Key Considerations for Duo Singing in Virtual Reality and Videoconferencing: An Exploratory Study with Bigscreen and Zoom

Ben Loveridge
The University of Melbourne
Melbourne, Australia
b.loveridge@unimelb.edu.au

***Abstract*—This study investigates the key considerations of duo singing in virtual reality (VR) and videoconferencing (VC) settings. The COVID-19 pandemic disrupted in-person musical collaborations, leading performers to explore online alternatives. While VR offers a novel visual medium for musical collaboration, its potential in networked music performance (NMP) has been underexplored compared to VC. In this study, ten participants from Australia and the United States, equipped with internet access and VR headsets, engaged in leader-follower singing sessions in both VR (Bigscreen) and VC (Zoom) environments. A thematic analysis of semi-structured interviews revealed five key themes: presence, performance, affect, usability, and usefulness. Participants reported feeling more co-present and less inhibited in VR, fostering a greater sense of enjoyment and creativity, despite issues with facial expression fidelity and physical comfort. Conversely, while VC offered more realistic visual cues, it heightened self-consciousness in less experienced singers. Overall, participants favoured singing in VR, noting its ability to reduce inhibition while providing a novel environment. This study expands our understanding of visual communication in online music collaboration and offers insights for performers and designers of immersive musical experiences.**

Keywords—virtual reality, videoconferencing, networked music performance, duo singing

I. Introduction

With the wide availability of consumer videoconferencing (VC) technology, communication across the globe has become a reality. Immersive technologies such as virtual reality (VR) can allow us to feel present with others, despite physical separation. These advances have facilitated worldwide creative collaborations, with the underlying complexity often hidden from users, making communication appear instant. However, realistic musical interactions are often hindered by latency, an issue exacerbated by the COVID-19 pandemic, which forced musicians to find online collaboration solutions.

Improvements in broadband connectivity have enabled remote musical collaboration, also known as networked music performance (NMP) [1]. These systems allow geographically separated musicians to rehearse, perform, and teach in real-time over the internet [2]. Yet, strategies are required to compensate for network latency, due to geographic distance, network performance, musical synchronisation, and bandwidth [3]. These strategies might involve realistic interactions between performers or leader-follower styles, however managing out-of-sync visual cues often results in ignoring the video image to maintain rhythm [4]. Achieving low-latency video requires significant bandwidth and specialist networks, increasing setup and costs.

Determining the amount of necessary visual communication is a complex and underexplored area [5]. When geographical distance restricts physical interactions, appropriate ways to capture, transmit, and reconstruct musical interactions are needed. To date, VC technology has been the most common approach for visual communication online. However, recent advances have enabled new musical expressions using a range of technologies such as VR, Augmented Reality (AR) and Mixed Reality (MR), often classified using the umbrella term known as Extended Reality (XR). Through these explorations, new types of musical styles and experiences in immersive environments have emerged, giving rise to phrases such as Musical XR [6] and the Musical Metaverse [7,8].

While prior NMP studies have focused on the latency effects on singing [9, 10, 11], and the impact of video links [3, 5], few studies have examined VR and VC together in NMP contexts. Exploring both VR and VC can help guide platform selection for online music collaboration [12]. A detailed analysis by [6] underscored the need for further research in key areas such as collaboration, communication, sound perception, presence, virtual bodies, and virtual environments.

This study aims to advance research in the field by examining the key factors influencing duo singing in VR and VC environments. It seeks to explore how these platforms affect participant experience, performance quality, and overall usability, addressing the central question: What are the key considerations for duo singing in VR and VC settings?

This research contributes to the broader field of NMP by providing insights into the affordances and challenges of using VR and VC for musical collaboration. The findings can help inform creators of immersive musical worlds in experience design, as well as educators and performers exploring new platforms for music training and collaboration.

II. METHOD

Natural interaction in NMP, which mimics musicians playing together in the same physical space, typically requires low-latency audio and an echo-free connection [13]. This study was initially designed with performers located in the same building, but the COVID-19 pandemic forced a shift to a fully remote format. As a result, the originally planned realistic interaction model became impractical due to the complexity of running a dedicated NMP audio solution like Jacktrip [14]. Instead, the study adopted a leader-follower approach (LFA), where a remote performer follows the leader's musical output [15]. While this method created an out-of-sync experience for the leader (in this case the researcher), it allowed for a more affordable and accessible setup for participants. The follower experienced no apparent

network latency, enabling a synchronised, one-way interaction that closely resembled an in-person collaboration.

A. Participants

Ten participants were recruited for this study from Australia and the United States, (five males and five females, mean age of 38.25 ± 10.08). Based on the definition of a musician as someone with at least six years of musical expertise [16], the group comprised of seven musicians and three non-musicians. Including the perspectives of less experienced musicians was considered valuable for obtaining a broader a range of insights.

Five participants reported no prior experience with online singing, while two indicated having very little, and the remaining three had some degree of prior experience. Seven participants had minimal experience with VR wheras three reported a moderate level of prior experience. Due to the impact of the COVID-19 pandemic, the study was conducted entirely online in 2020 and 2021. Participants were required to have access to the internet and a VR headset. Table I lists the devices and equipment each participant used in the study. The study was approved by the Fine Arts and Music Human Ethics Advisory Group at The University of Melbourne.

TABLE I. *PARTICIPANT DEVICE AND EQUIPMENT INFORMATION*

ID	VR device	Own VR headset	VR audio headphones	VC device type	VC audio headphones
1	Quest 1	No	No	Laptop	Yes
2	Quest 1	Yes	Yes	Laptop	Yes
3	Quest 1	No	No	Laptop	Yes
4	Quest 2	Yes	No	Phone	No
5	Quest 2	Yes	No	Phone	No
6	Rift S	Yes	Yes	Laptop	Yes
7	Quest 2	Yes	No	Laptop	Yes
8	Vive	No	Yes	Laptop	No
9	Quest 1	No	No	Laptop	Yes
10	Quest 1	No	No	Laptop	No

B. Procedure

Upon recruitment, participants were invited to participate in duo singing sessions with the researcher. Each participant engaged in two sessions: one in a VC environment and one in a VR environment. The sequence of each setting was varied between participants to minimise potential order effects. Inspired by the idea of keeping development effort to a minimum to facilitate an efficient research process [17], pre-existing, off-the-shelf software applications were selected. Bigscreen, a VR application that allows customisable avatars in virtual environments, and the Zoom video conferencing tool. The familiar nursey rhyme *Twinkle Twinkle Little Star* was chosen to minimise any potential learning curve for participants.

In the VC setup depicted in Fig.1, a participant and the researcher are shown using the Zoom platform. In the VR setup shown in Fig.2, the participant wears a head-mounted display (HMD) and interacts with the researcher in the Bigscreen virtual campfire environment, designed to simulate a casual, co-present singing experience.

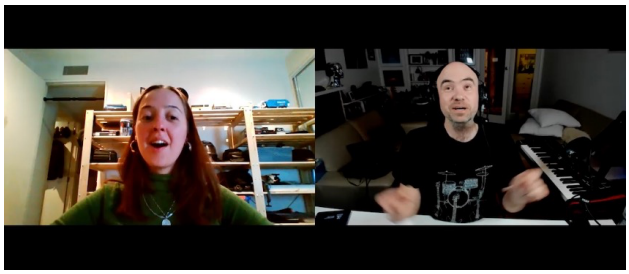

Fig. 1. *Participant (left) and researcher (right) singing together in Zoom.*

Fig. 2. *Participant (top left) and researcher (top right) singing together in Bigscreen*. Campfire view (bottom left) and participant's perspective (bottom right).

C. Data Collection and analysis

After each singing session, participants took part in a semi-structured interview to discuss their experiences across both settings. The interview was guided by the following questions:

- Can you describe your experience of singing during the session?
- Please describe any performance-related challenges or distractions?
- Can you describe any memorable experiences you had?
- Where did you look while you were singing?
- What was it about the experience that made you feel present or not present?
- Under what musical circumstance would you recommend this setting as an appropriate choice for musical collaboration?
- Do you have any other recommendations, improvements or comments about the setting?

This paper presents the qualitative findings from a broader mixed-methods study. While the quantitative data is not directly reported here, the preferred visual settings are referenced to complement and support the insights drawn from the interview data. The qualitative data from the interviews were analysed using thematic analysis [18], which involved coding the data to identify recurring themes and patterns related to the participant experiences in both the VR and VC environments.

III. RESULTS

Thematic analysis of the interview data identified 25 sub-themes, which were organised into five key themes: presence, performance, affect, usability, and usefulness (as shown in Fig.3). Each theme is discussed below, highlighting the similarities and differences between the VR and VC settings.

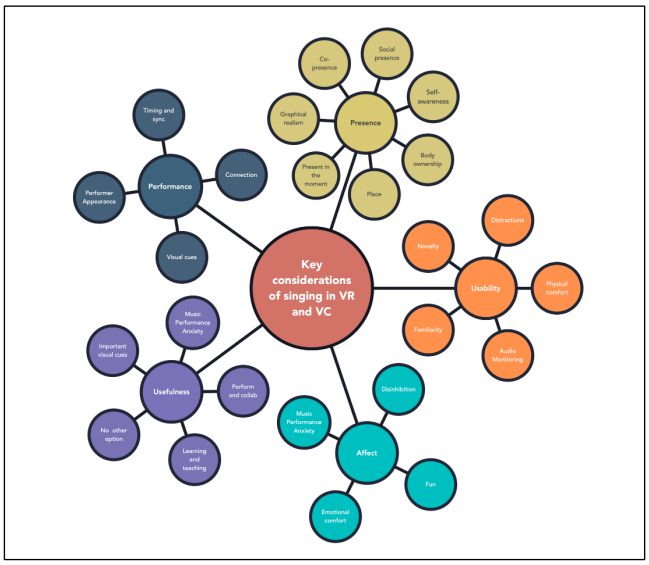

Fig. 3. *Key Considerations of Duo Singing in Virtual Reality and Video Conferencing*

TABLE II. *CHARACTERISTICS OF THEMES GENERATED THROUGH THE INTERVIEW CODING*

Theme	Characteristics
Presence	Related to sense of place, co-location, and self-awareness
Performance	Related to performative aspects with the other person
Affect	Emotional impact felt by participant
Usability	Affordances and issues with hardware and general usability
Usefulness	Related to future use cases and recommendations

A. Presence

Participants were asked to describe their experience of presence in each setting, with presence defined as the sensation of being in a different place while aware of their actual location. Data analysis identified seven sub-themes that characterised the experience of presence in both the VC and VR environments: place, copresence, social presence, body ownership, self-awareness, being present in the moment, and graphical realism.

The concept of 'place' surfaced as a significant theme in the interviews, encompassing both the physical and virtual locations where participants felt situated in each setting. Eight participants noted the sensation of being in a specific place within the VR context, with P1 remarking, "VR really does trick my brain into feeling like I'm somewhere else".

In contrast, only two participants referred to the VC setting as a distinct place. P8 reflected on this by stating: "Being just in a random room, not in a performance designated space…it was memorable in that, it's…not a memorable room to be performing in."

Participants reported a stronger sense of co-presence in the VR environment compared to the VC setting. In the interviews, eight participants in the VR setting described performing in the Bigscreen environment as if they were in the presence of the other avatar, while only two participants in the VC setting referred to copresence. As P9 noted about VR, "Feeling like I'm in the same room with you singing is gratifying. It's satisfying."

Half of the participants in the study mentioned aspects of body ownership in the VR setting, contrasting with none in the VC environment. In reference to Bigscreen, P10 noted "I was super aware of my hands, but like, not the rest of my body." Although participants could see their hands from a first-person view, they were not connected to fully rendered arms in Bigscreen. Additionally, when looking down in the VR environment, the absence of a virtual body or legs was described by one participant as being in a "dream".

Half of the participants reported reduced self-awareness in the VR setting, though generally describing it as positive experience that facilitated greater freedom for self-expression. P7 noted that being represented as an avatar in VR contributed to the feeling: "Because of the VR, you have that that degree of separation…you can either be yourself or be somebody else." Similarly, P2 commented, "There's something about wearing a mask…that…allows you to express yourself a bit more".

In contrast, participants who viewed their own faces on screen, due to Zoom's default self-view mode, became acutely aware of their own presence. Two participants specifically mentioned experiencing heightened self-awareness in the VC setting. When asked about his sense of presence, P9 commented: "In terms of presence, well, there's two parts. There's one of being aware of other's presence. And there's also a kind of awareness of your own presence or self-presentation."

Presence emerged as a key theme in both VR and VC settings, highlighting a complex relationship between body ownership, identity, and self-awareness. While co-presence with other avatars in the Bigscreen environment offered benefits, issues with body ownership detracted from the experience. Several participants expressed discomfort with seeing themselves mirrored in Zoom, while others expressed a desire for increased visibility of their own avatars in Bigscreen. The sense of invisibility in VR, enhanced by wearing both a virtual and physical mask, contributed to a greater sense of freedom and self-expression.

B. Performance

The second major theme identified in the interview data analysis relates to the performative aspects of the session. The term performance describes the musical interaction between the researcher and participant, even in the absence of a conventional audience. The sub-themes within this theme include performer appearance, visual cues, timing and synchronisation, and connection.

With the study implementing a leader-follower singing approach, participants could stay focused on the visual aspects of the experience without latency distractions. With the researcher leading the tune, participants could stay in time without worrying about the delayed audio returning from the other side. Participants felt that seeing the researcher's facial expressions and movements in real-time helped maintain a steady rhythm. From a performer appearance perspective, P3 commented: "When I can see you in video conferencing…it's

much more easy to be in time with someone when you can physically see them instead of in virtual reality".

In the VC setting, comments related to synchronisation and timing were mentioned by eight participants. P7 noted the ease of singing together in the leader-follower format: "There was no like, latency or hearing issues from my end. So it worked well for me." P3 mentioned that having access to the real-life person at the other end was beneficial: "I found it easier to kind of keep in time with you because I could see exactly what you were doing at that time."

A total of nine participants made comments relating to timing and synchronisation in the VR setting, P4 explaining: "I had the experience of being in-sync, whether it was or not." The importance of keeping time with the researcher was also mentioned by several participants. However, the VR environment presented challenges with timing due to the limitations of the avatar expressions with P3 remarking: "It's just the dead face…there's nothing going on, you don't see lips moving".

During the interviews, two participants in the VR setting described feelings of connection. P10 described this as: "There was more of a sense of connection in a different way to what one might get if it we were physically in the same room." P3 described their experience by commenting: "you are in the same location, experiencing the same things, but there isn't that same connection."

In the VC setting, four participants referred to a sense of connection during the session. P5 commented that singing with the researcher felt like "a real person connecting on a real person level." P3 described his sense of connection in the following way: "I feel like there is more of a connection visually with video conferencing, because…I can see you."

Overall, participants found it musically challenging signing with the VR avatar using the in-built facial cues, however they did find the hand movements useful. The detailed and true-to-life expressions present in the VC setting made the singing experience easier for timing, however the formality of the Zoom medium was noted as a negative aspect in a musical context.

C. Affect

The interview data revealed several overlapping factors influencing participants' feelings in both the VR and VC settings. These emotional states were categorised into the overall theme of affect and divided into the sub-themes: anxiety, emotional comfort, inhibition, and fun.

In the VR environment, half of the participants reported feeling less inhibited, attributing this to factors such as a sense of invisibility, surrealness of the environment and a lack of physical presence with the other performer. P10 described their time in VR as "a bit less inhibited, maybe because I couldn't see myself."

Responding to the nature of the VR campfire setting, P5 stated: "That environment is a little bizarre for me in the first place, but it allowed me, I believe, to be more uninhibited." P7 described the feeling of the VR setting by stating: "It was very, very homey and there's a lot more chance to make mistakes and not feel, not feel judged by it." In summing up his VR experience, P2 described the setting as: "it's just a bit more adventurous, it's fun."

Conversely, the Zoom environment made some participants feel self-conscious, particularly those with less singing experience: "I think I was maybe more nervous (in the Zoom setting). It felt more like a more formal situation" (P4). Alternatively, one participant (P1, non-musician), described how the absence of their physical presence allowed for more confidence: "I feel like without having someone immediately present in the same room as me, it does give you a bit more confidence to sing."

Only one participant mentioned the Zoom setting as being fun. This compared with the VR setting where six participants talked of the creativity and playfulness that the setting provided, explained by P6: "So singing in virtual reality seemed like, just being in a virtual environment kind of had the context of a little more like playfulness, more creativity in a sense."

D. Usability

The usability of the VC equipment and VR headset encompassed the following sub-themes: familiarity, audio, physical comfort, distractions, and novelty. As a remotely run study during the COVID-19 pandemic, participants were required to setup the equipment and run the software. Even with support available from the researcher online, challenges relating to this process were identified.

A lack of familiarity with using VR was mentioned by six of the participants, P6 expressing: "I was less familiar with Bigscreen and so I had more trouble figuring out the controls." Regarding their familiarity with Zoom, P6 said: "I didn't find as many distractions, but I don't know if that's simply because I'm familiar with Zoom."

In VR, physical comfort was an issue for some participants, with P10 commenting: "Physically, I find the headset a bit uncomfortable". The novelty of the VR experience, however, often outweighed these discomforts, P9 stating: "I just kind of felt like…this is the future".

The VC setting was found to be familiar to most participants, making it easier to use, but the constant on-screen visibility in Zoom was mentioned as off-putting with P8 stating: "It's a bit distracting…having my own face in front of me".

E. Usefulness

During the interviews, participants were asked about the circumstances under which they would recommend either setting for musical collaboration. Their responses indicated that both mediums could be suitable for singing when used in the right context. Suggestions were categorised into the following sub-themes: performance anxiety, performance and collaboration, lack of alternatives, essential visual cues, and learning and teaching. Participants mentioned preferring the campfire VR environment for casual, creative collaborations, while the VC setting was deemed more suitable for scenarios where precise visual cues were essential.

More than half of the participants recommended the VR setting for musical contexts where individuals might feel nervous or experience music performance anxiety. As P8 noted, "For someone who has really severe performance anxiety...I think it could be incredible because it felt like a really safe, calm space". The VC setting was mentioned by one participant as more appropriate for people who were already confident: "I think for professional musicians, it would be better on Zoom" (P9).

Findings suggest that singing around a VR campfire may be suitable for informal contexts, or for individuals who feel self-conscious or nervous when performing. On the other hand, Zoom may be ideal for situations where more precise visual cues are important, such as one-on-one music lessons, or for those who are already comfortable singing online.

F. Comparitive Analysis

Although all participants preferred VR for its enjoyable and fun atmosphere, both VR and VC were considered suitable for duo singing in different contexts. VR was particularly beneficial for less confident singers, as it helped reduce self-consciousness. However the limited expressive facial cues of the Bigscreen avatars posed challenges for musical synchronisation. In contrast, Zoom provided more realistic visual cues via the camera, but the direct on-screen appearance of the researcher was perceived as intimidating for less experienced singers. Additionally, experienced singers noted difficulties maintaining proper singing technique due to the physical constraints imposed by the VR headset. Participants suggested future improvements, such as enhanced audio monitoring in both settings and the inclusion of a virtual stage and microphone in VR.

IV. Discussion

This study aimed to explore the key considerations for duo singing in virtual reality (VR) and videoconferencing (VC). The findings reveal distinct advantages and challenges associated with each platform, offering valuable insights for musicians and technology facilitators alike.

A. The Influence of Performance Setting

A notable finding across both platforms was the impact of the environmental setting on participants' sense of performance anticipation and excitement. One experienced musician noted the absence of a pre-performance adrenaline rush in the Zoom setting, attributing it to a lack of resemblance to a traditional performance space. Similarly, the VR campfire setting, though a casual departure from a conventional stage, was seen as beneficial for those prone to performance anxiety. These findings align with existing research [19], suggesting that incorporating diverse environments into the design of immersive musical experiences can enhance performance training and foster creativity.

B. The Influence of Self-Awareness on Inhibition

The VR environment helped reduce participants' self-awareness and inhibition, leading to a more playful and creative experience. The anonymity offered by avatars allowed participants to express themselves more freely, which was especially beneficial for less confident singers. This aligns with the concept of the online disinhibition effect (ODE) [20], where individuals display reduced restraint and increased self-expression in online environments. The findings of this study suggest that the ODE was more present in the VR setting, as reflected in participants' comments about self-expression, fun and creativity.

One key factor contributing to participants' feelings of anonymity and disguise in VR was their embodiment as an avatar. Participants reported feeling less visible in the VR setting, as the limited visibility of their own virtual face and body reduced their self-awareness. While this enhanced feelings of fun and enjoyment in Bigscreen, the inability to see their own body posed challenges from a musical perspective. This finding aligns with research on 360-degree video choir singing, which also identified difficulties with body rendering for musicians [21].

In contrast, the VC setting made some participants feel self-conscious due to the constant visibility of their image on-screen. This was particularly noticeable among less experienced singers, who reported feeling scrutinised and anxious. These findings suggest that users might benefit from utilising platform features to mitigate virtual meeting fatigue, such as turning off self-view mode [22] or using fully digital avatars to reduce self-focused attention [23].

C. The Importance of Visual Cues for Timing, Synchronization, and Musical Connection

In NMP sessions, latency often poses challenges for musical synchronisation, particularly when video is part of the interaction. However, due to the leader-follower nature of the duo-singing style in this study, participants were largely unaware of the inherent latency in the connection. Without the distracting effects of network delays, participants' feedback during the interviews centered more on the visual aspects of the collaboration.

Across both the VR and VC settings, participants emphasised the importance of seeing facial expressions and bodily movements for effective musical interaction and connection. This finding aligns with prior research highlighting the role of visual communication in achieving synchronised musical performances [24]. Additionally, it expands on the work by [5], which describes video as enhancing social connection and communication between performers.

In the Bigscreen setting, participants initially looked for visual cues in the eyes of the other performer. When they were unable to see detailed movement in the face, they shifted their attention to other parts of the body, such as the hands. This prioritisation of facial expressions over body movements suggests that incorporating more detailed face capture technology in future musical contexts could enahnce the accuracy of visual cues.

Participants expressed mixed feelings about their feeling of connection across both setting. However, their interpretations of the word 'connection' varied. For some, it referred to the sense of being in the same room as the other avatar, while for others, connection related more to the visual synchronisation of the musical performance.

Zoom was seen as providing easier visual cuing as it provided a 'real person' view of the other participant. However, since VR was preferred overall as the medium, it suggests that the sense of co-presence and body ownership may have outweighed any limitations related to visual cues.

D. The Influence of Physical Comfort and Novelty on User Experience

Physical comfort was a notable issue in the VR setting, with some participants experiencing discomfort from the use of the HMD. Previous research has shown that both the weight and fit of a VR HMD can affect a participants ability to sing [19]. These factors highlight the need for ergonomic improvements in VR hardware to enhance user comfort. Despite these challenges, the novelty of VR and the immersive sense of being in another environment may have offset some of the discomfort experienced.

In contrast, few comfort issues were reported in the VC setting, likely due to the increased the familiarity of using Zoom during the pandemic for meetings and social activities. To reduce the novelty factor in future studies involving VR, providing participants with more extensive familiarisation sessions beforehand could prove beneficial.

E. Implications for Future Research and Practice

The study's findings have several implications for future research on networked music performance using immersive technologies. First, further exploration is needed to assess the potential of VR in various musical collaboration scenarios, including group performances and rehearsals. Testing the impact of different virtual environments such as concert halls, along with physiological measures, could advance research into areas like music performance anxiety within immersive settings [25]. Additionally, examining the use of VR in music education and therapy could provide valuable insights into its broader applications.

Advancements in avatar technology and VR hardware could address current limitations, making VR a more viable option for precise musical collaboration. Improvements such as more detailed facial expressions, greater physical comfort, and customisable virtual environments could enhance overall user experience. Designers might also consider adding a virtual mirror backstage, enabling performers to see themselves before a show. This feature, shown to foster a strong sense of avatar body ownership [26], could be particularly beneficial in musical performance settings. Additionally, designers must also be mindful of the privacy and ethical implications surrounding data capture in music performance, ensuring that appropriate safeguards are implemented within the system [27].

F. Limitations and Future Research

This study faced several limitations. First, the leader-follower singing style, necessitated by COVID-19 constraints effectively eliminated noticeable participant latency, limiting the findings' applicability to effectively a one-way musical interaction. Future research could explore realistic interaction approaches (RIA) where latency is an inherent factor in the connection. Second, inconsistencies in participants' equipment configuration, also due to pandemic-related constraints, may have influenced their individual experiences and responses. Future studies should aim for a more standardised setup to ensure consistency across participants.

The small sample size of 10 participants, influenced by recruitment challenges, limits the generalisability of the findings. Future studies should aim for a larger sample size, which now should be more achievable in a post-pandemic context with greater access to participants. Additionally, focusing on more experienced musicians could provide a more targeted approach for exploring this area in greater depth.

Technical and practical limitations, due to the reliance on home setups, restricted data collection to self-reported responses. Future research could incorporate real-time measurement such as eye tracking, heart rate monitoring, and other psychological assessments to capture participants' responses during sessions. Furthermore, investigating technological features like avatar fidelity, body tracking, and biometric capture, and their impact on the online disinhibition effect, is recommended.

Future research should also explore more complex musical content including instrument use and group singing activities, within diverse virtual environments. Integration of VR and NMP technology into a single system, alongside factors such as social interaction, audio integration, and platform standardisation [7], remains an area for future investigation.

V. Conclusion

This study examined the key considerations of duo singing in virtual reality (VR) and videoconferencing (VC) environments. All participants expressed a preference for the VR setting over VC for singing. Thematic analysis identified five key themes: presence, performance, usability, affect, and usefulness. VR enhanced the sense of place, body ownership, and co-presence, contributing to a more enjoyable and playful experience. However, while visual cues were essential for timing and synchronisation in both settings, the limited facial expressions in VR presented challenges.

The results suggest that the ideal platform for online singing depends on the context and specific nuances of each medium. A campfire-style VR setting may be useful for casual, creative collaborations, particularly for less confident singers. In contrast, VC may be more appropriate for performances requiring a greater realism and clear visual communication between performers. As hardware and software continues to improve, the fidelity gap between VR and VC is likely to narrow in the coming years.

The study contributes to the field of NMP and immersive technologies by highlighting the unique affordances and challenges of VR and VC for duo singing. By understanding these key considerations, experience designers, musicians, and educators can make informed decisions about creating and using online musical collaboration tools.

Acknowledgment

This research was supported by an Australian Government Research Training Program (RTP) Scholarship. Additional support was provided by Learning Environments, Student and Scholarly Services, The University of Melbourne.

References

[1] G Carôt, A., Rebelo, P., & Renaud, A. (2007, March 1). *Networked Music Performance: State of the Art*. Audio Engineering Society Conference: 30th International Conference: Intelligent Audio Environments. http://www.aes.org/e-lib/browse.cfm?elib=13914

[2] Iorwerth, M., & Knox, D. (2019). The application of Networked Music Performance to access ensemble activity for socially isolated musicians. *Proceedings of the Web Audio Conference 2019 – Diversity in Web Audio*. https://researchonline.gcu.ac.uk/en/publications/the-application-of-networked-music-performance-to-access-ensemble

[3] Rottondi, C., Chafe, C., Allocchio, C., & Sarti, A. (2016). An Overview on Networked Music Performance Technologies. *IEEE Access*, *4*, 8823–8843. https://doi.org/10.1109/ACCESS.2016.2628440

[4] Cáceres, J.-P., Hamilton, R., Iyer, D., Chafe, C., & Wang, G. (2008). To the Edge with China: Explorations in Network Performance. *ARTECH 2008: Proceedings of the 4th International Conference on Digital Arts*, 61–66.

[5] Iorwerth, M., & Knox, D. (2021). The Playing Together, Apart Framework: A framework for communication in Networked Music Performance. *Journal of Music, Technology and Education*. https://pure.uhi.ac.uk/en/publications/the-playing-together-apart-framework-a-framework-for-communicatio

[6] Turchet, L., Hamilton, Rob, & Çamci, Anil. (2021). Music in Extended Realities. *IEEE Access*, *9*, 15810–15832. https://doi.org/10.1109/ACCESS.2021.3052931

[7] Turchet, L. (2023). Musical Metaverse: Vision, opportunities, and challenges. *Personal and Ubiquitous Computing*. https://doi.org/10.1007/s00779-023-01708-1

[8] Boem, A., Tomasetti, M., & Turchet, L. (2024). *Harmonizing the Musical Metaverse: Unveiling needs, tools, and challenges from experts' point of view*.

[9] Bartlette, C., Headlam, D., Bocko, M., & Velikic, G. (2006). Effect of Network Latency on Interactive Musical Performance. *Music Perception: An Interdisciplinary Journal*, *24*(1), 49–62. https://doi.org/10.1525/mp.2006.24.1.49

[10] Olmos, A., Brulé, M., Bouillot, N., Benovoy, M., Blum, J. R., Sun, H., Lund, N., & Cooperstock, J. (2009). Exploring the role of latency and orchestra placement on the networked performance of a distributed opera. *Undefined*. https://srl.mcgill.ca/publications/2009-PRESENCE.pdf

[11] Schuett, N. (2002). *The Effects of Latency on Ensemble Performance* [Honors Thesis, Stanford University]. https://ccrma.stanford.edu/groups/soundwire/publications/papers/schuett_honorThesis2002.pdf

[12] Loveridge, B. (2020). Networked Music Performance in Virtual Reality: Current Perspectives. *Journal of Network Music and Arts*, *2*(1). https://commons.library.stonybrook.edu/jonma/vol2/iss1/2

[13] Drioli, C., Allocchio, C., & Buso, N. (2013). Networked Performances and Natural Interaction via LOLA: Low Latency High Quality A/V Streaming System. In P. Nesi & R. Santucci (Eds.), *Information Technologies for Performing Arts, Media Access, and Entertainment* (pp. 240–250). Springer Berlin Heidelberg.

[14] Cáceres, J.-P., & Chafe, C. (2010). JackTrip: Under the Hood of an Engine for Network Audio. *Journal of New Music Research*, *39*(3), 183–187. https://doi.org/10.1080/09298215.2010.481361

[15] Carôt, E., & Werner, C. (2007). Network music performance—Problems, approaches and perspectives. *In International School of New Media, Institute of Telematics, University of Lübeck. Music in the Global Village - Conference*.

[16] Zhang, D. J., Susino, M., McPherson, G. E., & Schubert, E. (2020). The definition of a musician in music psychology: A literature review and the six-year rule. *Psychology of Music*, *48*(3), 389–409. https://doi.org/10.1177/0305735618804038

[17] Schober, M. F. (2006). Virtual Environments for Creative Work in Collaborative Music-making. *Virtual Real.*, *10*(2), 85–94. https://doi.org/10.1007/s10055-006-0049-z

[18] Braun, V., & Clarke, V. (2022). *Thematic Analysis: A Practical Guide*. SAGE Publications Ltd.

[19] Ppali, S., Lalioti, V., Branch, B., Ang, C. S., Thomas, A. J., Wohl, B. S., & Covaci, A. (2022). Keep the VRhythm going: A musician-centred study investigating how Virtual Reality can support creative musical practice. *CHI Conference on Human Factors in Computing Systems*, 1–19. https://doi.org/10.1145/3491102.3501922

[20] Suler, J. (2004). The Online Disinhibition Effect. *CyberPsychology & Behavior*, *7*(3), 321–326. https://doi.org/10.1089/1094931041291295

[21] Daffern, H., Camlin, D. A., Egermann, H., Gully, A. J., Kearney, G., Neale, C., & Rees-Jones, J. (2019). Exploring the potential of virtual reality technology to investigate the health and well being benefits of group singing. *International Journal of Performance Arts and Digital Media*, *15*(1), 1–22. https://doi.org/10.1080/14794713.2018.1558807

[22] Bailenson, J. N. (2021). Nonverbal Overload: A Theoretical Argument for the Causes of Zoom Fatigue. *Technology, Mind, and Behavior*, *2*(1). https://doi.org/10.1037/tmb0000030

[23] Ratan, R., Miller, D. B., & Bailenson, J. N. (2022). Facial Appearance Dissatisfaction Explains Differences in Zoom Fatigue. *Cyberpsychology, Behavior, and Social Networking*, *25*(2), 124–129. https://doi.org/10.1089/cyber.2021.0112

[24] Bishop, L., & Goebl, W. (2015). When they listen and when they watch: Pianists' use of nonverbal audio and visual cues during duet performance. *Musicae Scientiae*, *19*(1), 84–110. https://doi.org/10.1177/1029864915570355

[25] Osborne, M., Glasser, S., & Loveridge, B. (2022). 'It's not so scary anymore. It's actually exhilarating': A proof-of-concept study using virtual reality technology for music performance training under pressure. *ASCILITE Publications*, e22116-1-e22116-8. https://doi.org/10.14742/apubs.2022.116

[26] Slater, M., Spanlang, B., Sanchez-Vives, M. V., & Blanke, O. (2010). First Person Experience of Body Transfer in Virtual Reality. *PLOS ONE*, *5*(5), e10564. https://doi.org/10.1371/journal.pone.0010

[27] Sparrow, L. A., Galwey, C., Loveridge, B., Glasser, S., Osborne, M. S., & Kelly, R. M. (2024). Heart and Soul: The Ethics of Biometric Capture in Immersive Artistic Performance. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–23. https://doi.org/10.1145/3613904.3642309