Networking AI-Driven Virtual Musicians in Extended Reality

Torin Hopkins*
ATLAS Institute
University of Colorado Boulder

Rishi Vanukuru†
ATLAS Institute
University of Colorado Boulder

Suibi Che Chuan Weng‡
ATLAS Institute
University of Colorado Boulder

Chad Tobin§
ATLAS Institute
University of Colorado Boulder

Amy Banic¶
Interactive Realities Lab
University of Wyoming

Mark D. Gross‖
ATLAS Institute
University of Colorado Boulder

Ellen Yi-Luen Do**
ATLAS Institute
University of Colorado Boulder

Figure 1: Guitar player playing with an AI-driven virtual musician (VMAI) and a friend drummer (Chad) using the networked system in XR

ABSTRACT

Music technology has embraced Artificial Intelligence as part of its evolution. This work investigates a new facet of this relationship, examining AI-driven virtual musicians in networked music experiences. Responding to an increased popularity due to the COVID-19 pandemic, networked music enables musicians to meet virtually, unhindered by many geographical restrictions. This work begins to extend existing research that has focused on networked human-human interaction by exploring AI-driven virtual musicians' integration into online jam sessions. Preliminary feedback from a public demonstration of the system suggests that despite varied understanding levels and potential distractions, participants generally felt their partner's presence, were task-oriented, and enjoyed the experience. This pilot aims to open opportunities for improving networked musical experiences with virtual AI-driven musicians and informs directions for future studies with the system.

Index Terms: Augmented Reality—Networked Music—Artificial Intelligence—

*e-mail: Torin.Hopkins@colorado.edu
†e-mail: Rishi.Vanukuru@colorado.edu
‡e-mail: Suibi.Weng@colorado.edu
§e-mail: chto7865@colorado.edu
¶e-mail: abanic@cs.uwyo.edu
‖e-mail: mdgross@colorado.edu
**e-mail: Ellen.Do@colorado.edu

1 INTRODUCTION

For generations music has served as a connecting force between people and cultures around the world. Traditionally, musical play is experienced as an in-person, social activity. With developments in audio networking and computational resources, musicians have been empowered to experience musical connection at a distance [3, 5, 9, 11, 15, 16, 27]. Research in this domain has focused mostly on auditory experience, with the capability for video sharing [5, 13], point cloud networking [24,25], and networked musician avatars [18] only appearing recently. Visual content has been shown to enhance co-presence, enhance communication capability, and provide interactions that are otherwise not possible using auditory information alone [18]. However, over the last few decades, one such transformation has been brought about by the inclusion of shared virtual spaces in musical interaction, enabling musicians to experience networked music that more closely resembles the in-person experience [17, 18, 21, 25].

More recently, integration of Artificial Intelligence (AI) in music experiences have enabled musicians to expand their creativity [2, 5, 22]. This has been explored in many facets of music technology, including generated music [2, 4, 6, 10], computer-supported composition [8, 20], embedded AI in instruments [28], integration of AI into musical devices [23], and even AI-driven musical improvisation [5, 22].

Currently, some networked music platforms afford the ability to generate music while jamming over the internet [5]. However, we see opportunity to create a networked experience where players not only

2771-1110/23/$31.00 ©2023 IEEE
DOI 10.1109/ISMAR-Adjunct60411.2023.00162

can play with generated backing tracks, but can include AI-driven virtual musicians (AIVMs) into their jam sessions. Networked AIVMs are more than backing tracks, defined as prerecorded and unchanging generated music, they are companions in the creative music making process.

Throughout this paper we propose an AIVM networked system that enables musicians can play music with AIVMs and human musical companions simultaneously. We briefly describe related research that has directly informed our work, an overview of the system architecture, preliminary feedback we received from a public showcase of the system, and future directions and improvements of the AIVMs.

2 Related Work

2.1 Networked Music Experiences

Near-instantaneous networked music experiences represent a pivotal shift in how musicians collaborate and perform, overcoming many geographical limitations [29]. Work in this area is predominantly focused on enabling high-quality audio sharing within a distance of around 1000 kilometers [5, 9, 16, 27]. 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 as video sharing [5, 13], point cloud networking—evident in projects like 'Wish You Were Here' [25]—and networked musician avatars, as implemented in Augmented Reality Drum Circle [18], 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.

2.2 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 [2, 4, 6, 10]. These developments represent a potential revolution in music composition, bringing flexibility and diversity to the process.

Alongside music generation, AI has also found a place in computer-supported composition [8, 20]. These platforms incorporate AI to assist in the music creation process, providing suggestions and improvements based on complex algorithms. Additional advancements include the embedding of AI into musical instruments and hardware [23, 28]. This integration opens up novel possibilities for musical interaction, performance, and composition.

2.3 Embodied AI Music

The realm of musical improvisation has also seen the integration of AI, with innovations like Shimon, an improvising robotic marimba player [22]. Robotic drummers, keyboardists, marimba, and flute players demonstrate the breadth of robotic instrumentalists created by various research institutes [19, 26]. 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 AI-driven virtual musicians into networked jam sessions. This exploration transcends backing tracks, aiming to provide musicians with dynamic, interactive partners in the creative process.

3 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. 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).

3.1 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 1 and 3).

The system leverages the Photon Engine Unity plugin [7] 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 [7]. 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 [5] server optimizers and a dedicated audio system for a low latency experience among remote players within roughly 1000 kilometers [5, 16, 27]. Together, these components create a comprehensive system that combines augmented reality with near-instantaneous audio collaboration.

3.2 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 [1]. 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. Ablton Live 11 also provides opportunity for integration with Max for Live, which can be further development for nuanced control of output for the purposes of improvising.

4 Preliminary Feedback

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 5 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.

4.1 Emergent Themes

We used inductive and deductive thematic methods to distill information from the player responses [12, 14]. We had limited feedback

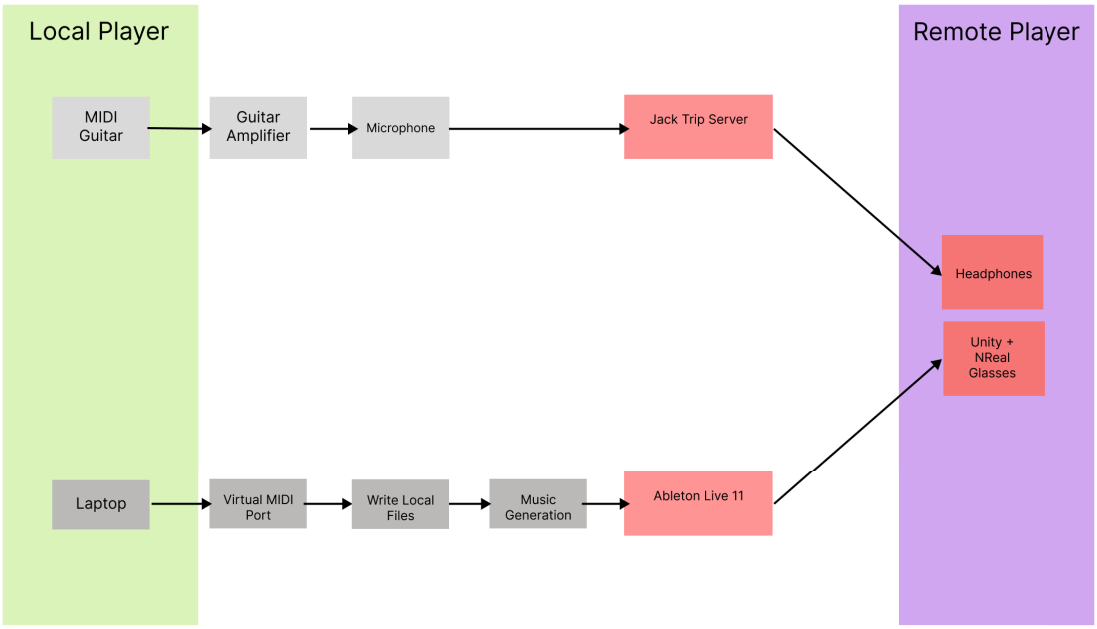

Figure 2: System diagram outlining the signal flow.

so we focused more on deductive logical strategies where emergent themes are more relevant. **These statements revolve around several central themes:**

4.1.1 Music and Instrumental Skill

Each of these statements reflects the players' respective experience with music and their instrument skills. They evaluate the quality and synchrony of the drum beats, the guitar sound, and their interplay. The references to the "Pink Floyd drummer" and the comparison of the beats to "Ozzy in his prime" suggest an appreciation for musical skills and icons.

4.1.2 Technology Integration in Music

The players are discussing a system - potentially an AI - that responds to and matches their music. They mention things like syncing, feedback responses, and sound delivery. They also comment on the potential for improvement and seamless integration. This theme indicates the integration of technology in music and its potential to improve networked music creation by optimizing musical timing.

4.1.3 Interactive Experience

The players describe their experiences as interactive - the system is not just playing music, but they are influencing it and being influenced in return. They note that the system guided them, and one speaker imagined playing with the system for hours. This suggests a level of engagement and interaction that goes beyond traditional musical experiences.

4.1.4 Potential and Development

The players saw potential in the system despite acknowledging its current limitations and challenges (e.g., difficulties in sound delivery, switching beats at odd times). They are optimistic about future improvements, indicating a theme of potential and continuous development.

Figure 3: A guitar player jamming with an AIVM guitar player.

4.1.5 Emotional Response

The players express their feelings in response to the music and the technology. They mention feeling good, being excited, and having fun. They appreciate the innovative approach and look forward to future advancements. This theme underscores the emotional impact of music and technology.

4.2 Aggregated Co-Presence

The aggregated co-presence scores from the data was assessed based on the questionnaire answers of 5 people (ages between 22-30 years old, with an average of 7 years playing guitar). We did not attempt to draw any conclusions from this data, rather it serves as a preliminary probe for assessing how players feel about the embodied aspects of the AIVM. Scores were given on a scale of 1-7, whereby 7 represented the most, 4 neutral, and 1 the least with respect to the question.

Several discernible themes emerge, providing insights into the dynamics of co-presence with the AIVM in the context of this study.

A theme of the partner's presence and noticeability surfaced. Participants generally reported that they were aware of their partner's

presence during their interaction, as indicated by average scores of 5.25 and 5.00 on "I noticed my partner" and "My partner's presence was obvious to me", respectively.

Understanding and clarity form another significant theme. While participants found it relatively easy to understand their partner with an average score of 4.25, other aspects such as discerning their partner's thoughts and focus showed more varied results, with average scores of 2.50 and 2.50 respectively. This disparity suggests potential challenges in communication or differences in communication styles.

The theme of attention and focus emerged as well. While some participants were easily distracted, reflected by an average score of 3.75 on "I was easily distracted from my partner when other things were going on", others reported maintaining focus on their partner throughout the interaction, with an average score of 4.75. This variation suggests a range of experiences among participants in terms of focus and attention management.

Moreover, participants generally indicated that they were able to effectively focus on their task activities, as indicated by an average score of 5.5. This indicates that, despite potential distractions and varying levels of understanding, task orientation remained relatively high.

Another critical theme pertains to enjoyment, with participants generally reporting a positive experience (average score of 6.25 of 7), signifying a favorable overall interaction despite the challenges encountered in some areas.

These findings offer useful pointers for potential improvements in collaborative task or environmental design to enhance mutual understanding and limit distractions. Further studies might delve deeper into individual differences that contribute to the variations in social presence experiences observed here.

5 Future Directions

In light of the preliminary feedback and thematic analysis, we propose several avenues for further research and development in the context of this AI-based music system. Foremost among these is the pressing need for technical refinement. Players noted several issues, such as timing of beat transitions, abrupt mid-phrase changes in speed, and complications with sound delivery. By addressing these aspects, we can ensure a more seamless and intuitively musical interaction with the AI system.

There are limitations in-built to the system, such as the inability for both players to influence the AIVM. In the current system only one player has control of the AIVM and is passing the local generated sound to the remote player using the near-instantaneous audio networking service, JackTrip. It is assumed that both players are in sync during the music jam and that it is tenable for the AIVM to only listen to one of the two players. In future studies we plan to investigate the interactions between the remote player and the AIVM as the exposition-style feedback did not lend itself to in-depth exploration of the dynamics between multiple players and the AIVMs.

The interactive nature of our AI system has received positive feedback. However, players indicated a somewhat prolonged period before the interaction truly engaged them. Future work should thus focus on enhancing the system's responsiveness and to bolster the sense of real-time interaction. Concurrently, the user interface could be improved to make the initial experience more accessible and engaging for players. This might entail a more straightforward interface or the provision of intuitive tutorials to streamline the user's entry into the system.

Machine learning optimization could significantly enhance the system's ability to adapt to each musician's unique rhythm, speed, and style. Personalizing the interaction in this manner could amplify the user satisfaction and lead to a more rewarding musical experience. We also suggest considering the integration of emotional feedback mechanisms. As our feedback highlighted, players experienced an emotional response to their interactions with the system. Therefore, developing features that can recognize and respond to the emotional tone of the music might lead to even more immersive and gratifying player experiences.

By addressing these areas, we envision our system evolving into a more seamless, intuitive, and enjoyable tool for musical expression and exploration for collaborative networked AIVMs. This research holds promise for a future where AI significantly enriches the landscape of musical interaction and creativity.

Acknowledgments

This work was supported in part by a grant from Ericsson Research. The authors wish to thank the Ericsson Research team for their consistent and helpful feedback throughout the project. The authors would also like to thank Sasha Novack, Emma Wenzel for their support in development and testing throughout the building process. Additionally the authors would like to thank the participants for donating their time to improving the system.

References

[1] Ableton. https://www.ableton.com/.

[2] Bandlab songstarter. https://www.bandlab.com/songstarter. Accessed on 6/20/2023.

[3] Jamkazam: Live, in-sync music jamming over the internet. https://jamkazam.com/.

[4] Magenta. https://magenta.tensorflow.org/.

[5] Make music together online with jacktrip. https://www.jacktrip.com/.

[6] Openai research, musenet, jukebox. https://openai.com/research/.

[7] Photon engine. https://www.photonengine.com/. Accessed on [Date].

[8] Create amazing music with hooktheory. https://www.hooktheory.com/, 2022.

[9] Jamulus. https://jamulus.io/, 2022.

[10] 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.

[11] N. Bouillot. Njam user experiments. In *Proceedings of the 7th International Conference on New Interfaces for Musical Expression - NIME '07*, 2007.

[12] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[13] R. Carlson and S. Hanna-Weir. Conducting during covid. *The Choral Journal*, 61(9):65–73, 2021.

[14] V. Clarke, V. Braun, and N. Hayfield. Thematic analysis. *Qualitative psychology: A practical guide to research methods*, 3:222–248, 2015.

[15] J.-P. Cáceres and C. Chafe. Jacktrip: Under the hood of an engine for network audio. *Journal of New Music Research*, 39(3):183–187, 2010.

[16] G. Davies. The effectiveness of lola (low latency) audio-visual streaming technology for distributed music practice. https://www.academia.edu/28770528/The_effectiveness_of_LOLA_LOw_LAtency_audiovisual_streaming_technology_for_distributed_music_practice, 2016.

[17] T. Hopkins, S. C.-C. Weng, R. Vanukuru, E. Wenzel, A. Banic, M. D. Gross, and E. Y.-L. Do. Studying the effects of network latency on audio-visual perception during an ar musical task. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 26–34. IEEE, 2022.

[18] 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.

[19] A. Kapur. A history of robotic musical instruments. In *ICMC*, vol. 10, p. 4599, 2005.

[20] E. Łukasik. Creative intersections of computing and music composition. In *Knowledge, Information and Creativity Support Systems: Recent Trends, Advances and Solutions: Selected Papers from KICSS'2013-8th International Conference on Knowledge, Information, and Creativity*

Support Systems, November 7-9, 2013, Kraków, Poland, pp. 553–563. Springer, 2016.

[21] L. Men and N. Bryan-Kinns. Lemo. In *Proceedings of the 2019 on Creativity and Cognition*, 2019. doi: 10.1145/3325480.3325495

[22] 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.

[23] T. Pelinski Ramos, R. Diaz Fernandez, A. Benito Temprano, et al. Pipeline for recording datasets and running neural networks on the bela embedded hardware platform. 2023.

[24] R. Schlagowski et al. Jamming in mr: Towards real-time music collaboration in mixed reality. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2022. doi: 10.1109/vrw55335.2022.00278

[25] 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*, pp. 1–16, 2023.

[26] 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.

[27] 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*, 2(2):1–18, 2021.

[28] L. Turchet, A. McPherson, and M. Barthet. Real-time hit classification in a smart cajón. *Frontiers in ICT*, 5:16, 2018.

[29] G. Weinberg. Interconnected musical networks: Toward a theoretical framework. *Computer Music Journal*, 29(2):23–39, 2005. doi: 10.1162/0148926054094350