

XR Jam: Design Considerations for Music Networking with AI in XR

1st Torin Hopkins

ATLAS Institute

University of Colorado, Boulder

Torin.Hopkins@colorado.edu

2nd Suibi Che Chuan Weng

ATLAS Institute

University of Colorado, Boulder

Suibi.Weng@colorado.edu

3rd Rishi Vanukuru

ATLAS Institute

University of Colorado, Boulder

Rishi.Vanukuru@colorado.edu

4th Sasha Novack

ATLAS Institute

University of Colorado, Boulder

Sasha.Novack@colorado.edu

5th Chad Tobin

ATLAS Institute

University of Colorado, Boulder

chto7865@colorado.edu

6th Emma Wenzel

ATLAS Institute

University of Colorado, Boulder

Emma.Wenzel@colorado.edu

7th Amy Banic

Department of Computer Science

University of Wyoming

abanic@cs.uwyo.edu

8th Mark Gross

ATLAS Institute

University of Colorado, Boulder

MDGross@colorado.edu

9th Ellen Yi-Luen Do

ATLAS Institute

University of Colorado, Boulder

Ellen.Do@colorado.edu

Abstract—The XR Jam project presents an integrated system designed to enhance musical interaction and spatial awareness in Extended Reality (XR) environments. Leveraging AI-driven Virtual Musical Agents (AIVMs), the platform elevates traditional musical experiences. Built on an XR framework, the system pushes conventional boundaries by enabling novel forms of interaction, such as object manipulation and social drumming with AIVMs and remote players. Furthermore, avatar-based interfaces help overcome social inhibitions, offering players a unique, uninhibited interactive experience. Pilot studies revealed the possibility for effective use of haptics in spatial localization. While haptic feedback enhanced spatial orientation in specific scenarios, front and back directions remained challenging to discern. The system promotes musical interaction, social engagement, and spatial orientation, although further refinements are needed for synchronization and player experience. Overall, this research opens new avenues for immersive musical experiences through a fusion of AI, XR, and haptic technologies.

Index Terms—component, formatting, style, styling, insert

I. INTRODUCTION

In recent years, digital technology has undergone a profound transformation, reshaping the landscape of human-computer interaction, networked experiences, extended reality (XR), and the integration of artificial intelligence (AI). This paper explores the interplay between these domains by introducing a networked music experience in XR for playing music with AI-Driven Virtual Musicians (AIVMs) and other players offering novel insights into considerations for designing such systems.

We introduce XR Jam, a system for playing music in real-time with AIVMs and other players offering a fusion of technology and creativity (see Figure 1). This innovation extends traditional musical play, which has historically been an

in-person, social activity. With the advent of fast and reliable networks, AIVMs, and XR-capable devices musicians now have the power to bridge geographical gaps and experience musical connection in entirely new ways. Our exploration of networked music experiences in XR provides valuable insights into how AI can enhance and augment the musical creative process within networked environments.

This paper represents the dynamic intersection of XR, networked musical experiences, and AIVM interactions. By presenting our prototype and the results of our pilot studies, we contribute not only to the academic discourse but also to the practical development of technology that empowers musicians to create, collaborate, and connect. We hope that our findings will inspire further exploration and innovation at the nexus of technology and music.

II. RELATED WORK

A. Networked Music Experiences

Near-instantaneous networked music experiences represent a pivotal shift in how musicians collaborate and perform, overcoming many geographical limitations [1]. Work in this area is predominantly focused on enabling high-quality audio sharing within a distance of around 1000 kilometers [2]–[5]. Though many services were available prior to the pandemic, such platforms served as essential avenues for artists to convene and create when the onset of the COVID-19 pandemic necessitated physical distancing. The increased demand for virtual collaboration during this period spotlighted the importance and potential of networked music experiences.

The field has witnessed advancements that go beyond auditory sharing, extending to the inclusion of visual components for a more enriched musical experience. Innovations such



Fig. 1. A) XR Jam system overview and components. B) XR Jam through AR glasses. C) Guitar player jamming with guitar playing AIVM. D) View of the haptic mat top down. E) View of haptic mat motor arrangement. F) Spatial sound location pilot setup. G) Just noticeable difference pilot with avatar onscreen.

as video sharing [2], [6], point cloud networking—evident in projects like ‘Wish You Were Here’ [7]—and networked musician avatars, as implemented in Augmented Reality Drum Circle [8], have emerged. Studies have shown that these visual components can significantly enhance co-presence and communication among musicians, thus elevating the level of interaction and immersion in these networked music experiences.

There has also been growing interest in networking spatial sound, haptics, and syncing these various sensory modalities over a network [9]–[11]. Therefore, we conduct pilot studies with haptics to gain a better understanding spatial perception in networked musical contexts.

B. AI Generated Music

Generated music encompasses the integration of artificial intelligence in various aspects of music technology, expanding the scope of music creation and interpretation. Tools which leverage AI to generate music, provide a new form of creativity for musicians to explore [12]–[15]. These developments represent a potential revolution in music composition, bringing flexibility and diversity to the process.

C. Embodied AI Music

The realm of musical improvisation has also seen the integration of AI, with innovations like Shimon, an improvising robotic marimba player [16]. Robotic drummers, keyboardists, marimba, and flute players demonstrate the breadth of robotic instrumentalists created by various research institutes [17], [18]. While the domain of embodied AI music is relatively new to digital embodiment, it harbors substantial potential for exploration. This research seeks to delve into this exciting intersection, examining the possibilities of integrating AIVMs [19], [20] into networked jam sessions. This exploration transcends backing tracks, aiming to provide musicians with dynamic, interactive partners in the creative process.

III. SYSTEM DESIGN

The system design section is separated into two subsections to explain how the system can provide a networked musical experience between multiple players and AIVMs (see Figure 1). The first section describes networking the AR content,

and the second describes the AI pipeline and how networking occurs between players and AIVMs (see Figure 2).

A. The AR Experience

The system incorporates several key components to deliver its functionality. It utilizes the Unity game engine, a versatile platform for creating interactive experiences in XR. NReal AR glasses are leveraged for augmented reality overlays on the real world, enabling players and AIVMs to be embodied as virtual avatars. The NReal AR glasses also provide front-facing cameras that track the hands as players are playing an instrument, enabling musical gestures to be networked for communication among musicians (see Figure 3).

The system leverages the Photon Engine Unity plugin [21] for seamless networking capabilities, facilitating multiplayer functionality and collaborative experiences. Photon provides servers and server distance optimization to ensure messages are sent and received in under a second [21]. This enables a more seamless experience for networking hand-tracking information, as well as musical gestures of the AI avatar.

Near-instantaneous audio is provided by Jacktrip [2] server optimizers and a dedicated audio system for a low latency experience among remote players within roughly 1000 kilometers [2], [4], [5]. Together, these components create a comprehensive system that combines augmented reality with near-instantaneous audio collaboration.

B. The AI Pipeline

The AI pipeline is built using a combination of python, Ableton, and Jacktrip. Python, running Long Short-Term Memory Recurrent Neural Networks (LSTM RNNs) that are trained on databases such as the Magenta Groove MIDI database, enable fast generation and output of music (see Figure 2).

The generated music is then passed to a virtual port that is connected to the digital audio workstation, Ableton Live 11 [22]. Ableton Live 11, a renowned music production software, is integrated to further control sound quality of a player or an AIVM. Though Ableton is not required to make the system function, it does render high quality sound and enable virtual port management better than most digital audio workstations. Ableton Live 11 also provides opportunity for integration with

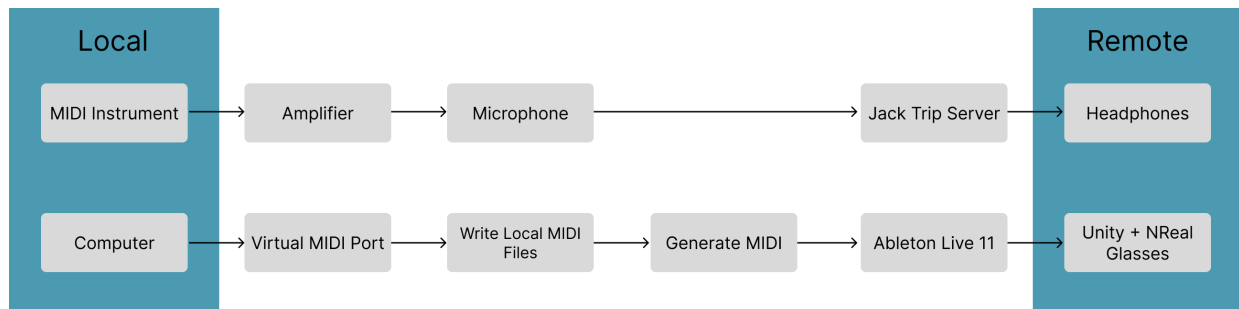


Fig. 2. System diagram outlining the signal flow.

Max for Live, which can be further development for nuanced control of output for the purposes of improvising.

C. Networked Haptics

We developed a communication pipeline to link a local player's MIDI instrument with a remote haptic device, using Python, OSC, and the Photon Unity plugin. The script on the local computer captures MIDI data and sends it via OSC to a Photon server. The server forwards the message to a remote computer connected to an Arduino, which triggers the vibration motors on the remote haptic device.

The haptic mat is made of a kitchen mat, which was chosen for its rigidity and comfort (see Figure 1). 20 vibration motors are grounded in series along the edges of the mat, and each is individually addressed by the microcontroller for independent control. We used a Teensy 4.0 to provide a slim profile that was embedded in the mats floor-facing side. The Teensy also provides a sufficient number of the necessary pins to drive all of the motors (see Figure 1).

To provide spatial awareness in the XR environment, we created a Haptic Mat with vibration motors arranged in a circle around the player's feet. The motors correspond to the direction of sound in the XR space, enabling the local player to sense the locations of remote participants. Preliminary tests were conducted to measure the effectiveness of the haptic feedback for spatial orientation. Details of these tests are provided in the following sections.

IV. PRELIMINARY FEEDBACK INTERACTING WITH AI

We assessed preliminary feedback based on an exposition-style showcase of the technology. Nearly 200 people viewed the system, and roughly 15 interacted with a virtual drummer in a staged networked environment. Our goal was to receive preliminary feedback about the usability of the system, explore the possibilities of the technology for use in several settings, and to gauge future directions for development. Participants were asked to provide feedback in the form of a 150-word free response and a questionnaire gauging co-presence using an aggregated measure, and five chose to leave detailed responses. We are not treating results as conclusive, rather are probing for themes and points of improvement for future iterations as well as experiments with the system.

We used inductive and deductive thematic methods to distill information from the player responses [23], [24]. We had limited feedback so we focused more on deductive logical strategies where emergent themes are more relevant.

Themes Identified: **Musical Proficiency** emerged prominently, with players lauding the system's capabilities and drawing parallels to iconic musicians like the Pink Floyd drummer" and Ozzy in his prime." **Technological Synergy** was another key theme, as feedback highlighted the system's adeptness at synchronizing with players' music, delivering sound effectively, and adapting to player input, thereby signifying the potential for a seamless technological integration in musical settings. **Interactivity** was emphasized by players describing the experience as highly interactive, noting the system's responsiveness to their musical inputs, thus enhancing engagement beyond traditional musical experiences. **Developmental Potential** was evident, despite some limitations like sound delivery issues and timing inconsistencies, with players optimistic about the system's future and its scope for improvement. **Emotional Resonance** was also noted, as the system elicited positive emotional responses from players, including feelings of excitement, enjoyment, and anticipation for future advancements. A noteworthy addition was **Social Comfort**, where players reported feeling less socially anxious and concerned when interacting with the AIVM compared to human players.

These themes collectively suggest that the system offers a promising avenue for enhancing musical experiences through technological integration, while also highlighting areas that require further refinement.

V. EXPLORING NETWORKING HAPTICS IN XR

A. Locating Objects in Space Using Sound and Haptics

We connected x,y coordinates of a sound source to the vibration motors on the haptic mat. Assuming the player is facing forward on the mat, the location of the sound source can be felt on the mat with respect to the position of the player (see Figure 1). Our pilot study was designed to determine what effect, if any, a haptic ring around one's feet had on spatialized audio perception. 12 people in our immediate environment participated in the pilot study. As the participants watched a planetoid object rotate around them in space, it produced an A440 tone which they heard through stereo headphones.



Fig. 3. A) Hand tracking as seen on remote avatar B) Hand tracking as seen through AR glasses. C) Removing pieces of instruments. D) Playing the instrument of another person. E) Deforming and manipulating virtual objects for play or sound. F) Customize the appearance and size of the AIVM or friends with gestures.

The audio was directional, so when the object was to their right, left, back, or front, the headphones output the audio accordingly. We compared their qualitative perception of the moving planetoid with and without haptics. We then had them close their eyes and guess where the planetoid for both conditions. These results are intended to provide insight for further exploration rather than provide statistically significant figures. The results are summarized in Table I.

TABLE I
COMPARATIVE DIFFERENCES BETWEEN CONDITIONS WITH AND WITHOUT HAPTICS

Direction	Without Haptics (%)	With Haptics (%)
Forward	43	71
Left	100	100
Right	100	100
Back	57	57

VI. SUMMARY OF XR-ENABLED MUSICAL INTERACTIONS WITH OTHER PLAYERS

We have used hand-tracking and inverse kinematics to control the motion of avatar hands while playing instruments, as well as translated head motion from the NReal to match the head orientation of the player's avatar. Other interactions that are unique to AR interaction such as stealing parts of player's instruments and changing their size are also detailed. These functionalities can be categorized into the following key ideas:

- **Object Manipulation:** Hand-tracking for avatar control, and unique AR interactions like resizing instruments.
- **Object Interactions:** Enables playful interactions such as throwing objects and stealing drumheads from other players.
- **Mesh Manipulation:** Enables players to play with virtual objects, deform exiting ones, overall interact with objects in a way not observed in jam sessions without XR.
- **Collaborative Drumming:** Facilitates social drumming experiences by enabling players to play each other's

virtual drums. Potentially teaching others drum beats or pushing social boundaries.

- **Player Customization:** Enables size and custom adjustments of drumming partners through hand gestures.

These functionalities not only add to the immersive experience but also open up possibilities for social interaction and collaboration among players not experienced in jam sessions without XR.

VII. DISCUSSION

Our integrated system presents a comprehensive platform for interactive musical experiences and spatial orientation within an XR environment. The introduction of AIVMs notably elevates traditional musical engagements, offering players the potential of virtually collaborating with iconic musicians or non-judgmental jam partners.

The XR framework extends the boundaries of conventional interactions, enabling players to explore an extensive array of novel possibilities. These range from object and mesh manipulations to social drumming, thereby overcoming the limitations inherent in live, physical settings. Additionally, the avatar-based interface serves as a tool for overcoming social inhibitions and anxieties, facilitating interaction.

Sound localization appeared to be relatively effective through audio cues alone, although haptic feedback showed potential for enhancing spatial orientation in specific circumstances. Front and back directions were found to be the most challenging for players to discern solely through audio. These findings suggest the possibility of enhancing sound localization in the system by further exploring the relationship between spatial haptics and audio.

Our system opens up new avenues for musical interaction, social engagement, and spatial orientation through a blend of AI, XR, and haptic technologies. While preliminary feedback and studies have been positive, they also indicate areas for further research and refinement, especially in terms of synchronization issues and player experience.

REFERENCES

- [1] G. Weinberg, "Interconnected musical networks: Toward a theoretical framework," *Computer Music Journal*, vol. 29, no. 2, pp. 23–39, 2005.
- [2] "Make music together online with jacktrip." <https://www.jacktrip.com/>.
- [3] "Jamulus." <https://jamulus.io/>, 2022.
- [4] L. Turchet and C. Fischione, "Elk audio os: an open source operating system for the internet of musical things," *ACM Transactions on Internet of Things*, vol. 2, no. 2, pp. 1–18, 2021.
- [5] G. Davies, "The effectiveness of lola (low latency) audiovisual streaming technology for distributed music practice," 2016.
- [6] R. Carlson and S. Hanna-Weir, "Conducting during covid," *The Choral Journal*, vol. 61, no. 9, pp. 65–73, 2021.
- [7] R. Schlagowski, D. Nazarenko, Y. Can, K. Gupta, S. Mertes, M. Billinghamurst, 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*, pp. 1–16, 2023.
- [8] 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*, vol. 3, p. 847284, 2022.
- [9] T. Hopkins, S. C.-C. Weng, R. Vanukuru, E. Wenzel, A. Banic, and E. Y.-L. Do, "How late is too late? effects of network latency on audiovisual perception during ar remote musical collaboration," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 686–687, IEEE, 2022.
- [10] C. Chafe, M. Gurevich, G. Leslie, and S. Tyan, "Effect of time delay on ensemble accuracy," in *Proceedings of the international symposium on musical acoustics*, vol. 31, p. 46, ISMA Nara, 2004.
- [11] N. Schuett and C. Chafe, "The effects of latency on ensemble performance - stanford university," 2002.
- [12] "Openai research, musenet, jukebox." <https://openai.com/research/>.
- [13] "Magenta." <https://magenta.tensorflow.org/>.
- [14] A. Agostinelli, T. I. Denk, Z. Borsos, J. Engel, M. Verzetti, A. Caillon, Q. Huang, A. Jansen, A. Roberts, M. Tagliasacchi, *et al.*, "Musiclm: Generating music from text," *arXiv preprint arXiv:2301.11325*, 2023.
- [15] "Bandlab songstarter." <https://www.bandlab.com/songstarter>. Accessed on 6/20/2023.
- [16] E. R. Miranda *et al.*, "Shimon sings-robotic musicianship finds its voice," in *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity*, Springer, 2022.
- [17] A. Takanishi, M. Sonehara, and H. Kondo, "Development of an anthropomorphic flutist robot wf-3rii," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS'96*, vol. 1, pp. 37–43, IEEE, 1996.
- [18] A. Kapur, "A history of robotic musical instruments," in *ICMC*, vol. 10, p. 4599, 2005.
- [19] 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*, vol. 3, 2022.
- [20] T. Hopkins, A. Jude, G. Phillips, and E. Y.-L. Do, "Virtual AI Jam: Ai-Driven Virtual Musicians for Human-in-the-Loop Musical Improvisation," *AIMC 2023*, aug 29 2023. <https://aimc2023.pubpub.org/pub/9064kmvw>.
- [21] "Photon engine." <https://www.photonengine.com/>. Accessed on [Date].
- [22] "Ableton." <https://www.ableton.com/>.
- [23] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006.
- [24] V. Clarke, V. Braun, and N. Hayfield, "Thematic analysis," *Qualitative psychology: A practical guide to research methods*, vol. 3, pp. 222–248, 2015.

ACKNOWLEDGMENT

The authors would like to thank Ericsson Research for supporting our work with funding and guidance throughout the project. The authors would also like to thank, Hooman Hedayati, Colin Soguero, Dan Szafir, Peter Gyory, and Darren Sholes for their thoughts and technical contributions.