Jamming in MR: Towards Real-Time Music Collaboration in Mixed Reality

Ruben Schlagowski*
HCAI-Lab, University of Augsburg

Kunal Gupta†
Empathic Computing Lab, University of Auckland

Silvan Mertes‡
HCAI-Lab, University of Augsburg

Mark Billinghurst§
Empathic Computing Lab, University of Auckland

Susanne Metzner¶
LMZ, University of Augsburg

Elisabeth André‖
HCAI-Lab, University of Augsburg

ABSTRACT

Recent pandemic-related contact restrictions have made it difficult for musicians to meet in person to make music. As a result, there has been an increased demand for applications that enable remote and real-time music collaboration. One desirable goal here is to give musicians a sense of social presence, to make them feel that they are "on site" with their musical partners. We conducted a focus group study to investigate the impact of remote jamming on users' affect. Further, we gathered user requirements for a Mixed Reality system that enables real-time jamming and developed a prototype based on these findings.

Keywords: Mixed Reality, Remote Collaboration, Music

Index Terms: Human-centered computing—Visualization—Visualization design and evaluation methods

1 INTRODUCTION

For many creative people, music allows them to express themselves and connect with others. However, during the global COVID pandemic, many musicians have been deprived of sharing their music with others in person. While remote solutions exist for a variety of social activities in other application domains (e.g. Zoom for online video conferences), solutions for real-time collaborative music-making are still sparse. Existing low-latency solutions, such as Jamulus [1], use a high bandwidth to reduce latency. This prevents musicians from seeing each other, via a parallel video stream and so limits the sense of another person's presence. This feeling of social presence, or of "being there with another person" [1] has been achieved elsewhere by integrating motion-tracked avatars representing users in Mixed Reality (MR) environments such as Augmented Reality (AR) or Virtual Reality (VR) [3,4].

To prototype a system that supports Social Presence for real-time music collaboration, we used focus group interviews to collect musicians' requirements for a MR based solution. As part of our focus group, we had musicians jam online using an existing browser application and report how they felt before, during, and after the jam session. This was to understand the impact of remote music collaboration on musicians' emotional state and affect.

From the results of the focus group, we conceptualized and developed a prototype that enables musicians to jam with each other's 3D point clouds in MR while using their preferred physical musical instruments. Our prototype makes use of traditional analogue low-latency audio signals between adjacent rooms and video see-through AR technologies. We regard it as a glimpse into the future when network expansion is sufficiently advanced to better enable comparable low-latency communications over the internet.

*e-mail: ruben.schlagowski@uni-a.de
†e-mail: kgup421@aucklanduni.ac.nz
‡e-mail: silvan.mertes@uni-a.de
§e-mail: mark.billinghurst@auckland.ac.nz
¶e-mail: susanne.metzner@uni-a.de
‖e-mail: elisabeth.andre@uni-a.de

[1]https://jamulus.io/

2 FOCUS GROUP STUDY

We recruited 15 musicians (7 female, 8 male, Average Age: 35.9 (SD=12.4)) from Germany and New Zealand through mailing lists. Their musical expertise ranged from professional to beginner. The participants were paid for their efforts. Our study goals were to (1) identify requirements for a MR system that enables remote real-time jamming and (2) to investigate the impact of a short remote jams session on the positive/negative affect of participants. For the latter, we let participants jam together online with PLINK![2]

2.1 Study Procedure

The moderated group interviews (three Zoom-sessions lasting approx. one hour) started with questions on the impact of the pandemic on the musician's musical activities and wellbeing. Subsequently, a online jam session of about 10 minutes took place, during which participants used PLINK! to make music together.

After this stimulus, participants were asked how they felt during and after the jam session. Finally, there was a short presentation introducing the basics of VR, a video about an existing VR jamming app [3], and participants were asked about their opinions and requirements for such a system. All interviews were recorded, transcribed and a thematic analysis was conducted using the MaxQDA software [4]. From the resulting quantification of reoccurring themes (codes), code clouds were generated (Figure 1 and 2).

2.2 Findings

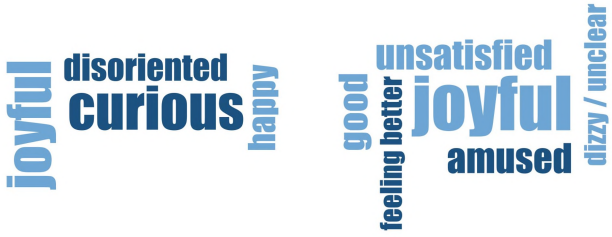

Figure 1: Code clouds of the thematic analysis of emotions participants felt during (left) and after the jam (right). Code size reflects the amount of times it was mentioned during focus group interviews.

[2]https://plink.in/
[3]https://store.steampowered.com/app/732480/Jam_Session_VR/
[4]https://www.maxqda.com/

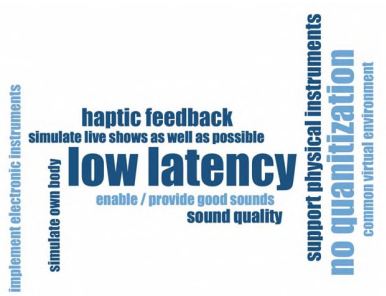

Figure 2: Code cloud of the requirements for the MR system.

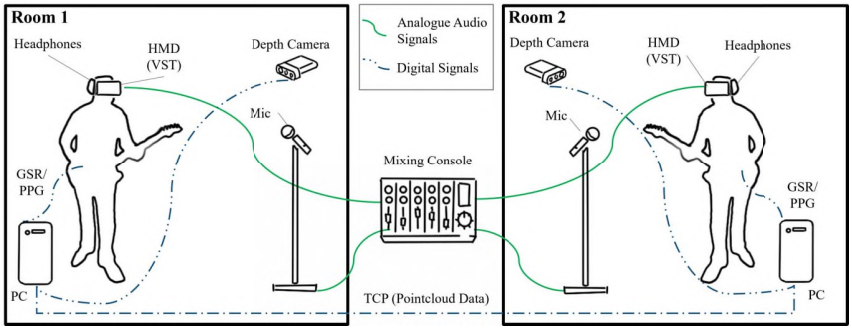

Figure 3: Technical Setup of the MR prototype that was developed based on the user requirements.

After the jam session, participants reported that they felt curious, joyful and happy, but also disoriented, which was largely caused by their inexperience with the application (cf. left code cloud in Figure 1). When being asked to describe their feelings after the jam-session (cf. right code cloud in figure 1), six participants said that they felt joyful, others said they were feeling better, good or even amused. Three participants said that the browser-based interface was not in line with their expectations for a real-time jamming application and emphasized that such interaction "cannot be compared to playing with real physical instruments".

The greatest user requirement for a MR based jamming application was low latency (mentioned 14 times). Another popular requirement was not to temporally quantize the audio data or the notes played, as is the case with some existing applications. Four participants stated that they would like to perform with their actual physical instruments and five participants emphasized the importance of haptic feedback while playing musical instruments. Another reoccurring theme was the importance of good sound quality, either in general (three mentions) or regarding simulated instruments (two mentions). There was a lively discussion regarding the visual representation of the jamming partner. Some participants had concerns regarding the use of virtual avatars resembling the jamming partner, which could be perceived as being "a little creepy". On the other hand, some participants undermined the importance of visual representations of their jamming partners. For instance, one participant stated, that if she "could find herself standing amongst other musicians [...]" this would "improve [her] experience" and "simulate a feeling of interpersonal relationships".

3 MR Prototype

We addressed the most predominant requirements raised by the participants in our prototype design (see Figure 3 for the technical setup). Low latency and high sound quality were achieved by using analogue signal chains and a classical mixing console, connecting two adjacent and isolated rooms. As such, each participant was able to have a custom mix of two microphones capturing both the physical instruments and voices on their ears (via bodypacks & headphones). We used the stereoscopic front-facing cameras of two HTC Vive pro VR headsets to provide musicians with a real-time view of their surroundings including their physical instruments. These video streams were augmented with colorized point clouds (no avatars) of their respective jamming partners. The point cloud data was captured by Microsoft Azure Kinect cameras, transmitted via TCP to the other room and rendered in real-time with a shader (see Figure 4) in a game engine (Unity3D). During first tests of the prototype, we achieved a stable point cloud frame rate of 15 frames per second without any noticeable latency fluctuations in jam sessions lasting up to 20 minutes. Participants of these early tests gave positive feedback and stated that seeing point clouds of their jamming partners were "beneficial to the overall experience" and increased their likeliness to "get into the flow".

Figure 4: Two jamming partners' views during a remote jam session with the prototype.

4 Conclusion and Future Work

This poster gives a summary of the results of a focus group, aiming to (1) investigate the emotional benefits of remote music collaboration and (2) to gather user requirements for a MR application that fosters social presence (co-presence) in remote music collaboration scenarios. Based on collected user requirements, we conceptualized and implemented a prototype that received positive feedback from musicians during early tests. In the future, we will investigate the impact of our prototype on user affect (positive and negative), user feeling of Social Presence with their partner, and the likeliness of experiencing the (desirable) state of flow [2]. We also want to use recorded physiological (GSR/PPG), body tracking and audio data to investigate correlations with the participants' sensation of flow during the jam sessions.

Acknowledgments

This research was partly funded by the Bundesministerium für Bildung und Forschung (BMBF, Ministry of Education and Research) in the DIGISTA project (no. 01UO1820A).

References

[1] F. Biocca and C. Harms. Defining and measuring social presence: Contribution to the networked minds theory and measure. *Proceedings of PRESENCE*, 2002:1–36, 2002.

[2] M. Csikszentmihalyi and M. Csikzentmihaly. *Flow: The psychology of optimal experience*, vol. 1990. Harper & Row New York, 1990.

[3] P. Heidicker, E. Langbehn, and F. Steinicke. Influence of avatar appearance on presence in social vr. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 233–234. IEEE, 2017.

[4] B. Yoon, H.-i. Kim, G. A. Lee, M. Billinghurst, and W. Woo. The effect of avatar appearance on social presence in an augmented reality remote collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 547–556. IEEE, 2019.