



AR Drum Circle: Real-Time Collaborative Drumming in AR

Torin Hopkins^{1*}, Suibi Che Chuan Weng¹, Rishi Vanukuru¹, Emma A. Wenzel¹, Amy Banic^{1,2}, Mark D. Gross¹ and Ellen Yi-Luen Do¹

¹ATLAS Institute, University of Colorado Boulder, Boulder, CO, United States, ²Interactive Realities Lab, University of Wyoming, Laramie, WY, United States

AR Drum Circle, is an augmented reality (AR) platform we developed to facilitate a collaborative remote drumming experience. We explore the effects of virtual avatars that are rendered in a player's view to provide the joy and sensation of co-present music creation. AR Drum Circle uses a head mounted AR display, providing players with visual effects to assist drummers in coordinating musical ideas while simultaneously using a latency-optimized remote collaboration service (JackTrip). AR Drum Circle helps overcome barriers in music collaboration introduced when using remote collaboration tools on their own, which typically do not support real-time video, lack spatial information, and volume control between players, all of which are important for in-person drum circles and music collaboration. This paper presents the results of several investigations: (1) analysis of an in-depth, free-form response survey of drumming communication provided by a small group of expert drummers, (2) observations and analysis of 20 videos of live drum circles, (3) a case study using the AR Drum Circle application in pairs of remotely located participants, and (4) a collaborative drumming case study using a popular networked conferencing application (Zoom). Findings suggest that substantial communicative information including facial expressions, hand-tracking, and eye-contact is lost when using AR alone. Findings also indicate that the AR Drum Circle application can make players less self-conscious and more willing to participate while playing drums improvisationally with others.

Keywords: augmented reality, remote collaboration, network latency, drum circle, user studies

OPEN ACCESS

Edited by:

Mary C. Whitton,
University of North Carolina at Chapel
Hill, United States

Reviewed by:

Chris Rhodes,
University College London,
United Kingdom
Ben Loveridge,
University of Melbourne, Australia

*Correspondence:

Torin Hopkins
torin.hopkins@colorado.edu

Specialty section:

This article was submitted to
Technologies for VR,
a section of the journal
Frontiers in Virtual Reality

Received: 01 January 2022

Accepted: 22 June 2022

Published: 19 August 2022

Citation:

Hopkins T, Weng SCC, Vanukuru R,
Wenzel EA, Banic A, Gross MD and
Do EY-L (2022) AR Drum Circle: Real-
Time Collaborative Drumming in AR.
Front. Virtual Real. 3:847284.
doi: 10.3389/frvir.2022.847284

1 INTRODUCTION

Musicians of all stripes enjoy—and benefit from—playing together. In particular, jamming (defined as informal collaborative music creation) is an essential feature of many forms of music, such as drumming, all over the world. However, limitations in data transmission have required most networked musical collaboration to occur asynchronously or in co-located environments (Table 1). Amidst a global pandemic, networked musical collaboration has become popular because of the social and emotional benefits of bringing people from far distances to play together (Kokotsaki and Hallam, 2011; Mastnak, 2020; Antonini Philippe et al., 2020; Signiant, 2021). Few networks, even in cities with high internet connectivity, support the speeds needed for near-instantaneous data transmission (Simsek et al., 2016). To minimize the inevitable latency, current real-time networked music systems eschew video, and other information present in jamming that enable people to feel co-present, defined as the presence of other actors that shape behavior in social interaction (Jung et al., 2000), and enjoy the socially collaborative experience of playing music (Drioli

et al., 2013). The resulting experiences lack the richness of in-person interaction and collaboration. Some systems, such as LoLa, do provide high-quality audio and visual experiences on-screen but have utilized dedicated, high-performance networks to accomplish this (Drioli et al., 2013). Though audio-visual quality is provided in such systems, spatial communication and sensation, which are important to collaborative music activities, are not present in 2-dimensional, on-screen experiences (Krueger, 2011; Bishop, 2018). Augmented Reality (AR) is positioned to address the shortcomings of current systems by providing 3-dimensional visual information, networked communication, and engaging with the local environment. While AR has previously been explored in the form of novel musical instruments (Billinghurst et al., 2004) and composition experiences using tangible cards (Poupyrev, 2000; Berry et al., 2003), little research has been done to explore how to augment and enhance a remote real-time collaborative drumming experience.

We utilize immersion and avatars to increase co-presence and non-verbal communication important for drumming collaboration. Visual information is extremely important for in-person drumming, because of the reliance on cues and other visual communication techniques to collaborate (Thompson et al., 2005; Eaves et al., 2020). We utilize proven methods to increase co-presence, based on two findings: 1) AR has been shown to increase co-presence for remote communication and augmented video conferencing (Billinghurst et al., 2004; Jo et al., 2016), 2) Avatars with life-like animations can provide social cues which have been shown to increase co-presence (Lee et al., 2020).

An early example of a system that provided a shared, immersive, real-time, musical experience was explored in a desktop-based, virtual environment through PODIUM (Jung et al., 2000). While researchers found co-presence elements, such as avatars, were ineffective, the experiment was limited to two subjects each playing a simple tune on a keyboard (either the melody or accompaniment) guided by a conductor. As a result, there is still more to learn about the role of immersion, co-presence, and augmented elements for other instruments, musical contexts, and social dynamics. Our investigation addresses this gap in the research and explores AR in a novel way by designing and investigating a platform that supports a more complex social environment, as well as the dynamic composition elements of drumming improvisation. To address the limitations of existing systems and explore the utility of AR technologies in supporting collaborative music-making, we developed “AR Drum Circle,” a platform that uses AR enhanced real-time and non real-time strategies to create a satisfying remote musical collaboration, specifically improvisational drumming. We approached this challenge by designing a system that uses avatars as a mediating technology to communicate between remote drumming partners.

The past decade has seen much progress in AR technology—experiences are more spatially stable, and the general availability of hardware that supports high-fidelity AR experiences, both on mobile devices (Nowacki and Woda, 2020) and head-worn displays (Lin et al., 2021; Xu, 2021), has greatly

increased. These advancements, together with constantly improving internet connectivity speeds, make the possibility of immersive remote collaboration for creative activities all the more real and exciting. In this paper we extend recent developments in AR and focus on the Drum Circle for a unique and collaborative musical experience.

This paper is organized into six sections. **Section 2** discusses observations of drum circles gathered through video analysis and interviews with drummers. **Section 3** addresses preliminary findings from the observations. **Section 4** outlines a case study investigating the differences between the AR Drum Circle and the Zoom conferencing application. **Section 5** describes results of the case study conducted. **Section 6** discusses results and future directions of the AR Drum Circle platform. **Section 7** provides a concluding statement and outlook for the use of the application.

2 MATERIALS AND METHODS

2.1 Preliminary Drum Circle Investigation

As the name suggests, Drum Circles are gatherings of people playing percussion instruments (often hand drums like the Djembe) while facing each other and arranged in a circle. They are informal sessions where people with a wide range of musical abilities improvise and play rhythms together. The main focus of a Drum Circle is not musical composition, but rather enjoyment and fostering a sense of togetherness. Drum Circles have continued to be popular in communities around the world, and are also used as a framework for music education and therapy (Vinnard, 2018). Given the relative simplicity of the musical instruments involved, the low barrier to entry in terms of skill, and the social dynamics unique to a Drum Circle, this context seemed ideal to base our explorations.

To gather preliminary data we conducted a survey with domain experts and analyzed videos of expert drummers participating in an informal drum circle. In the following section we describe the methods used for capture and analysis, as well as synthesize preliminary results that were incorporated into the design of our application.

2.1.1 Survey

We began by asking three expert musicians (each with more than 15 years experience playing different instruments) to respond to a detailed online questionnaire, in order to learn more about their experience with musical collaboration, drum circles, and their expectations from AR applications that aim to support remote jamming.

2.1.1.1 Participants

Two of the expert musicians (E1 and E3) specialized in drums, while one (E2) mainly played the guitar. E1 and E2 were formally trained, while E3 was primarily self-taught. E1 was most familiar with drum circles, having participated in and organized many sessions themselves. E2 and E3 had both occasionally participated in drum circles as a social activity. All three musicians found

drum circles to be enjoyable because of the ability to “have a musical conversation with multiple people” (E1), produce music “complimented by a talented group” (E2), and the “primal nature” of the experience (E3).

2.1.1.2 Remote Musical Collaboration

All respondents mentioned that despite the pandemic, they rarely attempted to remotely collaborate with other musicians in real-time. This was mainly due to internet connectivity issues that cause audio and video to be out of sync. When composing music as a group, E1 mentioned that they used to record a track locally, send the audio file to their bandmates over the internet, and have them add to that track asynchronously.

2.1.1.3 Communication in a Jam Session

When playing music together, all respondents mentioned that they used a combination of speech, eye contact, facial expressions, and bodily movements to coordinate the session and influence the kind of music being played. Talking or shouting seemed more appropriate in social settings (E1) while focused jams called for more subtle indications such as eye or hand gestures (E2).

Physical cues and gestures, such as body movements, head nods, eye contact, and facial expressions, were rated by all respondents as a particularly important part of musical collaboration. According to E1, novice drummers rely more on facial expressions to convey musical intention, while more experienced drummers focus on each others’ hands as they strike the drum in different ways. Eye contact was also used to pass on the leadership of the drum circle, and signal to other musicians about an upcoming change in tempo or rhythm.

2.1.1.4 Features for an AR Jam Session

2.1.1.4.1 Real vs. Virtual Drum. All respondents mentioned that they would prefer playing on a real drum as opposed to a virtual or holographic one. This is because of the haptic feedback received from a real drum that would otherwise be missing, making it harder to understand when a hit takes place and when a player does a “swing and miss” (E2).

2.1.1.4.2 Avatar Appearance. Respondents stressed on the importance of representing remote players as humanoid avatars in order to not detract from the collaborative experience. Facial similarity was brought up as an important requirement (E1). The ability to customize one’s own avatar by changing their gender, clothes, and accessories, was also mentioned as a feature that would be nice to have, but not essential.

2.1.1.4.3 Design Ideas. The musicians noted that it would be good to have some way to control the volume of music coming from the various remote players, either using AR buttons and sliders or knobs on a physical interface (E1). They also suggested visual elements to help keep all players in sync, such as a “. . . giant holographic metronome in the middle with flashing lights keeping the base beat going” (E2).

2.1.2 Observations of Drum Circle Videos

To investigate how drummers communicate in a drum circle we captured and analyzed twenty videos of drum circles, ranging from 8–20 min. Participants ranged between two and twenty experienced (regularly participate in drum circles) drummers.

Though many types of drums were used in the recorded drum circles, the Djembe (a traditional West African drum) was referenced for the types of drum strikes and rhythmic patterns. Our observations revealed crucial interaction dynamics, such as body language, eye-contact, gestures, and verbal communication that participants use in a drum circle. Using Anvil video annotation software (Kipp, 2012) we marked and counted the occurrences of the various communication techniques (Figure 1). We labelled and counted the instances when players looked at their own hands, or their drum, versus looking at the other player’s face, hands, or drum, in a single session. We concluded that verbal communication, head motion, facial expression, and eye-contact play a key role in communication while jamming (Table 2). These observations were used to inform the behavior of the representations of remote player avatars, by controlling their attention and mannerisms.

Table 2 lists the body motion looked for in the videos and how many times a motion was made to make the communicative action.

2.2 Preliminary Finding Implementation

Building upon our analysis of interaction dynamics in drum circles, we explored how the strengths of AR could overcome the challenges of networked music collaboration tools. Our main design goals were to recreate and expand on some of the dynamics of in-person drum circles (physical, social cues).

To address these goals, we developed the AR Drum Circle, an application that enables musicians to participate in a remote drum circle with their friends. Remote participants are rendered as avatars in the local player’s field of view using the Nreal Light AR glasses. To accommodate the demands of low latency communication in musical collaboration, body language is conveyed through compressed messages. Once received over the network, the compressed messages initiate pre-recorded motion of the avatar to minimize the effects of latency. Audio is transmitted at near-instantaneous speeds using an audio communication service called JackTrip (Cáceres and Chafe, 2010). This audio feed, in conjunction with the visuals rendered by the AR glasses, results in a collaborative experience with near-instantaneous audio transmission as well as a reactive AR avatar.

2.3 Case Study

2.3.1 Goal

The goal of this study was to test the ability of the AR Drum Circle application to create a sense of co-presence. Additionally, we gather data about experience with a questionnaire between trials to measure a sense of co-presence. As shown by Beacco et al. (2021), this type of analysis can yield further insights into the experience of a virtual environment. Including a written analysis of the collaborative drumming task over Zoom enabled us to compare sentiment scores between Zoom and the AR Drum Circle application.

TABLE 1 | Summary of latency/delay perception and tolerance in common music systems.

Type of delay (latency)	Delay perceptible below (msec)	Delay tolerance (upper limit msec)	Reference
Strike-sound	10	20	Chew et al. (2005)
Network sound delay in music activities causing tempo dilation (tempo dependant, \pm 2% of tempo)	20	100 (or \pm 2% of tempo interval)	Schuett. (2002); Kleimola. (2006)

TABLE 2 | Summary of communication techniques.

Communication technique	Number of occurrences
Eye Contact	31
Head motion for signaling	25
Head motion rhythmic	6
Head/Body Facing	3
Facial expression	13
Head motion rhythmic	6
Looking at own instrument/hands	29
Looking at another's instrument/hands	13

TABLE 3 | Cluster 1.

Themes average semantic analysis score: 0.6705	Example phrase from text
General +	-
General -	"It was difficult"
Perception of Latency	"It was difficult to sync up"
Animation of Avatar	-

TABLE 4 | Cluster 2.

Themes average semantic analysis score: 0.1027	Example phrase from text
General +	"I think we both had fun"
General-	-
Perception of Latency	"There was a tiny bit of delay, but wasn't too bad"
Animation of Avatar	-

TABLE 5 | Cluster 3.

Themes average semantic analysis score: 0.79	Example phrase from text
General +	"The whole experience was very cool"
General -	-
Perception of Latency	"We talked to sync and exchange leads"
Animation of Avatar	"I think as the drumming went on I stopped following the avatar"

To test co-presence we administered the commonly used networked minds measure of social presence (NMM) questionnaire (Biocca and Harms, 2002). The NMM questionnaire

consists of 16 questions to measure the perceived level of co-presence, described in further detail in the methods section.

To ensure proper counterbalancing between trials we changed the order in which participants experienced the NReal and Zoom tasks for every other pair of participants. All other variables remained the same for the two trials.

2.3.2 Methods

This study posed minimal to no risk to participants and was exempted from IRB review. Written informed consent was obtained from the individuals for the publication of any potentially identifiable images or data included in this article.

2.3.2.1 Participants and Experimental Procedures

We recruited 13 participants (seven Females and six Males) from the University of Colorado Boulder. Twelve participants were ages 18–32 (min: 18, max: 32, median: 23, mean: 24.18, SD: \pm 4.66), and one participant was 59 years of age. We conducted a within-subject design study, whereby participants exhibited every test condition.

Before beginning the test, participants were given a drum improvisation tutorial consisting of a 10 min period of instructional collaborative drum improvisation, led by an expert percussionist.

In each task, players were asked to have a drum improvisation that reflects how they would improvise with others in a social drumming experience. This was done for both tasks using the AR Drum Circle application and Zoom, in conjunction with real-time audio provided by JackTrip.

The two tasks were counterbalanced to reduce effects of learning and comfort in the drum improvisation tasks throughout the test. All players participated in both tasks for a duration of 5 min, which was pre-determined to be an optimal period of time for minimizing mind-wandering and maximizing comfort in a drum improvisation task. Following each task participants filled out a questionnaire to gauge co-presence and wrote a 150-word free response statement regarding their experience with the task.

The experiment was conducted in pairs of players located in separate, sound-isolated, rooms. Each player was connected to a local network through an ethernet cable to reduce latency between the players as much as possible. We did this to avoid substantial differences in visual latency between experimental procedures as latency was not the focus of this study.

2.3.2.2 Procedure 1: AR Drum Circle Application

The experiment was conducted in pairs of players located in separate rooms. To reduce visual latency a local server was established to transmit messages corresponding to motion.

TABLE 6 | Cluster 4.

Themes average semantic analysis score: -0.6705	Example phrase from text
General +	"Experience was good, the drummer looked really cool"
General -	"After the session I felt a little bit of spatial orientation issues"
Perception of Latency	"The latency is acceptable definitely"
Animation of Avatar	"The avatar made me feel less self-conscious"

TABLE 7 | Cluster 1.

Themes average semantic analysis score: 0.1548	Example phrase from text
General +	-
General -	"It was way harder to use Zoom than the avatar"
Perception of Latency	"I expected perfection on Zoom and it was really disappointing."
Animation of Avatar	"With the avatar, I was more willing to accept some of the delay/image problems"

TABLE 8 | Cluster 2.

Themes average semantic analysis score: 0.4404	Example phrase from text
General +	"The experience was good"
General -	"times, it felt a bit plane just to see the face of my partner"
Perception of Latency	"I didn't feel a delay in the jamming process."
Animation of Avatar	"I only saw her face which make me notice body language but I felt I was missing the actual hand-movement"

TABLE 9 | Cluster 3.

Themes Average semantic analysis score: 0.719	Example phrase from text
General +	"This was a better experience since I could actually see her and connect"
General -	"camera had inverted everything so I wasn't sure what part of the drum to hit"
Perception of Latency	"I was able to follow my other partner"
Animation of Avatar	"probably a little bit better because of the use of his hands as visual cues to follow"

Each player was connected to a local network through an ethernet cable to reduce visual latency between the players as much as possible because latency was not the focus of this study (Figure 2).

TABLE 10 | Cluster 4.

Themes average semantic analysis score: 0.929	Example phrase from text
General +	"The experience was much more fun"
General -	"wasn't sure if my jam partner and I were hearing the same thing due to the Zoom-mediated experience"
Perception of Latency	"This was a very fun experience as making music over zoom settings is often frustrating and unproductive."
Animation of Avatar	"I can see his hand very nicely in the video that makes it quite easy to follow"

We built the AR Drum Circle application to the NReal Light Developers kit head-mounted display AR glasses (52° field of view, 1000 nit brightness, 45.32hFOV, 25.49vFOV, 1920 × 1080 resolution, 16:9 aspect ratio, 42.36 pixels per degree). We constructed the AR Drum Circle application using the Unity game engine (version 2020.3.0f1) in conjunction with Nreal's own tracking engine for the NReal Light headset, which enables the recognition of planar surfaces upon which remote players are displayed.

The avatar appeared as a humanoid robot. We chose this avatar model because it is non-distracting and would not distract the players from the improvisational activity. Because our avatars are representing the drumming partner and not used for self-representation, we did not feel an accurate human model needed to be used. Limitations with the application prohibited us from directly applying a reliable form of non-distracting eye-gaze, so we did not include it in this version of the application (Figure 3).

JackTrip Virtual Studio (a service that hosts servers and connects to a home network) was used to establish a low-latency (under 15 msec) audio streaming environment for each player. Both players connected to a nearby server provided by JackTrip (located in Denver, CO. Round trip distance of approximately 60 miles). Latency was verified by recording the sound at the source and output location using two microphones connected to an additional computer. A Raspberry Pi 4, was used to create a bridge connection directly between an audio interface (M-Audio, M-Track Duo) and a localized router. Each player was fitted with a Shure SM57 microphone (to afford speech communication between players) and a MIDI drum that produced its own sound heard by the other player over the network. Headphones were used to isolate the players from surrounding noise and the audio interface was used to control the volume of each player. Players used a MIDI drum (KAT KTMP1 MIDI controller) and animations were triggered by MIDI data (drum hits) sent to and from jam partners.

The MIDI forwarding device runs a Python script to read MIDI messages and send each message to the NReal light AR glasses. MIDI messages are a sequence of bytes that represent drum "notes" or "hits". Each message contains: a note type ("note on" or "note off"), a note number (0–127) representing a unique drum sound (e.g., snare, kick, etc.), a note velocity (0–127) representing how hard the drum was hit, and a value that represents the time between successive messages.

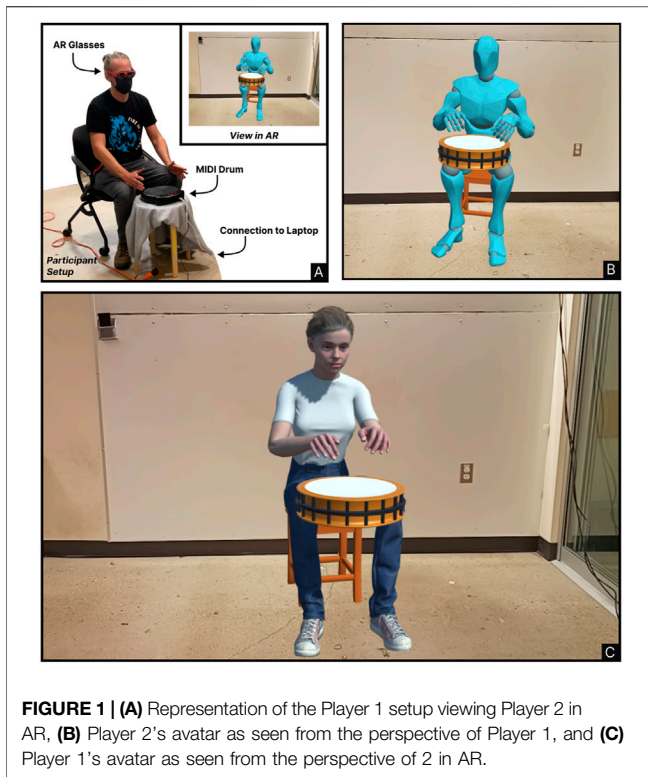


FIGURE 1 | (A) Representation of the Player 1 setup viewing Player 2 in AR, **(B)** Player 2's avatar as seen from the perspective of Player 1, and **(C)** Player 1's avatar as seen from the perspective of 2 in AR.

2.3.2.3 Procedure 2: Zoom

To compare a popular video conferencing application (Zoom) to our AR Drum Circle application we set up a laptop with a Zoom meeting open and encouraged players to manipulate the position of the laptop camera. Players were then connected to the JackTrip servers in the same manner as with the AR Drum Circle application case study, enabling real-time audio streaming between players. As a result, players were able to see each other through Zoom and hear the drumming and speech of the other players in real time through the JackTrip server system.

2.3.3 Analysis Techniques

We collected a 150 word free-form statement and several questionnaires. The free-form statement was analyzed using sentiment analysis described in section [4.2.1]. The questionnaires included an NMM standard co-presence questionnaire and a questionnaire that aimed to specifically compare the AR application to zoom with regard to the collaborative drumming experience.

2.3.3.1 Sentiment Analysis

To analyze the results from free-form written responses, we employed the natural language processing technique of sentiment analysis. Because of the variable nature of the responses we chose the basic criterion of positive and negative feelings about both applications, and how these sentiments relate in the context of the written response.

The sentiment analysis method we used provided a score for each of the written passages between -1 and 1, designed to reflect

positive and negative feelings present in the written passages of each participant. We used the open source Natural Language Processing Toolkit (NLTK) in Python, to perform the sentiment analysis. The NLTK package contains a supervised model of pre-trained algorithms. A corpora of approximately one million words are used to train the sentiment analysis module by the NLTK team led by Steven Bird and Liling Tan.

Standard natural language processing data cleaning methods were used. This includes the removal of “stopwords” (commonly used conjunctions and words such as, “and”, “the”, and “but”), punctuation, and word-level tokenization. This data-cleaning process was used to ensure that the sentiment scores accurately reflect the positive and negative feelings present in the written passages for each participant associated with each application.

2.3.3.2 NMM Co-Presence Questionnaire

After each improvisational task (AR Drum Circle and Zoom), in addition to a 150 word statement, we gave participants a networked minds measure of social presence (NMM) questionnaire (Biocca and Harms, 2002). This questionnaire consists of 16 questions on a likert 7-point scale used to measure co-presence, defined as the degree to which an individual believes they are not alone, maintain focal awareness, and the degree to which they feel their partner is also aware of their presence (Biocca and Harms, 2002).

2.3.3.3 Comparison Interview

We administered a comparison questionnaire to investigate aspects of the AR Drum Circle application missing from the more common networked experiences. Zoom, and other networked conferencing applications, have been utilized for networked audio/visual communication. Video, though delayed from real-time audio, has been used for showing facial expressions and other visual information important to conversation. We administered a comparative questionnaire to assess the qualitative differences in the experience of the two applications and what, if anything, Zoom provided that could be useful to include in the AR Drum Circle application.

3 RESULTS

3.1 Sentiment Scores

We grouped the responses into clusters based on the sentiment scores using ‘two-step clustering’ in IBM SPSS Statistics Package. The two-step cluster analysis is designed to cluster categorical and continuous. The analysis calculates distances between variables assuming independence of the variables and Gaussian as well as multinomial distribution for both variable types. The analysis is conducted in two steps: 1) a cluster features tree is made by associating the closeness of nodes from a root or starting node on the tree, and 2) the nodes of the tree are then grouped to determine the “best” clusters based on the Schwarz Bayesian Criterion or the Akaike Information Criterion.

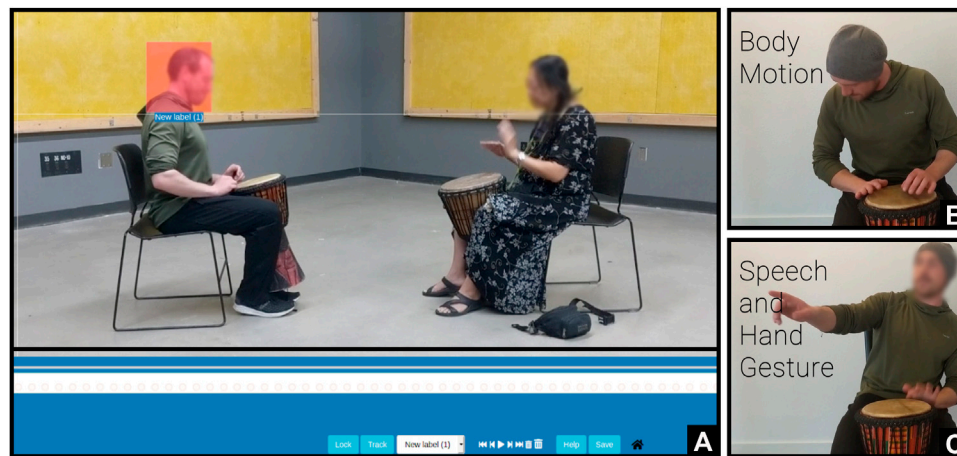


FIGURE 2 | Left (A): Analysis of two djembe drummers during an improvisation session. Eye-glances, facial expressions, and other body motion were recorded and analyzed. Researchers placed a pink square and labeled it for “looking at his drum head” among other communication modalities (e.g., looking at the other drummer’s face, or hands, or drum heads, etc.) Right (B) and (C): We captured motion from a live drummer using IR-based motion capture. The motion data was converted into animations that control the avatars. Written informed consent was obtained from the individuals for the publication of any potentially identifiable images or data included in this article.

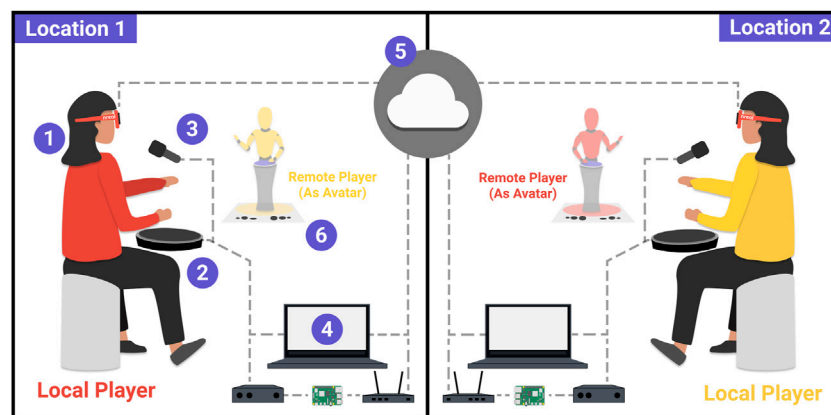


FIGURE 3 | AR Drum Circle system overview schematic. (1) AR Headset (Nreal Light), (2) MIDI Drum Pad or acoustic drum, (3) Microphone for speech communication, (4) MIDI forwarding device, (5) Local UDP server and JackTrip server in Denver, (6) Room Environment (via plane detection). MIDI notes from User 1 are routed to User 2 to trigger animation of the avatar in the field of view of User 2 using the Local UDP server. Live audio is sent to nearby JackTrip servers to create near-instantaneous audio transmission.

The number of clusters for the output of the two-step clustering were chosen based on the optimization of the silhouette coefficient and the elbow method. The silhouette coefficient contrasts the average distance to elements in the same cluster and identifies whether outliers are present within each cluster. This was further confirmed using the elbow method—a visual heuristic for determining cluster number in a data set. The elbow method maximizes the cluster number with respect to variance for each possible cluster number in the set. Four clusters were identified as the optimal number of clusters for each of the analyzed data sets.

Figures 4, 5 show summaries of the 4 clusters, demonstrating clear separations between them. Cluster 2 is

referred to as “Average” indicating the mean of the sentiment scores.

3.2 Thematic Analysis

We conducted thematic analysis on the free-form responses used to derive sentiment scores. We separated the results by theme to further investigate why sentiments were reported as positive or negative by participants according to sentiment analysis. We calculated and sorted the most used words and phrases by frequency to inform our decisions for selecting general themes. We then located the words and phrases within the responses for context and decided on four main themes that adequately

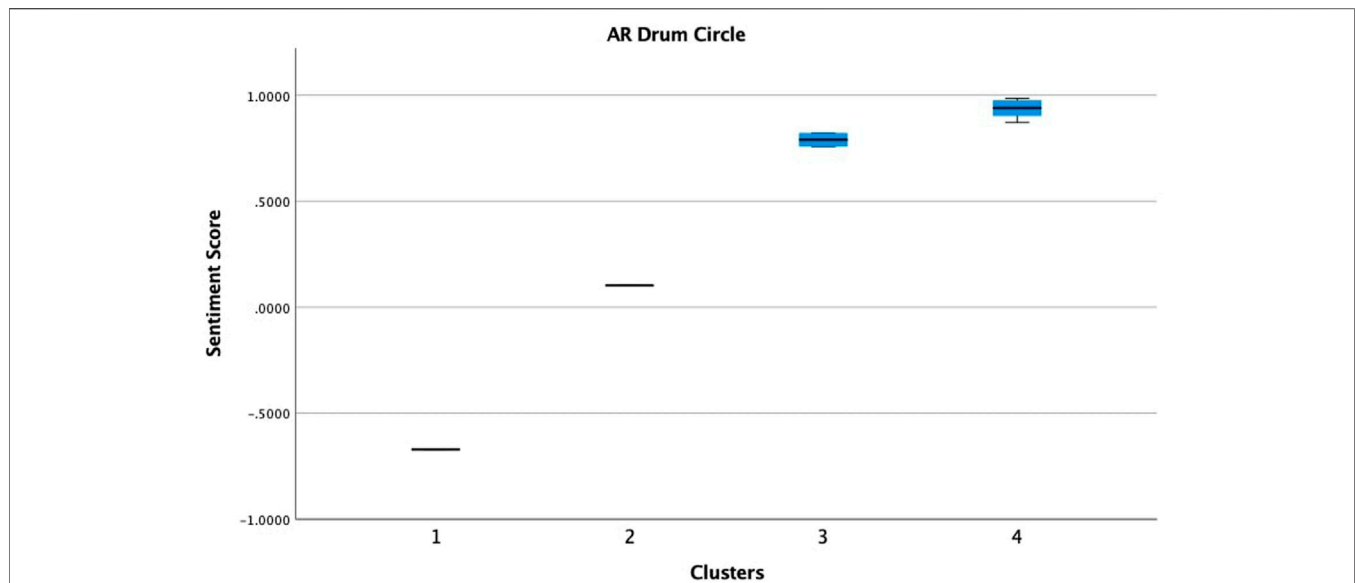


FIGURE 4 | The sentiment scores for the written responses to the AR Drum Circle application ranged from -0.671 to 0.986 , with mean of $0.727 \pm \text{SD}: 0.480$, and median = 0.903 . The sentiment scores are not significantly different between females and males, and are not correlated with age.

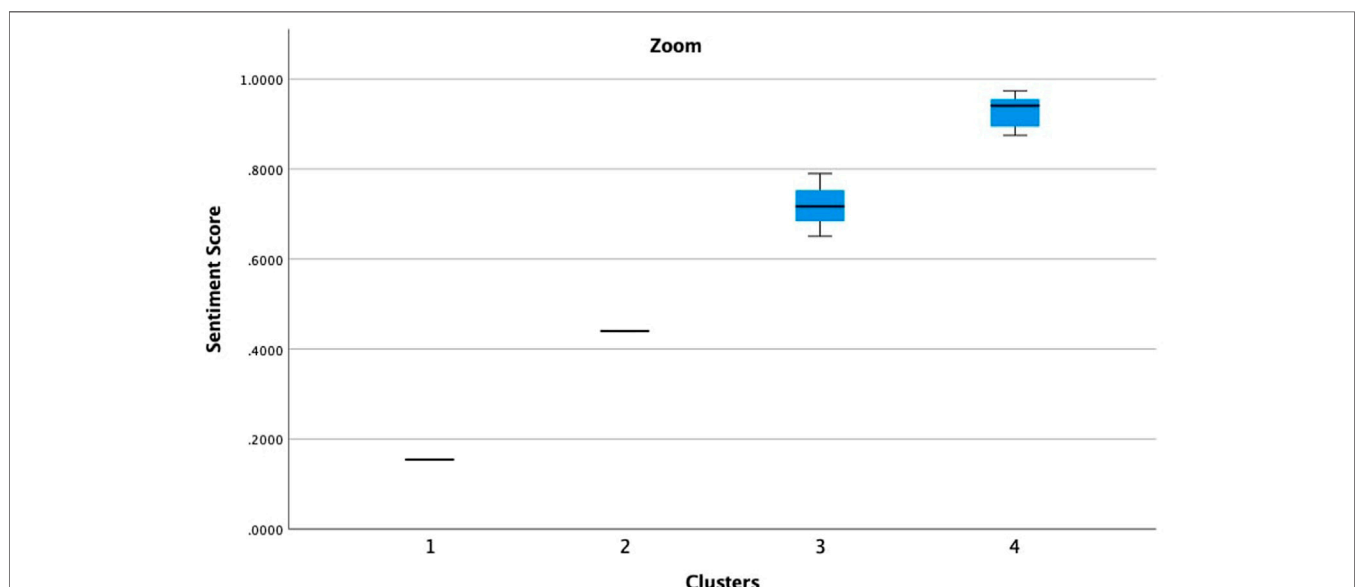


FIGURE 5 | The sentiment scores for the written responses to the Zoom application ranged from 0.1548 to 0.9741 , with mean of $0.783 \pm \text{SD}: 0.244$, and median = 0.891 . The sentiment scores are not significantly different between females and males, and are not correlated with age.

described all of the written content. We applied the themes and clustered the data for the AR and Zoom applications. Below are examples that represent the sentiments of the participants for each application by cluster and theme.

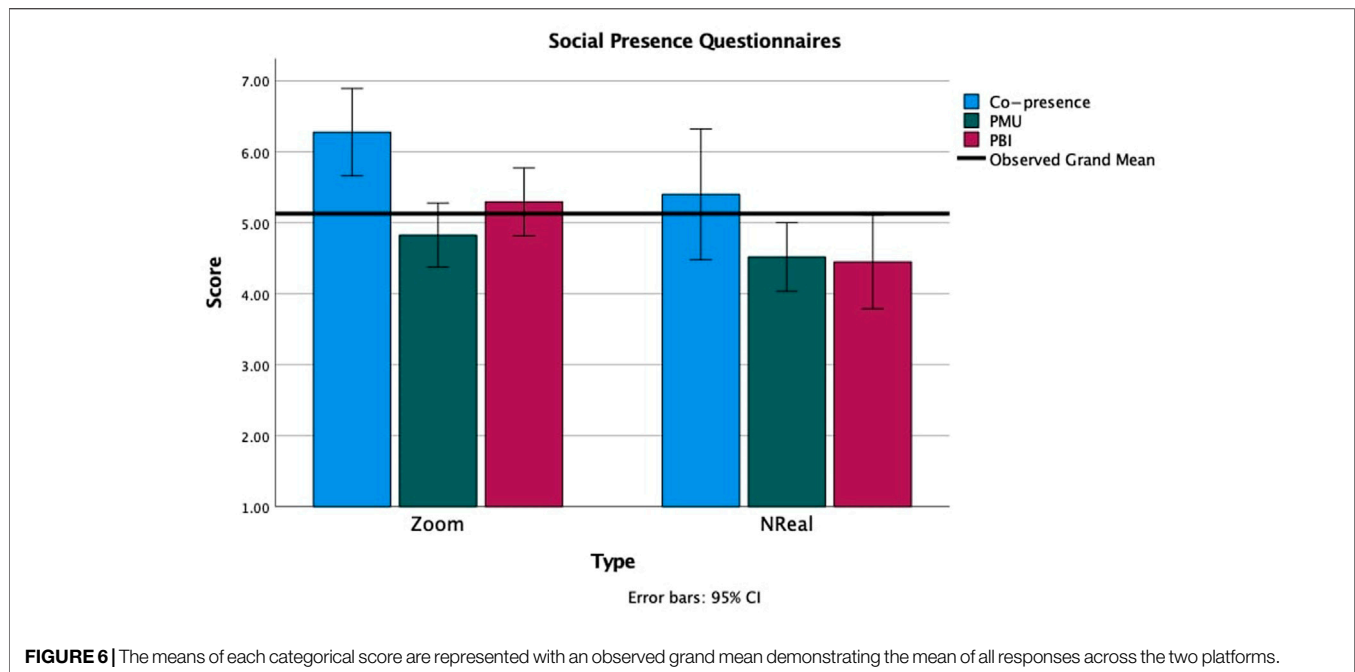
3.2.1 AR Drum Circle Thematic Analysis

The tables in this section are in reflection of the AR Drum Circle organized by theme and cluster (Tables 3–6). Several example phrases for the given theme are provided to give context to how

the themes were chosen. The mean semantic analysis score is calculated by averaging the semantic analysis scores from each of the passages in the cluster.

3.2.2 Zoom Thematic Analysis

Similar to the tables in the sections above, the tables in this section are the reflection of participants organized by theme and cluster (Tables 7–10). Several example phrases for the given theme are provided to give context to the themes. The mean semantic



analysis score is calculated by averaging the semantic analysis scores from each of the passages in the cluster.

3.2.3 NMM Co-Presence Questionnaire–AR Drum Circle/Zoom

The NMM questionnaire used to gauge co-presence yielded better, yet comparable results to the AR Drum Circle Application run on the NReal device. No significant differences were observed between the two platforms for any of the categories used to measure co-presence, including perceived message understanding (PMU) and perceived behavioral interdependence (PBI). The means of each categorical score are represented in **Figure 6**, with an observed grand mean demonstrating the mean of all responses across the two platforms.

4 DISCUSSION

Using AR we aim to enhance co-presence in remote drumming collaborations, traditionally accomplished without visuals. Preliminary observations including video analysis, motion capture, and drummer interviews afforded insight into what features were the most necessary to include. We selected several animations that represented a natural idle drum state (head motion, eye-gaze, and foot tapping). We also used compressed music messages that come directly from the drum pad. Results yielded several key insights: 1) the AR Drum Circle application made players feel more comfortable playing with others because they became less self-conscious, 2) visual issues are apparent in Zoom, because of limited camera angles, 3) latency appears to be more apparent in the AR Drum Circle app,

but when latency is noticed in Zoom it tends to be more distracting.

Through experimentation we explore the differences in 3D spatial configurations in AR using an avatar with the 2-dimensional representations afforded by the commonly used networked conferencing application, Zoom. Results indicated that there are benefits to the 3-dimensional representations, as well as a sense of comfort that can be gained by limiting visual cues. This discovery was made possible through written and interview feedback, and would not have been so easily obtained through traditional quantitative methods. Future versions of the application that include multiple players and video in the AR experience will enable players to select desirable features that are attuned to their comfort level. They will also enable players to communicate amongst themselves spatially by more accurately representing hand gestures and head movements related to musical communication in 3-dimensions.

4.1 Limitations and Future Work

Exploring musical or non-musical forms of avatar animation is important to us moving forward. One way to enhance avatar animation is to embed sound processing into the application. By analyzing sound, we can assess whether musicians are playing together, keeping a beat, or communicating other aspects of the jam. This data can then inform head motion, foot motion, or other forms of body language.

Tangible extensions of the application, such as a foot pedal, can also fill this role. For example, we noticed during user testing that head motion in collocated jamming environments, was used to indicate turn-taking. We could use a foot pedal to indicate the intention of turn-taking, enabling players to more seamlessly jam together.

We also noticed that talking (as an auditory and visual experience) is important in jam sessions; therefore, we would like to add an option to animate the avatars when players talk to each other. Future versions of the application will include the ability to send audio messages that move the lips and head of the avatar instead of solely moving the arms and torso.

The use of video can also accomplish this goal. In future versions of the application, we would like to explore the use of video to show facial expressions and hand gestures through one or more video window. The use of video could enhance the ability for players to communicate with each other. The ability to turn on and off features in the application may improve the experience of playing music over a network and should also be investigated in future versions of the AR Drum Circle application.

Lastly, because of limitations in shipping and availability, many NReal devices were difficult to obtain. For our test, we only had 2 NReal devices. When more headsets are available, we can also test spatial aspects of a drum circle with more than two participants.

5 CONCLUSION

Music improvisation is a complex form of human communication. Real-time remote music collaboration requires high network speeds and is limited by distance. Therefore, exploring mixed methods to support real-time as well as non real-time collaboration tools can help mitigate the effects of latency and enhance collaborative musical experiences. AR Drum Circle contributes to the space of networked musical collaboration tools by using AR to address critical shortcomings of remote jamming. With more and more collaborative musical activities occurring remotely, tools that foster musical interactions in the comfort of one's home are of high importance. The AR Drum Circle expands what is possible in musical collaboration by retaining a sense of co-presence, increasing comfort, and reducing self-consciousness in shared, remote experiences.

REFERENCES

- Antonini Philippe, R., Schiavio, A., and Biasutti, M. (2020). Adaptation and Destabilization of Interpersonal Relationships in Sport and Music during the Covid-19 Lockdown. *Heliyon* 6, e05212. doi:10.1016/j.heliyon.2020.e05212
- Beacco, A., Oliva, R., Cabreira, C., Gallego, J., and Slater, M. (2021). "Disturbance and Plausibility in a Virtual Rock Concert: A Pilot Study," in *2021 IEEE Virtual Reality and 3D User Interfaces (VR)* (IEEE), 538–545. doi:10.1109/vr50410.2021.00078
- Berry, R., Makino, M., Hikawa, N., and Suzuki, M. (2003). "The Augmented Composer Project: the Music Table," in *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings*, 338–339. doi:10.1109/ISMAR.2003.1240749
- Billinghurst, M., Kato, H., and Poupyrev, I. (2004). *Tangible Augmented Reality*. Biocca, F., and Harms, C. (2002). "Defining and Measuring Social Presence: Contribution to the Networked Minds Theory and Measure," in *Proceedings of PRESENCE 2002*, 1–36.
- Bishop, L. (2018). Collaborative Musical Creativity: How Ensembles Coordinate Spontaneity. *Front. Psychol.* 9. doi:10.3389/fpsyg.2018.01285

DATA AVAILABILITY STATEMENT

The anonymized data supporting the conclusions of this article will be made available by the authors upon request.

ETHICS STATEMENT

Written informed consent was obtained from the individuals for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

TH, AB, and EY contributed to the conception of the project. TH, SC, EW, and RV developed the prototypes. TH, SC, and EW and conducted the experimental study. TH, AB, SC, and EW performed the statistical analysis. TH wrote the first draft of the manuscript. SC, EW, RV, AB, and EY wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This study received funding from Ericsson Research. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

ACKNOWLEDGMENTS

We would like to acknowledge Hooman Hedayati, Peter Gyory, Darren Sholes, and Colin Seguro for their assistance in developing the experimental prototype, our collaborators at Ericsson Research for their support and feedback, the drum players whose videos were used for analysis, and the participants involved in the surveys and experimental study.

- Cáceres, J.-P., and Chafe, C. (2010). Jacktrip: Under the Hood of an Engine for Network Audio. *J. New Music Res.* 39, 183–187. doi:10.1080/09298215.2010.481361
- Chew, E., Sawchuk, A., Tanoue, C., and Zimmermann, R. (2005). "Segmental Tempo Analysis of Performances in User-Centered Experiments in the Distributed Immersive Performance Project," in *Proceedings of the Sound and Music Computing Conference* (Salerno, Italy).
- Drioli, C., Allocchio, C., and Buso, N. (2013). "Networked Performances and Natural Interaction via Lola: Low Latency High Quality A/v Streaming System," in *International Conference on Information Technologies for Performing Arts, Media Access, and Entertainment* (Springer), 240–250. doi:10.1007/978-3-642-40050-6_21
- Eaves, D. L., Griffiths, N., Burridge, E., McBain, T., and Butcher, N. (2020). Seeing a Drummer's Performance Modulates the Subjective Experience of Groove while Listening to Popular Music Drum Patterns. *Music. Sci.* 24, 475–493. Publisher: SAGE Publications Ltd. doi:10.1177/1029864919825776
- Jo, D., Kim, K.-H., and Kim, G. J. (2016). "Effects of Avatar and Background Representation Forms to Co-presence in Mixed Reality (MR) Tele-Conference Systems," in *SIGGRAPH ASIA 2016 Virtual Reality Meets Physical Reality: Modelling and Simulating Virtual Humans and Environments* (New York, NY, USA: Association for Computing Machinery), 1–4. SA '16. doi:10.1145/2992138.2992146

- Jung, B., Hwang, J., Lee, S., Kim, G. J., and Kim, H. (2000). "Incorporating Co-presence in Distributed Virtual Music Environment," in *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (New York, NY, USA: Association for Computing Machinery), 206–211. VRST '00. doi:10.1145/502390.502429
- Kipp, M. (2012). Multimedia Annotation, Querying and Analysis in Anvil. *Multimed. Inf. Extr.* 19. doi:10.1002/9781118219546.ch21
- Kleimola, J. (2006). *Latency Issues in Distributed Musical Performance*. Telecommunications Software and Multimedia Laboratory.
- Kokotsaki, D., and Hallam, S. (2011). The Perceived Benefits of Participative Music Making for Non-music University Students: a Comparison with Music Students. *Music Educ. Res.* 13, 149–172. Publisher: Routledge _eprint. doi:10.1080/14613808.2011.577768
- Krueger, J. (2011). Extended Cognition and the Space of Social Interaction. *Conscious. Cognition* 20, 643–657. doi:10.1016/j.concog.2010.09.022
- Lee, S., Lee, N., and Sah, Y. J. (2020). Perceiving a Mind in a Chatbot: Effect of Mind Perception and Social Cues on Co-presence, Closeness, and Intention to Use. *Int. J. Human-Computer Interact.* 36, 930–940. Publisher: Taylor & Francis _eprint. doi:10.1080/10447318.2019.1699748
- Lin, G., Panigrahi, T., Womack, J., Ponda, D. J., Kotipalli, P., and Starner, T. (2021). "Comparing Order Picking Guidance with Microsoft Hololens, Magic Leap, Google Glass Xe and Paper," in *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications* (New York, NY, USA: Association for Computing Machinery), 133–139. HotMobile '21. doi:10.1145/3446382.3448729
- Mastnak, W. (2020). Psychopathological Problems Related to the COVID-19 Pandemic and Possible Prevention with Music Therapy. *Acta Paediatr.* 109, 1516–1518. eprint. doi:10.1111/apa.15346
- Nowacki, P., and Woda, M. (2020). "Capabilities of Arcore and Arkit Platforms for Ar/vr Applications," in *Engineering in Dependability of Computer Systems and Networks*. Editors W. Zamojski, J. Mazurkiewicz, J. Sugier, T. Walkowiak, and J. Kacprzyk (Cham: Springer International Publishing), 358–370. doi:10.1007/978-3-030-19501-4_36
- Poupyrev, I. (2000). "Augmented Groove: Collaborative Jamming in Augmented Reality," in *SIGGRAPH 2000 Emerging Technologies Proposal*.
- Schuett, N. (2002). "The Effects of Latency on Ensemble Performance," in *Bachelor Thesis* (CCRMA Department of Music, Stanford University).
- Signiant (2021). *VR and AR Are Surging under COVID-19 | Signiant*.
- Simsek, M., Aijaz, A., Dohler, M., Sachs, J., and Fettweis, G. (2016). 5G-Enabled Tactile Internet. *IEEE J. Sel. Areas Commun.* 34, 460–473. Conference Name: IEEE Journal on Selected Areas in Communications. doi:10.1109/JSAC.2016.2525398
- Thompson, W. F., Graham, P., and Russo, F. A. (2005). Seeing Music Performance: Visual Influences on Perception and Experience. *Semiotica*, 2005, 203–227. Number: 156 Publisher. De Gruyter Mouton. doi:10.1515/semi.2005.2005.156.203
- Vinnard, V. (2018). The Classroom Drum Circle Project: Creating Innovative Differentiation in Music Education. *Delta Kappa Gamma Bull.* 85, 43–46.
- Xu, C. (2021). Nreal: Ready-To-Wear Mixed Reality Glasses. *SPIE AVR21 Ind. Talks II* 11764, 1176409. doi:10.1117/12.2597454
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2022 Hopkins, Weng, Vanukuru, Wenzel, Banic, Gross and Do. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.