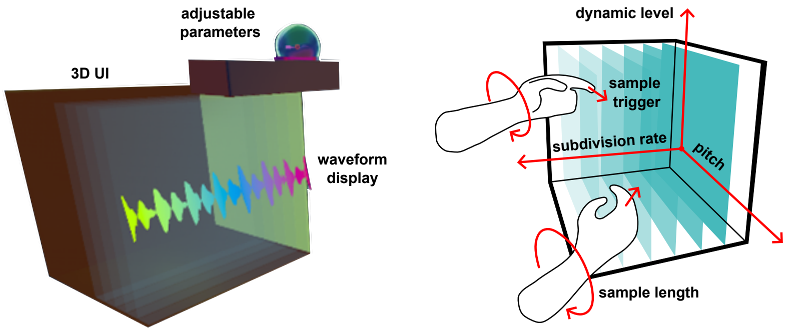

Figure 3: The *Universal Rhythm Box* resulting from the first workshop. On the left, the instrument composed of the main 3D UI, the container of the adjustable parameters, and the waveform display. On the right, the main mapping between the 3D space of the UI and the actions of the musicians.

5.1 Procedure

Each group completed three 60-minute sessions over two months, following an iterative three-phase structure [4].

Phase 1 - Exploratory Interaction: Participants tested six instrument prototypes (each presented as multiple identical instances for simultaneous interaction). Instruments were placed across six vertical levels that participants navigated together, enabling experience of both constrained and open-ended interaction paradigms.

Phase 2 - Group Debrief and Iteration: After testing each instrument, participants paused for approximately five minutes of discussion, reflecting on components of the experience such as perceived latency, responsiveness, feedback clarity, and ensemble cohesion. Pre-planned variations (toggling quantization, introducing click tracks) were tested based on participant suggestions.

Phase 3 - Selection and Refinement: Participants cast two votes each using virtual markers to indicate the preferred prototypes. The top three instruments were replicated for deeper evaluation in subsequent sessions incorporating scenario-based tasks (e.g., constructing 16-bar drum loops, coordinating call-and-response patterns) to test prototypes under realistic performance conditions [57]. Participants continued articulating design suggestions for enhancing expressivity, usability, and collaborative potential.

5.2 Results

Across two co-design groups (Group A and Group B), participants evaluated six distinct virtual rhythmic instrument prototypes. The evaluation process revealed marked differences in perceived responsiveness, expressivity, and collaborative usability, with participants ultimately converging on a subset of designs that balanced simplicity with expressive depth.

5.2.1 Participant Feedback. The prototypes elicited distinct responses, revealing clear patterns in what musicians valued for collaborative VR music-making.

P1 was unanimously rejected despite its familiar drum-pad metaphor. Participants found it sluggish, even with quantization disabled, and the absence of tactile feedback proved problematic—one noted there was "*no feeling of ricochet or impact.*" Its single-sample output further limited its usefulness in ensemble settings, as participants struggled to maintain steady rhythms under network latency.

P3 emerged as the most favored prototype due to its balance of simplicity, expressivity, and support for multi-hand interaction. Both experts and novices appreciated its clear visual feedback (waveform overlays) and haptic cues, which reinforced action–sound congruence. Participants described it as "*less of a mind sore,*" requiring minimal cognitive effort while still offering substantial expressive depth.

P2 improved on P1 in terms of responsiveness but shared a key limitation: it lacked a mechanism for sustaining foundational rhythms, making it unsuitable for rhythm-section roles.

The remaining prototypes revealed trade-offs between accessibility and sophistication. The pre-programmed patterns in P4 appealed to novices but frustrated experts seeking real-time control. Both groups ultimately deprioritized it for collaborative use. P5 offered enhanced control over rhythmic texture, but the separation of pitch into an adjacent interface was viewed as a drawback compared to the integrated design of P3. The algorithmic complexity of P6 impressed some expert users, but participants struggled to recognize their own contributions within the layered output—highlighting that perceptual clarity is essential in ensemble contexts.

Across all evaluations, three design principles emerged: (1) successful prototypes integrated multiple axes of expression within a single interface rather than distributing controls; (2) effective designs balanced expressive capability with immediate comprehensibility; and (3) visual and haptic feedback were essential for attribution, enabling users to recognize individual contributions within complex ensemble soundscapes.

5.2.2 Voting Results and Selection. After initial exploration, participants voted for preferred prototypes. Group A selected P2, P3, and P5; Group B selected P3, P5, and P6. Notably, P3 and P5 were common to both groups despite different musical backgrounds.

All participants rejected P1, confirming that simple velocity-sensitive interfaces do not adequately support beat-making in collaborative VR contexts. Participants showed minimal support for P4, indicating that pre-programmed approaches, while useful for solo beginners, fail in ensemble contexts

In subsequent sessions, the instruments selected by each group were instantiated for deeper evaluation, revealing further convergence. Group A largely abandoned P2; Group B discarded P6. Both groups consolidated around P3 for its balance of expressive depth and cognitive accessibility. This convergence across independent groups suggests the Pattern Grid approach identified design principles transcending individual preferences, a point we examine in the Discussion.

[4]Selected recordings of the workshops are available at: https://youtu.be/xBt8Y5zIaQ0?si=ciodltOivOdm_sfv (accessed: 12/04/2026)

Figure 4: Four screenshots taken from the recordings of the co-design session of the second workshop.

5.3 Consolidated Instrument Design

Synthesizing participant feedback, the R&D team of PatchWorld in collaboration with two authors implemented the *Universal Rhythm Box* (see Figure 3). While participants did not directly construct the instrument, its design parameters (four control axes, two-hand support, waveform display, haptic reinforcement, and expert mode) were derived from the structured workshop feedback. The resulting instrument was subsequently deployed in collaborative jam sessions within PatchWorld, providing informal validation of the key design decisions.

The *Universal Rhythm Box* is a single semi-transparent box with four controllable axes (pitch via horizontal position, subdivision rate via depth, dynamic level via vertical position, sample length via wrist rotation) and strike detection on the surface itself [5]. Key features informed directly by participant feedback include:

- **Unified interaction surface:** Four controllable axes within a single interface, addressing participants' preference for integrated rather than distributed controls.
- **Two-hand support:** Simultaneous dual-sample triggering, identified as essential for bilateral expressivity and complex beat creation.
- **Real-time waveform display:** Immediate visual feedback visible to both the player and other musicians, addressing attribution concerns in ensemble contexts.
- **Adjustable haptic feedback:** Reinforces user agency and helps musicians identify their contributions within complex soundscapes.
- **Optional expert mode:** Hidden by default to prevent information overload for novices. When enabled via Edit button, exposes advanced customization. This includes form factor, such as enlarging the interface to room-scale (to "play from within"), duplicating into multiple pad-style surfaces, or rotating into unconventional geometries; and customization of Mapping. Any axis or hit behavior can be reassigned to alternative parameters (filter cutoff, LFO depth, effect sends).

6 Workshop 2: Spatial Layout Co-Design

Following the first workshop, participants explored ideal spatial arrangements for VR jam sessions that should accommodate musicians with different ranges of skill levels [6]. During initial sketch sessions in a blank virtual world, participants arranged color-coded markers representing skill levels and used draggable shapes

[5]An explanation of the *Universal Rhythm Box* is available at: https://youtu.be/q3uD5X-UTVQ?si=6tEAkea7puMKX-Ej (accessed: 25/04/2026)

[6]Selected recordings of the workshop are available here: [Part 1] https://youtu.be/Tc3Sxoilg6c?si=YlHfhvL4iCsstgeh, [Part 2] https://youtu.be/VdBQhaGkSxY?si=5rBmad0tXv6db7G4, [Part 3] https://youtu.be/eixU0u7PRYA?si=0X6UXeRJ-NIiu94_ (accessed: 25/04/2026)

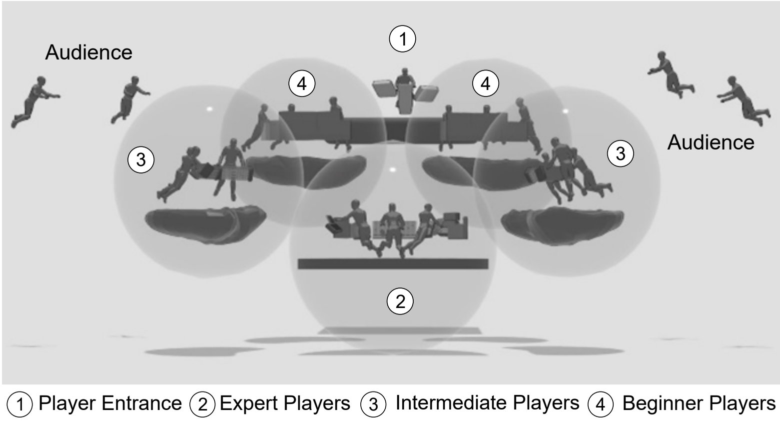

Figure 5: A rendering of the space layout co-designed during the second workshop.

to design spatial layouts. These sessions generated a range of solutions, including tiered layouts and concentric restriction zones (see Figure 4). Subsequently, the co-design session transitioned to recreating a realistic 3D VR environment for music-making that can accommodate both situations.

6.1 Procedure

Spatial layout workshops employed low-to-high fidelity prototyping [55, 58] across three phases:

Phase 1Scenario Definition and Sketching: Participants sketched spatial layouts using an empty virtual world with neutral blocks and colored tokens representing stations and roles, sketching arrangements and annotating instrument placements.

Phase 2 Low-Fidelity Prototyping in VR: Groups spent approximately 10 minutes arranging gray blocks and colored tokens, using virtual paint-buckets and pencils to sketch concepts and define spatial relationships. Participants annotated elements to indicate instrument and equipment placement (e.g., drums, bass, harmonics, mixer). This exercise revealed that small-scale sketches proved insufficient for evaluating sight lines and proximity until instantiated at life-size, but provided the basis for informative discussion.

Phase 3 High-Fidelity Instantiation and Task-Driven Evaluation: The facilitator imported actual virtual instruments into the VR world. The groups tested the preferred layouts through structured tasks (navigating to stations, executing call-and-response exercises) while the moderator recorded observations on visual access, crowding, and navigation ease. Participants' feedback prompted iterative refinements: station circles were reduced from five to three for peripheral visibility, and dome-enclosed *stage units* were introduced to demarcate expert areas while preserving audience engagement.

6.2 Emergent Spatial Concepts

Through the staging layout co-design process, four dominant spatial concepts emerged from the workshop sessions:

(1) **Tiered Circles:** A central and restricted expert area dedicated to expert players is surrounded by concentric rings for intermediate and beginners, resembling amphitheater-like tiers.
(2) **Glass Dome Enclosure:** Semi-transparent domes should be used for signaling access control, reinforcing visual hierarchy while preserving visibility.
(3) **Reproducible Stage Units:** Modular stage islands hosting instrument-specific clusters (e.g., percussive or melodic instruments), enabling flexible configurations through locked/unlocked behaviors and dynamic entry gating via mini-tasks (e.g., sequence a beat to gain entry).
(4) **Circular Player Pods:** A ring of cuboidal objects placed at chest height, with players facing inward in a close-knit circle, yielding strong peer-to-peer engagement and clear audience separation when enclosed in a transparent glass-like orb.

6.3 In-Situ Spatial Trials

In the concluding session, a modular *player pod* concept was developed and instantiated using three to five instrument instances (see Figure 5), and participants physically occupied the space (see Figure 6). Unlike the low-fidelity sketching phase, occupying the space at actual scale revealed some practical constraints that participants had not anticipated during the planning phase. Key observations included:

Optimal Ensemble Size. Systematic testing revealed clear performance differences based on ensemble size. With **five musicians**, the pod became crowded, causing avatar occlusion and peripheral vision loss. Participants reported difficulty tracking their peers' gestural cues. Reducing to **four musicians** improved visibility considerably but still produced occasional occlusion at the edges of the pod. **Three musicians** proved to be the optimal configuration: all performers could simultaneously see the avatars of the other participants and the shared waveform visualizations.

Semicircular Alignment. Positioning three instruments in a semicircle of approximately 120–150 degrees, facing the expert host area, preserved mutual visibility among all performers while maintaining a clear orientation toward the audience-facing zone. Several participants spontaneously remarked on the improved sense of ensemble cohesion.

Elevational Tweaks. Slight elevational offsets of approximately 20–30 cm for intermediate stations proved unexpectedly consequential. Raising intermediate stations modestly resolved this, ensuring clear line-of-sight to both peer musicians and the expert host.

6.4 Results

6.4.1 Finalized Spatial Layout. The spatial layout for mixed-skill groups devised by participants integrates the spatial concepts, ensemble-size findings, and elevational refinements developed across the workshop sessions into a coherent three-zone configuration:

- **Central host platform:** For session host and expert performers. This zone provides direct access to multiple advanced instruments organized in a semicircle. Its central position ensures that the gestural activity of these musicians remains visible to all ensemble members simultaneously, reinforcing their role of musical anchors and coordinators.
- **Intermediate performer stations:** Arranged in opposing semicircles at a slight elevational offset, these stations ensure clear visibility across the ensemble and toward the central zone. Multiple stations are envisaged, each instrument cluster focused on a particular musical section (e.g., rhythm or melody), allowing participants to specialize while remaining spatially integrated with the broader ensemble.
- **Peripheral beginner and viewer zone:** A surrounding ring with transparent barriers signals non-interactive or limited-interaction areas without breaking spatial immersion. This approach enables novices to observe expert performers at close range, supporting informal skill acquisition without requiring active participation.

This configuration addresses two design imperatives such as *inclusivity*, enabling novices to enter and observe without disrupting the session, and *workspace integrity*, ensuring that expert performers retain focused access to complex instruments and configurations.

6.4.2 Thematic Insights from Qualitative Analysis. Thematic coding of session transcripts revealed four key concerns related to collaborative VR music-making.

Latency and Timing: Delays between triggering sounds and hearing results caused frustration and reduced immersion. Participants reported difficulties maintaining rhythmic coherence even with quantization enabled, suggesting that latency compensation must go beyond beat alignment to include predictive cues and anticipatory feedback.

Collaboration and Coordination: Beyond technical issues, participants identified the lack of shared rhythm tools and sync indicators as barriers to musical alignment. Most critically, they frequently could not determine whether timing problems originated from their own actions, system delays, or collaborators' performances –an ambiguity that, if left unresolved, can undermine confidence and inhibit experimentation.

Instrument Design: While participants valued the ability to manipulate different parameters simultaneously, they found it confusing to manage too many parameters under performance pressure. Novices preferred instruments capable of producing interesting sounds with minimal configuration, while experts appreciated access to deeper parameter control.

Technical Considerations: Experienced participants requested features such as loopers, presets, MIDI integration, and the ability to save and share configurations.

7 Discussion

7.1 Understanding Prototype Success and Failure

The contrasting reception of the different prototypes of virtual instruments reveals an issue inherent in VRMIs. While musical expertise develops through sensorimotor coupling between gestures, haptic feedback, and sonic outcome, the familiar drum-pad metaphor of P1 promised this coupling while systematically violating embodied predictions resonating with findings that users prefer physical sticks for inertial feedback [42]. Our results extend this to collaborative contexts, where negative transfer from familiar interaction paradigms appeared to be even more disruptive.

Conversely, P3 succeeded by embracing the actual affordances of VR rather than simulating physical instruments. This resonates with the Piivert framework [8], which suggested that VR instrument design should be guided by 3D UIs principles, such as separating interaction into selection, modulation, and excitation gestures. This is what P3 implements through the combination of spatial positioning, wrist rotation, and hand strikes respectively. The opposite reception of P2 and P4 further illuminates these dynamics: P2 provided responsiveness without mechanisms for sustaining foundational rhythms, whereas the presets in P4 appealed to beginners but frustrated experts seeking real-time control.

Our findings suggest VR instruments should balance immediate accessibility with sustained depth, a challenge P3 addressed through simple core interaction combined with multi-axis expressivity.

Figure 6: Two screenshots of one of the jam sessions where the final layout and instruments were tested.

7.2 Visual Feedback as Constitutive

An initial reading might conclude that visual feedback is used as a way to compensate for the limited haptic feedback available in commercial VR systems. However, participant behavior suggests something more fundamental: VR music-making is establishing a new sensory hierarchy where vision becomes primary rather than supplementary.

We noticed that participants did not use the waveform display of P3 simply as a substitute or enhancement for haptic feedback, but they actively played *onto* the visualization as the primary performative target, adjusting hand positions based on visual parameter changes rather than sound alone. This aligns with findings that visual feedback significantly influences agency and control in VR [5, 7], extending the guideline for clear sensory mapping proposed by Serafin et al. [62]: visual feedback appears to function as the locus of action itself.

This enabled peculiar capabilities for group playing. All musicians could see the same real-time visualizations, creating shared perceptual anchors. Hand positions became readable as musical parameters, making intentions visible before becoming audible, echoing collaborative surfaces like the Reactable [38] and Jamo-Drum [15]. This suggests that designers of VR instruments should prioritize interactive visualization as a primary design dimension, as already proposed by Berthaut et al. [9].

However, this raises concerns. If VR music-making evolves as vision-primary, blind and visually impaired musicians face exclusion [30]. Yet recent work found that musicians maintained task focus under VR degradation if audio remained stable [16], suggesting a nuanced hierarchy: audio stability is primary for temporal coordination, while visual feedback is primary for spatial attribution.

7.3 Social Aspects and Progressive Complexity

While Dobrian and Koppelman [31] argue that lack of virtuosity in NIMEs inhibits expression, in multi-user VR environments collective musicality may matter more than individual virtuosity. This was observed also by Blaine and Fels: "*the quality of social interaction often takes precedence over the musical output produced*" [13]. Participants' convergence on semicircular arrangements (prioritizing group coherence over individual expressivity) supports this view. Expression in collaborative contexts emerges from relationships between performers, instruments, spatial arrangement, and visual feedback [35].

Both instrument and spatial design converged on progressive complexity (the gradual disclosure of advanced controls and spaces as user expertise develops) as an architectural principle. While HCI literature establishes progressive disclosure [50, 64], our findings identify three interdependent axes: *temporally*, features unlock as expertise develops (optional expert modes); *spatially*, tiered layouts position advanced instruments centrally and observation peripherally; *socially*, users occupy dynamically shifting expertise hierarchies across instrument types and contexts. This extends Blaine and Forlines' scaffolding strategies [14] for shared and collaborative VR environments.

Critically, these dimensions cannot be designed independently. The waveform displays of the *Universal Rhythm Box* required a semicircular arrangement of the players to ensure mutual visibility. While expert mode addresses instrument complexity, the spatial zones address workspace access. In shared virtual worlds, the effectiveness of instruments, spaces, and social protocols emerges from their interaction, suggesting that Musical Metaverse research requires systemic frameworks that address entire ecologies rather than isolated artifacts.

7.4 The Spatial Discovery

The finding that three musicians represent optimal minimal ensemble size emerged through systematic trials, not theoretical prediction. Participants found that five musicians created avatar occlusion and peripheral vision loss; four provided only marginal improvement.

This extends prior work on VR collaboration territoriality. Men and Bryan-Kinns [46] identified personal and group territories in dyadic VR music-making but did not examine how these dynamics can scale. The proposed three-person configuration may reflect a threshold at which territorial negotiation remains manageable—beyond this it may exceed what current VR interfaces support given field-of-view and resolution limitations.

The semicircular arrangement minimizes territorial overlap while maximizing mutual visibility, supporting both personal territory (every player is immediately in front of each instrument) and group territory (the shared central visual field). Elevation offsets (20-30cm adjustments) address a VR-specific challenge: unlike physical ensembles where musicians occupy distinct locations with clear presence, VR avatars and instruments can occlude each other. These seemingly minor adjustments proved crucial for maintaining line-of-sight.

The tiered access zones resonate with the finding by Men and Zhao [47] that gradient privacy mechanisms outperform rigid barriers. This design achieves analogous gradation through physical distance and visual transparency: novices observe experts without entering their workspace, creating continuous engagement gradients rather than hard boundaries.

7.5 Attribution Ambiguity

Beyond well-documented network latency challenges [28, 29], the findings of the two workshops reveal a distinct problem: musicians cannot determine whether timing issues originate from their own performance, system latency, network delays, or collaborators' errors. This *attribution ambiguity* (the inability to determine the source of timing errors in networked performance) breaks the feedback loop that can enable skill development and expressive control. One participant articulated this: "*I don't know if I'm playing badly or if the system is making me sound bad or if we're just not synced.*" In VR contexts, attribution ambiguity might also operates socially. When musicians hear something going off-beat, they face a series of questions: is it this error coming from their settings? Is it because of a lag? Is it a problem of the network? The lack of clarity can undermine trust and mutual adjustment essential to ensemble performance [60], making performers hesitant to provide feedback when they cannot identify error sources. The problem extends debates about liveness in technologically-mediated performance [1]: networked VR creates a hybrid mode, where spatial co-presence is simulated, temporal immediacy can be limited, and individual agency remains ambiguous. This situation might require new analytical frameworks.

While latency reduction remains important but insufficient on its own, design must explicitly address attribution through intention signaling, latency visualization, adaptive quantization, and social protocols for acknowledging errors. Yaseen et al. [71] propose a synchronization wheel that provides visual feedback of collective timing through coupled oscillators, representing one promising direction.

7.6 Design Guidelines

Drawing from the preceding discussion, we distill five empirically-grounded design principles:

Embrace 3D UIs principles and VR-native interactions: Abstract geometric interfaces—grids, volumes, fields—outperform physical instrument replicas lacking multimodal feedback. Visual feedback should serve as an expressive musical dimension in its own right, not merely a substitute for haptic feedback, with individual real-time visualizations visible to all ensemble members.

Implement progressive complexity across multiple dimensions: Complexity management must operate temporally (features unlocking through expert modes), spatially (tiered access zones with graduated proximity), and socially (accommodating shifting expertise hierarchies).

Optimize spatial arrangements for mutual visibility: Semicircular configurations (120-150 degrees) with graduated elevation offsets prevent peripheral vision loss. Design for three-musician baseline ensembles, scaling through multiple pods rather than large circles prone to occlusion.

Address attribution ambiguity explicitly: Provide mechanisms helping musicians distinguish self-generated timing issues from system latency or ensemble coordination problems.

Validate through full-scale embodied trials: Low-fidelity prototyping generates concepts but cannot validate spatial configurations requiring actual-scale occupation.

These guidelines reflect a core finding: VR music-making succeeds by leveraging the actual affordances of the medium such as visual primacy, spatial malleability, and computational flexibility (the ability to dynamically reconfigure mappings and instrument behaviors in ways that are impossible with fixed physical instruments), rather than by replicating physical practice.

7.7 Limitations and Future Work

Our study has notable limitations. Remote workshop delivery constrained hardware compatibility (Meta Quest only) and network latency affected rhythmic precision. The small sample size (n = 9) limits the generalizability of the results, and we present these findings as promising design directions that require further validation rather than as universal principles. We focused exclusively on rhythmic instruments. Future work should extend these principles to melodic and harmonic contexts with larger, and more diverse participants as well. Additionally, only longitudinal studies can validate whether our designs support skill progression as participants anticipated. Nevertheless, the convergence observed across independent groups suggests the findings reflect important design requirements rather than idiosyncratic preferences.

8 Conclusion

Through systematic co-design involving iterative prototyping and qualitative analysis, we developed and validated two key contributions to VR musical interface research: (1) the *Universal Rhythm Box* instrument emphasizing multi-hand expressivity and intuitive feedback, and (2) a spatial layout promoting ensemble cohesion across mixed-skill groups.

The findings presented here–including optimal three-musician configurations, semicircular alignments, and elevational offsets–provide actionable guidelines for future VR music system design. The identification of persistent challenges in latency, coordination tools, and interface intuitiveness highlights areas requiring continued research and development.

The success of participatory co-design suggests extending such methodologies to other Musical Metaverse aspects, including audience integration and hybrid physical-virtual performance. Broader questions remain: can musicians develop expert-level virtuosity with VR instruments? What does virtuosity mean when instruments are infinitely customizable? VR-native musical practice may fully emerge only when a generation grows up with VR as their primary musical experience—the Musical Metaverse timeline may extend beyond technical development to encompass generational cultural shifts.

Future work should also examine long-term usage patterns, cross-cultural applications, and integration with emerging VR technologies. We invite the NIME community to build upon these foundations in advancing the state of collaborative virtual music-making.

Acknowledgments

We acknowledge the support of the MUSMET project funded by the EIC Pathfinder Open scheme of the European Commission (grant agreement n. 101184379) and by the Swiss State Secretariat for Education, Research and Innovation (SERI). Views and opinions expressed are however those of the author(s) 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.

Ethics Statement

All participants provided informed consent for workshop participation and as well for audio and video documentation for research purposes. The study was conducted in accordance with the University of Trento ethical guidelines.

References

[1] Philip Auslander. 2022. *Liveness: Performance in a mediatized culture.* Routledge.

[2] Hanif Baharin, Noris Norowi, Khatriza Ahmad Saffian, and Yap Eng Sim. 2023. Participatory Design Workshop to Create a Virtual Reality Musical Instrument Based on Tumbuk Kalang. In *International Visual Informatics Conference.* Springer, 142–151.

[3] Jonathan Bell. 2023. Networked music performance in patchxr and flucoma. In *International Computer Music Conference (ICMC) 2023.*

[4] Florent Berthaut. 2020. 3D interaction techniques for musical expression. *Journal of New Music Research* 49, 1 (2020), 60–72.

[5] Florent Berthaut. 2021. Musical exploration of volumetric textures in mixed and virtual reality. In *International Conference on New Interfaces for Musical Expression (NIME 2021).*

[6] Florent Berthaut and Luke Dahl. 2022. The Effect of Visualisation Level and Situational Visibility in Co-located Digital Musical Ensembles. In *Proceedings of the International Conference on New Interfaces for Musical Expression.* Auckland, New Zealand.

[7] Florent Berthaut, Myriam Desainte-Catherine, and Martin Hachet. 2010. DRILE : An Immersive Environment for Hierarchical Live-Looping. In *New interface for musical expression.* 192–197.

[8] Florent Berthaut, Myriam Desainte-Catherine, and Martin Hachet. 2010. Pivert: Percussion-Based Interaction for Immersive Virtual Environments. In *Proceedings of the IEEE Symposium on 3D User Interfaces.* IEEE, 15–18.

[9] Florent Berthaut, Myriam Desainte-Catherine, and Martin Hachet. 2011. Interacting with 3D reactive widgets for musical performance. *Journal of New Music Research* 40, 3 (2011), 253–263.

[10] Florent Berthaut and Martin Hachet. 2016. Spatial interfaces and interactive 3d environments for immersive musical performances. *IEEE computer graphics and applications* 36, 5 (2016), 82–87.

[11] Florent Berthaut, Victor Zappi, and Dario Mazzanti. 2014. Scenography of Immersive Virtual Musical Instruments. In *Proceedings of the Workshop on Sonic Interactions in Virtual Environments.* 19–24.

[12] Mark Billinghurst, Pablo Cesar, Mar Gonzalez-Franco, Katherine Isbister, Julie Williamson, and Alexandra Kitson. 2024. Social XR: The Future of Communication and Collaboration (Dagstuhl Seminar 23482). *Dagstuhl Reports* 13, 11 (2024), 167–196.

[13] Tina Blaine and Sidney Fels. 2003. Contexts of collaborative musical experiences. In *Proceedings of the 2003 conference on New interfaces for musical expression.* 129–134.

[14] Tina Blaine and Clifton Forlines. 2002. Jam-O-World: Evolution of the Jam-O-Drum Multi-Player Musical Controller into the Jam-O-Whirl Gaming Interface. In *Proceedings of the International Conference on New Interfaces for Musical Expression.* Dublin, Ireland, 1–6.

[15] Tina Blaine and Tim Perkis. 2000. The Jam-O-Drum interactive music system: a study in interaction design. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques.* 165–173.

[16] Alberto Boem, Ovidiu Costin Andrioaia, Domenico Stefani, Alessandra Micalizzi, and Luca Turchet. 2025. A Decoupled VR and Real-time Audio System for Distributed Musical Collaboration. In *2025 IEEE 6th International Symposium on the Internet of Sounds (IS2).* 1–10.

[17] 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)*. Association for Computing Machinery, New York, NY, USA, Article 279, 20 pages.

[18] Alberto Boem, Damian Dziwis, Matteo Tomasetti, Sacha Etezazi, and Luca Turchet. 2024. "It Takes Two": Shared and Collaborative Virtual Musical Instruments in the Musical Metaverse. In *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*. IEEE, 1–10.

[19] 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.

[20] Alberto Boem, Matteo Tomasetti, and Luca Turchet. 2024. Harmonizing the Musical Metaverse: unveiling needs, tools, and challenges from experts' point of view. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Utrecht, Netherlands, Article 33, 9 pages.

[21] Alberto Boem and Luca Turchet. 2023. Musical Metaverse Playgrounds: Exploring the Design of Shared Virtual Sonic Experiences on Web Browsers. In *2023 4th International Symposium on the Internet of Sounds*. IEEE.

[22] Alberto Boem and Luca Turchet. 2024. Selection as Tapping: An Evaluation of 3D Input Techniques for Timing Tasks in Musical Virtual Reality. *International Journal of Human-Computer Studies* 185 (2024), 103231.

[23] Virginia Braun and Victoria Clarke. 2019. Reflecting on Reflexive Thematic Analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (2019), 589–597.

[24] Virginia Braun and Victoria Clarke. 2022. *Thematic Analysis: A Practical Guide*. SAGE Publications.

[25] Leonard Bruns, Benedict Saurbier, Tray Minh Voong, and Michael Oehler. 2024. Presence and flow in virtual and mixed realities for music-related educational settings. In *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*. IEEE, 1–7.

[26] Michel Buffa, Ayoub Hofr, and Dorian Girard. 2024. WAM Jam Party: Using Web Audio Modules in the Musical Metaverse. In *IEEE International Symposium on the Internet of Sounds 2024*.

[27] Michel Buffa, Marco Winckler, Quentin Escobar, Samuel Demont, Ayoub Hofr, and Adam Mir-Sadjadi. 2025. Interactive Audio Sculpting: Plugin Customization and UI Affordances in Immersive Environments. In *IS 2025-IEEE International Symposium on the Internet of Sounds*.

[28] 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.

[29] Chris Chafe and Michael Gurevich. 2004. Network time delay and ensemble accuracy: Effects of latency, asymmetry. In *Audio Engineering Society Convention*, Vol. 117. 6208.

[30] Khang Dang, Hamdi Korreshi, Yasir Iqbal, and Sooyeon Lee. 2023. Opportunities for Accessible Virtual Reality Design for Immersive Musical Performances for Blind and Low-Vision People. In *Proceedings of the 2023 ACM Symposium on Spatial User Interaction*. 1–21.

[31] Christopher Dobrian and Daniel Koppelman. 2006. The E in NIME: Musical Expression with New Computer Interfaces. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Paris, France, 277–282.

[32] 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.

[33] 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.

[34] Gregorio Andrea Giudici, Domenico Stefani, Alberto Boem, and Luca Turchet. 2025. Connecting MIDI Interfaces to the Musical Metaverse. In *2025 IEEE 6th International Symposium on the Internet of Sounds (IS2)*. 1–9.

[35] Ian Hattwick and Marcelo Wanderley. 2012. A dimension space for evaluating collaborative musical performance systems. In *Proceedings of the international conference on New Interfaces for Musical Expression*.

[36] Robert Jack, Tony Stockman, and Andrew McPherson. 2017. Navigation of Parameter Spaces in Audio Mixers: Effect of Proprioceptive Feedback Through Distance and Force-Sensing. In *Proceedings of the International Conference on New Interfaces for Musical Expression*.

[37] Sergi Jordà. 2005. Multi-user instruments: models, examples and promises. In *Proceedings of the 2005 conference on New interfaces for musical expression*. 23–26.

[38] Sergi Jordà, Günter Geiger, Marcos Alonso, and Martin Kaltenbrunner. 2007. The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In *Proceedings of the 1st international conference on Tangible and embedded interaction*. 139–146.

[39] Alexandra Kitson, Alissa N Antle, and Petr Slovak. 2023. Co-designing a virtual reality intervention for supporting cognitive reappraisal skills development with youth. In *Proceedings of the 22nd annual ACM interaction design and children conference*. 14–26.

[40] 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*. 95–102.

[41] Ben Loveridge. 2023. An Overview of Immersive Virtual Reality Music Experiences in Online Platforms. *Journal of Network Music and Arts* 5, 1 (2023).

[42] Teemu Mäki-Patola. 2005. User interface comparison for virtual drums. In *Proceedings of the 2005 conference on New interfaces for musical expression*. 144–147.

[43] Dario Mazzanti, Victor Zappi, Darwin G. Caldwell, and Andrea Brogni. 2014. Augmented Stage for Participatory Performances. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. 29–34.

[44] Yanni Mei, Jie Li, Huib De Ridder, and Pablo Cesar. 2021. Cakevr: A social virtual reality (vr) tool for co-designing cakes. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–14.

[45] Liang Men and Nick Bryan-Kinns. 2018. LeMo: Supporting Collaborative Music Making in Virtual Reality. In *2018 IEEE 4th VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*. 1–6.

[46] Liang Men and Nick Bryan-Kinns. 2019. LeMo: exploring virtual space for collaborative creativity. In *Proceedings of the 2019 Conference on Creativity and Cognition*. 71–82.

[47] Liang Men and Danqi Zhao. 2021. Designing privacy for collaborative music making in virtual reality. In *Proceedings of the 16th International Audio Mostly Conference*. 93–100.

[48] Michael J. Muller. 2003. Participatory Design: The Third Space in HCI. In *The Human-Computer Interaction Handbook*. L. Erlbaum Associates, 1051–1068.

[49] Teemu Mäki-Patola, Juha Laitinen, Aki Kanerva, and Tapio Takala. 2005. Experiments with Virtual Reality Instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. 11–16.

[50] Donald A. Norman. 2013. *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books.

[51] Marcos Novak. 1991. Liquid architectures in cyberspace. In *Cyberspace: first steps*. 225–254.

[52] Chaeryeong Oh, Dayoung Lee, and Alexandria Smith. 2024. Doongdoong.club: A Web-Based Metaverse Music Sequencer With Korean Onomatopoeic and Mimetic Words. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Utrecht, Netherlands, Article 91, 5 pages.

[53] Sophia Ppali, Ethan Cheung, Alexandra Covaci, Wan-Jou She, and Chee Siang Ang. 2025. Creating with Care: Co-Designing Immersive Experiences through Art-Making with People Living with Dementia. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–18.

[54] 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*. 1–19.

[55] Marc Rettig. 1994. Prototyping for Tiny Fingers. *Commun. ACM* 37, 4 (1994), 21–27.

[56] Toni Robertson and Jesper Simonsen (Eds.). 2012. *Routledge International Handbook of Participatory Design*. Routledge.

[57] Mary Beth Rosson and John M Carroll. 2007. Scenario-based design. In *The human-computer interaction handbook*. CRC Press, 1067–1086.

[58] Jim Rudd, Ken Stern, and Scott Isensee. 1996. Low vs. High-Fidelity Prototyping Debate. *Interactions* 3, 1 (1996), 76–85.

[59] Elizabeth B.-N. Sanders. 2002. From User-Centered to Participatory Design Approaches. *Design and the Social Sciences* (2002), 1–8.

[60] R Keith Sawyer. 2006. Group creativity: Musical performance and collaboration. *Psychology of music* 34, 2 (2006), 148–165.

[61] Douglas Schuler and Aki Namioka (Eds.). 1993. *Participatory Design: Principles and Practices*. CRC Press.

[62] Stefania Serafin, Cumhur Erkut, Juraj Kojs, Niels C. Nilsson, and Rolf Nordahl. 2016. Virtual Reality Musical Instruments: State of the Art, Design Principles, and Future Directions. *Computer Music Journal* 40, 3 (2016), 22–40.

[63] Syed Hammad Hussain Shah, Anniken Susanne T Karlsen, Mads Solberg, and Ibrahim A Hameed. 2023. A social VR-based collaborative exergame for rehabilitation: codesign, development and user study. *Virtual Reality* 27, 4 (2023), 3403–3420.

[64] Ben Shneiderman, Catherine Plaisant, Maxine Cohen, Steven Jacobs, Niklas Elmqvist, and Nicholas Diakopoulos. 2016. *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (6th ed.). Pearson.

[65] 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*. 2150–2161.

[66] Luca Turchet. 2023. Musical Metaverse: Vision, Opportunities, and Challenges. *Personal and Ubiquitous Computing* 27 (2023), 1811–1827.

[67] Luca Turchet and Mathieu Barthet. 2018. Co-design of Musical Haptic Wearables for electronic music performer's communication. *IEEE Transactions on Human-Machine Systems* 49, 2 (2018), 183–193.

[68] Luca Turchet, Rob Hamilton, and Anil Çamci. 2021. Music in Extended Realities. *IEEE Access* 9 (2021), 15810–15832.

[69] Hung Vo and Richard Boulanger. 2024. Csound in the MetaVerse – From Cabbage to CsoundUnity and Beyond: Developing a Working Environment for SoundScapes, SoundCollages, and Collaborative SoundPlay. In *Proceedings*

of the 7th International Csound Conference. ICSC 2024, Vienna, Austria, 115–119.

[70] R.C. Waters and J.W. Barrus. 1997. The rise of shared virtual environments. *IEEE Spectrum* 34, 3 (1997), 20–25.

[71] Azeema Yaseen, Sutirtha Chakraborty, and Joseph Timoney. 2022. A cooperative and interactive gesture-based drumming interface with application to the Internet of Musical Things. In *International Conference on Human-Computer Interaction*. Springer, 85–92.

[72] Victor Zappi, Florent Berthaut, and Dario Mazzanti. 2022. From the Lab to the Stage: Practical Considerations on Designing Performances with Immersive Virtual Musical Instruments. In *Sonic Interactions in Virtual Environments*. Springer, 385–424.

[73] Victor Zappi, Marco Gaudina, Andrea Brogni, and Darwin Caldwell. 2010. Virtual sequencing with a tactile feedback device. In *International Workshop on Haptic and Audio Interaction Design*. Springer, 149–159.

[74] Victor Zappi and Andrew McPherson. 2012. Design and Use of a Hackable Digital Instrument. In *Proceedings of the International Conference on Live Interfaces*.