Architecting the Musical Metaverse: Lessons from 5G and Emerging Technologies

Architecting the Musical Metaverse: Lessons from 5G and Emerging Technologies

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

Krishnendu S Tharakan
KTH Royal Institute of Technology
Stockholm, Sweden

Luca Turchet
University of Trento
Trento, Italy

Cristina Rottondi
Politecnico di Torino
Torino, Italy

Carlo Centofanti
University of L'Aquila
L'Aquila, Italy

Carlo Fischione
KTH Royal Institute of Technology
Stockholm, Sweden

***Abstract*—The Musical Metaverse envisions immersive, interactive environments where geographically distributed users co-create and experience music in real-time. These scenarios impose demanding constraints on communication and computation infrastructures, requiring ultra-low latency, deterministic audio delivery, and synchronized multimodal feedback. This paper presents a critical investigation into the capabilities and limitations of 5G and emerging technologies in enabling such scenarios. Building upon empirical evaluations and architectural studies, it identifies key bottlenecks in public and non-standalone 5G deployments and explores the role of private standalone infrastructures enhanced with Mobile Edge Computing. Furthermore, it assesses the potential of complementary paradigms—such as Reconfigurable Intelligent Surfaces, mmWave communications, AI-driven orchestration, and Digital Twins—for supporting scalable and expressive musical applications rooted in shared creative expression, embodied interaction, and remote co-presence. The analysis shows that current technologies, while promising, remain insufficient to fully meet the stringent requirements of real-time musical interaction. It identifies key technological gaps and outlines future directions toward intelligent, adaptive, and musically coherent infrastructures that can support the experiential and collaborative nature of the Musical Metaverse.**

***Index Terms*—Musical metaverse, 5G, Internet of musical things, quality of service.**

I. Introduction

The Musical Metaverse (MM) represents a transformative vision of how musicians, listeners, and interactive systems converge within virtual and mixed reality environments to facilitate real-time, immersive musical experiences [1]. Central to this vision is the ability to seamlessly connect geographically distributed users, allowing them to interact naturally through networked musical performance (NMP) systems [2] and immersive audio rendering technologies [3] with the addition of multimodal data streams. Such interactive scenarios impose stringent performance requirements on the underlying communication infrastructures, particularly regarding ultra-low latency, high reliability, and substantial bandwidth capabilities to support rich multimedia streams and synchronous collaboration.

In recent years, 5G technology has emerged as a critical enabler of advanced multimedia applications, promising significant enhancements in ultra-reliable low-latency communication (URLLC), network slicing, and mobile edge computing (MEC). These technological advancements position 5G as a promising candidate for supporting MM applications by meeting their unique and demanding real-time performance needs [4], [5]. Network slicing allows the creation of isolated, virtualized network instances tailored specifically to application requirements, potentially providing optimized Quality of Service (QoS) for latency-sensitive musical interactions [6]. Meanwhile, MEC facilitates computational offloading, enabling intensive audio processing tasks—such as head-related impulse response (HRIR)-based rendering, Ambisonics encoding/decoding, and real-time audio mixing—to be executed closer to end users, significantly reducing overall latency and computational burdens on client devices [7].

From a broader Metaverse perspective, technologies such as mmWave communications, Reconfigurable Intelligent Surfaces (RIS), Digital Twins (DT), and Machine Learning (ML) algorithms are particularly appealing due to their potential to enhance bandwidth availability, reduce interference and latency, augment the sense of presence, and improve the overall performance of immersive scenarios [8], [9], [10].

Despite these promising features, current deployments and preliminary studies indicate that the practical applications of these technologies do not still achieve desired results, and especially the use of 5G in MM scenarios faces significant limitations. Empirical assessments have highlighted persistent latency spikes, jitter inconsistencies, incomplete slicing implementations, and limitations related to MEC deployment, particularly in non-ideal radio conditions [11], [12]. Addition-

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. This work was also 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" CUP F83C22001690001.

ally, recent analyses of Extended Reality (XR)—encompassing augmented reality (AR), virtual reality (VR), and mixed reality (MR)—underscore further complexities. XR scenarios introduce highly demanding requirements regarding data throughput, ultra-low latency, synchronization accuracy, and strict energy constraints, adding further challenges to effective integration with existing 5G infrastructures [13].

This paper critically examines the state-of-the-art in enabling technologies for the MM, with a particular focus on their capacity to meet the unique demands of musical interactivity. We analyze experimental evidence, architectural models, and standardization efforts related to 5G and complementary innovations such as RIS, mmWave communications, AI-driven orchestration, and DTs. Our aim is to identify key gaps and inform future research directions toward robust, scalable infrastructures for musical extended realities.

II. Background and Related Work

When the fifth generation (5G) wireless communication standard emerged, researchers faced the critical challenge of identifying "killer applications" that would justify the necessity of this advanced technology [14]. Over time, numerous compelling application scenarios have been proposed, along sectors as mobility, industry, smart cities, healthcare [15], as well as diverse multimedia-related applications such as DTs and Multimedia-Internet of Things (Multimedia-IoT). Within Multimedia-IoT, a distinct differentiation arises based on the medium involved, and when sound is specifically considered, the Internet of Sounds (IoS) domain becomes relevant [16]. Further specialization within the IoS leads to the Internet of Musical Things (IoMusT) [17], a technological framework characterized by interconnected musical devices, smart instruments, wearables, and systems designed explicitly for musical applications. Building upon the IoMusT infrastructure, the concept of the MM emerges as a specialized and immersive digital application domain [1]. The MM envisions virtual and mixed reality environments leveraging XR technologies to create deeply immersive and interactive musical experiences. By integrating IoMusT-supported musical devices within these advanced virtual environments, the MM significantly expands the scope of musical engagement, offering novel interaction paradigms, economic models, and creative expressions. Thus, the MM can be viewed as a sophisticated, future-oriented application domain deeply embedded in, and reliant upon, the technological foundations provided by the IoMusT. A recent survey further expands this framework by exploring current implementations of the MM, emphasizing immersive spatial audio, virtual venues, embodied interaction, and collaborative XR environments as foundational elements for future applications [18]. These developments consolidate MM as a transdisciplinary domain at the intersection of audio engineering, HCI, and networked digital art.

In recent years, extensive research has addressed various aspects of these domains, spanning technological challenges, immersive interaction techniques, new business models, ethical considerations, and applications such as NMP. The latter gained additional attention due to the increased need for remote musical collaboration, particularly emphasized during the COVID-19 pandemic [19], [20]. Alongside these developments, a significant portion of the literature has examined the role of 5G architectures in enabling real-time, interactive musical systems [12]. Two key technological enablers frequently analyzed in this context are *network slicing* and MEC, which have been evaluated for their potential to meet the stringent performance requirements of the MM [6], [7].

Network slicing refers to the creation of virtualized, application-specific partitions of the 5G infrastructure, each configured with tailored QoS parameters. In the context of MM, slicing enables the isolation and prioritization of latency-sensitive streams, such as high-resolution audio, haptic feedback, and gesture data [6]. MEC, instead, brings computational resources closer to end users by deploying processing capabilities at the network edge. This allows latency-critical tasks, such as spatial audio rendering, real-time signal processing, or multimodal synchronization to be offloaded from user devices, reducing round-trip delays and enabling more responsive interaction in immersive music scenarios [7]. Beyond real-time offloading, recent architectural work introduced the concept of *Digital Signal Processing as a Service* (DSPaaS), which envisions modular and orchestrable audio processing chains executed across MEC and cloud infrastructures [21]. This approach allows for scalable deployment and adaptive configuration of audio services, reinforcing the role of MEC not only as a latency-reducing layer but also as a service management plane tailored to musical interaction contexts.

III. Technical Requirements for MM Applications

While the concept of the Metaverse is still evolving, recent surveys have identified common technical demands underlying immersive and interactive applications across a range of use cases. Among these, studies emphasize real-time interaction, ultra-low latency, high-resolution rendering, multimodal data integration, and scalable connectivity as foundational enablers of Metaverse experiences [22]–[24]. In particular, latency is consistently highlighted as a core limiting factor: excessive end-to-end delays degrade user immersion and induce motion sickness, especially in XR contexts [19], [25]. Synchronization across heterogeneous devices and the ability to seamlessly offload computational tasks to edge resources are likewise critical to sustain responsive, high-fidelity environments [26].

In addition to these general constraints, the MM vision introduces domain-specific requirements rooted in the psychoacoustic sensitivity of musical interaction and the real-time coordination of distributed users. Notably, MM scenarios are inherently multi-user and involve simultaneous participation of musicians, audiences, and virtual agents within shared XR environments. This imposes stringent demands on system scalability: infrastructures must maintain latency guarantees, synchronization accuracy, and Quality of Service (QoS) even as the number of concurrent users increases. Supporting such scalability entails robust orchestration mechanisms, dynamic

resource allocation, and predictable performance under varying network loads—especially in shared MEC or slicing-based deployments.

These constraints become particularly severe in real-time musical collaboration—whether in NMP, mixed-reality jams, or shared spatial audio experiences—which requires precise temporal alignment and deterministic audio transmission. Experimental studies on 5G-enabled NMP systems report that end-to-end latencies must consistently remain below 30 ms, with jitter constrained under 10 ms to maintain perceptual coherence [7], [11]. Packet loss must be minimized, as burst errors can irreparably disrupt musical timing. Throughput demands increase significantly in immersive settings where multi-channel, high-resolution audio and positional metadata are transmitted continuously.

Moreover, the MM envisions real-time rendering of spatial audio based on user movement and head tracking. This implies additional computation, often offloaded to MEC-enabled architectures. Such workloads must be served within ultra-reliable and low-latency budgets to ensure perceptual stability in binaural or Ambisonic scenes [3], [7]. Beyond audio transport, the infrastructure must support feedback loops for gesture, haptic cues, and avatar synchronization, adding further constraints on uplink/downlink symmetry and consistent bandwidth availability.

To technically support MM applications effectively, a communication infrastructure must satisfy several interdependent requirements:

- Latency and jitter: end-to-end audio latency must remain consistently below 30 ms, with jitter under 10 ms. These thresholds ensure temporal alignment and prevent rhythmical desynchronization during interactive performances [19].
- Packet loss and reliability: the system must maintain extremely low packet loss rates, with mechanisms to mitigate burst losses (e.g., real-time packet loss concealment), as such disruptions are especially harmful to musical coherence [11], [27].
- Throughput and resolution: high-throughput links are required to accommodate multi-channel, high-resolution audio streams and spatial metadata for immersive rendering [8], [11].
- Bidirectional interactivity: uplink and downlink capacities must be balanced to enable not just audio transport, but also real-time multimodal feedback (e.g., gesture tracking, head movement, haptics), which are integral to XR-based music systems [19].
- Computational offloading: MEC integration is essential for offloading tasks such as spatial audio rendering and real-time signal processing, preserving performance on lightweight user devices, and can be extended through modular orchestration of DSP pipelines, as proposed in recent DSPaaS paradigms [7], [21].
- Synchronization and clock precision: accurate timing synchronization across distributed clients is critical for maintaining phase coherence and alignment of audio-visual events [11], [19].
- Scalability and concurrent orchestration: MM infrastructures must handle simultaneous participation of multiple users interacting in shared XR sessions. This entails horizontal scalability of MEC resources, dynamic orchestration of audio and sensor flows, and QoS preservation under increasing system load [28].

Given these constraints, 5G technology, with its support for URLLC, is theoretically well-suited to enable MM applications. However, as the following sections demonstrate, practical deployments reveal a range of limitations and trade-offs that must be critically evaluated.

This paper builds upon both general Metaverse surveys and domain-specific experiments to define a refined set of technical thresholds against which current 5G deployments are critically assessed.

IV. Enabling technologies for the MM

Meeting the stringent requirements of the MM involves the interplay of multiple enabling technologies across networking, computation, orchestration, and immersive content modeling. While 5G architectures provide foundational support—such as URLLC, network programmability, and MEC integration—additional paradigms are critical to address the broader demands of MM applications.

In particular, beyond ensuring individual interaction fidelity, MM infrastructures must scale to support concurrent users and distributed sessions within shared XR environments. This requires horizontally scalable systems that preserve latency guarantees, interactivity, and perceptual coherence across multiple geographically dispersed participants.

This section reviews the main technological enablers underpinning MM scenarios, highlighting both their theoretical capabilities and their applicability to real-time, expressive, and scalable musical interactions.

Ultra-Reliable Low-Latency Communications (URLLC): URLLC is a core 5G capability designed to provide extremely low latency and high reliability, critical for MM applications requiring instantaneous responses and minimal packet loss [8].

Network Slicing: Network slicing allows the creation of isolated, virtualized network instances (slices) tailored specifically to the unique requirements of different applications. Each slice can be configured with optimized QoS parameters, such as guaranteed bandwidth, low latency, and high reliability [29], [30].

For the MM, network slicing can provide dedicated, optimized network resources for latency-sensitive musical interactions. This ensures that high-resolution audio streams, positional metadata, and real-time feedback (e.g., gesture tracking, head movement, haptics) receive the necessary priority and performance, even in congested network environments. This isolation helps in mitigating unpredictable jitter and providing consistent low-latency paths, which are often limitations in public Non-Standalone (NSA) 5G infrastructures [31].

Private 5G Networks: Private 5G networks, also referred to as non-public networks (NPNs), represent a compelling solution for MM deployments requiring stringent latency, reliability, and orchestration guarantees. These networks are owned and operated by individual organizations, such as concert venues, music academies, or XR performance spaces, granting full control over network configuration, traffic prioritization, and access to on-premise edge resources.

Unlike public NSA architectures, which suffer from shared infrastructure limitations and variable performance, private Standalone (SA) 5G deployments can natively support features like dedicated slices and local MEC integration. This allows NMP systems to maintain sub-30 ms end-to-end latencies and sub-1% packet loss even under complex bidirectional interactions [12]. Moreover, private networks facilitate rapid deployment of context-specific optimizations, such as local orchestration policies, QoS tuning for spatial audio, and secure federation across multi-site artistic ecosystems.

Mobile Edge Computing (MEC): MEC facilitates computational offloading by bringing computing resources closer to the end-users, at the edge of the mobile network. This significantly reduces the round-trip time for data processing [30].

MM applications often involve computationally intensive tasks such as spatial audio rendering (e.g. HRIR-based rendering, Ambisonics encoding/decoding) and real-time audio mixing [7]. Offloading these tasks to MEC servers alleviates the computational burden on client devices (like VR headsets or smartphones) while maintaining immersive audio quality. This allows to preserve performance on lightweight user devices and ensuring perceptual stability in binaural or Ambisonic scenes within ultra-reliable and low-latency budgets [3]. MEC also supports bidirectional interactivity by enabling rapid processing of feedback loops for gesture, haptic cues, and avatar synchronization [32].

Moreover, by distributing workloads across geographically distributed MEC nodes, such architectures inherently support horizontal scalability, enabling concurrent participation of multiple users without overloading central servers or introducing prohibitive latencies. This is particularly relevant in MM scenarios involving shared XR spaces or synchronous multi-performer sessions, where maintaining perceptual consistency and interactivity at scale is critical.

These immersive applications also benefit from standardized APIs that ensure compatibility across platforms and devices. The WebXR Device API defines a unified interface for accessing XR functionalities within web browsers, supporting multimodal rendering and sensor data integration [33].

Reconfigurable Intelligent Surfaces (RIS) and mmWave Extensions: RIS are engineered metasurfaces capable of dynamically manipulating electromagnetic waves to control propagation environments. By adjusting the phase and amplitude of reflected signals in real time, RIS can reshape wireless channels to improve coverage, extend Line-of-Sight (LoS) conditions, and mitigate interference or blockage effects [34]. Similarly, millimeter-wave (mmWave) communication leverages higher frequency bands (30–300 GHz) to achieve multi-gigabit throughput and low-latency transmission, albeit with increased sensitivity to obstacles and signal degradation [35].

In MM scenarios characterized by dynamic user movement, dense urban layouts, or indoor environments with significant obstructions, RIS-assisted mmWave links can enhance the stability and reliability of audio-visual streams. This is particularly relevant in virtual performance halls or networked rehearsal studios, where preserving LoS and ensuring ultra-low-latency are critical for real-time spatial audio streaming and synchronized interactions. RIS-enabled architectures can dynamically adapt to user topology and environmental changes, helping to maintain phase coherence and reduce jitter in immersive audio experiences.

Joint Optimization and Edge-Orchestrated Intelligence: Joint optimization frameworks aim to concurrently manage communication and computation resources to minimize service latency and maximize quality of experience. These approaches typically rely on AI-driven controllers, such as reinforcement learning agents or meta-learning algorithms that adjust system parameters in real time. Techniques like alternating optimization and model-based orchestration, including Multi-Objective Soft Actor-Critic (MO-SAC) strategies, enable dynamic adaptation of offloading decisions, edge caching policies, and transmission configurations according to contextual changes [8], [36]. Within MM ecosystems, such intelligent orchestration mechanisms can sustain real-time constraints under variable network and device conditions. For example, adaptive scheduling of spatial audio rendering workloads and dynamic control of synchronization pipelines can be achieved by continuously monitoring user topology, computational demand, and latency budgets.

A particularly relevant abstraction in this context is DSPaaS, which encapsulates audio-specific operations, such as reverberation, spatialization, mixing, or transformation, into deployable, remotely callable processing units. Rather than relying on monolithic client-side pipelines, DSPaaS enables modular distribution of real-time audio tasks across edge and cloud infrastructures, supporting low-latency interaction while simplifying orchestration. In MM scenarios, this model promotes shared access to standardized audio engines, dynamic reconfiguration of musical signal paths, and context-aware instantiation of processing modules. Moreover, DSPaaS decouples signal processing functionality from hardware constraints, facilitating deployment across heterogeneous devices and fostering reproducibility in collaborative performance environments [21].

Machine Learning Strategies: One of the main challenges in the MM is dealing with the variety of data types that are exchanged between users. These data types each have their own timing and bandwidth needs. For example, audio must be transmitted with very little delay, video requires a large amount of bandwidth, and EEG signals must be kept accurate to make meaningful interpretations. Sending all of this data over a single 5G slice is very demanding. Issues like uneven delays (i.e., jitter), lost packets, and poor synchronization between data types can affect the overall experience. ML can help address these problems by using techniques like deep learning-

based Packet Loss Concealment (PLC) algorithms, which allow systems to recreate missing audio or video segments [10], [27], [37]. It can also help prioritize more important data during network congestion, ensuring that core elements of the performance are preserved.

In a typical MM system, the architecture comprises two interconnected layers: the virtual musical space, which simulates performances, venues, and interactions in immersive digital environments, and the physical musical space, where real-world musicians, instruments, and sensors operate. These two layers continuously interact to optimize performance delivery, synchronize musical actions, and support user engagement. Within this system, three key operations are essential: intelligent learning, real-time sensing, and user service fulfillment [38]. In the virtual space, dynamic models are required to instantly adapt to user's actions, such as live improvisation or collaborative rehearsals. To support such responsiveness, federated learning (FL) offers a promising solution, especially when compared to centralized ML. FL enables on-device training across user instruments or wearables and aggregates only the learned models into the virtual environment, preserving user privacy and minimizing communication load.

Digital Twins: In the MM, DTs serve as virtual counterparts of real-world musical entities such as instruments, performers, concert venues, or entire ensembles. These digital replicas operate in continuous synchronization with their physical counterparts, enabling real-time monitoring, control, and adaptation of musical processes and interactions. Through integration with IoMusT devices, any change in the physical musical environment like a performer's gesture, instrument tuning, or audience movement is reflected almost instantly in its DT within the virtual space. For example, a DT of a violin can simulate bowing techniques or sound pressure changes based on real-time sensor data, offering immersive training environments or enabling virtual collaboration across distances.

While originally developed within Cyber-Physical Systems (CPS) to provide deterministic models of physical assets, DTs face intrinsic limitations when applied to socio-musical domains, where human behavior, expressivity, and emergent interactions are central [9]. To address this, MM scenarios require an evolution of DTs beyond passive mirroring, toward adaptive representations capable of capturing and responding to artistic intent.

Thus advanced DTs also incorporate deep learning models to learn from sensor data (e.g., audio, motion, EEG), enabling features such as gesture interpretation, performance analysis, and intelligent control of virtual instruments and environments.

Figure 1 schematizes the possible contributions of all the actors discussed above.

V. Empirical Evaluations of Network Architectures and Audio Processing Strategies for MM

This section consolidates key experimental findings from recent literature that address technological enablers of the MM, spanning both network infrastructures and audio signal processing techniques. The analysis is organized into two parts: first, evaluations focused on MM-specific scenarios; second, broader investigations that inform optimization strategies, continuity mechanisms, and distributed orchestration.

A. Evaluations in MM Scenarios

Turchet et al. [6] evaluated the impact of network slicing in a private SA 5G setup involving a 10-node audio-only NMP system. The study reported a marginal latency increase (from 23.95 ms to 24.24 ms) due to MEC overhead, but achieved a 20% reduction in packet loss and burst errors, validating the benefits of flow isolation without compromising timing stability.

In [12], a comparative assessment between private SA and public NSA architectures for audio-only NMP applications revealed that only SA deployments consistently met NMP latency requirements (latency < 23 ms, packet loss $< 1\%$), while NSA networks experienced severe instability, including latency peaks over 800 ms and loss bursts exceeding 3000 packets.

Martusciello et al. [7] experimentally validated MEC-enabled spatial audio rendering using two binaural techniques—HRIR convolution and virtual loudspeakers—offloaded to edge servers. Subjective evaluations and objective measurements confirmed stable perceptual quality, with delays under 50 ms in both indoor (47 ms) and outdoor (44 ms) environments, reinforcing the role of edge processing for immersive audio delivery.

Turchet and Casari [39] evaluated the feasibility of using Starlink's satellite infrastructure for audio-only NMP in rural areas. Two scenarios were tested—Starlink-to-Starlink and hybrid satellite-wired—using Elk LIVE devices for continuous bidirectional audio transmission. Analysis of over 450,000 packets revealed high variability in latency (mean 168 ms, peaks over 3 seconds) and large burst losses (up to 2835 packets), confirming that current LEO satellite configurations do not meet MM performance targets. However, the study provides a critical baseline for exploring Non-Terrestrial Networks (NTNs) as a future avenue for inclusive and geographically distributed music interaction, paving the way for the design of 6G architectures.

Turchet et al. [11] further analyzed the component-wise latency contributions in a 5G-enabled IoMusT system. Their findings quantify the deterministic delay of the audio processing chain (14.32 ms, including analog-to-digital conversion, DSP, packetization, jitter buffer), and highlight that only 15.68 ms remains available for network transport. These values offer a concrete budget for MM system designers and reinforce the importance of minimizing transmission and buffering delays.

B. Architectural Optimization and Audio Continuity Strategies

Beyond MM-specific deployments, broader research highlights key technologies for optimizing network architectures and enhancing audio continuity. Huynh et al. [36] proposed a

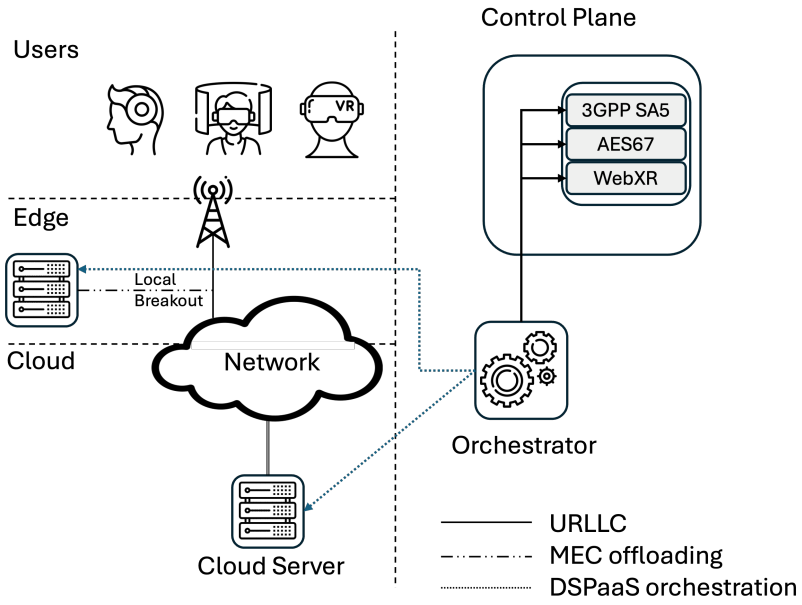

Fig. 1. Functional stack of enabling technologies for the Musical Metaverse, highlighting the layered orchestration of communication, computation, intelligence, and immersive interaction.

DT architecture leveraging MEC and URLLC. Their model integrates joint optimization of communication and computation, showing that task caching at the edge can reduce end-to-end latency by approximately 10 ms under constrained energy and caching conditions.

Gao et al. [8] introduced a multi-objective optimization framework that combines RIS, mmWave/THz communication, and intelligent orchestration via reinforcement learning. The system achieved significant improvements in latency, resource allocation, and user experience for immersive services, establishing a strong foundation for future MM infrastructure.

Turchet and Krstulović [21] proposed and experimentally validated the DSPaaS paradigm using a testbed based on Elk LIVE devices, Raspberry Pi platforms, and 5G SA connectivity. They compared three orchestration strategies—asynchronous (cloud-only), reactive (edge-triggered), and proactive low-latency (edge+cloud coordinated). Results indicated that while cloud-based DSP suffers from high latency and instability, hybrid strategies can meet MM constraints under 30 ms with improved resource efficiency. The evaluation also revealed that sequential packet loss patterns have a stronger perceptual impact than random ones. This calls for more expressive metrics when evaluating continuity in musical contexts as recently highlighted in [40].

At the signal processing level, Verma et al. [27] proposed a deep learning-based PLC model that outperformed classical autoregressive methods. The model operated 10x faster than real time on GPU but was 5x slower on CPU, highlighting the need for hardware acceleration in real-time deployments. Sacchetto et al. [37] focused on low-complexity AR models, achieving superior Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) metrics compared to traditional PLC techniques. Listening tests confirmed improved perceptual fidelity, especially for solo instrument contexts. Mezza et al. [10] proposed PARCnet, a hybrid AR+DL architecture trained on the MAESTRO dataset. It achieved state-of-the-art performance in both objective (NMSE, PEAQ, PLCMOS) and subjective (MUSHRA [41]) evaluations, while maintaining real-time processing capability on CPU, making it a viable candidate for embedded or mobile MM applications.

Taken together, these results demonstrate the trade-offs between latency, reliability, and perceptual fidelity in audio-only NMP as well as MM systems, and underscore the importance of aligning infrastructure choices with the specific performance constraints of expressive, time-sensitive musical interaction.

A summary of the outcomes from the discussed papers is reported in table I.

VI. Limitations and Technical Bottlenecks of Current Technologies for MM

Despite promising advances in networking, edge computing, and audio signal processing, several structural and operational limitations hinder the deployment of MM applications at scale. These issues span multiple layers, from infrastructure to signal-level processing and orchestration, and emerge clearly in recent experimental studies.

NSA Deployments and Limited Edge Coverage: Public NSA architectures, often still reliant on legacy 4G cores, consistently fail to meet MM requirements for latency, jitter, and uplink stability. Empirical data report latency peaks exceeding 800 ms and high burst loss under congested or mobile conditions [12]. Even when edge nodes are theoretically available,

TABLE I
SUMMARY OF KEY RESULTS ON INFRASTRUCTURE AND AUDIO PROCESSING STRATEGIES FOR MM APPLICATIONS

Study	Scenario	Findings	Implications
Turchet et al. [6]	10-node NMP with private SA 5G and slicing	Slight latency increase (+0.29 ms); packet loss and long burst errors reduced by 20%	Slicing improves flow isolation and reliability without affecting timing stability
Turchet et al. [12]	Private SA vs. public NSA 5G in NMP	SA: stable <23 ms latency and <1% packet loss; NSA: >800 ms spikes, high burst loss	SA + MEC is required for real-time IoMusT; NSA lacks reliability due to WAN and missing edge support
Martusciello et al. [7]	MEC-enabled 5G with spatial audio	Delay: 44–47 ms; under perceptual threshold	MEC offloading is viable for immersive spatial audio services
Turchet et al. [3]	P2P vs. cloud spatial audio streaming	P2P: <30 ms; Cloud: >60 ms	P2P 5G better supports low-latency interaction than centralized models
Turchet and Casari [39]	NMP via Starlink (LEO satellite)	Mean latency: 168 ms; up to 2835 packet loss burst	Current satellite NTN not suitable for MM; foundational for inclusive rural scenarios
Turchet et al. [11]	Latency breakdown in 5G-enabled IoMusT system	Audio chain delay: 14.32 ms; network budget: 15.68 ms	Highlights hard latency constraints for real-time MM performance
Turchet and Krstulović [21]	DSPaaS testbed (Elk LIVE, Raspberry Pi, 5G SA)	Cloud: unstable latency; hybrid edge-cloud: <30 ms, high efficiency	DSPaaS viable for modular MM signal processing; orchestration affects performance
Huynh et al. [36]	DTs with MEC + URLLC	Edge caching reduced e2e latency by 10 ms	Optimized MEC-based DTs reduce latency in metaverse services
Gao et al. [8]	RIS + mmWave/THz for URLLC	Improved latency/cost and localization	Advanced tech enables scalable, high-performance immersive environments
Verma et al. [27]	DL-based PLC for NMP	GPU: 10× faster than real time; CPU: 5× slower	DL-PLC improves audio recovery; GPU acceleration recommended
Sacchetto et al. [37]	Auto-regressive model for PLC	Lower error, better perceptual quality for solos	Auto-regressive-based PLC offers efficient real-time concealment
Mezza et al. [10]	PARCnet (AR+DL hybrid)	Real-time on CPU; best perceptual scores	Hybrid PLC is accurate and efficient for musical content

lack of orchestration or integration prevents effective workload distribution, undermining the potential of MEC.

Inconsistent QoS and Slicing Overhead: While slicing offers strong theoretical benefits, its practical deployment introduces signaling overhead and exposes inconsistencies across infrastructure providers. Studies show that isolated slices may not guarantee end-to-end QoS unless managed within private SA contexts [6]. Moreover, dynamic service chaining is rarely supported in public deployments, limiting the feasibility of adaptive MM workflows [21].

DSPaaS Limitations on Commodity Hardware: Recent evaluations of DSPaaS architectures [21] reveal performance trade-offs when running low-latency audio effects on embedded devices. Models such as reverberation, dynamic range control, or spatialization require DSPs or GPU acceleration to meet sub-30 ms budgets. When deployed on CPU-only platforms (e.g., Raspberry Pi), added delays due to codec latency (e.g., Opus $\sim$26.5 ms) and network jitter often break MM constraints. Furthermore, Audio over IP protocols like AES67, [42], or Dante, [43], are not natively compatible with unmanaged networks, making professional-grade routing and timing difficult to replicate outside controlled environments.

Packet Loss and Perceptual Degradation: MM systems remain highly sensitive to burst losses, especially in bidirectional audio exchanges. Despite improvements from hybrid PLC methods [10], perceptual degradation remains non-negligible when loss patterns are temporally correlated. Evaluations stress that traditional metrics (MAE, RMSE) may underestimate musical degradation, calling for more reliable objective metrics [40].

Clock Drift and Multimodal Desynchronization: XR-based musical scenarios require strict time alignment between audio, visual, and gestural data. Yet clock drift and asymmetric latencies across transmission channels persist as unresolved challenges. In consumer-grade deployments, timestamping inconsistencies and lack of synchronized protocols limit the fidelity of joint interactions, particularly under mobile or low-bandwidth conditions [7], [44].

Uplink Bottlenecks in Interactive Scenarios: MM interactions involve not only playback but continuous upstream feedback (audio, gestures, haptics). Public networks, however, often prioritize downlink bandwidth, making the uplink a bottleneck. This imbalance particularly affects improvisation, avatar synchronization, and haptic feedback, which require stable upstream throughput [11].

Non Terrestrial Networks (NTN) Latency and Stability Issues: Evaluations of Low Earth Orbit (LEO) satellite networks like Starlink show that, despite theoretical latencies under

60 ms, real-world conditions exhibit average delays of 168 ms and packet bursts exceeding 2800 lost units [39]. Contributing factors include beam switching, environmental interference, and limited uplink shaping. While promising for inclusive coverage, NTNs in their current form remain unsuitable for interactive musical scenarios.

DTs and Behavioral Adaptivity: Most DT frameworks deployed in MM are passive mirrors of the physical domain. Their inability to adapt to user behavior or musical context limits their use in real-time co-creative environments. Efforts to integrate behavior-adaptive ML models are ongoing [21], but practical deployments remain rare and often lack sufficient temporal resolution or cross-modal integration.

Cross-Domain Fragmentation: Finally, the MM landscape suffers from a lack of interoperable standards bridging audio signal processing, XR rendering, network control, and user-level semantics. This fragmentation hampers integration across system layers, leading to brittle prototypes and non-replicable deployments [1], [21], [39]. Without unified APIs or interface layers, the scalability of MM remains confined to bespoke or tightly controlled setups.

VII. Discussions and Future Directions

As our analysis has shown, current infrastructures and protocols still face several limitations when applied to latency-sensitive and interaction-rich applications such as the MM. To move toward more robust, scalable, and inclusive deployments, we outline key directions for future technological research and development.

Towards 6G-Enabled MM Architectures: The evolution from 5G to 6G is expected to introduce paradigm shifts across communication, computation, and sensing. Visionary works foresee an integrated network ecosystem based on ubiquitous wireless intelligence, where sensing, communication, and computing are co-designed to support applications with ultra-low latency, semantic awareness, and spatial context-awareness [28], [45]. Such capabilities align with the real-time synchronization and QoS constraints of MM scenarios, particularly when enhanced by emerging paradigms such as joint communication and sensing and holographic-type services.

Edge-Centric AI and Lightweight Inference: Future MM systems will increasingly benefit from edge-deployed AI models capable of handling dynamic, multimodal data with strict latency and privacy requirements. Recent evaluations show that commercial edge devices (e.g., Jetson Orin, EdgeTPU) can sustain real-time AI inference while preserving energy budgets and ensuring data locality [46]. However, the trade-off between model complexity and responsiveness remains a challenge, especially in decentralized musical environments with multiple interaction points. Research must further explore scalable model compression, decentralized orchestration, and context-aware adaptation.

Global Accessibility through STIN Architectures: Ensuring equitable access to MM experiences requires overcoming terrestrial infrastructure constraints, especially in rural or underdeveloped regions. Satellite-Terrestrial Integrated Networks (STIN) offer a promising solution for reliable, differentiated service provisioning across heterogeneous environments [47]. These architectures can potentially support scalable deployments of distributed musical XR applications with adaptive capacity control and game-theoretic resource allocation, however these avenues have not yet been explored.

Reinforced Privacy and Security for Immersive Environments: As MM systems increasingly rely on AI-driven interactions, real-time telemetry, and personal data streams, privacy and security become central enablers rather than auxiliary concerns. The evolution toward endogenous network security, with cryptographic and trust mechanisms in the communication stack, must be prioritized. Efforts should focus on protecting XR devices from protocol-level attacks, enabling user-centric identity management, and enforcing low-latency privacy-preserving inference [28].

Interoperability and Standardization: Finally, realizing the potential of MM ecosystems depends on interoperability across platforms, devices, and network layers. Fragmentation across metaverse applications hinders cross-domain collaboration and collective intelligence. Future directions should prioritize open standards and reference architectures for the integration of audio-visual content, orchestration APIs, and distributed computation, following the principles proposed for the Open Metaverse [48].

Toward Multimodal QoE Evaluation Frameworks: A further direction involves the development of integrated multimodal Quality of Experience (QoE) evaluation frameworks. While current assessments typically address each modality in isolation (for example, using MUSHRA for audio or latency-based metrics for visual rendering) these approaches fail to capture the perceptual interplay between sensory channels that characterizes immersive MM scenarios. Future work should explore composite methodologies that combine auditory fidelity evaluation, questionnaires targeting cybersickness and visual comfort (such as the SSQ), and qualitative user studies on immersion and co-presence. Such frameworks would enable user-centered validation of system performance and support iterative design cycles for MM environments.

VIII. Conclusion

This work presented a comprehensive investigation into the technological foundations and limitations of current communication and computation infrastructures for enabling the MM. By conducting a multi-layered analysis encompassing technical requirements, enabling paradigms, empirical validations, and observed constraints, we identified critical gaps between the performance targets of interactive musical applications and the capabilities of state-of-the-art 5G systems.

Our review demonstrates that only private SA 5G deployments, augmented with MEC and intelligent orchestration mechanisms, consistently satisfy the stringent constraints imposed by MM use cases, including sub-30 ms end-to-end latency, sub-millisecond synchronization accuracy, and robust support for bidirectional multimodal interaction. In parallel, signal-level strategies such as DSPaa and hybrid PLC models

exhibit the potential to address expressivity and continuity requirements at the application layer.

Nonetheless, limitations remain evident across infrastructure availability, protocol integration, uplink asymmetry, satellite-induced instability, and the absence of cross-domain interoperability frameworks. These bottlenecks suggest that incremental refinement of 5G deployments alone will be insufficient to support scalable and inclusive MM ecosystems.

Future research should converge toward 6G native architectures that feature integrated sensing, communication and computation; edge-resident AI for context-aware adaptation; satellite-terrestrial convergence for global reach; and standardized orchestration protocols to ensure semantic interoperability across platforms. Such advancements are essential not only to fulfill the real-time performance needs of MM applications but also to foster sustainable, culturally inclusive, and co-creative digital environments rooted in musical interaction.

ACKNOWLEDGMENT

We acknowledge the support of the MUSMET project funded by the EIC Pathfinder Open scheme of the European Commission (grant agreement n. 101184379). 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.

REFERENCES

[1] L. Turchet, "Musical Metaverse: vision, opportunities, and challenges," *Personal and Ubiquitous Computing*, vol. 27, no. 5, pp. 1811–1827, Oct. 2023.

[2] C. Rottondi, C. Chafe, C. Allocchio, and A. Sarti, "An overview on networked music performance technologies," *IEEE Access*, vol. 4, pp. 8823–8843, 2016.

[3] L. Turchet, C. Rinaldi, C. Centofanti, L. Vignati, and C. Rottondi, "5G-Enabled Internet of Musical Things Architectures for Remote Immersive Musical Practices," *IEEE Open Journal of the Communications Society*, pp. 1–1, 2024, conference Name: IEEE Open Journal of the Communications Society.

[4] I. Akyildiz and H. Guo, "WIRELESS COMMUNICATION RESEARCH CHALLENGES," vol. 3, pp. ITU Journal on Future and Evolving Technologies (ITU J–FET), 04 2022.

[5] M. Gapeyenko, V. Petrov, S. Paris, A. Marcano, and K. I. Pedersen, "Standardization of Extended Reality (XR) over 5G and 5G-Advanced 3GPP New Radio," *IEEE Network*, vol. 37, no. 4, pp. 22–28, Jul. 2023.

[6] 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.

[7] F. Martusciello, C. Centofanti, C. Rinaldi, and A. Marotta, "Edge-Enabled Spatial Audio Service: Implementation and Performance Analysis on a MEC 5G Infrastructure," in *2023 4th International Symposium on the Internet of Sounds*. Pisa, Italy: IEEE, Oct. 2023, pp. 1–8.

[8] X. Gao, W. Yi, Y. Liu, and L. Hanzo, "Multi-objective optimization of urllc-based metaverse services," *IEEE Transactions on Communications*, vol. 71, no. 11, pp. 6745–6761, 2023.

[9] X. Wang, J. Yang, J. Han, W. Wang, and F.-Y. Wang, "Metaverses and demetaverses: From digital twins in cps to parallel intelligence in cpss," *IEEE Intelligent Systems*, vol. 37, no. 4, pp. 97–102, 2022.

[10] A. I. Mezza, M. Amerena, A. Bernardini, and A. Sarti, "Hybrid packet loss concealment for real-time networked music applications," *IEEE Open Journal of Signal Processing*, vol. 5, pp. 266–273, 2024.

[11] L. Turchet and P. Casari, "Latency and Reliability Analysis of a 5G-Enabled Internet of Musical Things System," *IEEE Internet of Things Journal*, vol. 11, no. 1, pp. 1228–1240, Jan. 2024.

[12] ——, "Assessing a Private 5G SA and a Public 5G NSA Architecture for Networked Music Performances," in *2023 4th International Symposium on the Internet of Sounds*. Pisa, Italy: IEEE, Oct. 2023, pp. 1–6.

[13] J. K. Sundararajan, H.-J. Kwon, O. Awoniyi-Oteri, Y. Kim, C.-P. Li, J. Damnjanovic, S. Zhou, R. Ma, Y. Tokgoz, P. Hande, T. Luo, K. Mukkavilli, and T. Ji, "Performance evaluation of extended reality applications in 5g nr system," in *2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2021, pp. 1–7.

[14] M. Massaro and S. Kim, "Why is south korea at the forefront of 5g? insights from technology systems theory," *Telecommunications Policy*, vol. 46, no. 5, p. 102290, 2022.

[15] E. Commission, "5G Observatory Quarterly Report 8," Online, July 2020, available: https://5gobservatory.eu/wp-content/uploads/2020/07/90013-5G-Observatory-Quarterly-report-8_1507.pdf.

[16] L. Turchet, G. Fazekas, M. Lagrange, H. S. Ghadikolaei, and C. Fischione, "The internet of audio things: State of the art, vision, and challenges," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 10 233–10 249, 2020.

[17] L. Turchet, C. Fischione, G. Essl, D. Keller, and M. Barthet, "Internet of musical things: Vision and challenges," *IEEE Access*, vol. 6, pp. 61 994–62 017, 2018.

[18] C. Rinaldi and C. Centofanti, "The Musical Metaverse: Advancements and Applications in Networked Immersive Audio," in *Proceedings of the 2024 International Symposium on the Internet of Sounds (IS2)*. IEEE, 2024, pp. 56–63.

[19] B. Loveridge, "Networked music performance in virtual reality: Current perspectives," *Journal of Network Music and Arts*, vol. 2, no. 1, 2020. [Online]. Available: https://commons.library.stonybrook.edu/jonma/vol2/iss1/2

[20] E. M. Morgan-Ellis, "Covid-19 and participatory music-making," *Encyclopedia*, vol. 4, no. 2, pp. 709–719, 2024. [Online]. Available: https://www.mdpi.com/2673-8392/4/2/44

[21] L. Turchet and S. Krstulović, "DSP as a Service: Foundations and Directions," in *Proceedings of the 2024 International Workshop on the Internet of Sounds (IWIS)*. IEEE, 2024, pp. 1–5.

[22] M. Hatami, Q. Qu, Y. Chen, H. Kholidy, E. Blasch, and E. Ardiles-Cruz, "A survey of the real-time metaverse: Challenges and opportunities," *Future Internet*, vol. 16, no. 10, 2024. [Online]. Available: https://www.mdpi.com/1999-5903/16/10/379

[23] F. Shi, H. Ning, X. Zhang, R. Li, Q. Tian, S. Zhang, Y. Zheng, Y. Guo, and M. Daneshmand, "A new technology perspective of the metaverse: Its essence, framework and challenges," *Digital Communications and Networks*, vol. 10, no. 6, pp. 1653–1665, 2024.

[24] M. Xu, W. C. Ng, W. Y. B. Lim, J. Kang, Z. Xiong, D. Niyato, Q. Yang, X. Shen, and C. Miao, "A full dive into realizing the edge-enabled metaverse: Visions, enabling technologies, and challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 1, pp. 656–700, 2023.

[25] H. Lin, S. Wan, W. Gan, J. Chen, and H.-C. Chao, "Metaverse in education: Vision, opportunities, and challenges," in *2022 IEEE International Conference on Big Data (Big Data)*, 2022, pp. 2857–2866.

[26] H. Wang, H. Ning, Y. Lin, W. Wang, S. Dhelim, F. Farha, J. Ding, and M. Daneshmand, "A survey on the metaverse: The state-of-the-art, technologies, applications, and challenges," *IEEE Internet of Things Journal*, vol. 10, no. 16, pp. 14 671–14 688, 2023.

[27] P. Verma, A. I. Mezza, C. Chafe, and C. Rottondi, "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.

[28] M. Christopoulou, M. Giannakou, K. Frantzeskakis, A. Panagopoulos, G. Pierris, and G. Xylomenos, "5G/6G Architecture Evolution for XR and Metaverse: Feasibility Study, Security and Privacy Challenges for Smart Culture Applications," *IEEE Access*, vol. 13, pp. 103 077–103 095, 2025.

[29] 3GPP, "5G; Management and Orchestration; Architecture Framework (3GPP TS 28.533 Version 17.0.0 Release 17)," 3GPP Technical Specification 28.533, Sep. 2021, available online: https://www.3gpp.org/ftp/Specs/archive/28_series/28.533/28533-h00.zip (accessed on 14 July 2025).

[30] S. Karunarathna, S. Wijethilaka, P. Ranaweera, K. T. Hemachandra, T. Samarasinghe, and M. Liyanage, "The role of network slicing and edge computing in the metaverse realization," *IEEE Access*, vol. 11, pp. 25 502–25 530, 2023.

[31] 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.

[32] M. Maier, A. Ebrahimzadeh, and M. Chowdhury, "The tactile internet: Automation or augmentation of the human?" *IEEE Access*, vol. 6, pp. 41 607–41 618, 2018.

[33] W3C, "WebXR Device API," W3C Candidate Recommendation, Immersive Web Working Group, 2021, available online: https://immersive-web.github.io/webxr/ (accessed on 14 July 2025).

[34] X. Zhang, H. Zhang, K. Sun, K. Long, and Y. Li, "Human-centric irregular ris-assisted multi-uav networks with resource allocation and reflecting design for metaverse," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 3, pp. 603–615, 2024.

[35] O. Abari, "Enabling high-quality untethered virtual reality," in *Proceedings of the 1st ACM Workshop on Millimeter-Wave Networks and Sensing Systems 2017*, ser. mmNets '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 49.

[36] D. Van Huynh, S. R. Khosravirad, A. Masaracchia, O. A. Dobre, and T. Q. Duong, "Edge intelligence-based ultra-reliable and low-latency communications for digital twin-enabled metaverse," *IEEE Wireless Communications Letters*, vol. 11, no. 8, pp. 1733–1737, 2022.

[37] M. Sacchetto, Y. Huang, A. Bianco, and C. Rottondi, "Using autoregressive models for real-time packet loss concealment in networked music performance applications," in *Proceedings of the 17th International Audio Mostly Conference*, ser. AM '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 203–210.

[38] L. U. Khan, Z. Han, D. Niyato, M. Guizani, and C. S. Hong, "Metaverse for wireless systems: Vision, enablers, architecture, and future directions," *IEEE Wireless Communications*, vol. 31, no. 4, pp. 245–251, 2024.

[39] L. Turchet and P. Casari, "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)*, 2024, pp. 1–8.

[40] L. Vignati and L. Turchet, "On the lack of a perceptually-motivated evaluation metric for Packet Loss Concealment in Networked Music Performances," *Journal of the Audio Engineering Society*, 2025.

[41] I. T. Union, "Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems," International Telecommunication Union, Radiocommunication Sector (ITU-R), Geneva, Recommendation ITU-R BS.1534-3 BS.1534-3, October 2015, also known as MUSHRA: MUlti Stimulus test with Hidden Reference and Anchor. [Online]. Available: https://www.itu.int/rec/R-REC-BS.1534-3-201510-I/en

[42] "AES67-2018: AES standard for audio applications of networks - High-performance streaming audio-over-IP interoperability," Audio Engineering Society, New York, NY, Standard, Apr. 2918.

[43] DanteAudinate. (2024) GetDante.com. [Online]. Available: https://www.getdante.com/

[44] S. Giacomelli, C. Centofanti, J. Santos, M. Galbiati, T. Salvi, F. Graziosi, and C. Rinaldi, "Remote immersive audio production: State of the art implementation, challenges, and improvements," in *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*, 2024, pp. 1–10.

[45] J. Wang, X. Li, R. Zhang, L. Xiao, and H. V. Poor, "On the Road to 6G: Visions, Requirements, Key Technologies, and Testbeds," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 886–921, 2023.

[46] L. Xiong, J. Malik, and V. Ramesh, "Edge Artificial Intelligence for Real-Time Decision Making Using NVIDIA Jetson Orin, Google Coral Edge TPU, and 6G for Privacy and Scalability," in *Proc. of the IEEE Conference on Smart Edge Networks*, 2025.

[47] F. Jameel, M. Bennis, A. Elgabli, and M. Debbah, "Towards Global Metaverse Accessibility with RSMA-based Satellite-Terrestrial Integrated Networks: A Game Theoretic Approach," *IEEE Open Journal of the Communications Society*, vol. 6, pp. 1453–1470, 2025.

[48] A. Kulkarni, P. Loiseau, R. Vieri, and L. Turchet, "Interoperability is a Fundamental Requirement for the Open Metaverse," *IEEE Access*, vol. 13, pp. 76 452–76 463, 2025.