Harmonizing the Musical Metaverse: unveiling needs, tools, and challenges from experts' point of view

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

Matteo Tomasetti
Dept. of Information Engineering and Computer Science
University of Trento
Trento, Italy
matteo.tomasetti@unitn.it

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

ABSTRACT

The Musical Metaverse (MM) represents an innovative frontier for the field of New Interfaces for Musical Expression (NIME). The MM holds the potential to redefine areas such as musical composition and performance via immersive environments based on technologically mediated social interactions. Despite substantial research on single-user immersive systems, the intersection of NIME and the MM remains largely unexplored. In this paper, we systematically explore this domain by examining previous and current approaches, alongside conducting interviews with eleven experts who have created multi-user immersive musical environments and authored publications on this topic. The goal is to map such an uncharted territory by collecting valuable insights and leveraging the perspective of experts to provide an understanding of the potentials and challenges inherent in creating immersive social environments for musical activities. Our results reveal that existing multi-user immersive environments make use of diverse implementation approaches but face challenges due to the absence of standardized technology stacks, particularly in networking and data synchronization.

Author Keywords

NIME, Musical Metaverse, Musical XR, Networked Music Performance

CCS Concepts

•**Applied computing** → **Sound and music computing;** Performing arts; •**Human-centered computing** → *Empirical studies in HCI;*

1. INTRODUCTION

The Metaverse has been referred to as a "next-generation Internet" where different interoperable virtual worlds will converge and blend with the physical world to create new avenues for various human activities [24]. It represents a future-facing concept that was welcomed with both excitement and skepticism. Moreover, the characteristics of the Metaverse are still ill-defined [65], and to be realized, it will ultimately require a convergence of next-generation telecommunication technologies [11], immersive systems [46], and even legislation [30].

Among early explorations, immersive multi-user environments appear to be the example of prototypical examples of the Metaverse, where use cases, technologies, and interactions have been explored. Sometimes referred to as Shared Virtual Environments (SVEs) [82] or Social eXtended Reality (Social XR) [51], multi-user immersive environments enable distant users to interact and collaborate in shared, immersive, multisensory environments. Such environments reflect some, entire, or even none aspects of the physical world, through the use of Augmented, Mixed, and Virtual Reality (AR, MR, VR) technologies, where communication happen over the internet and is digitally mediated. In these environments, three-dimensional embodied avatars are commonly used as the means through which users manifest and perform their actions [32], but volumetric live videos can be used as well [79].

Even though the term "Metaverse" has existed since the VR prime time [26], the progressive digitalization of our daily life, the outbreak of the COVID-19 pandemic, and the increasing interest of tech corporations brought about a refreshed attention to multi-user and social immersive environments. In this context, the Metaverse emerges as an avenue for new musical experiences, which has led to the proposal of the so-called Musical Metaverse (MM) [73]. Here, geographically-displaced musicians can interact -as well as engage with their audience- through immersive networked environments, with ultra-low latency audio, overcoming both geographical and physical limitations.

While Social XR and SVEs have been studied by the Human-computer Interaction (HCI) and networked multimedia research communities [45, 45, 33, 70, 44], thus far such areas have received comparatively less attention by the music technology community. Despite the recent interest of the NIME community on Musical XR (see e.g., [15, 85, 10, 35, 16, 81, 80, 68]), thus far the focus has mostly been on single-user experiences. As of now, there is no comprehensive overview of existing multi-user immersive musical systems. While different scholars have devoted their attention to virtual concerts as social events [48, 43], existing studies mostly looked at the MM from the point of view of the audience [63, 72, 59]. Moreover, to our best knowledge, the perspective of artists and developers who created these social immersive worlds for musical expression has not yet been formalized.

NIME'24, 4–6 September, Utrecht, The Netherlands.

Table 1: A summary of the literature survey.

	Immersive Experience			Type of Application		Musical Instruments		Audio Rendering		Networking		Data Streaming	
Multi-user environment	Type of Experience	Number of users	Embodiment	Custom-made	Commercial	Physical	Virtual	3D	2D	Co-located	Remote	Control data	Audio data
"Avatar Orchestra Metaverse" [53]	VR	> 2	3D Avatars		○		○	○			○	○	
"VR Interface for collaborative performance" [60]	VR	2	-	○			○	○		○			
"SpectraScore VR" [23]	VR	> 2	-	○			○		○	○		○	
"LeMo" [55]	VR	2	3D Avatars	○			○	○		○		○	
"Multi-modal musical environments" [38]	MR	> 2	-	○		○	○	○				○	○
"Carillon" [39]	VR	> 2	-	○			○	○		○		○	
"Trois Machins de la Grâce Aimante" [37]	VR	> 2	-	○			○		○	○		○	
"Kilqore" [25]	MR	2	-	○		○	○	○		○		○	
"Game Over" [66]	MR	2	2D Avatars	○		○	○		○	○		○	
"Immersive Dreams" [49]	VR	2	-	○			○	○		○		○	
"NeuralDrum" [64]	VR/MR	2	Point Clouds	○		○			○	○		○	
"Framework for assessing social presence in VR" [77]	VR	2	3D Avatars	○		○			○	○		○	
"The Entanglement" [27]	VR	2	Point Clouds 3D Avatars	○		○		○		○		○	○
"Orchestra" [29] [28]	VR	> 2	3D Avatars	○			○	○			○	○	
"Recording in the Metaverse" [20] [21]	VR/MR	4	3D Avatars	○		○		○		○		○	○
"Networked XR Music" [75]	MR	2	Point Clouds	○		○			○	○			○
"AR Drum Circle" [40]	AR	2	3D Avatars	○		○			○	○		○	○
"The Virtual Drum Circle" [78]	MR	2	3D Avatars	○		○	○		○	○		○	○
"Musical Metaverse Playgrounds" [18]	VR	> 2	3D Avatars	○			○	○			○	○	
"Wish You Were Here" [69]	MR	2	Point Clouds	○		○			○	○		○	
"Avatar representation in XR for Immersive NMP" [41]	VR	4	3D Avatars	○		○		○		○		○	○
"NMP in PatchXR and FluCoMa" [14]	VR	> 2	3D Avatars		○		○	○			○	○	

To bridge these gaps, we conducted: 1) an overview of existing musical multi-user immersive environments using XR and networked technologies as instances of the MM; 2) an in-depth series of interviews with 11 experts regarding their motivations, and technical choices behind the creation of such environments. With the present study, we aim to contribute in shaping a research agenda for the MM, and inspire NIME practitioners to further explore this topic.

2. DEFINING THE MUSICAL METAVERSE

The MM can be viewed as the convergence of different technologies such as Networked Music Performance (NMP) systems [67, 31], Musical XR [76], and Internet of Musical Things [74]. However, the MM has its own peculiarities. In the MM, differently from NMP systems, embodiment, presence, and immersion occupy an equal position compared to elements such as low-latency streaming and audio quality [36]. Differently from single-user XR, multi-user experiences are shaped by social interactions mediated through the use of Head-Mounted Displays (HMDs) and the internet. Moreover, in the MM, connected users can play with electric, acoustic, or purely 3D virtual instruments at all ends of the networked environments.

2.1 Multi-user Immersive Environments

While networked music performance was explored since the early age of computer music, it was mostly approached through collaborative controllers as well as screen and desktop-based systems for supporting collaborative and social music creation. For a detailed overview of this field, see [83, 12, 22].

While previous work have discussed summarized some instances of networked immersive systems for music [76, 84], they were not systematically approached as a research area. In 2006 Naeff and Collicut described a system used for co-located networked immersive music performance using stereoscopic projections and spatial audio [60]. Nevertheless, one of the early examples of social, virtual and remote music practice can be found in the activities of the "Avatar Orchestra Metaverse" [2], a collective that created collaborative networked music practices inside Second Life since 2007 [6]. Here, compositions are conceived as spaces, with sound generated by virtual instruments, and 3D avatars used as sound sources [53]. While Second Life is a virtual social platform, it does not provide support for modern HMDs.

However, following the resurgent interest in VR hardware after the release of the Oculus DK 1 in 2012, several researchers and composers started exploring multi-user immersive music creation systems. At the time of writing, VR systems using HMDs represent the majority of works in musical XR. In VR, different musical practices have been explored, such as groups of musicians playing virtual instruments together [23, 37, 23, 49, 14], or web-based live coding environments [28]. Most of the work in multi-user VR focuses on collaborative musical instruments where connected musicians can perform and co-create using shared virtual instruments [39, 55, 18]. While these examples focus on virtual instruments, performances with electric and acoustic systems have been explored. Cairns et al. described a study on a rock band playing inside a virtual replica of the BBC Maida Vale Studios [20, 21]. Dziwis and Von Coler created a performance for two distant violinists that can be experienced by the audience through XR devices [27].

Only a handful of works explored the use of MR and AR displays, as a way to provide a more realistic type of interactions between distant musicians [75, 69, 78], or providing tools to practice simple instrument learning tasks [40]. Composers have also explored MR systems where virtual environments become a score for performers to play with [38, 25, 66]. Moreover, dedicated tools for sound generation in multi-user environments have also been recently developed [29].

Table 1 summarizes the main characteristics of the environments surveyed in literature.

Apart from the examples described above, some multi-user immersive musical applications were recently made available to the public on stores such as the one of Quest and

SteamVR. Among these, VRROOM [8] lets users attend virtual concerts and music events using HMDs. PatchWorld [4] is a virtual environment for musical creation based on a patching paradigm similar to the one of Cycling '74 Max and equipped with a custom audio engine, which provides among many features a multiplayer mode.

3. EXPERTS' INTERVIEWS

In recent years, for better understanding emerging practices and trends, several scholars have started to ponder on the NIME practice [52, 42, 54] through reflexive activities with experts [58, 17]. This approach can also effectively explore complex issues and yield insights into specialized practices [13]. Inspired by these works, we conducted a series of interviews involving experts, selected from whom had previously developed SVEs and and networked XR musical systems, such as the ones reviewed in 2.1. We selected 24 projects where XR technologies were used (see Table 1). Specifically, we considered projects where two or more musicians were connected together, and where networked technologies play a fundamental role.

3.1 Participants

We conducted in-depth interviews with 11 experts. We contacted 18 authors from the 24 selected projects (some authored two or more publications). They were contacted through public emails and the authors' professional network. Out of 18, 11 experts answered to our request. The average age was 39.9 (SD = 15.6). All respondents had an interdisciplinary background in music composition, music technology, telecommunication, and visual arts. Two are, respectively, founders and CTOs of companies dedicated to XR-based music platforms; five are active as composers and developers; and four are researchers in fields such as network music performance and Musical XR. Eight were from EU countries, two from the UK, and one from the USA. They were all male.

The interviews were conducted online, using platforms such as ZOOM and Google Meet, and recorded as audio and video. Participants were instructed in advance about the purpose of the interview and its procedure. The interviews were conducted by two of the authors, with one discussing with the interviewee and the other acting as a note-taker. Ten interviews were conducted in English and one in Italian. The latter was translated into English to be integrated into the analysis.

3.2 Methodology

During the interviews, we asked interviewees to:

1. elaborate on their experience linked to the development of their works, including motivations and usage;
2. describe the technical stack employed in their projects and provide insights into how sound was treated, as well as how the networking element was managed;
3. think about the main issues and challenges they found in developing such systems.

After completing all interviews, we transcribed each recording and integrated them with our notes. Subsequently, we conducted a reflexive thematic analysis [19], applying principles of grounded theory [34] to analyze the collected material. After three rounds of analysis, we reached a consensus on emergent themes that best summarized the ideas expressed by the interviewees. We organized the themes into five categories: motivations and outcomes, modes of experience, tools for creation, network and streaming, and opportunities and challenges.

4. RESULTS

In this section, we report the results of the analysis, organized into five main themes. The interviewees are referred to as P, numbered from 1 to 11, and their quotes are indicated in italics.

4.1 Motivations and Outcomes

At first, we asked participants to discuss their motivations for creating multi-user immersive environments for musical creation. Seven participants (P2, P4, P6, P7, P8, P10, P11) were motivated by their interest in exploring novel compositional practices. P2 and P7 were interested in applying multiplayer computer game mechanics to composition and improvisation, e.g., *"What happens to music if a composer applies game inputs? What, then, are the musical consequences of the performance?"* (P2). Then, P4 was interested in creating audiovisual pieces in terms of *"world building"*. Differently, P8 cited the need to explore collaborative compositional strategies, e.g., *"As a composer, I was no longer interested in writing and memorizing notes but in finding new collaborative strategies for making music together"*. Instead, five participants (P1, P3, P4, P6, P9) mentioned the need to create systems for research purposes, on topics like networked and immersive audio technologies, social presence, and musical interactions, e.g., *"Doing research regarding the social presence in the MM is essential to ensuring accessibility"* (P1).

4.2 Modes of Experience

We asked participants to describe the technologies that a musician (or audience member, if present) should use to experience their immersive environments, especially regarding the visual and the audio components. Our findings indicate that the choice of a specific device (for visual or auditory stimuli) depends on the desired level of immersion should be conveyed. Most of the interviewees developed environments for HMDs (P1, P3, P5, P6, P7, P9, P10, P11). Among these, standalone HMDs were used, such as Meta Quest 2 for VR applications and Microsoft Hololens for MR. In most of these environments, the HMDs were used either by the musicians for playing together (P1, P3, P5, P6, P10, P11), or by audience members to attend a virtual concert remotely (P4, P9).

Conversely, four interviewees (P2, P4, P7, P8, P11) developed systems that do not require performers or audience members to wear any device, since the environment is projected or presented with monitors in the context of a performance, in a concert hall or similar spaces. However, the interviews revealed that combining these two modes of experience is also feasible (P4, P7, P11), e.g., *"Two performers wear HMDs in the theater's physical space while the audience is seated in the hall and sees the performers, and what the performers see in the HMDs is projected on two big screens in front of the audience"* (P11).

We then inquired about how the users (whether musicians or audience members) and the environments were visually designed and presented. Almost all interviewees used 3D graphics, except P7 who used 2D.

When the experience was based on purely VR, the environments were the most diverse, such as music clubs (P9,

P10), concert halls (P4, P8, P9), or imaginary open worlds (P4, P7, P10). Such environments also include audio-reactive 3D objects (i.e., lights, props, architectural elements), which are used as *"dynamic scenography"* (P4, P9). Users were presented as three-dimensional embodied avatars (P4, P7, P8, P9, P10, P11). We found that interviewees employed different types of avatars, from minimal (i.e., with only head and hands (P4)), to full-body ones (P8, P9, P10, P11). Only P4 explored the use of monochromatic volumetric point-cloud representations of the remote musicians.

Conversely, in MR environments, the physical environment was either the physical one inhabited by musicians and the audience (P6) or a mix of the physical and the virtual one (P1, P3, P5). These environments were designed to connect musicians playing physical instruments together. Here, musicians were embodied into full-body semi-realistic avatars, controlled through motion capture (P6), or half-bodies virtual avatars controlled by the position and rotation of the HMDs (P1, P3, P5).

An interesting aspect that emerged in this context, is the role of spatial audio, being it positional or binaural. In Figure 1, we summarize the main techniques used for audio rendering found in literature. Experts considered spatial audio important but not essential in multi-user environments (P8, P9, P10, P11). What was considered essential was the quality of audio, which according to them should be prioritized.

4.3 Tools for Creation

We surveyed the software tools used by the experts to design their multi-user immersive environments. The majority of experts used Unity [7] (P1, P2, P3, P5, P6, P7, P9, P10) for VR and MR experiences, e.g., *"I use Unity because in my opinion it is very flexible, and also I love it because it has a large library of assets and resources developed by the community, and also it is easily integrated with other languages that I use"* (P6).

However, we found two exceptions. First, two interviewees developed and implemented new instruments and customized solutions using consumer-ready platforms such as Second Life (P8) and PatchWorld (P11), which can offer integrated multi-user capabilities and allow for partial customization (e.g., importing and loading external 3D assets). Second, P4 developed a series of environments using Networked A-Frame [3], a multi-player extension of A-Frame [1], which is a popular framework used for creating immersive applications running on web browsers through the WebXR API [9]. In terms of tools, the experts reflected the results of the analysis of the literature, as summarized in Figure 2.

We then explored how audio was acquired, generated, and processed. Only a few interviewees, specifically P1, P3, P4, and P5, used physical instruments like drums, guitars, or violins. For P1 and P3, audio was captured on a host PC and then processed further using a series of plugins in a Digital Workstation.

The majority of interviewees developed 3D virtual instruments, with the sound generated and rendered locally, directly in the immersive environment. While some instruments allow the triggering of audio samples (P2, P6, P8, P11), most interviewees used real-time sound synthesis running on a host PC (P2, P6, P7) through software such as Supercollider and Ableton Live. P4 used audio generation based on Web Audio, leveraging specialized libraries and scripting languages such as PdWebParty and Bytebeat. P10 and P11 used the sound engine available inside PatchWorld.

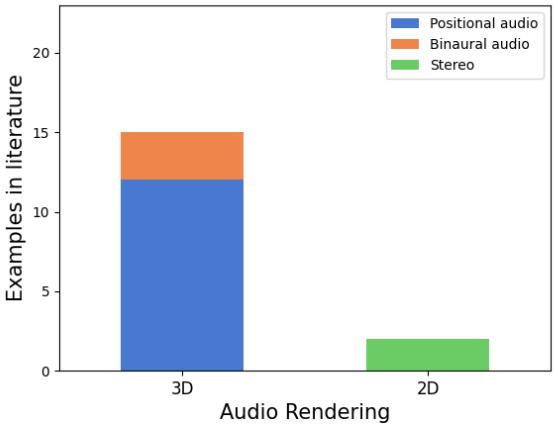

Figure 1: Breakdown of the literature analysis: typologies of audio rendering.

4.4 Network and Streaming

An important aspect we examined was how data were transmitted and synchronized in the immersive environments. The majority of interviewees developed co-located experiences using local networks (P1, P2, P3, P4, P5, P7). This allowed them to reduce latency and jitter, especially for systems requiring tight timing. However, the majority of them transmitted only data used for controlling and synchronizing virtual 3D instruments. These control data were transmitted using protocols such as OSC or the WebRTC data channels. Only a handful of them made use of real-time audio streaming (P1, P3, P4, P6). However, this approach required the development of custom systems capable of integrating XR devices and NMP software, such as Sonobus and JackTrip, e.g., *"I mainly use JackTrip because, in my opinion, it is the best and can keep the latency stable, avoiding too much packet loss concealment."* (P3).

Some interviewees (P8, P9, P10, and P11) had designed and used remote environments potentially accessible from nearly anywhere on Earth. Apart from the ones based on platforms such as Second Life and PatchWorld, for custom environments interviewees used commercial systems based on relay servers such as Photon Fusion [5]. Photon Fusion is a widely used solution for real-time multi-user networking and synchronization in games and XR applications, especially the ones made with Unity. Photon was chosen by the interviewees for its scalability, which proved useful for virtual concerts (P9) However, with such systems only control data were streamed in real-time, such as the ones for synchronizing avatar position in space and animations. The only audio supported by tools such as Fusion is speech, captured from the microphones embedded in HMDs, which are not optimal for musical interaction, in both terms of quality and latency.

In Figure 3, we summarize the main tools used for control and audio data streaming and synchronization as found in the literature.

4.5 Opportunities and Challenges

At the end of each interview, we asked participants to discuss the challenges they encountered while developing their multi-user immersive environments and to relate their experiences to the vision of the Musical Metaverse.

The majority of the interviewees identified the networking component as the most challenging part of the develop-

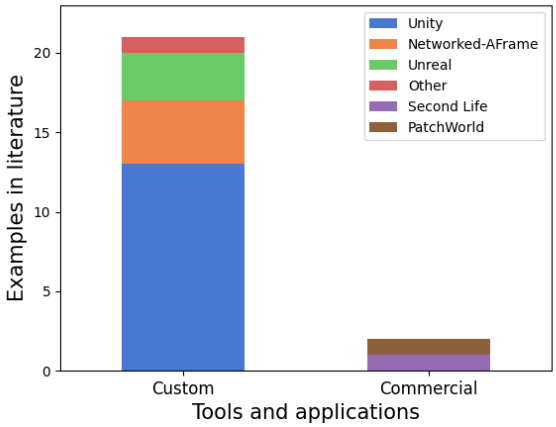

Figure 2: Breakdown of the literature analysis: software tools and applications used for the development of immersive environments.

ment process (P1, P3, P4, P5, P6, P7, P9, P10, P11). A main issue is linked to the necessity of having stable and high-quality audio streaming, e.g., *"Network bandwidth and the network infrastructure still represent the most significant challenges because they do not guarantee the ability to have stable and quality streaming"* (P3). Interviewees also identified network latency and jitter as a major element of concern, which represents a significant barrier for true real-time music playing in distance. According to them, this calls for more efficient tools and systems that can alleviate problems especially the ones related to network transmission of audio data.

However, as several interviewees pointed out, improving audio streaming alone will not solve the challenges that characterize multi-user immersive environments for music making. An issue that differentiates immersive networked environments from regular NMP systems is the synchronization between the audio data and all the control data shared between the connected users. Differently from synchronizing audio to a video stream, here audio has to be synchronized between several 3D assets such as avatars and virtual objects (i.e., controllers, UIs). What emerged from the discussion with experts is a lack of tools dedicated to solving this issue. At the moment, audio and control data streaming are performed in parallel, requiring different systems (i.e., Sonobus for audio, and OSC through Wi-Fi for control data). However, participants deemed that synchronization is always the result of a custom implementation (P1, P3, P6, P9, P10, P11).

A related challenge is the difficulty of integrating existing software and musical hardware with all tools such as game engines, and NMP systems, used to create multi-user immersive environments (P1, P2, P3, P5, P9), e.g., *"Suppose one tries to make a standalone application using the combination of Unity and Wwise [...] one will not reach the complexity in terms of sound synthesis that can be achieved with Unity and an audio programming language, such as Supercollider"* (P2).

Another challenge identified is the testing phase of multiplayer systems, particularly for small teams or individuals working alone as composer/developers. According to interviewees, testing these applications effectively requires at least two people to be available simultaneously. As P1 noticed: *"For multi-player app testing, it is necessary to be at least in two people to see if the sync messages, data, and audio components work correctly. It is not simple at all"*. Interviewees also emphasized the difficulties encountered in deploying these immersive environments (P1, P3, P4, P6, P7, P10). These issues are currently viewed as the primary barriers to achieving the Musical Metaverse from a technological perspective.

Two interviewees (P9, P10), who manage commercial platforms for Musical XR, highlighted the problem of sustainability of multi-user immersive environments. These are complex ecosystems that must be maintained not only from a technical point of view. In light of the MM, how such environments can sustain themselves without relying solely on advertisements? How musicians can monetize their involvement in these immersive environments? At the moment, there are no standardized ticketing systems, and main sources of revenue have not been identified. These are important issues that might prevent several stakeholders (i.e., artists, music industry, investors) from spending their energy, money, and time in such environments.

Lastly, another issue that emerged concerns audience participation. Four interviewees, primarily composers (P2, P8, P10, P11), brought up this point. According to them, although involving the audience presents an interesting opportunity in the MM, enabling audience members or any connected user to influence a performance or musical event is seen as problematic, e.g., *"What are the differences between playing a piece and playing a game? [...] when one plays a piece, one has to communicate something to the audience, whereas usually, when one plays a game, one is basically playing for oneself and not for an audience. What are the musical consequences? Unpredictable things often happen with these dynamics in these multi-player environments within a performative situation"* (P2). In addition, as mentioned by P8, because these environments revolve around social interactions, problems could arise from a lack of mutual understanding between users sharing a common virtual space. As a result, implementing some form of moderation and procedures for conflict resolution might be necessary. However, it remains unclear who should govern these processes.

5. DISCUSSION

In this study, we first reviewed previous work on multi-user social immersive musical environments. These works were not systematically approached before, not only as a cohesive body of work, but also as instances of the MM. We then interviewed 11 expert developers to gather information about their practices, tools, challenges, and opportunities to help guide and inspire future research and design efforts toward the MM.

Taken together, both the survey performed on previous work and experts interviews reveal:

- a variety of approaches are used for the implementation of multi-user immersive environments, depending on their purpose and motivations;
- the lack of a standardized technology stack, especially for networking and data synchronization.

The first observation may not be surprising for the NIME community since it recalls the domain-specific nature of Digital Musical Instruments (DMIs) [50]. However, there is a critical aspect that distinguishes the systems we surveyed from general DMIs: they function not merely as musical tools but as environments where social and musical interactions unfold. While these environments bear some resemblance to collaborative and social music practices as the

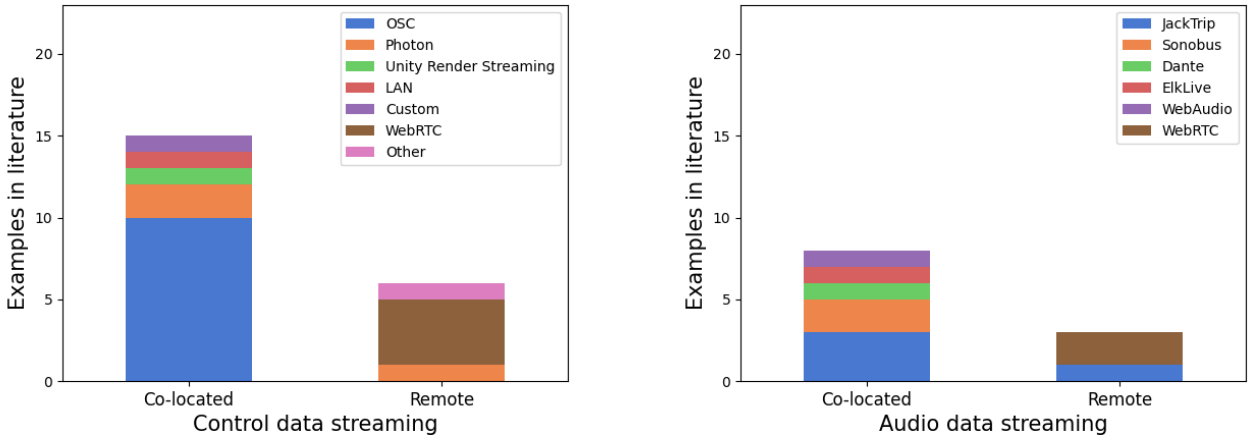

Figure 3: Breakdown of the literature analysis: control and audio data streaming protocols and tools.

ones of networked smartphones [62] and laptop orchestras [61], the embodied and immersive characteristics of multi-user music environments were thus far not being systematically explored in NIME literature [47]. Defining multi-user immersive musical environments as a specific research area is crucial for advancing the MM field, in its technological, artistic, and social characteristics.

The second observation highlights the challenge of integrating XR software and hardware with audio and networking components of multi-user environments. As noted in the interviews, the only current method to achieve this integration is through the development of custom solutions. However, these solutions are difficult to thoroughly document and replicate. Furthermore, latency and jitter issues limit their effectiveness to local networks or short distances.

While previous work in NMP explored techniques for synchronizing audio and video [67], there are very few dedicated studies on the synchronization of full-body avatar movements with real-time audio data streams for immersive systems (i.e., [21, 78, 40]). Some interviewees noted that especially synchronizing control data between audio and 3D elements among connected peers in a multi-user environment poses several challenges, particularly for beat-based music, which requires precise quantization to prevent unpleasant experiences during collaborative music-making. However, this can result in an unrealistic playing experience. Consequently, experts strongly desire protocols and standards that could simplify development and testing, and also facilitate the creation of more stable and scalable networked immersive environments. However, no particular solutions have been suggested.

It is important to note that the current state of multi-user environments has certain limitations for hosting large-scale events. Popular social VR applications like VRChat (which uses Photon Fusion) and Mozilla Hubs (based on WebRTC) typically support only about 50-80 users in the same environment. Although these platforms have hosted concerts and music festivals, they were not originally designed for truly networked music experiences but rather as environments where a single data stream (from live video or sound recordings) is broadcast to all users simultaneously. While it is possible to extend these limits using custom servers and protocols, creating large immersive concerts where multiple musicians play together from different locations remains challenging.

While there is abundant literature on Musical XR systems, these were mostly conceived for single-user experiences (i.e., [71, 76, 15]). Such systems usually do not require stable, fast internet connection and reliable transmission protocols. However, the success of multi-user environments heavily depends on these aspects.

As some experts commented, the quality of bandwidth of a network varies depending on country, and internet providers. Therefore, accessibility is not guaranteed, resulting in poor performance, and ultimately turning many users (including musicians) away from such systems. While this issue goes beyond the reach of creators and developers of musical experiences, synchronization, and networking, are perceived as the main technical bottlenecks for achieving a truly interoperable and functional Musical Metaverse.

Although the MM should support a broad spectrum of musical practices, the specific benefits and values these environments offer to different types of music-making remain unclear. To date, research on multi-user XR systems has been limited, mainly focusing on the experiences of two musicians connected simultaneously [69, 18, 75, 78, 55]. In cases where more participants were included, the sample sizes were notably small, such as studies involving four participants [20, 41, 21]. Additionally, it is important to notice that these studies have been performed only within local networks. As a result, we see a pressing need to thoroughly understand the user experience and acceptance of such environments. Key questions about the primary use cases and needs of musicians remain unanswered: Are these environments suitable for group rehearsals? Recording sessions? Collective improvisation, or performances? Should they accommodate the largest number of concurrent users?

According to some of the experts we interviewed, this lack of compelling use cases and characteristics of the user experience is one of the main push-back for the musical industry, and one of the reasons of its perceived lack of interest in the idea of the MM. NIME researchers and practitioners may have an important role in finding and validating convincing use cases and applications.

Our literature review showed that multi-user immersive environments are distributed across the entire spectrum of the Reality-Virtuality Continuum [56], being mostly composed of MR and VR experiences. While at the moment there is a strong division between VR (experienced with HMDs) and MR systems (experienced with goggles but also with projections), several experts pointed out the importance of giving to the MM a "fluid" identity, where users can move across several shades of realities, as a way to avoid the experience of immersive environments tight to a specific

and normative set of devices. However, at the moment, it is not clear if musicians need total or partial immersion. Moreover, it is unclear if social interactions among musicians should happen in complete virtual environments with embodied avatars or in an environment that mixes virtual and physical elements.

A final note on limitations. The findings discussed in this paper were derived from the analysis of an homogeneous pool of participants, both in terms of gender and geographical location. While our study aimed at mapping the state of the art of multi-user immersive music systems, with a focus on technology, further endeavors should broaden the investigation to include diverse perspectives on the topic, and extend the analysis to social and ethical aspects that have not been trated in this work.

6. CONCLUSION

Multi-user musical environments hold significant potentials for music-making, yet their exploration within NIME and HCI has been limited. Our literature analysis and interviews with experts have identified the main challenges related to this scarcity and suggested directions for further research. We have also provided a detailed overview of the tools and practices used to create these environments and discussed their role and limitations as well.

To fully realize a truly immersive and interoperable Musical Metaverse that accommodates various musical practices, it is crucial to address the existing technical challenges already visible in current multi-user environments and develop compelling use cases that attract audiences, composers, and musicians.

Our work is not intended to offer definitive conclusions about multi-user immersive environments for music. Instead, we view it as starting a dialogue within the NIME community that includes composers, performers, and researchers working with XR and NMP systems.

7. ACKNOWLEDGMENTS

We deeply thank all the interviewees who took the time to share their expertise with us.

8. ETHICAL STANDARDS

This paper complies with the ethical standard of the NIME conference [57]. We acknowledge the support of the MUR PNRR PRIN 2022 grant, prot. n. 2022CZWWKP, funded by Next Generation EU. Participants were formally contacted by the authors. All of them volunteered and provided verbal informed consent.

9. REFERENCES

[1] AFrame. `https://aframe.io/`. Last Accessed: 2024-01-15.

[2] Avatar Orchestra Metaverse (AOM). `http://www.avatarorchestra.org/`. Last Accessed: 2024-01-15.

[3] Networked AFrame. `hhttps://github.com/networked-aframe/networked-aframe`. Last Accessed: 2024-01-15.

[4] PatchWorld. `https://patchxr.com/`. Last Accessed: 2024-01-15.

[5] Photon Fusion. `https://www.photonengine.com/fusion`. Last Accessed: 2024-01-15.

[6] Second Life. `https://secondlife.com/`. Last Accessed: 2024-01-15.

[7] Unity. `https://unity.com/`. Last Accessed: 2024-01-15.

[8] VRROOM. `https://vrroom.world/`. Last Accessed: 2024-01-15.

[9] WebXR Device API. `https://immersiveweb.dev/`. Last Accessed: 2024-01-15.

[10] C. Arslan, F. Berthaut, A. Beuchey, P. Cambourian, and A. Paté. Vibrating shapes: Design and evolution of a spatial augmented reality interface for actuated instruments. In *NIME 2022*, 2022.

[11] A. M. Aslam, R. Chaudhary, A. Bhardwaj, I. Budhiraja, N. Kumar, and S. Zeadally. Metaverse for 6g and beyond: The next revolution and deployment challenges. *IEEE Internet of Things Magazine*, 6(1):32–39, 2023.

[12] Á. Barbosa. Displaced soundscapes: A survey of network systems for music and sonic art creation. *Leonardo Music Journal*, 13(1):53–59, 2003.

[13] K. L. Barriball and A. While. Collecting data using a semi-structured interview: a discussion paper. *Journal of Advanced Nursing-Institutional Subscription*, 19(2):328–335, 1994.

[14] J. Bell. Networked Music Performance in PatchXR and FluCoMa. In *International Computer Music Conference (ICMC) 2023*, 2023.

[15] F. Berthaut. 3d interaction techniques for musical expression. *Journal of New Music Research*, 49(1):60–72, 2020.

[16] S. Bilbow. Evaluating polaris˜ - An Audiovisual Augmented Reality Experience Built on Open-Source Hardware and Software. In *NIME 2022*, 2022.

[17] A. Boem, G. M. Troiano, , G. Lepri, and V. Zappi. Non-Rigid Musical Interfaces: Exploring Practices, Takes, and Future Perspective. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 17–22, May 2021.

[18] 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*, pages 1–9. IEEE, 2023.

[19] V. Braun and V. Clarke. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health*, 11(4):589–597, 2019.

[20] P. Cairns, A. Hunt, J. Cooper, D. Johnston, B. Lee, H. Daffern, and G. Kearney. Recording music in the metaverse: a case study of xr bbc maida vale recording studios. In *Audio Engineering Society Conference: AES 2022 International Audio for Virtual and Augmented Reality Conference*. Audio Engineering Society, 2022.

[21] P. Cairns, A. Hunt, D. Johnston, J. Cooper, B. Lee, H. Daffern, and G. Kearney. Evaluation of Metaverse Music Performance With BBC Maida Vale Recording Studios. *Journal of the Audio Engineering Society*, 71(6):313–325, 2023.

[22] C. Çakmak, A. Çamci, and A. G. Forbes. Networked virtual environments as collaborative music spaces. In *NIME*, pages 106–111, 2016.

[23] B. E. Carey. Spectrascore vr: Networkable virtual reality software tools for real-time composition and performance. In *International conference on New Interfaces for Musical Expression (NIME)*, 2016.

[24] R. Cheng, N. Wu, S. Chen, and B. Han. Will

metaverse be nextg internet? vision, hype, and reality. *IEEE Network*, 36(5):197–204, 2022.

[25] M. Ciciliani. Virtual 3D environments as composition and performance spaces. *Journal of New Music Research*, 49(1):104–113, 2020.

[26] J. D. N. Dionisio, W. G. B. III, and R. Gilbert. 3d virtual worlds and the metaverse: Current status and future possibilities. *ACM Comput. Surv.*, 45(3), jul 2013.

[27] D. Dziwis and H. von Coler. The Entanglement: Volumetric Music Performances in a Virtual Metaverse Environment. *Journal of Network Music and Arts*, 5(1):3, 2023.

[28] D. Dziwis, H. von Coler, and C. Porschmann. Live Coding in the Metaverse. In *2023 4th International Symposium on the Internet of Sounds*, pages 1–8. IEEE, 2023.

[29] D. Dziwis, H. Von Coler, and C. Pörschmann. Orchestra: a toolbox for live music performances in a web-based metaverse. *Journal of the Audio Engineering Society*, 71(11):802–812, 2023.

[30] B. Egliston, M. Carter, and K. E. Clark. Who will govern the metaverse? examining governance initiatives for extended reality (xr) technologies. *New Media & Society*, page 14614448231226172, 2024.

[31] L. Gabrielli, S. Squartini, L. Gabrielli, and S. Squartini. *Wireless networked music performance*. Springer, 2016.

[32] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. A. Sasse. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 529–536, 2003.

[33] R. K. Ghamandi, Y. Hmaiti, T. T. Nguyen, A. Ghasemaghaei, R. K. Kattoju, E. M. Taranta, and J. J. LaViola. What and how together: A taxonomy on 30 years of collaborative human-centered xr tasks. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 322–335, 2023.

[34] B. Glaser and A. Strauss. *Discovery of grounded theory: Strategies for qualitative research*. Routledge, 2017.

[35] M. Graf and M. Barthet. Mixed reality musical interface: Exploring ergonomics and adaptive hand pose recognition for gestural control. In *NIME 2022*, 2022.

[36] G. Hajdu. Embodiment and disembodiment in networked music performance. In *Body, Sound and Space in Music and Beyond: Multimodal Explorations*, pages 257–278. Routledge, 2017.

[37] R. Hamilton. Trois Machins de la Grâce Aimante: A virtual reality string quartet. In *Proceedings of the 2019 International Computer Music Conference, New York*, 2019.

[38] R. Hamilton, J.-P. Caceres, C. Nanou, and C. Platz. Multi-modal musical environments for mixed-reality performance. *Journal on Multimodal User Interfaces*, 4:147–156, 2011.

[39] R. Hamilton and C. Platz. Gesture-based collaborative virtual reality performance in carillon. In *Proceedings of the 2016 international computer music conference*, pages 337–340, 2016.

[40] T. Hopkins, S. C. C. Weng, R. Vanukuru, E. A. Wenzel, A. Banic, M. D. Gross, and E. Y.-L. Do. AR Drum Circle: Real-Time Collaborative Drumming in AR. *Frontiers in Virtual Reality*, 3:847284, 2022.

[41] A. Hunt, H. Daffern, and G. Kearney. Avatar representation in extended reality for immersive networked music performance. In *Audio Engineering Society Conference: AES 2023 International Conference on Spatial and Immersive Audio*. Audio Engineering Society, 2023.

[42] A. R. Jensenius and M. J. Lyons. Trends at NIME–Reflections on Editing "A NIME Reader".

[43] T. Kaneko, H. Tarumi, K. Kataoka, Y. Kubochi, D. Yamashita, T. Nakai, and R. Yamaguchi. Supporting the sense of unity between remote audiences in vr-based remote live music support system ksa2. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pages 124–127. IEEE, 2018.

[44] N. Krome and S. Kopp. Towards real-time co-speech gesture generation in online interaction in social xr. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*, IVA '23, New York, NY, USA, 2023. Association for Computing Machinery.

[45] M. E. Latoschik, F. Kern, J.-P. Stauffert, A. Bartl, M. Botsch, and J.-L. Lugrin. Not alone here?! scalability and user experience of embodied ambient crowds in distributed social virtual reality. *IEEE transactions on visualization and computer graphics*, 25(5):2134–2144, 2019.

[46] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley Professional, 2017.

[47] B. Loveridge. Networked music performance in virtual reality: current perspectives. *Journal of Network Music and Arts*, 2(1):2, 2020.

[48] B. Loveridge. An overview of immersive virtual reality music experiences in online platforms. *Journal of Network Music and Arts*, 5(1), 2023.

[49] A. MacLean and D. Ogborn. Immersive Dreams: A Shared VR Experience. In *NIME*, pages 380–381, 2020.

[50] T. Magnusson and E. H. Mendieta. The acoustic, the digital and the body: A survey on musical instruments. In *Proceedings of the 7th international conference on New interfaces for musical expression*, pages 94–99, 2007.

[51] S. Mann, Y. Yuan, F. Lamberti, A. El Saddik, R. Thawonmas, and F. G. Prattico. extended meta-uni-omni-verse (xv): Introduction, taxonomy, and state-of-the-art. *IEEE Consumer Electronics Magazine*, 2023.

[52] A. Marquez-Borbon and P. Stapleton. Fourteen years of nime: the value and meaning of'community'in interactive music research. In *NIME*, pages 307–312, 2015.

[53] G. Martín. Social and psychological impact of musical collective creative processes in virtual environments; Te Avatar Orchestra Metaverse in Second Life. Musica/Tecnologia Music. *Technology*, 75:75–87, 2018.

[54] R. Masu, A. P. Melbye, J. Sullivan, and A. R. Jensenius. NIME and the environment: toward a more sustainable NIME practice. In *NIME 2021*. PubPub, 2021.

[55] L. Men and N. Bryan-Kinns. LeMo: exploring virtual space for collaborative creativity. In *Proceedings of the*

2019 on Creativity and Cognition, pages 71–82. 2019.

[56] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994.

[57] F. Morreale, N. Gold, C. Chevalier, and R. Masu. Nime principles & code of practice on ethical research. 2023.

[58] F. Morreale, A. P. McPherson, and M. Wanderley. Nime identity from the performer's perspective. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 168–173, July 2018.

[59] A. Munoz-Gonzalez, S. Kobayashi, and R. Horie. A multiplayer vr live concert with information exchange through feedback modulated by eeg signals. *IEEE Transactions on Human-Machine Systems*, 52(2):248–255, 2022.

[60] M. Naef and D. Collicott. A vr interface for collaborative 3d audio performance. In *Proceedings of the 2006 conference on New interfaces for musical expression*, pages 57–60, 2006.

[61] D. Ogborn. Live coding in a scalable, participatory laptop orchestra. *Computer Music Journal*, 38(1):17–30, 2014.

[62] J. Oh, J. Herrera, N. J. Bryan, L. Dahl, and G. Wang. Evolving the mobile phone orchestra. In *NIME*, pages 82–87, 2010.

[63] K. E. Onderdijk, L. Bouckaert, E. Van Dyck, and P.-J. Maes. Concert experiences in virtual reality environments. *Virtual Reality*, pages 1–14, 2023.

[64] Y. S. Pai, R. Hajika, K. Gupta, P. Sasikumar, and M. Billinghurst. Neuraldrum: Perceiving brain synchronicity in xr drumming. New York, NY, USA, 2020. Association for Computing Machinery.

[65] S.-M. Park and Y.-G. Kim. A metaverse: Taxonomy, components, applications, and open challenges. *IEEE Access*, 10:4209–4251, 2022.

[66] A. Pirchner. Ergodic and emergent qualities of realtime scores. anna and marie and gamified audiovisual compositions. In *Proceedings of the International Conference on Technologies for Music Notation and Representation–TENOR*, volume 20, pages 189–197, 2020.

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

[68] G. Santini. Augmented piano in augmented reality. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 411–415. Birmingham City University, July 2020.

[69] R. Schlagowski, D. Nazarenko, Y. Can, K. Gupta, S. Mertes, M. Billinghurst, and E. André. Wish You Were Here: Mental and Physiological Effects of Remote Music Collaboration in Mixed Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2023.

[70] J.-H. Schröder and H.-C. Jetter. Towards a model for space and time in transitional collaboration. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 223–227, 2023.

[71] S. Serafin, C. Erkut, J. Kojs, N. C. Nilsson, and R. Nordahl. Virtual reality musical instruments: State of the art, design principles, and future directions. *Computer Music Journal*, 40(3):22–40, 2016.

[72] M. Slater, C. Cabriera, G. Senel, D. Banakou, A. Beacco, R. Oliva, and J. Gallego. The sentiment of a virtual rock concert. *Virtual Reality*, 27(2):651–675, 2023.

[73] L. Turchet. Musical Metaverse: vision, opportunities, and challenges. *Personal and Ubiquitous Computing*, pages 1–17, 2023.

[74] L. Turchet, C. Fischione, G. Essl, D. Keller, and M. Barthet. Internet of musical things: Vision and challenges. *IEEE Access*, 6:61994–62017, 2018.

[75] 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*, pages 226–230, 2022.

[76] L. Turchet, R. Hamilton, and A. Çamci. Music in extended realities. *IEEE Access*, 9:15810–15832, 2021.

[77] B. Van Kerrebroeck, G. Caruso, and P.-J. Maes. A methodological framework for assessing social presence in music interactions in virtual reality. *Frontiers in Psychology*, 12:663725, 2021.

[78] B. Van Kerrebroeck, K. Crombé, S. M. de Leymarie, M. Leman, and P.-J. Maes. The virtual drum circle: polyrhythmic music interactions in mixed reality. *Journal of New Music Research*, pages 1–21, 2024.

[79] I. Viola, J. Jansen, S. Subramanyam, I. Reimat, and P. Cesar. Vr2gather: A collaborative social vr system for adaptive multi-party real-time communication. *IEEE MultiMedia*, 2023.

[80] Y. Wang and C. Martin. Cubing Sound: Designing a NIME for Head-mounted Augmented Reality. In *NIME 2022*, 2022.

[81] Y. Wang, M. Xi, M. Adcock, and C. P. Martin. Mobility, space and sound activate expressive musical experience in augmented reality. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 128–133, May 2023.

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

[83] G. Weinberg. Interconnected musical networks: Toward a theoretical framework. *Computer Music Journal*, 29(2):23–39, 2005.

[84] V. Zappi, F. Berthaut, and D. Mazzanti. From the lab to the stage: Practical considerations on designing performances with immersive virtual musical instruments, 2022.

[85] K. C. Zellerbach and C. Roberts. A framework for the design and analysis of mixed reality musical instruments. In *NIME 2022*, 2022.