The Musical Metaverse Research Project a MusicTech Survey

Stefano Giacomelli
DISIM, University of L'Aquila, Italy
L'Aquila, Italy
stefano.giacomelli@graduate.univaq.it

Agostino Di Scipio
MeAQ, Conservatory of Music "A. Casella"
L'Aquila, Italy
a.discipio@consaq.it

Claudia Rinaldi
DISIM, University of L'Aquila / CNIT
L'Aquila, Italy
claudia.rinaldi@univaq.it

***Abstract*—The *Musical Metaverse* pilot project is a multidisciplinary research initiative funded under the Italian PRIN 2022 program, aiming to develop an inclusive, low-latency eXtended Reality platform for networked musical collaboration. Participatory studies involving musicians, educators, and researchers inform the development of XR interfaces, multisensory feedback, and Human-Computer interface tools. Technological challenges in Networked Music Performance are addressed via 5G slicing, edge computing, satellite links, and embedded audio processing. Real-world validations span remote concerts, VR composition tools, and inclusive pedagogical use cases. This survey critically reviews the project outcomes, structured across three interrelated pillars: *User-Centered Design*, *Technological Innovation*, and *Application & Evaluation*. By integrating social, artistic, and engineering perspectives, the project outlines a comprehensive framework for distributed, immersive, and ethically-aware musical interaction.**

***Index Terms*—Musical Metaverse (MM), Music Technology (MusicTech), Networked Music Performance (NMP), Musical XR, Accessibility, Multi-sensory Interactive systems, Internet of Musical Things (IoMusT)**

I. Introduction

Recent advances in music technology, immersive media, and networked systems have enabled new paradigms for musical expression, education, and performance. Among these, the notion of a *Musical Metaverse* has gained traction as a socio-technical vision that integrates Extended Reality (XR), real-time (RT) interaction, and inclusive design into cohesive frameworks for collaborative music-making. Realizing this vision requires not only technological development but also a deep understanding of artistic practices, accessibility, and Human-Computer Interaction (HCI) models, especially as infrastructures like 5G, edge computing, and the Internet of Musical Things (IoMusT) offer unprecedented opportunities for distributed, expressive environments.

The Musical Metaverse: an inclusive Extended Reality platform for networked musical interactions (MM pilot project) is a 24-month research Project of Relevant National Interest[1] (PRIN), aimed at developing an XR-based platform for collaborative musical interaction through the convergence of Virtual Reality (VR) and Augmented Reality (AR) technologies. Started in October 2023 and ending in February 2026[1], the project aligns with two European Research Council (ERC) domains: Social Sciences and Humanities (SH5_5: Music and musicology) and Physical Sciences and Engineering (PE6_9, PE7_8: HCI, visualization, and communication networks).

The MM pilot project addresses key XR challenges (particularly communication latency, a core limitation for remote collaboration) while embedding inclusivity (e.g.: gender balance, disability access) and ethical principles into its methodology. It involves four Research Units (RUs): the University of Trento with the Creative, Intelligent and Multisensory Interactions Laboratory (CIMIL - UniTN)[2] and Polytechnic University of Turin (PoliTO)[3], leading development in Musical XR, HCI, algorithmic ethics, telecommunications, and networked performance; the Conservatory of L'Aquila (ConsAQ)[4] and the School of Audio Engineering (SAE) Italy (Milan)[5], focusing on artistic and pedagogical aspects in classical, electroacoustic, and popular music.

This paper surveys the MM project research activities with a structured overview aimed at technologists, musicians, and contemporary musicologists. Section III analyzes how user-centered principles shape inclusive XR systems. Section IV reviews enabling technologies for low-latency musical interaction, while Section V discusses real-world applications and evaluation strategies. The paper concludes by outlining possible limitations, open challenges, and future directions for inclusive and distributed Musical Metaverse environments.

II. Musical Metaverse Foundations and Key Challenges

The Metaverse is generally described as a digital counterpart to the physical world, where users interact through avatars [1]. While this vision has influenced many domains, its application to musical contexts remains largely unexplored. The MM pilot project was launched to address three foundational gaps: *(1)* the lack of formalized frameworks for music-specific XR

This work was partially supported by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART") and by the the MUR NRRP PRIN 2022 grant, prot. n. 2022CZWWKP, funded by NextGenerationEU.

[1]https://prin.mur.gov.it/Home

[2]https://www.cimil.disi.unitn.it/projects

[3]https://www.polito.it/impatto-sociale/polito-per-il-sociale/la-progettualita/le-esperienze-di-impatto-sociale/musical-metaverse

[4]https://www.consaq.it/eventi-e-attivita/prin-musical-metaverse-presentazione.html

[5]https://www.sae.edu/ita/progetti-di-ricerca

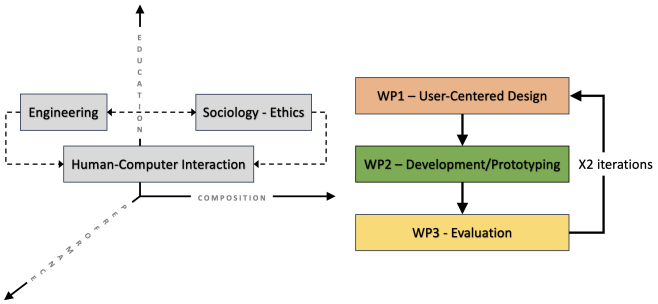

Fig. 1. Musical Metaverse – Research Project Iterative Methodology

practices; *(2)* the limited capacity of current XR infrastructures to meet RT and expressive demands of networked collaboration and education; and *(3)* the absence of inclusive design strategies tailored to higher music education and *avantgarde* performance. Together, these issues call for a socio-technical ecosystem in which expressive freedom, accessibility, and low-latency interaction converge.

The project central question is: *how can XR and IoMusT technologies be directed toward inclusive, expressive, and low-latency musical collaboration?* To address this, the project research spans three interdependent domains: *Needs & Design* (eliciting user requirements via participatory methods), *Technological Foundations* (developing XR-enabling technologies), and *Empirical Validation* (testing and refining prototypes through feedback and evaluation). Each output adheres to open-science principles, starting from systematic state-of-the-art (SoA) analysis and proceeding through user-centered and ethically-aware methodologies to reproducible implementations.

A. Project Research Methodology

The MM project adopts a *doubly iterative* workflow composed of three interconnected phases — *Design*, *Development*, and *Evaluation* — repeated over two 12-month cycles to incrementally increase technology readiness and deliver working *proof-of-concepts* (Figure 1). Each phase feeds into the next, while evaluation results recursively inform both design and implementation.

The *Design* phase involves rapid prototyping of XR interfaces, multimodal data acquisition, and preliminary User eXperience (UX) metrics under controlled network settings. Participatory and user-centered approaches involving musicians, technologists, and educators are employed to extract both conceptual and technical requirements. The *Development* phase translates these requirements into software and networking prototypes, such as low-latency architectures, spatial audio rendering, audio Digital Signal Processing (DSP) modules, and distributed interaction systems. In the *Evaluation* phase, prototypes are tested for real-world applicative scenarios, with both objective metrics (latency, jitter, accuracy) and subjective indicators (presence, cybersickness, social cohesion) assessed. This iterative structure ensures that each development cycle

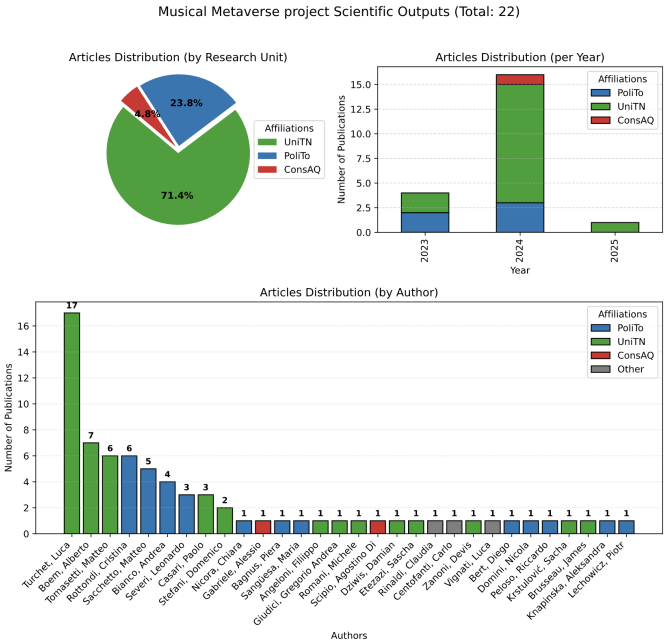

Fig. 2. Musical Metaverse project: Overall Publications (RU co-authorship counts as 1, per *Affiliation*)

aligns with actual use-case demands and can be refined accordingly.

This methodology is structured around three foundational pillars:

- ***Pillar I – User-Centered Design (UCD)***: applies sociological, cognitive, and ethical frameworks through co-design processes. It emphasizes transparency, trust, equity, gender balance, privacy, and security using an *ethics-by-design* approach [2], aligned with gendered innovation principles [3]. Accessibility for disabled professional and student musicians is a key focus.
- ***Pillar II – Technological Innovation***: develops XR and networking solutions for composition, performance, and pedagogy. Targets include sub-30 ms End-to-End (E2E) latency for realistic NMP [1], [4], [5], and Machine Learning (ML)-based real-time Packet Loss Concealment (PLC) [6] and traffic prediction for enhanced Quality of Experience (QoE).
- ***Pillar III – Application & Evaluation***: integrates user and technical assessments of the systems, feeding findings back into design and development cycles.

B. Survey Methodology

To offer a structured and interconnected overview of the ongoing research activities, this survey was conceived as a self-assessment exercise aligned with the open-science principles of the MM pilot project. It aims to provide a critical snapshot of the current integration achieved across technical, epistemic, cultural, and application domains, particularly from the perspective of music and technology (MusicTech) students as end users. The first author serves as an external collaborator for RU-ConsAQ, with a hybrid academic, artistic, and

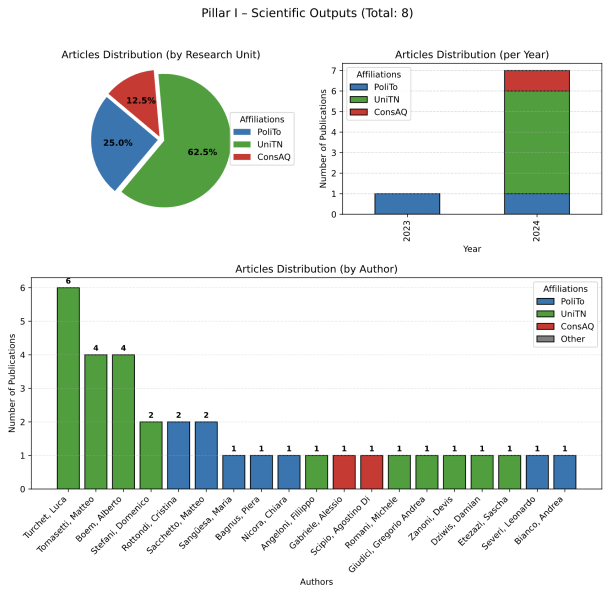

Fig. 3. Musical Metaverse project: Pillar I Publications (RU co-authorship counts as 1, per *Affiliation*)

technical background. To preserve neutrality, the review was conducted independently under the supervision of CNIT and the University of L'Aquila, which are not directly involved in the project execution.

Following the criteria outlined in the original PRIN proposal[1], all peer-reviewed and open-access publications explicitly acknowledging the MM pilot project grant (prot. `2022CZWWKP`) were collected (Figure 2). The search included all works authored by at least one RU-leader and indexed in Google Scholar or Scopus, between 2023 and 2025 (last update: 11/04/2025). Each publication was examined and categorized — non-exclusively — according to its contribution to the three research pillars.

III. Pillar I: User-Centered Design

The selected studies (Figure 3) align with the UCD framework of the MM project and investigate how musicians and stakeholders envision or interact with XR-enhanced music systems. Table I summarizes their methods and findings.

Boem et al. [7] conducted semi-structured interviews with 11 experts involved in 24 XR-music projects. Thematic analysis revealed motivations centered on immersive collaboration, networked presence, and artistic exploration. Most environments relied on Head-Mounted Display (HMD)-based VR, MR, or projection mapping. Spatial audio was appreciated but often secondary to Hi-Fi audio. Preferred development tools included Unity [11], A-Frame [12], and PatchWorld [13]; audio processing involved SuperCollider and Ableton Live. Latency, jitter, and lack of cross-platform standards were recurring technical barriers.

Complementing this, a workshop with 14 electroacoustic practitioners [8] explored envisioned XR use in composition, performance, and education. Participants valued MR over full VR, favoring hybrid tools aligned with embodied musical practices. They emphasized modular workflows, spatial notation, and integration with familiar environments.

Sacchetto et al. [9] quantitatively examined sound-to-visual/tactile mappings among 56 Deaf and Hard of Hearing (DHH)-focused students and educators. Results identified robust cross-modal correspondences used to design a multisensory interface supporting inclusive learning.

Collectively, these studies converge on a shared vision of the MM as a platform for collaborative creation, immersive education, and sensorial enrichment. They highlight priorities including low-latency networking, customizable XR interfaces compatible with common Digital Audio Workstations (DAWs), and support for tactile feedback. Technologically, modular environments and reliable synchronization across modalities were seen as essential.

Beyond implementation, participants framed the MM as a socio-cultural space where co-presence and critical reflection intersect. Concerns about the limits of full virtuality, economic accessibility, and lack of open standards persist. Many favored MR systems where physical presence and expressive nuance remain central.

A. The Vision and Architecture of the Musical Metaverse

Following early user-driven explorations, the MM project has progressively evolved into a structured framework blending technical and theoretical foundations. Two key contributions, by Rottondi et al. [14] and Turchet [15], define its architectural and conceptual direction.

Despite differing scopes, both works emphasize inclusivity, accessibility, and scalability, highlighting the synergy between technological innovation and UCD. Rottondi et al. [14] propose a modular architecture based on *building blocks*: accessible HMIs for users with sensory/motor impairments, motion tracking, immersive spatial audio engines, and haptic devices. These components are coordinated through low-latency streaming and edge/cloud infrastructures to support Hi-Fi remote performances across heterogeneous networks.

Conversely, Turchet [15] frames the MM within the Internet of Musical Things and People (IoMusTP), a socio-technical mesh of human/non-human agents co-constructing musical meaning. Rather than technical modules, he outlines ten ethical design dimensions (e.g. privacy, diversity, transparency, sustainability, and access) critiquing techno-solutionist assumptions and advocating for a reflexive, *more-than-human* approach.

Common priorities emerge. Both call for *ubiquitous access*, pursued via network-aware architectures in [14], and embedded in pervasive paradigms in [15]. Similarly, *accessibility and inclusiveness* are addressed through assistive technologies and ethically grounded design practices. Both works promote *multi-modal interaction*: Rottondi et al. emphasize immersive audio and haptics; Turchet advocates cultural and perceptual adaptability.

Together, they outline two complementary visions: one engineering-focused and system-oriented, the other normative

TABLE I
SUMMARY OF USER-CENTERED PRELIMINARY SURVEY STUDIES

Study	Stakeholders	Collection Methodology	Domains and Key Results
[7]	XR developers, researchers in NMP and immersive media	*Qualitative*: semi-structured interviews with 11 experts	**Motivations**: immersive collaboration, Hi-Fi audio, networked presence **Tools**: Unity, SuperCollider, Ableton, OSC, WebRTC, Sonobus, JackTrip **Challenges**: latency, synchronization, lack of standards
[8]	Electroacoustic musicians, students and teachers	*Workshop-based* inquiry with 14 musicians	**Scenarios**: MM for composition, performance, pedagogy **Outcomes**: spatial workflows, XR-enhanced concerts, historical reenactments **Critique**: expressivity limits in VR, MR preferred, latency concerns
[9]	Music students and educators (DHH focus)	*Quantitative*: survey with 56 participants	**Focus**: perceptual diversity, accessible XR pedagogy **Goal**: map sound–visual–tactile correspondences **Results**: consistent cross-modal associations, informing haptic interface design
Total participants: 81 + 30 in [10]			

and socio-culturally reflexive. In synthesis, they suggest that MM ecosystems must merge robust infrastructures with ethical, sustainable, and user-driven principles to enable inclusive, adaptable, and musically grounded XR environments.

B. Musician-Centric Design Examples

Several technical contributions take a musician-centric approach, enabling performers to shape XR systems for greater expressivity, collaboration, and accessibility, in line with UCD principles.

Boem et al. [16] explored shared Virtual Musical Instruments (VMIs) for distributed performance via WebXR. Three prototypes were developed: the Spatialized Performance Interface (SPI), Sonification of Interactions (SOI), and Behavioral Data Interface (BDI), addressing spatial co-creation, relational sonification, and embodied interaction. A user study assessed these through System Usability Scale (SUS), Creativity Support Index (CSI), and Networked Minds Social Presence Inventory (NMSPI) metrics. Results (Table II) showed that embodied, symmetrical designs (SOI, BDI) fostered creativity and presence, while SPI asymmetry hindered mutual engagement.

Romani et al. [17] introduced BCHJam[6]: an IoMusT ecosystem integrating Brain-Computer Interfaces (BCIs), smart instruments, and MR headsets. Electroencephalogram (EEG) signals — Event-Related Potentials (ERPs) and Alpha/Beta waves — were mapped to Open Sound Control (OSC)-based control of audio effects and visuals in Unity. Iterative testing revealed a trade-off between responsiveness and false positives. A 1s focus window with 90% confidence optimized performance, enhancing audience-performer immersion.

Sacchetto et al. [9] designed a multisensory prototype for DHH students using TanvasTouch devices to deliver coordinated visual, audio, and tactile stimuli. Cross-modal mappings (Table III) translated instrument properties into perceptible textures and colors, enabling sonically grounded, inclusive music education.

Stefani et al. [18] presented *Esteso*[7], a Max/MSP-based AI co-performer for double bass improvisation. It integrates gesture recognition via IRCAM-RAVE Neural Networks model and K-Nearest Neighbours (K-NN) classifier, an interaction engine, and RT sound synthesis (granular processing, timbre transfer, reverb). Three sessions revealed Esteso capacity for extended technique recognition and dialogue, but also limits in temporal awareness, soft dynamic handling, and improvisational independence (Table IV).

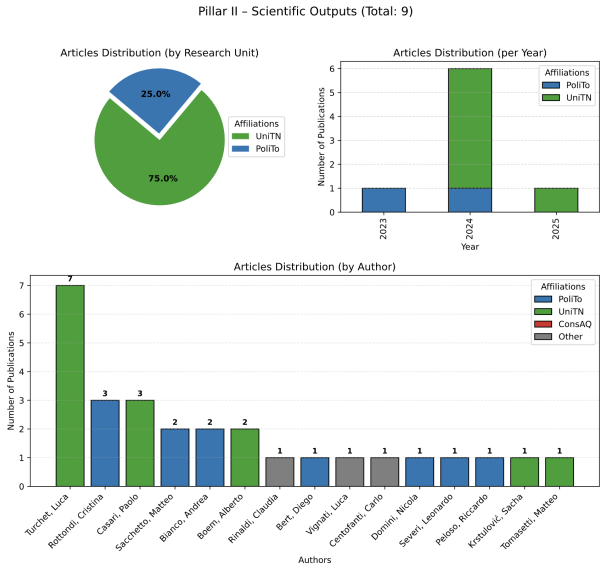

Fig. 4. Musical Metaverse project: Pillar II Publications (RU co-authorship counts as 1, per *Affiliation*)

Collectively, these examples show how musician involvement enhances expressive potential and inclusivity. Recurring design priorities are: *(1)* embodied multi-modal interaction across gesture, sound, and visuals; *(2)* modular, low-latency architectures for co-presence; *(3)* inclusive, cross-sensory interfaces for education and performance; *(4)* MR-preferred, hybrid configurations favoring physico-digital continuity.

These insights directly inform Pillar II strategies and the prototypical implementations of Pillar III.

IV. PILLAR II: TECHNOLOGICAL INNOVATION

Realizing the MM project vision requires robust network architectures and advanced technological frameworks capable

[6]https://github.com/BRomans/BCHJam

[7]This paper was excluded from the overall count of publications for Pillar I due to not fulfilling the acknowledgment requirements defined in our methodology. However, we decided to include it in the presentation given its strong temporal, contents and personnel-related correlation with the project's activities.

TABLE II
SUMMARY OF EVALUATION RESULTS FROM [16]

VMIs	Quantitative Evaluation	Qualitative Observations
SOI	CSI: **74.33**; SUS: **78.33**; NMSPI: high	Intuitive mappings; promoted creative exploration
SPI	CSI: 64.5; SUS: 71.45; NMSPI: lowest	Role ambiguity reduced engagement
BDI	CSI: 57.72; SUS: 68.54; NMSPI: **highest**	"Drum circle" dynamics; some latency issues

TABLE III
CROSS-MODAL ASSOCIATIONS INFORMING THE PROTOTYPE IN [9]

Instrument	Colors	Lightings	Shape/Texture
Harp	Yellow, White, Green	Solar, Hot	Curved, Smooth
Piano	Red, Brown, Black	Dazzling, Cold	Irregular, Hard

TABLE IV
MUSICOLOGICAL OBSERVATIONS FROM EVALUATION IN [18]

Theme	Observation
Time	AI missed silences as cues; lacked timing awareness
Dynamics/Timbre	Poor soft-dynamic response; timbrally strong but predictable
Comparative Feedback	S1: aggressive/stable; S2: clean/unpredictable; S3: over-controlled, less freedom

of supporting the stringent demands of RT-NMPs. As shown in Figure 4, this pillar spans three interdependent directions: *(Ia)* design and evaluation of *low-latency infrastructures, architectures*, and *enablers* for Hi-Fi remote musical interaction; *(Ib)* optimization of *network behavior* via slicing, orchestration, and adaptive control; and *(II)* deployment of *hardware-level* and *computational strategies*, including embedded DSP, RT co-processors, and distributed DSP services. The aim is to ensure consistent QoS and QoE across heterogeneous scenarios, from urban 5G to rural contexts, relying on satellite or edge infrastructures [14], [15].

The design choices are strongly influenced by the artistic and didactic needs highlighted in Pillar I — e.g.: stability in distributed performance, timing precision in composition interfaces, and inclusive, multi-modal learning systems. The technologies discussed thus align technical feasibility with expressive and educational relevance.

A. Networking Enablers: Infrastructures, Architectures & Optimization

Several MM studies addressed networking architectures for critical MM applications as Immersive NMPs (INMPs), leveraging 5G, Multi-acces Edge Computing (MEC), and Software-Defined Networks (SDN) to reduce jitter and enable predictable latency.Identified enablers are Virtualized Network Function (VNF) orchestration and ML-based traffic control. Non-Terrestrial Networks (NTNs) have also been explored to extend coverage to infrastructure-sparse areas.

Turchet et al. [19] proposed a dual-network architecture (Figure 5) integrating SDN, NFV, and MEC for INMPs within the IoMusT framework. Two configurations were tested: MEC-enabled RT spatial rendering, and embedded computing for reduced round-trip time (RTT). Both used binaural rendering, PLC, and ML-based traffic prediction. Field trials with 10

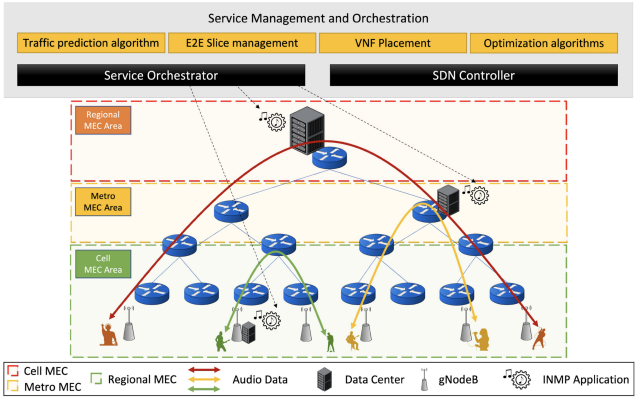

Fig. 5. MEC architecture proposal for INMPs, in [19].

TABLE V
QOS PARAMETERS: LATENCY AND PACKET LOSS

Study	Latency Mean	Latency STD	PL
[21] Starlink	168 ms	243 ms	16%
[20] MEC w. slicing	24.24 ms	0.39 ms	0.6%
[20] MEC no-slicing	23.95 ms	0.56 ms	0.64%
[22] indoor	33.6 ms	11.6 ms	2.4%
[22] outdoor	29.85 ms	9.44 ms	0.9%

nodes [20] confirmed E2E latencies below 30 ms. MEC offered scalability; embedded setups minimized delay. Server placement emerged as a key factor, and QoS degradation in heterogeneous networks remains a challenge.

Turchet and Casari [21] evaluated Starlink-based Low-Earth Orbit (LEO) connectivity for rural IoMusT (Figure 6) using Elk LIVE [23] with UDP audio. Two configurations — rural-to-rural (Starlink–Starlink) and rural–to-urban (Starlink–wired) — were tested. The former showed unacceptable delay (168.21 ms) and Packet Error Ratio (PER: 0.159), while the latter yielded moderate RTTs (45.68 ms, PER: 0.043). Both setups failed to meet NMP latency needs, indicating current LEOs are only viable for asynchronous use cases. Recommendations include predictive correction, buffering, and future exploration of 6G satellite paradigms.

To ensure scalable QoS, Turchet et al. [20], [24] explored 5G network slicing. Tests across slicing configurations of co-located/remote Core Networks (CNs) and MECs under stress conditions showed that MEC-User Plane Functions (UPF) setups — while slightly increasing local delay — reduced Wide Area Network (WAN) latency and improved isolation. A follow-up study involving 10 nodes [20] confirmed that slicing improves PER, burst loss, and missed packets, with only marginal latency impact. Linear Mixed Models (LMM)

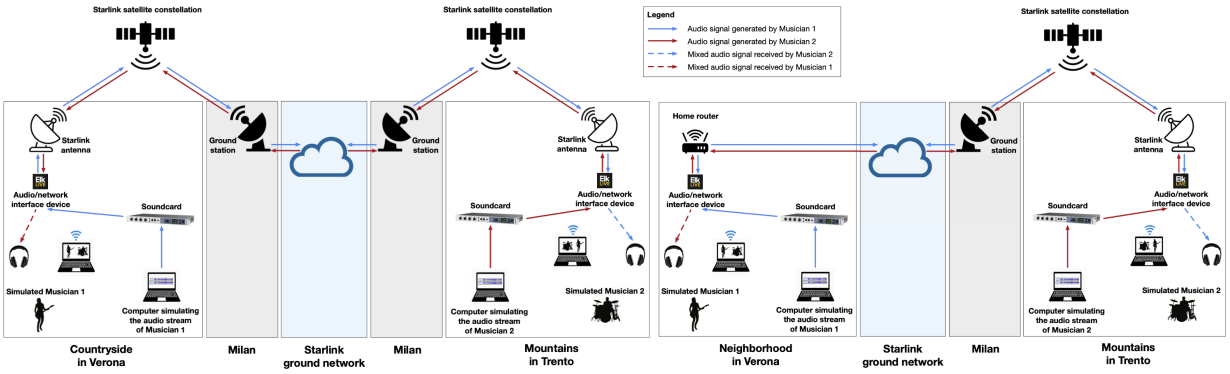

Fig. 6. LEO satellite scenarios in [21]: rural node-to-node (Left) and mixed rural-urban node-wired (Right).

and Analysis of Variance (ANOVA) validation (225.000 packets/session) revealed weak correlation between latency and reliability — emphasizing need for independent optimization.

Collectively, these studies validate slicing and MEC as key enablers of low-latency, scalable music networks. Despite NTN limitations, challenges persist in public WAN integration, multi-slice orchestration, and dynamic edge-cloud deployments.

B. Hardware & Optimization Solutions for DSP & TelCom

Beyond networking, RT musical systems require high-fidelity signal processing and efficient computation. Sacchetto et al. [25] proposed a RT-PLC method based on Burg's Auto-regressive (AR) model, optimized for Raspberry Pi 4B. Three variants — standard, recursive denominator, and hybrid — were tested. The hybrid method achieved sub-2.9 ms latency up to 128^{th} order, outperforming baselines (e.g.: silence substitution). Floating-point compensation improved stability. Future directions include incremental models and multichannel support.

Bert et al. [26] presented a co-designed Transport Triggered Architecture (TTA)-ASIP system on Field Programmable Gate Array (FPGA, Figure 7) for RT-DSP. Components include a custom I2S audio board, TTA processor (via OpenASIP), and a Raspberry Pi interfaced via SPI. Achieved latency was 0.859 ms, with $<$350 mW power and $<$25% FPGA resources usage. The system scales to 160+ audio sources and supports lightweight orchestration-ready deployments.

Turchet and Krstulović [27] introduced DSP-as-a-Service (DSPaaS): a cloud/edge paradigm for distributed audio. Three tiers — *asynchronous*, *reactive*, *real-time* — were defined. Only edge-tier matched perceptual latency, though packet loss remained problematic. Challenges include routing, PLC, and adaptive orchestration.

Boem et al. [28] critically reviewed audio limits in XR-based MM. Citing [8], they argue current XR systems lack support for deterministic audio timing, high-quality I/O, and standardized spatial audio. Their "Self vs Distant Reality" model highlights expressive latency layers, even in optimized embedded setups [26]. These gaps also hinder accessibility [9].

Recommendations include predictive buffering, hybrid deployments, and open standards (e.g.: MPEG, IVAS). A study by Boem and Turchet [10] illustrates UCD-aligned evaluation of five 3D input techniques in VR music scenarios. Using Meta Quest 2 HMDs, 30 users (musicians/non-musicians) performed rhythmic tapping via five paradigms: Gaze Point (GP), Controller Point (CP), Controller Touch (CT), Hand Point (HP), and Hand Touch (HT). CT showed best timing accuracy and user preference, especially in bi-manual mode. Gaze-/hand-based methods showed higher asynchrony and fatigue. No significant interaction was found with musical expertise, reinforcing generalizability.

This convergence of hardware, software, and evaluation demonstrates that expressivity in distributed music systems requires more than computational speed. Optimization must be guided by performative goals, inclusivity, and artistic values — redefining where and how audio computation takes place across the MM stack.

V. Pillar III: Applications & Evaluations

While technological infrastructures and DSP strategies are essential MM enablers, their true value emerges in concrete artistic and interactive settings. Pillar III encompasses applied research where immersive environments, musical interaction models, and compositional strategies are tested via real-world use cases and structured evaluations.

The studies span three application domains: *web-based musical prototyping* (browser-native XR/audio tools); *embodied spatial composition in VR* (locomotion as performative metaphor); and *infrastructure-integrated NMP systems* (remote rehearsals and telepresence with embedded computing). These contributions validate the frameworks of Pillars I and II, while serving as testbeds for iterative refinement between user needs, system design, and artistic goals. Despite different strategies, all emphasize immersive spatiality, embodied control, and performer-centered evaluation.

Boem and Turchet [29] introduced *Musical Metaverse Playgrounds*: browser-based shared environments for collaborative music-making using A-Frame, WebAudio APIs, and Resonance Audio SDK. Two playgrounds enabled 3D sound manipulation and voice-driven audiovisual modulation via signal

TABLE VI
SUMMARY OF 5G SLICING EVALUATION STUDIES

Study	Architectures / Scenario	Techniques	Metrics	Evaluation	Key Findings
[24]	Co-/Remote CN & MEC; 5G SA; 8x emulated NMP	Slicing, UPF placement	Latency, jitter, PER	Stress tests, UDP injection	MEC slicing reduces WAN latency and ensures isolation
[20]	10-node 5G SA testbed, dual NMP sessions	Slicing, MEC-UPF	Latency (23.95 → 24.24 ms), PER, missed packets, burst loss	LMMs, ANOVA, 225k packets/session	Improved reliability with minimal latency impact

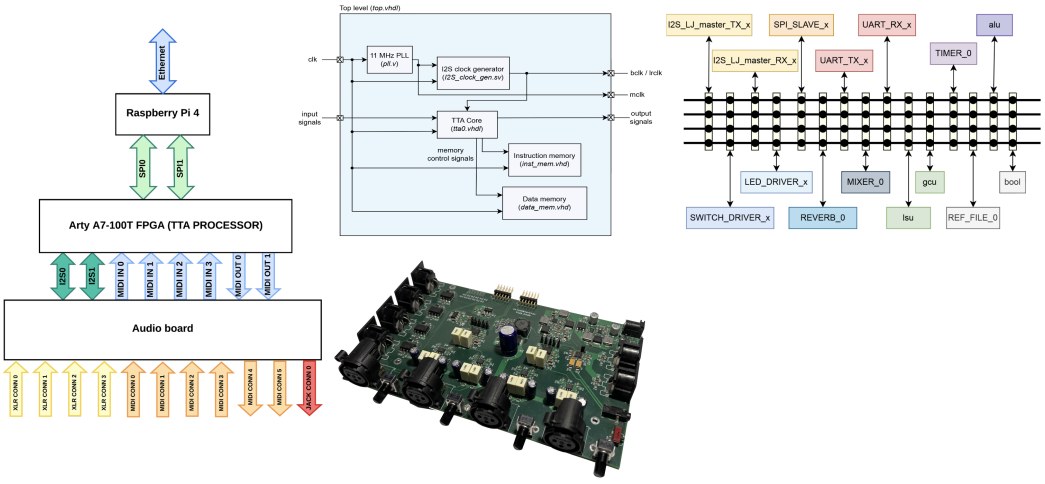

Fig. 7. Hardware & communication protocol in [26].

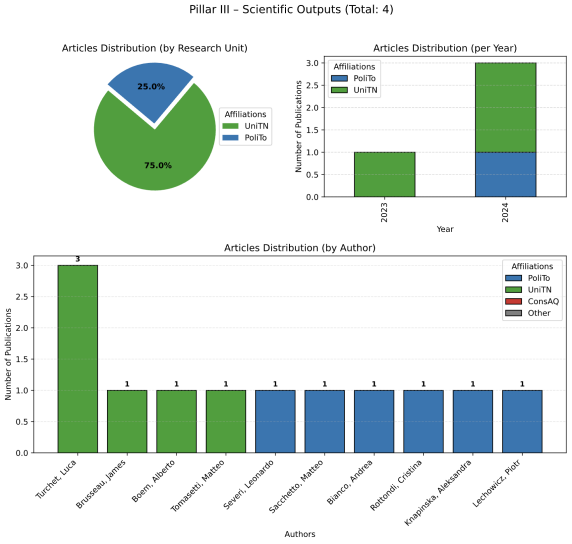

Fig. 8. Musical Metaverse project: Pillar III Publications (RU co-authorship counts as 1, per *Affiliation*)

descriptors. Eight users (Meta Quest 1 HMDs) participated in a think-aloud/user study. Results showed high appreciation for spatial sound and accessibility, with limitations in interaction granularity, visual feedback, and browser latency — echoing challenges of latency and AV sync outlined in [19], [20].

Tomasetti and Turchet [30] explored hand-held locomotion as compositional tool in a VR nautilus-shaped sound auditorium. Using Meta Quest 2 HMDs, users navigated via three modes: *continuous*, *teleportation*, and *mixed*. Sound triggering/modulation was tied to spatial position, rendered in RT via Unity and Wwise with Ambisonics algorithms. Fifteen acousmatic composers preferred the *mixed* mode for balance between flow and gesture; *continuous* favored immersive structuring while *teleportation* enabled spatial formal gestures. No major cybersickness emerged. Participants valued embodied movement as compositional device and suggested richer feedback and gesture controls. The work operationalizes the MM's performative ecology [15].

Severi et al. [31] presented *MEVO*, a modular NMP platform for telepresence and remote rehearsal. Combining smart instruments, 3D Unity visuals, and low-latency UDP audio over Raspberry Pi 4B with PREEMPT-RT Linux, MEVO supports P2P streaming with jitter buffers and Node.js control. A distributed concert between Turin and Wrocław (6 performers, 2.75h) served as evaluation. RTT remained below 59 ms; Mouth-to-Hear (M2E) latency ranged 32–61 ms; packet loss was low (<0.18%, no PLC). Musicians highlighted avatar cues, spatial feedback, and modular controls as key to presence and ensemble stability. MEVO extends the hardware optimizations of Pillar II [26] into real-world artistic scenarios, thus reflecting the inclusive ethos of MM [15].

Together, these studies demonstrate spatial reasoning, embodied control, and performer-centric design as core MM dimensions. Integration of musical affordances, RT responsiveness, and user evaluation emerge as crucial to advancing both artistic and technological fronts. Applications confirm and extend the system-level paradigms from Pillars I–II, acting as

a feedback loop between infrastructure and musical practice (Table VII).

A. An Ethical Evaluation Framework

Brusseau and Turchet [32] introduced the first ethics framework for IoMusT, merging general Artificial Intelligence (AI)/Information Technologies (IT) values with music-specific concerns. Beyond standard principles (e.g.: autonomy, dignity, fairness), it adds *originality* and *de-centralization*, structured across *individual*, *societal*, and *machine* domains (Figure 9).

The framework proposes a three-step process: *(1)* Preparation (interdisciplinary team, socio-technical narratives), *(2)* Evaluation (top-down and bottom-up), and *(3)* Narration & Recommendation (transparent articulation of ethical tensions). It was tested in a case study on vibro-tactile haptic feedback in musical performance, revealing trade-offs (e.g.: expressivity vs. control, privacy vs. inclusiveness). Bodily diversity challenged assumptions of fairness and agency.

Beyond abstract values, performance, interoperability, and explainability are treated as ethical necessities in expressive, latency-sensitive systems. This structured yet accessible framework enables designers and artists to embed ethics into MM development, operationalizing the socio-technical vision of [15].

VI. Observations & Discussion

The outcomes surveyed across the three pillars of the MM project depict a rich and layered research trajectory. Yet, when approached from a systemic perspective, several tensions and open questions emerge. This section highlights four thematic axes of reflection — each cutting across technical, artistic, and epistemological layers — in order to identify structural critical points and outline strategic directions for future development.

A. Project Visibility and Knowledge Accessibility

Despite its methodological coherence and forward-looking scope, the MM pilot project still lacks a centralized and accessible digital repository for its outputs. Although the original PRIN proposal envisioned such a portal, the only currently active website is linked to a related but separate HORIZON-funded project[8], and indexes only the outputs of that initiative (MUSMET). This fragmentation makes it harder for external researchers, educators, and practitioners to access publications, tools, and materials, limiting both dissemination and collaboration. Moreover, while this survey identifies more than a dozen peer-reviewed papers acknowledging MM funding, the actual number is likely higher — with additional works still under review, published in non-indexed venues, or pending open-access release. This underlines the need for improved traceability strategies and open-science infrastructures to ensure visibility beyond academic circles.

Another critical gap lies in participatory evaluation. While the MM project promotes inclusivity and openness, no large-scale surveys or open-form assessments have been conducted to explore public understanding of its vision or to identify user needs in a non-occasional structured way. Engagement events, such as those held at ConsAQ or SAE Milano, often remain poorly documented or accessible only through interpersonal exchanges, limiting both impact and reproducibility.

[8]https://cordis.europa.eu/project/id/101184379

B. Inclusion, Embodiment, and the MR/VR Design Tension

User studies in Pillars I and III consistently point to a clear preference for MR environments over fully immersive VR. This preference reflects not only ergonomic concerns — such as headset fatigue, limited expressivity, or unreliable motion tracking — but also deeper conceptual positions: many artists resist the de-materialization of musical gesture and favor embodied presence. MR setups allow for hybrid interaction grammars that resonate with performer intuition, educational practice, and inclusive design — particularly valuable when working with users who have sensory or cognitive diversity [9]. These preferences also relate to technical performance: many HMDs still struggle to offer stable audio I/O, precise synchronization, or low-jitter gesture detection [28], leading to a sense of mistrust or abstraction in latency-sensitive musical contexts.

Taken together, these findings suggest that further research should aim to more closely align MR/VR design choices with the full range of musicians' embodied knowledge and sensory expectations in real-world performance contexts. This alignment becomes particularly critical in scenarios where musical interaction relies on subtle physical cues, complex spatial dynamics, or shared attentional states. Additional insights from other recent studies further support these observations [33]–[35]. On the AI side, promising experiments such as *Esteso* [18] suggest new forms of co-creation between humans and machines. Yet, limitations remain: insufficient temporal awareness, overly predictable responses, and weak dynamic nuance. These constraints raise a more profound question — not only what kind of AI musicians need to improvise with, but also how to ethically structure those interactions [32]. Musical AI needs to go beyond reactivity to support genuine co-creativity.

C. Latency, Scalability, and Experimental Limitations

The network architectures discussed in Pillar II — ranging from MEC to slicing and embedded computing [20], [22] — are technically robust and well-engineered. However, most experimental validations have occurred under controlled or emulated conditions. Real-world deployments, particularly in educational or artistic institutions, remain limited.

The case study involving Starlink satellite connectivity [21] is a notable exception and highlights serious bottlenecks in latency and reliability. Even MEC-based setups, though promising in testbeds, have yet to be scaled in environments such as music conservatories — which often lack access to academic backbones (like GARR), as well as funding for advanced networking solutions. The cost and complexity of required hardware (e.g., SBCs for DSP [26], Elk LIVE [23]) further constrain scalability and adoption.

TABLE VII
SUMMARY OF APPLICATION-ORIENTED STUDIES IN PILLAR III

Study	Methodology & Scenario	Development Tools	Goals & Metrics	Participants	Key Findings
[29]	Web-based VR playgrounds; exploratory study with Meta Quest 1	A-Frame, WebAudio API, Node.js, Resonance Audio	Latency, AV feedback, accessibility	8 mixed	Positive reception of spatial audio; issues in control and sync
[30]	VR sound auditorium explored via three locomotion modes	Unity, Wwise, Ambisonics	Cybersickness, agency, expressive control	15 composers	Mixed mode favored; spatial movement as compositional gesture
[31]	Live remote concert via MEVO; telemetry and musician feedback	Unity, Raspberry Pi 4B, PREEMPT-RT, UDP stack	RTT, M2E latency, perceived stability	6 musicians	Stable use over 2.75h; spatial cues supported presence
Total participants (across studies): 29					

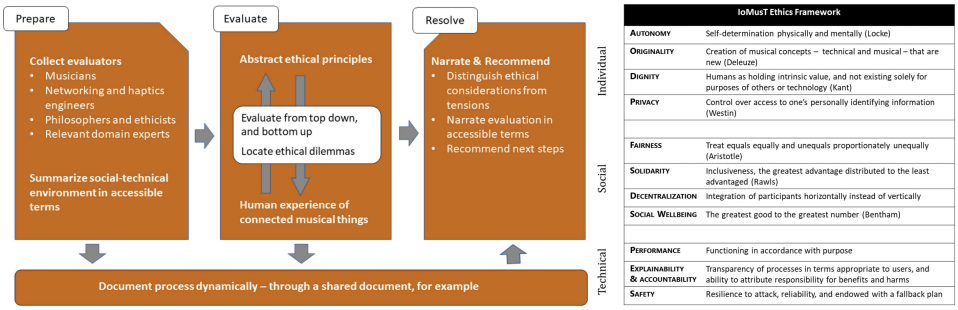

Fig. 9. The ethical framework of IoMusT proposed in [32]: applicative pipeline and principles.

Crucially, Network E2E latency alone does not capture the full UX. Instead, latency must be understood as a stack — encompassing sensor input, control logic, DSP rendering, audiovisual output, and human reaction as correctly evidenced in related works [28]. Addressing this requires a shift toward the concept of *musical latency budgets*, in which all system layers are designed with integrated timing constraints in mind.

D. Latent Potential: AI, Sustainability, and Cross-Domain Integration

Although the MM project foregrounds technological innovation, its integration of AI remains limited. This pilot project framework does not necessarily require pervasive use of AI technologies; however, several studies have highlighted the potential benefits of doing so in future iterations [15], [28]. Beyond *Esteso* [18] and the FPGA-optimized SBC [26], no current contribution leverages ML for DSP acceleration, network state prediction, or adaptive multi-modal control. This is notable in light of a growing body of research demonstrating how AI can enhance RT audio synthesis, interaction personalization, and QoE management. Within complex and expressive systems, AI agents and tools emerge not merely as a technical add-on, but as a cultural opportunity: they could offer support for fault-tolerance, interaction-aware learning, and dynamic optimization, through easy adaptation of *off-the-shelf* solutions that preliminarily bridge user needs with system-level requirements, while simultaneously opening new avenues for creative experimentation and performance optimization.

Environmental sustainability also remains under-addressed. Most systems rely on closed, energy-intensive platforms — such as Meta Quest headsets or desktop-based Unity pipelines — with little attention to power consumption, component modularity, or reuse. As XR systems become more widely adopted, especially in education, balancing immersion with ecological responsibility will become increasingly relevant.

VII. CONCLUSIONS

This brief survey examined the multidisciplinary outcomes of the *Musical Metaverse* pilot research project through its three interrelated pillars — *User-Centered Design* (UCD), *Technological Innovation*, and *Applications & Evaluation* — articulated across four cross-cutting axes: visibility and dissemination, embodied interaction and MR–VR design, latency and scalability, and latent opportunities in AI and sustainability. The project validated systems combining immersive XR, low-latency communication, and inclusive design practices. MR configurations consistently proved more expressive and ergonomic than fully immersive VR, while mixed-method evaluations revealed persistent limitations in technology responsiveness, hardware accessibility, and real-world deployment. Key challenges remain open: the absence of a centralized knowledge hub hinders reproducibility and outreach; systemic latency across sensing, networking, and rendering calls for holistic optimization; and technologies like MEC and embedded DSP are still confined to controlled scenarios. In parallel, unexploited potential in AI-driven DSP, ecological sustainability, and educational scalability represents both a challenge and an opportunity.

Strategic directions could thus be devised as follows. *Infrastructure and dissemination:* create a public MM repository aggregating publications, tools, and workshop documentation. *Latency-aware system design:* adopt musical latency budgets to guide the optimization of gesture capture, rendering, and communication pipelines. *AI integration:* leverage AI for RT

audio processing, adaptive QoE management, and co-creative improvisation frameworks. *Scalability and sustainability:* design energy-efficient XR and network systems tailored to institutional and low-budget deployment. *Critical-cultural contextualization:* engage academic institutions in music and sound arts to frame the MM activities within appropriate disciplinary and cultural contexts. Notably, several experiments rooted in electroacoustic and experimental art practices risk being overlooked unless interpreted through expert lenses.

Finally, while the MM project has piloted new methodologies and tools, its long-term impact relies on generalizing them across music, pedagogy, technology, and ethics. This means evolving from prototypes to design patterns, with dissemination, engagement, and policy alignment integrated into the research life-cycle. The MM initiative is poised to grow into a reproducible, scalable infrastructure for immersive and inclusive musical interaction — shaping future distributed ecosystems within and beyond the IoS community.

REFERENCES

[1] L. Turchet, "Musical metaverse: vision, opportunities, and challenges," *Personal Ubiquitous Comput.*, vol. 27, no. 5, p. 1811–1827, Jan. 2023.

[2] M. d'Aquin et al., "Towards an "ethics by design" methodology for ai research projects," in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, ser. AIES '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 54–59.

[3] M. Nielsen, C. Bloch, and L. Schiebinger, "Making gender diversity work for scientific discovery and innovation," *Nature Human Behaviour*, vol. 2, no. 10, pp. 726–734, 2018.

[4] C. Rottondi et al., "An overview on networked music performance technologies," *IEEE Access*, vol. 4, pp. 8823–8843, 2016.

[5] X. Jiang et al., "Low-latency networking: Where latency lurks and how to tame it," *Proceedings of the IEEE*, vol. 107, no. 2, pp. 280–306, 2019.

[6] P. Verma et al., "A deep learning approach for low-latency packet loss concealment of audio signals in networked music performance applications," in *2020 27th Conference of Open Innovations Association (FRUCT)*, 2020, pp. 268–275.

[7] A. Boem, M. Tomasetti, and L. Turchet, "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*. Zenodo, Oct. 2024, pp. 206–214.

[8] A. Boem et al., "User needs in the musical metaverse: a case study with electroacoustic musicians," in *Proceedings of the International Conference on New Interfaces for Musical Expression*. Zenodo, Oct. 2024, pp. 221–229.

[9] M. Sacchetto et al., "Collection of Design Directions for the Realization of a Visual Interface with Haptic Feedback to Convey the Notion of Sonic Grain to DHH Students," in *2023 4th International Symposium on the Internet of Sounds*. Pisa, Italy: IEEE, Oct. 2023, pp. 1–7.

[10] A. Boem and L. Turchet, "Selection as Tapping: An evaluation of 3D input techniques for timing tasks in musical Virtual Reality," *International Journal of Human-Computer Studies*, vol. 185, p. 103231, May 2024.

[11] Unity Technologies, "Unity," 2023. [Online]. Available: https://unity.com/

[12] A-Frame Community, "A-frame," 2025. [Online]. Available: https://aframe.io/

[13] PatchXR, "Patchworld," 2025. [Online]. Available: https://patchxr.com/

[14] C. Rottondi et al., "Toward an Inclusive Framework for Remote Musical Education and Practices," *IEEE Access*, vol. 12, pp. 173 836–173 849, 2024.

[15] L. Turchet, "Entangled Internet of Musical Things and People: A More-Than-Human Design Framework for Networked Musical Ecosystems," *IEEE Transactions on Technology and Society*, vol. 5, no. 4, pp. 355–367, Dec. 2024.

[16] A. Boem et al., ""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)*. Erlangen, Germany: IEEE, Sep. 2024, pp. 1–10.

[17] M. Romani et al., "BCHJam: a Brain-Computer Music Interface for Live Music Performance in Shared Mixed Reality Environments," in *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*. Erlangen, Germany: IEEE, Sep. 2024, pp. 1–9.

[18] S. Domenico et al., "Esteso: Interactive ai music duet based on player-idiosyncratic extended double bass techniques," in *Proceedings of the International Conference on New Interfaces for Musical Expression*. Zenodo, Oct. 2024, pp. 490–498.

[19] L. Turchet et al., "5G-Enabled Internet of Musical Things Architectures for Remote Immersive Musical Practices," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 4691–4709, 2024.

[20] L. Turchet and P. Casari, "Performance Analysis of Slicing on a 10-node 5G Architecture for Networked Music Performances," in *2024 IEEE Symposium on Computers and Communications (ISCC)*. Paris, France: IEEE, Jun. 2024, pp. 1–6.

[21] ——, "The Internet of Musical Things Meets Satellites: Evaluating Starlink Support for Networked Music Performances in Rural Areas," in *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*. Erlangen, Germany: IEEE, Sep. 2024, pp. 1–8.

[22] F. Martusciello et al., "Edge-enabled spatial audio service: Implementation and performance analysis on a mec 5g infrastructure," in *2023 4th International Symposium on the Internet of Sounds*, 2023, pp. 1–8.

[23] L. Turchet and C. Fischione, "Elk audio os: An open source operating system for the internet of musical things," *ACM Trans. Internet Things*, vol. 2, no. 2, Mar. 2021.

[24] L. Turchet and P. Casari, "On the Impact of 5G Slicing on an Internet of Musical Things System," *IEEE Internet of Things Journal*, vol. 11, no. 19, pp. 32 079–32 088, Oct. 2024.

[25] M. Sacchetto, C. Rottondi, and A. Bianco, "Implementation and optimization of Burg's method for real-time packet loss concealment in networked music performance applications," *Personal and Ubiquitous Computing*, vol. 28, no. 5, pp. 727–743, Oct. 2024.

[26] D. Bert et al., "FPGA-based Low-Latency Audio Coprocessor for Networked Music Performance," in *2023 4th International Symposium on the Internet of Sounds*. Pisa, Italy: IEEE, Oct. 2023, pp. 1–8.

[27] L. Turchet and S. Krstulović, "DSP as a Service: Foundations and Directions," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 6212–6226, 2024.

[28] A. Boem, M. Tomasetti, and L. Turchet, "Issues and Challenges of Audio Technologies for the Musical Metaverse," *Journal of the Audio Engineering Society*, vol. 73, no. 3, pp. 94–114, Apr. 2025.

[29] A. Boem and L. Turchet, "Musical Metaverse Playgrounds: exploring the design of shared virtual sonic experiences on web browsers," in *2023 4th International Symposium on the Internet of Sounds*. Pisa, Italy: IEEE, Oct. 2023, pp. 1–9.

[30] M. Tomasetti and L. Turchet, "Handheld controller-based locomotion in Virtual Reality as an approach to interactive music composition: insights from composers' preferences," *Digital Creativity*, vol. 35, no. 3, pp. 234–258, Jul. 2024.

[31] L. Severi et al., "Demonstration of a Networked Music Performance Experience with MEVO," Apr. 2024, arXiv:2404.09665 [cs].

[32] J. Brusseau and L. Turchet, "An Ethics Framework for the Internet of Musical Things," *IEEE Transactions on Technology and Society*, pp. 1–1, 2024.

[33] L. Turchet, N. Garau, and N. Conci, "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*, ser. AM '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 226–230. [Online]. Available: https://doi.org/10.1145/3561212.3561237

[34] A. Mancianti, "Artistic Strategies Towards a Possible Performative Approach to Embodiment in VR," in *Adjunct Publication of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video. TVX2018*. Seoul, Korea, Republic of: ACM, June 2018. [Online]. Available: https://doi.org/10.6084/m9.figshare.6526175.v1

[35] S. Dietz and C. Martin, "Luna: An ar musical instrument on the meta quest 2," in *Proceedings of the International Conference on New Interfaces for Musical Expression*. Zenodo, Jun. 2025, pp. 590–593. [Online]. Available: https://doi.org/10.5281/zenodo.15698970