

Heart Fire: A Smart Watch-based Musician-Listener Interaction System for Online Live-streamed Concerts: A Pilot Study

Tianyi Wang ¹

¹Affiliation not available

October 30, 2023

Abstract

Purpose: Build new musician-listener interaction using smart watch for online music.

Tools: Smart Watch, Kinect, Touchdesigner

Heart Fire: A Smart Watch-based Musician-Listener Interaction System for Online Live-streamed Concerts: A Pilot Study

Tianyi WANG, Shima OKADA

Abstract—Online live-streaming has become the new status quo in live music performances in the post-COVID-19 era. However, methods to enhance interaction between musicians and listeners at online concerts are yet to be adequately researched. In this pilot study, we propose Heart Fire, a system to promote musician-listener interaction that visualizes the listeners' mental states using a smart watch-based architecture. Accordingly, the listeners' heart rates are first measured using a Galaxy smart watch and then processed into a real-time animation of a burning flame, whose intensity is dependent on the heart rate, using Azure Kinect and TouchDesigner. The feasibility of the proposed system was confirmed using an experiment involving ten subjects. Each subject selected two types of music—cheerful and relaxing. The BPM and energy of each song were measured, and each subject was asked to answer a questionnaire about the emotions they experienced before, during, and after listening to the songs. The results demonstrated that the proposed system is capable of visualizing audience response to music in real-time.

Index Terms—Online Live-streaming, Smart Watch, Musician-Listener Interaction, Visualization

1 INTRODUCTION

STRICT quarantine and spatial distancing measures to contain the COVID-19 pandemic have altered people's lifestyles fundamentally on a global level. People are spending more time at home because they work and study online. Fink et al. ranked the daily domestic activities performed by people during the lockdown and found that listening to music ranked third, just behind staying up-to-date with the outside world (for example, talking to other people over the phone or watching the news) and domestic chores (for example, cleaning and cooking) [1]. This is a reasonable observation because people use music as an effective tool to cope with various forms of psychological distress such as anxiety, depression, loneliness, stress, and poor sleep quality [2], [3], [4], [5], which have been exacerbated by the pandemic. Indeed, several studies have demonstrated that listening to music aids relaxation, improves mood, reduces negative emotions, abates loneliness, and relieves daily stress [6], [7].

To prevent transmission of infections in crowded on-site concerts, artists have started to perform music differently—primarily via online live-streaming. 70% of artists were reported to have live-streamed since the COVID-19 breakout, with 40% of live-streaming once a month or more and over 80% of them willing to make live-streaming a permanent part of their performance plans even after in-

- Tianyi WANG is with Ritsumeikan Global Innovation Research Organization, Ritsumeikan University, Shiga, Japan.
E-mail: t-wang@fc.ritsumei.ac.jp
- Shima OKADA is with the Department of Robotics, Faculty of Science and Engineering, Ritsumeikan University, Shiga, Japan.
E-mail: s-okada@fc.ritsumei.ac.jp

Corresponding author: Tianyi WANG
Manuscript received ; revised .

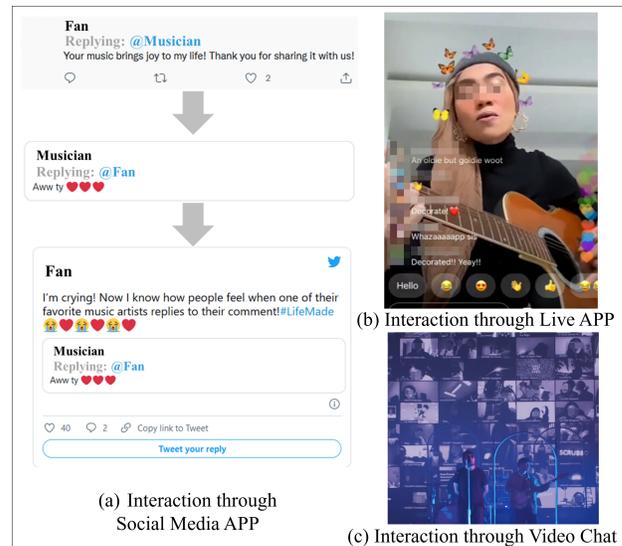


Fig. 1. Present Online Musician-Listener Interaction.

person events are resumed. Similarly, over 70% of music listeners have reportedly watched a live-stream at home, and 60% plan to continue live-streaming even after in-person concerts are resumed [8], [9]. Thus, online live-streaming of concerts can be expected to become the new status quo in the music industry for the foreseeable future.

Besides performances, interactions between musicians and listeners are also being increasingly conducted online. Fig. 1 depicts an example of musician-listener interactions on online platforms. By using social media apps such as Twitter (Fig. 1 (a)), music fans and artists can communicate with each other. Popular live-streaming apps (Fig. 1 (b)) enable audiences to directly comment and send emoticons

during a musical performance, which the artist can notice in real time. Online video chatting software is another popular tool used by artists to organize online concerts (Fig. 1 (c)). Although audiences can turn on their cameras and display their faces in such an architecture, it is difficult for artists to read the atmosphere because of the overabundance of faces on their screens. Thus, online music performances lack real-time and dynamic interaction. Concert-goers usually clap and wave their hands synchronously with the rhythm of the ongoing song, sing along, and cheer excitedly and loudly, which inspires the performing artists. In addition, musicians may occasionally improvise based on their listeners' responses. Thus, online concerts weaken the bond between musicians and listeners—if musicians are not able to experience the listeners' responses, their enthusiasm, in turn, is discouraged.

In this context, several studies have been conducted to improve musician-listener interaction in online concerts through different perspectives. Yang et al. proposed a smart light stick to improve interaction [10]. The stick is equipped with a triaxial accelerometer and LEDs, whose colors change when the stick is shaken at the correct frequency. Similar light-based interaction enabled using wireless motion sensors were also proposed in other studies [11], [12], [13]. Since the beginning of the smartphone era in 2007, musician-listener interaction has been increasingly carried out on smartphone-based systems through sound [14], text [15], and control interface [16]. Hödl et al. established that the utilization of a smartphone as a technical device can inspire creative work in artists and enhance their interaction with listeners [17]. An interesting application for musician-listener interaction was proposed in [16]. The researchers designed a smartphone app on which audience members could draw using brushes of different shapes, colors, sizes, and borders during musical events. These artworks were visible to the musician during the concert. However, Hödl et al. also argued that using smartphones during a musical performance may also serve as a distraction [18], because of the necessity of operating the device—whether by typing or handling it in particular ways to convey their responses to the performer.

In a recent study, heart rate variability (HRV) was analyzed as a metric of mental state while listening to music on different devices, for example, stationary speakers, headphones, and smartphone speakers [19]. The authors concluded that, in terms of HRV, live-streaming was less impressive than an on-site concert due to poorer audio quality on low-level audio devices. However, the study analyzed mental state using retrospective analysis rather than analyzing the mental state information continuously in real-time. Thus, it does not encapsulate the real-time state of mind of the audience members. Furthermore, the collection of electrocardiograph (ECG) data, which is necessary to apply traditional HRV analysis, of people attending online concerts is difficult.

Therefore, it is essential to develop a real-time and user-friendly method to visualize the mental states of online concert attendees. In this study, we propose Heart Fire, a smart watch-based system capable of visualizing the mental states (for example, calm or excited) of people attending online concerts. The right-hand figure in Fig. 2 depicts an

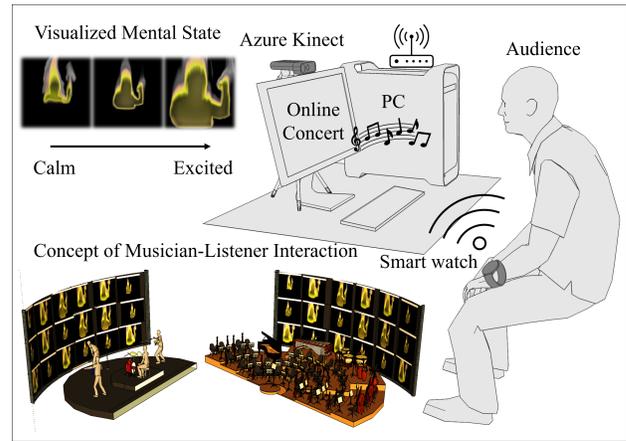


Fig. 2. Concept of Proposed System.

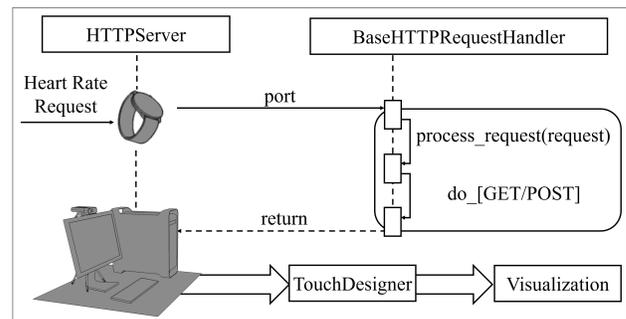


Fig. 3. Overview of Proposed System.

illustration of the proposed system. The audience member's mental state is detected using a smart watch and transmitted to a local computer via Wi-Fi. It is then visualized using Azure Kinect and processed using TouchDesigner. The output of Heart Fire is depicted at the top-left corner of Fig. 2. When the audience member is calm, the visualized fire burns slowly, whereas when they are excited, the fire burns intensely. We believe that the proposed system will inspire novel musician-listener interaction during online performances.

The remainder of this paper is organized as follows. In Section 2, the details of the proposed system are presented in addition to an experiment conducted to evaluate its performance. The results are presented in Section 3. Social aspects and technique notes are discussed in Section 4. Finally, the paper is concluded in Section 5.

2 METHODS

Fig. 3 depicts an overview of the proposed system. Briefly, the system consists of two parts—a smart watch to record real-time heart rate and a local computer equipped with Azure Kinect and TouchDesigner to process the recorded heart rate signal and convert it into visualized data.

2.1 Smart Watch

In this study, we used the SAMSUNG Galaxy Watch with a 42 mm body (GW42) to record the heart rate. GW42 is a Tizen-based wearable device with a 1.15 GHz Exynos 9110

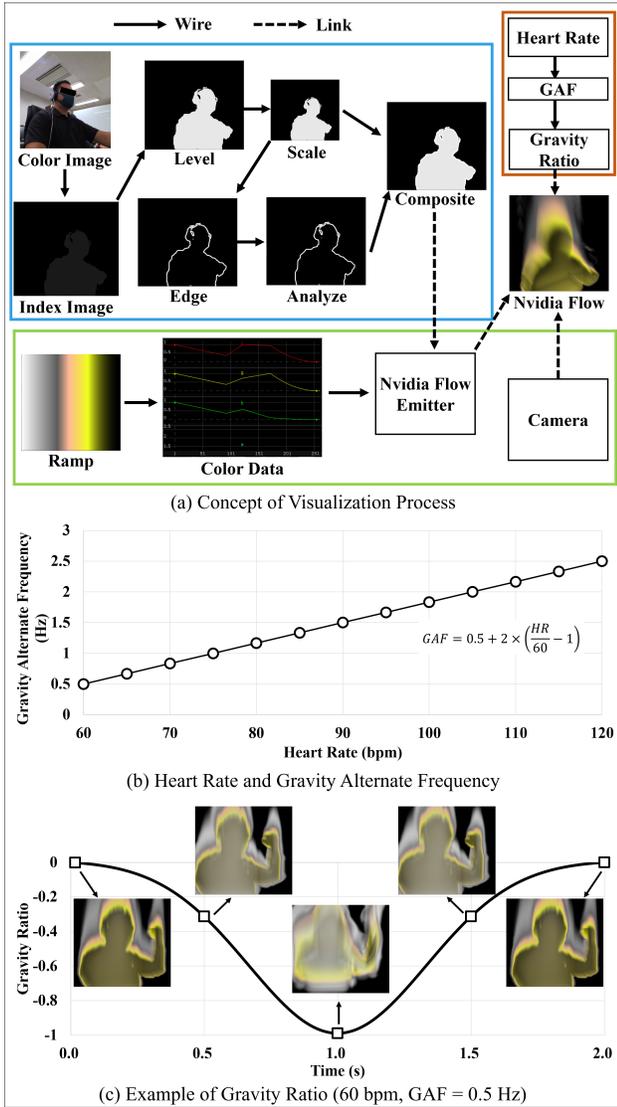


Fig. 6. Converting Heart Rate to Visualized Media.

smart watch and then converted to a parameter used for visualization.

Firstly, the index image was used to separate the user’s image from the background. Then, parameters such as brightness, gamma, and contrast of the video were adjusted to clarify the user’s silhouette. Following that, the edge of the user’s image was identified after adjusting the size of the image. The rescaled image and image of the edge were then combined into one file and used as one of the parameters of visualization. The ramp node was used with color data to stimulate the effect of fire, and the camera node was used to ensure a proper viewing angle. Finally, real-time recorded heart rate was used to simulate a burning fire. In Heart Fire, the heart rate is directly proportional to the intensity of the fire. Intensity was selected to be the variable property instead of size (e.g., fire size > 1) to represent the mental state owing to the constraint of window size to display the user’s video. Nvidia Flow was used to control the intensity of the flame by adjusting the gravity of the fire effect along the Y-axis (that is, in the vertical direction). Thus, a Gaussian

curve with maximum value of 1 and different frequencies was used to control the gravity alternate frequency (GAF). Fig. 6 (b) illustrates the GAF with respect to the heart rate. Assuming that the heart rate of a healthy person usually ranges from 60 to 120 beats per minute (bpm), the heart rate was fitted using the following equation to determine the interactive GAF:

$$GAF = 0.5 + 2 \times \left(\frac{HR}{60} - 1 \right) \quad (2)$$

When the heart rate lies between 60 and 120 bpm, the GAF lies between 0.5 and 2.5 Hz. Higher heart rates are visualized using more intensely burning fire. Fig. 6 (c) depicts an example of the temporal variation of the gravity ratio corresponding to a heart rate of 60 bpm and a GAF of 0.5 Hz. Details of each node are summarized in Table 1. A screenshot of the TouchDesigner interface is depicted in Fig. 7.

2.3 Experiment

The feasibility of the proposed system was experimentally verified. Ten subjects (eight males and two females, age: 25.2 ± 3.5 y.o.) participated in the experiment, and the contents of the study were explained to them orally before the experiment. Data that contains any information that may identify any individual person were not used in this research.

We asked all subjects to choose two pieces of music by themselves—one that cheers them up (Type 1) and another that relaxes them (Type 2). Then, the beats per minute (BPM) and energy of both types of music were investigated using a music analysis website, Musicstax [25]. BPM is defined as the tempo of a track in beats per minute, and energy is a measure of how intense a track sounds based on measurements of dynamic range, loudness, timbre, onset rate, and general entropy—0% indicates low energy and 100% indicates high energy.

During the experiment, each subject wore a smart watch and listened to their chosen music on YouTube. Heart rate data were recorded 1 min before playing the music (Pre), while playing the music (Music), and 1 min after the music had finished playing (Post), for reference. The corresponding GAFs were calculated using Eq. 2. All three GAFs were normalized using the average Pre-GAF. Based on the concept of the proposed interaction system, the following visualized effect was expected: 1. Subjects’ heart rates were expected to increase while listening to Type 1 music (Type 1 HR) and the GAF was expected to be higher (Type 1 GAF); 2. Subjects’ heart rates were expected to remain constant or decline while listening to Type 2 music (Type 2 HR) and the GAF was expected to be lower (Type 2 GAF).

After the experiment, a follow-up Semantic Differential scale (SDs)-based survey was conducted to record the emotions experienced by the subjects after listening to music. The SDs consist of eight pairs of factors that were designed to capture user attitude towards particular pieces of music [26]. Every pair of factors was rated on a 5-point bipolar rating scale (± 2 : strong, ± 1 : slight, 0: neutral), and subjects were asked to rate each factor. An example of SDs is depicted in Fig. 8.

TABLE 1
Parameters of TouchDesigner Node

Name	Node	Main Parameter	Value
Color Image	Kinect Azure (TOP)	Image	Color
Index Image	Kinect Azure (TOP)	Image	Player Index
Level	Level (TOP)	Brightness	2
		Gamma	3
		Contrast	5
Scale	Transform (TOP)	Scale	1 (according to need)
Edge	Edge (TOP)	Type	Luminance
		Black Level	0.7
		Strength	7
Analyze	Anti Alias (TOP)	Edge Detect Source	High Luminance
Composite	Composite (TOP)	Input	Scale + Analyze
Ramp	Ramp (TOP)	DAT	Ramp1 keys
		Hue	360
Color Data	TOP to (CHOP)	TOP	Ramp
Nvidia Flow Emitter	Nvidia Flow Emitter (COMP)	Type	Shape TOP
		Shape OP	Composite
		Material Color	Color Data
Cameraa	Camera (COMP)	Transform	Scale Rotate Translate
Heart Rate	Table (DAT)	Input	GW42
Gravity Alternate Frequency (GAF)	Evaluate (DAT)	Input	Heart Rate
		Algorithm	Self-Decided
Gravity Ratio	LFO (CHOP)	Type	Gaussian
Nvidia Flow	Nvidia Flow (TOP)	Frequency	GAF
		Gravity y	Gravity Ratio

TOP: Texture Operators
 CHOP: Channel Operators
 COMP: Components
 DAT: Data Operators

We used JASP (version 0.14.1.0, The Netherlands) to perform statistical analysis. The mean and standard deviation of BPM, Energy, SDs, and normalized GAF were calculated. The Shapiro-Wilk test was used as the normality test and Levene’s test was used to confirm the equality of variances. Based on the results of appropriate tests, the Mann-Whitney Test was used to analyze BPM, Energy, and SDs corresponding to both types of music. The appropriate effect size was calculated using Rank-biserial (r_B), where an effect size $r_B < 0.1$ indicates a trivial effect, $0.1 < r_B < 0.3$ indicates small effect, $0.3 < r_B < 0.5$ indicates medium effect, and $r_B > 0.5$ indicates large effect. Nonparametric two-way ANOVA (Kruskal-Wallis Test) was used to ascertain the presence of a significant difference in GAF between the different periods and the different types of music. In case a significant difference was observed, Dun’s Post Hoc Comparisons were tested. The significance level was defined to be 5%.

3 RESULT

Table 2 summarizes the results of BPM, Energy, and SDs corresponding to both types of music. Fig. 9 (a) and (b) depict the BPM and Energy of both types of music. On average, BPM of Type 1 songs was higher than those of Type 2 songs, 139.2 ± 26.9 bpm and 125.7 ± 37.2 bpm, respectively. However, there was no significant difference in bpm between the two types of music ($p = 0.38$).

On average, Type 1 music exhibited higher energy than Type 2 music ($87.4 \pm 9.3\%$ and $45.5 \pm 19.1\%$, respectively). The Mann-Whitney test revealed that Type 1 music exhibited significantly higher energy than Type 2 music, with $U = 98.00$ and $p < 0.01$. Rank-Biserial Correlation (0.96) also suggested that this effect was large.

TABLE 2
BPM, Energy, SDs Results for Two Types of Music

Factor	Type 1	Type 2	W	p	r_B
BPM (bpm)	139.2 (26.9)	125.7 (37.2)	62.0	0.38	0.24
Energy (%)	87.4 (9.3)	45.5 (19.1)	98.0	< 0.01	0.96
Dark-Bright	-1.0 (1.4)	-0.4 (1.1)	32.0	0.16	-0.36
Light-Heavy	1.1 (0.7)	-0.2 (1.6)	73.0	0.08	0.46
Soft-Hard	0.8 (1.3)	1.0 (1.4)	45.0	0.72	-0.10
Anxious-Peaceful	-0.5 (1.3)	-1.2 (1.0)	66.0	0.22	0.32
Muddy-Clear	-0.9 (1.4)	-1.5 (1.1)	67.5	0.15	0.35
Crisp-Smooth	0.4 (1.4)	-0.9 (1.6)	71.0	0.10	0.42
Gentle-Intense	-1.3 (0.9)	1.6 (0.7)	2.0	< 0.01	-0.96
Thin-Thick	0.0 (1.1)	-0.9 (0.6)	76.5	0.04	0.53

Data: mean (SD)

The radar chart depicted in Fig. 9 (c) presents the SD values. According to the subjective survey, on average, similar emotions were reported on four metrics (Dark-Bright, Soft-Hard, Anxious-Peaceful, Muddy-Clear) after listening to music. For both types of music, subjects felt relatively bright (both < 0), soft (both > 0), peaceful (both < 0), and clear (both < 0). However, their responses on the Light-Heavy, Crisp-Smooth, Gentle-Intense, and Thin-Thick scales differed. Type 1 music made subjects feel more light, crisp, intense, and thin emotions. Among these, signifi-

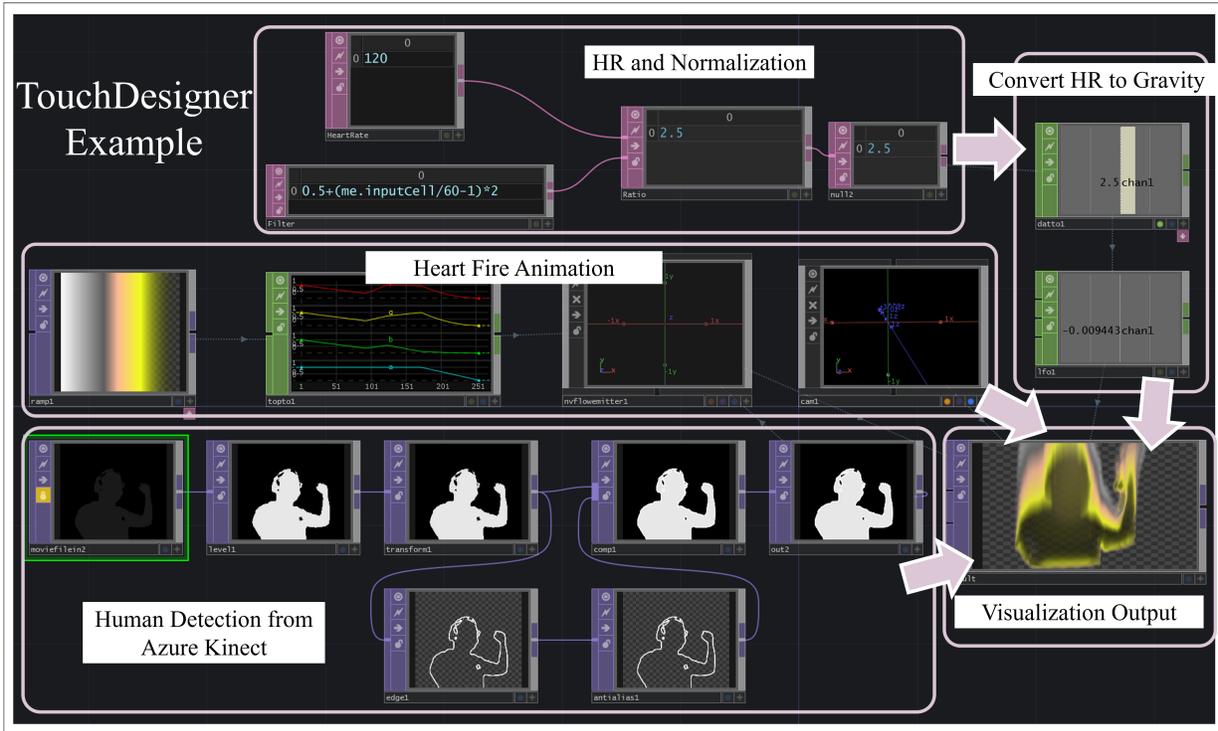


Fig. 7. Example of Node Design in TouchDesigner.

	Strong	Slight	Neutral	Slight	Strong	
	-2	-1	0	1	2	
Bright						Dark
Heavy						Light
Hard						Soft
Peaceful						Anxious
Clear						Muddy
Smooth						Crisp
Intense						Gentle
Thick						Thin

Fig. 8. Semantic Differential Scale Survey.

cant difference was observed corresponding to the Gentle-Intense emotion (Type 1: -1.3 ± 0.9 , Type 2: 1.6 ± 0.7 , $U = 2.00$, $p < 0.01$), and Thin-Thick emotion (Type 1: 0.0 ± 1.1 , Type 2: -0.9 ± 0.6 , $U = 76.5$, $p < 0.05$). Rank-Biserial Correlation (-0.96 and 0.53 , respectively) suggests that these are large effects.

TABLE 3
Results of Normalized GAF

Music Type	Period	Mean	SD
Type 1	Pre	1.000	0.000
	Music	1.029	0.029
	Post	1.003	0.029
Type 2	Pre	1.000	0.000
	Music	0.995	0.036
	Post	0.998	0.036

The normalized GAFs are summarized in Table 3 and plotted in Fig. 10. As expected, GAF increased while lis-

TABLE 4
Kruskal-Wallis Test: Normalized GAF

Factor	Statistic	df	p
Music Type	7.569	1	0.006
Period	1.029	2	0.598

TABLE 5
Dunn's Post Hoc Comparisons - Music Type

Comparison	z	W_i	W_j	p
Type 1-Type 2	2.751	36.700	24.300	0.003

tening to Type 1 music, while it decreased slightly while listening to Type 2 music. However, no significant difference was noticed between the normalized GAFs corresponding to different periods. Kruskal-Wallis tests revealed that different types of music significantly changed the normalized GAF, $H(1) = 7.569$, $p < 0.01$. Dunn's Post Hoc Comparisons demonstrated that both types of music significantly affected normalized GAF ($p < 0.01$).

4 DISCUSSION

In this exploratory study, we proposed a smart watch-based musician-listener interaction system that visualizes audiences' mental states using real-time animation. Presumably, this is the first attempt of visualize online musician-listener interaction using an app on a wearable device.

The experimental results demonstrate that the proposed Heart Fire architecture can successfully convert the audience's mental state into visualized information. The various emotions experienced by listeners of different types of music was faithfully reflected by the intensity of the displayed

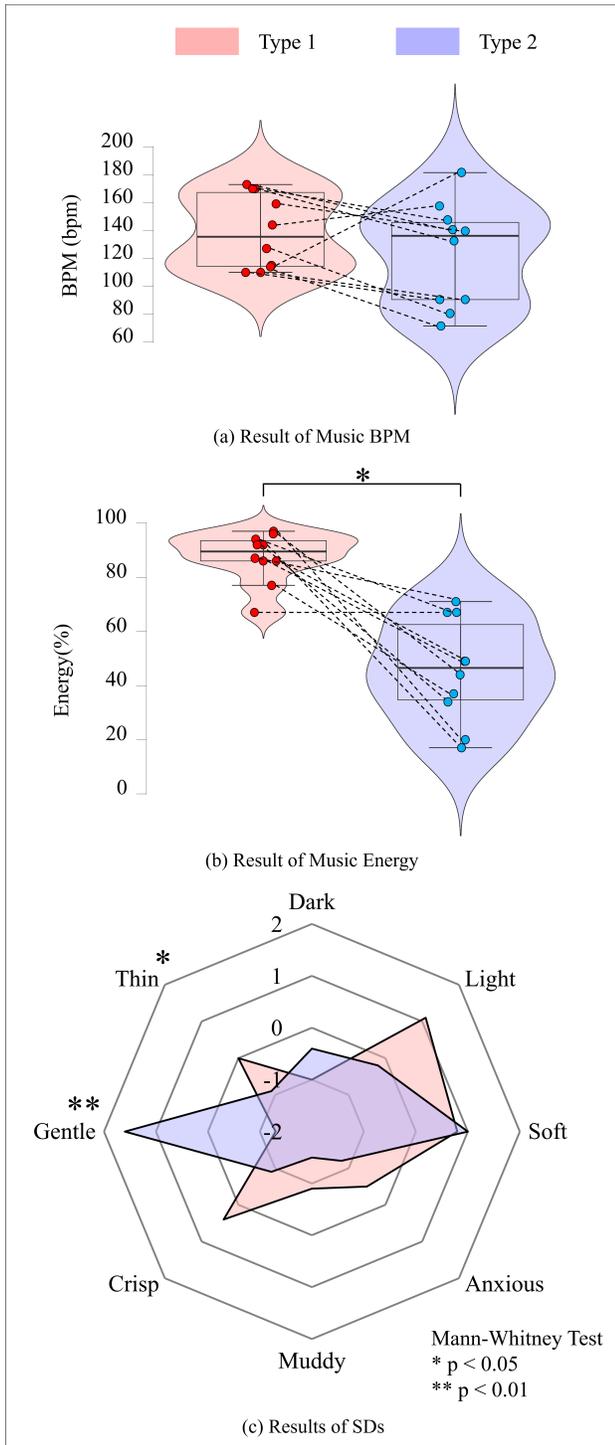


Fig. 9. Characteristics associated with the two types of selected music. Red data correspond to Type 1 music, Blue data correspond to Type 2 music. Violin plots depicted in (a) and (b) represent the music BPM and energy, and the dashed lines represent paired data. The radar chart depicted in (c) illustrates the subjective responses on the SDs, plotted using mean value. Mann-Whitney Test was performed, * indicates $p < 0.05$, ** indicates $p < 0.01$.

fire determined by continuously monitoring the heart rate using a smart watch. Although no significant difference was noticed during the periods before and after listening to the music, listeners generally felt excited while listening to cheerful songs and felt calm while listening to relaxing

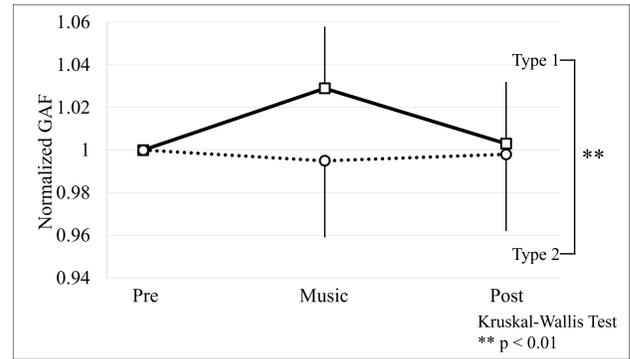


Fig. 10. Result of normalized GAF. The vertical axis represents the normalized GAF, and the horizontal axis represents time. The solid line with square markers represent the Type 1 music, and the dashed line with circle markers represent the Type 2 music. Line charts were plotted using mean value and standard deviation. Kruskal-Wallis test was run, ** indicates $p < 0.01$.

songs.

No significant difference was noticed in BPM among the two types of music. This can be attributed to the fact that the pieces of music were chosen by the listeners, introducing a degree of subjectivity into the selection process. However, it was also noticed that, for most of the listeners (8 out of 10), Type 1 music exhibited higher BPM than Type 2 music—this is evident from the dashed lines plotted in Fig. 9 (a). The energy of the two types of music, on the other hand, exhibited the expected trend—Type 2 music had lower energy compared to Type 1 music.

The SD-based survey revealed the differences in emotion experienced by listeners while listening to the two types of music—it was relatively more intense for Type 1 music and relatively gentler for Type 2 music. This could be one of the reasons behind the higher heart rate in the first case compared to the second. Although subjective emotions regarding different types of music is beyond the scope of this study, the proposed system was proved to be capable of visualizing audience members’ emotions by converting it into the animation of a burning fire.

In the following subsections, we further discuss this research from the two perspectives indicated by our experimental observations—social aspects that should be taken into account while considering musician-listener interaction during online live-streamed concerts and some notes on techniques related to hardware and software.

4.1 Social Aspects

Previous studies attempted to improve musician-listener interaction in different ways. Onderdijk et al. [27] enabled listeners to vote for the final song that they want to hear in a live-streamed concert, but they found the voting mechanism to be incapable of increase the feeling of agency in listeners by itself. This shortcoming was attributed to the one-off nature of the exercise—voting was only enabled during the last part of the concert. This indicates that inculcating a sense of agency in online concert attendees is ineffective if it is not implemented in a continuous manner. In this research, heart rates of concert attendees were visualized using the animation of a burning fire with a detection frequency of 2

Hz. Thus, it enables musicians to be aware of the listeners' mental states continuously. The current study is merely an exploratory one, with no feedback from actual musicians yet. But, it seems reasonable to expect that musicians will be able to use this information to revise the order of their setlists or even improvise based on the mental states of listeners. On the other hand, such improvisations and modifications to the setlist will correspond to the listeners' moods and improve their sense of agency by promoting the feeling that they are part of the performance. In future studies, we would like to enable collaborations between audience members and onstage performers, such as an interactive MTV, creating an experience that would be truly unique to each performance.

Application of virtual reality (VR) or 360° monitors can also augment the physical and spatial presence and involvement of listeners [28], [29]. However, although the short-term use of VR could enhance connectedness between listeners and musicians, watching live-streamed performances for longer duration is a more relevant contributor to enhanced musician-listener interaction—as the watching duration increases, the listeners' empathy with the performer grows stronger [30]. Whether visualized musician-listener interaction increases empathic concern and induces a greater social bond is a question that remains to be answered.

Other than video quality, which is an important factor in the aforementioned tools, audio quality is also a crucial factor to enhance the sense of involvement and presence in live-streamed performances and improving listener experience by promoting social connectedness with the performance, as shown in [31]. In this study, listeners used over-ear headphones (Beats Pro, frequency response = 20–20000 Hz, sound pressure level = 115 dB) and Hi-Fi stereo audio cables (MPS E-100, 1 m, 99.999997% OFC, with customized 24 K gold-plated plug) during the experiment to guarantee acceptable audio quality. The effect of the quality of headphones on mental state of attendees of live-streamed concerts is beyond the scope of this study. The minimum requirements of audio equipment in order to achieve an acceptable sense of presence in listeners should be investigated.

Popular tools for attending online concerts include social media platforms and conference software, where attendees can also see other audience members besides the performers. Although such tools induce a greater sense of social presence, viewing others or being watched by others may distract listeners from the performance, thereby compromising their sense of connectedness with the artist. Our survey revealed an interesting observation—when we asked a subject to choose a cheerful song, he first considered selecting one by his favorite band of female performers but hesitated to select it as he felt embarrassed in letting other people know that his favorite musicians are a female group. Although his final choice also cheered him up, we considered the initial choice to have been a better choice as it was his immediate response. Musical preferences comprise private information—it is essential to avoid personal data leaks especially on the internet. In this study, mental state data were transmitted directly to a local PC, instead of a cloud or a server. In fact, data is not recorded at all because

the proposed method is based on real-time information. Moreover, visualization using TouchDesigner only uses the silhouette of the listener's image and the animated fire masks all identifiable features of the body. This makes it difficult for third parties to discern a person's identity based on only their heart rate and a processed image. Thus, the proposed system adequately protects its users' privacy.

In a recent study, Yang et al. used an LED stick not only for interaction during a musical performance, but also as a social connection tool after the event [10]. Shaking the LED stick at similar frequencies indicated that different audience members shared similar levels of enthusiasm, after which they were free to add each other as friends. It would be interesting to apply this social function to the methodology of our study as well. This would enable users to connect with other who experience similar emotions or exhibit animated fires of similar intensity while listening to a particular song.

4.2 Technique Note

Kayser et al. established that emotional responses of listeners to pieces of music can be studied based on their facial expressions [32]. However, analysis of facial expressions is sensitive to illumination conditions and susceptible to occlusion by hair or spectacles of the user. Moreover, all emotions do not produce discernible facial expressions. Facial EMG (fEMG) is another method for estimating audience emotions during live-streams [33]. When listeners sing along with a sad song, corrugator supercilii, which are muscles used in frowning and are associated with negative valence, exhibit higher activation. In contrast, when they sing along with a happy song, zygomaticus major, which is a muscle used while smiling and is associated with positive valence, exhibits higher activation. However, although fEMG enhances musician-listener interaction, it may disrupt the listener's experience as it uses surface electrodes attached to the listener's face. In this study, we used only a Samsung Galaxy Watch to measure heart rate. There are several advantages to this approach over fEMG—smartwatches are more non-invasive devices, easier to use, and much cheaper. Further, the proposed method does not require calibration and is not affected by noise from the environment or by body movement. Alternative smart watches, such as Apple Watch Series, Fitbit, Garmin, Xiaomi Mi Band, may also serve the purpose [34]. However, further investigation is required to ensure if they can fulfil the requirements of direct transmission of heart rate to TouchDesigner or other software, without passing through an online server or a cloud.

Azure Kinect has been widely used in media enterprises [34], healthcare [35], [36], and robotics [37]. In this study, we used its depth sensor to identify the silhouette of the human body. However, Azure Kinect is also equipped with an RGB camera and an array of seven microphones, which offers the advantage of expanding the scope of the proposed architecture in the future. Further, TouchDesigner facilitates the implementation of Azure Kinect by removing the difficulty of programming using a special prefabricated TOP. A high-performance PC may be needed to ensure that the animation is of high quality. Thus, a 1-GB GPU (for desktop PC: Nvidia GeForce 600 Series or better, or AMD HD 7000 Series or better; for workstation: Nvidia Quadro K series or

AMD W or V series or better) is recommended. The Heart Fire effect is only available if the user is using an Nvidia graphics card.

In a previous study, Onderdijk et al. underlined the trend among musicians to use tools that they are already acquainted with, rather than to explore new platforms, while organizing online concert [38]. This unwillingness can be primarily attributed to the steep learning curve of learning to operate new tools and the general need for technical knowledge. However, the proposed architecture uses TouchDesigner, which supports codeless design of visualization. Thus, it is much easier to use than other popular programming languages, e.g., Python or C/C++. The free account (non-commercial version) provided by TouchDesigner is also an enabling factor for musicians on a tight budget.

Popular video-conferencing platforms, for example, Zoom and Skype, are not suitable for synchronized playing and have been generally reported to be incapable of dealing with latency issues [38]. Although a latency of 100 ms is acceptable in live-streamed concerts, usually latencies between 20 and 40 ms are considered optimal. The total time between frames and playback rate of TouchDesigner is 26.8 ms, which is optimal for visualization. This frame time can be further decreased by optimizing the TouchDesigner nodes using the Null TOP node to avoid complicated network updates.

This study suffers from certain limitations. Due to its exploratory nature, only ten subjects were selected for the experiment, because of which the results must be interpreted with caution. Moreover, the songs were selected by the subjects themselves, and variance in their types and lengths complicated the analysis. However, self-selected music was still preferred as it was deemed to be effective in motivating the subject's emotions [7]. Finally, music was played on YouTube and not live-streamed from any online concert due to the difficulty of organizing such a concert within the COVID-19 restrictions. In the future, we intend to investigate feedback from musicians. Further, this study did not implement VR or 360° monitors to enhance the sense of involvement and presence among the audience. Such realistic experiential environments must be considered in subsequent studies.

5 CONCLUSION

Even after the pandemic, attending on-site concerts will remain riskier compared to watching live-streamed concerts. This is because the lack of social distancing at such concerts enables the spread of COVID-19. In this study, a smart watch-based system was proposed to enhance musician-listener interaction in online concerts. The listeners' mental states were visualized based on their heart rates using Azure Kinect and TouchDesigner. Experiments confirmed that higher heart rates were visualized using larger and more intense fire and lower heart rate were visualized using smaller and less intense fire.

ETHICS STATEMENT

Ethical review and approval was not required for this research on participants in accordance with the local legisla-

tion and institutional requirements.

AUTHOR CONTRIBUTIONS

TW composed the first draft of the manuscript. TW contributed to the concept of the study, construction of the system, main part of the experimental design, data analysis, and the compilation of the results and discussion. SO contributed to the experimental design and revised the manuscript. All authors have approved the submitted version.

ACKNOWLEDGMENTS

This study is partly supported by Research Progressing Project Ritsumeikan University (B21-0132). We would like to thank Editage (www.editage.com) for English language editing.

REFERENCES

- [1] L. K. Fink, L. A. Warrenburg, C. Howlin, W. M. Randall, N. C. Hansen, and M. W. Fuhrmann, "Viral tunes: changes in musical behaviours and interest in coronamusic predict socio-emotional coping during COVID-19 lockdown," *Humanit. Soc. Sci. Commun.*, vol. 8, no. 18, pp. 1–11, 2021, [Online]. Available: <http://dx.doi.org/10.1057/s41599-021-00858-y>.
- [2] P. Dagnino, V. Anguita, K. Escobar, and S. Cifuentes, "Psychological Effects of Social Isolation Due to Quarantine in Chile: An Exploratory Study," *Front. Psychiatry*, vol. 11, 2020, doi: 10.3389/fpsy.2020.591142.
- [3] A. Dawel et al., "The Effect of COVID-19 on Mental Health and Wellbeing in a Representative Sample of Australian Adults," *Front. Psychiatry*, vol. 11, 2020, doi: 10.3389/fpsy.2020.579985.
- [4] C. Franceschini et al., "Poor sleep quality and its consequences on mental health during the COVID-19 lockdown in Italy," *Front. Psychol.*, vol. 11, pp. 1–15, 2020, doi: 10.3389/fpsyg.2020.574475.
- [5] S. K. Brooks et al., "The psychological impact of quarantine and how to reduce it: rapid review of the evidence," *Lancet*, vol. 395, no. 10227, pp. 912–920, 2020.
- [6] A. E. Krause, J. Dimmock, A. L. Rebar, and B. Jackson, "Music Listening Predicted Improved Life Satisfaction in University Students During Early Stages of the COVID-19 Pandemic," *Front. Psychol.*, vol. 11, 2021, doi: 10.3389/fpsyg.2020.631033.
- [7] D. Vidas, J. L. Larwood, N. L. Nelson, and G. A. Dingle, "Music Listening as a Strategy for Managing COVID-19 Stress in First-Year University Students," *Front. Psychol.*, vol. 12, 2021, doi: 10.3389/fpsyg.2021.647065.
- [8] Statista Research Department, "Musicians livestreaming performances during the COVID-19 pandemic as of March 2021," 2021. [Online]. Available: <https://www.statista.com/statistics/1231081/musicians-livestreaming-behavior/>.
- [9] Statista Research Department, "Attitudes toward livestreaming concerts and music events among fans worldwide as of August 2020," 2021. [Online]. Available: <https://www.statista.com/statistics/1221280/concert-livestream-attitudes-and-participation-worldwide/>.
- [10] J. Yang, Y. Bai, and J. Cho, "Smart Light Stick: An Interactive System for Pop Concert," in *Proceedings of the 2017 5th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering (ICMMCCE 2017)*, 2017, pp. 596–600, doi: 10.2991/icmmcce-17.2017.110.
- [11] M. Young, "Concert LED Wristbands," *Trendhunter*, 2015. <https://www.trendhunter.com/trends/led-wristband> (accessed Nov. 18, 2021).
- [12] J. Freeman, "Large audience participation, technology, and orchestral performance," in *Proceedings of the 2005 International Computer Music Conference*, 2005, pp. 757–760.
- [13] M. Feldmeier and J. A. Paradiso, "An interactive music environment for large groups with giveaway wireless motion sensors," *Comput. Music J.*, vol. 31, no. 1, pp. 50–67, 2007, doi: 10.1162/comj.2007.31.1.50.

- [14] N. Schnell, F. Bevilacqua, S. Robaszekiewicz, and D. Schwarz, "Collective Sound Checks: Exploring intertwined sonic and social affordances of mobile web applications," in *TEI 2015 - Proceedings of the 9th International Conference on Tangible, Embedded, and Embodied Interaction*, 2015, pp. 685–690, doi: 10.1145/2677199.2688808.
- [15] L. Dahl, J. Herrera, and C. Wilkerson, "TweetDreams: Making music with the audience and the world using real-time Twitter data," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2011, no. June, pp. 272–275, [Online]. Available: <http://www.nime2011.org/proceedings/papers/G16-Dahl.pdf>.
- [16] J. Freeman, S. Xie, T. Tsuchiya, W. Shen, Y. L. Chen, and N. Weitzner, "Using massMobile, a flexible, scalable, rapid prototyping audience participation framework, in large-scale live musical performances," *Digit. Creat.*, vol. 26, no. 3–4, pp. 228–244, 2015, doi: 10.1080/14626268.2015.1057345.
- [17] O. Hödl, C. Bartmann, F. Kayali, C. Löw, and P. Purgathofer, "Large-scale audience participation in live music using smartphones," *J. New Music Res.*, vol. 49, no. 2, pp. 192–207, 2020, doi: 10.1080/09298215.2020.1722181.
- [18] O. Hödl, F. Kayali, and G. Fitzpatrick, "Designing interactive audience participation using smart phones in a musical performance," in *ICMC 2012: Non-Cochlear Sound - Proceedings of the International Computer Music Conference 2012*, 2012, pp. 236–241.
- [19] Y. Hirata, M. Okamoto, and M. Miura, "Why is live streaming less impressive? Measurement of stress evaluation based on heart rate variability when listening to music," in *The Japanese Society for Music Perception and Cognition (2)-7*, 2021, pp. 25–30.
- [20] SAMSUNG Galaxy, "Galaxy Watch," SAMSUNG, 2021. <https://www.galaxymobile.jp/galaxy-watch/specs/> (accessed Sep. 22, 2021).
- [21] Loic2665, "HeartRateToWeb." GitHub, 2020, [Online]. Available: <https://github.com/loic2665/HeartRateToWeb>.
- [22] J. A. Albert, V. Owolabi, A. Gebel, C. M. Brahm, U. Granacher, and B. Arnrich, "Evaluation of the pose tracking performance of the azure kinect and kinect v2 for gait analysis in comparison with a gold standard: A pilot study," *Sensors (Switzerland)*, vol. 20, no. 18, pp. 1–22, 2020, doi: 10.3390/s20185104.
- [23] B. Sargeant, F. Mueller, and J. Dwyer, "Using HTC Vive and TouchDesigner to projection-map moving objects in 3D space: A playful participatory artwork," *CHI Play 2017 Ext. Abstr. - Ext. Abstr. Publ. Annu. Symp. Comput. Interact. Play*, pp. 1–11, 2017, doi: 10.1145/3130859.3131427.
- [24] TouchDesigner, "TouchDesigner," Derivative, 2017. <https://docs.derivative.ca/TouchDesigner> (accessed Sep. 23, 2021).
- [25] Musicstax, "Musicstax: Music Analysis," Musicstax, 2021. <https://musicstax.com/>.
- [26] I. Takeshi, K. Yoshinobu, and N. Yasuo, "Music Database Retrieval System with Sensitivity Words Using Music Sensitivity Space [in Japanese]," *Trans. Inf. Process. Soc. Japan*, vol. 42, no. 12, pp. 3201–3212, 2001.
- [27] K. E. Onderdijk et al., "Livestream Experiments: The Role of COVID-19, Agency, Presence, and Social Context in Facilitating Social Connectedness," *Front. Psychol.*, vol. 12, 2021, doi: 10.3389/fpsyg.2021.647929.
- [28] J. W. B. Elsey, K. van Andel, R. B. Kater, I. M. Reints, and M. Spiering, "The impact of virtual reality versus 2D pornography on sexual arousal and presence," *Comput. Human Behav.*, vol. 97, pp. 35–43, 2019, doi: 10.1016/j.chb.2019.02.031.
- [29] A. Dekker, F. Wenzlaff, S. V. Biedermann, P. Briken, and J. Fuss, "VR Porn as 'Empathy Machine'? Perception of Self and Others in Virtual Reality Pornography," *J. Sex Res.*, vol. 58, no. 3, pp. 273–278, 2021, doi: 10.1080/00224499.2020.1856316.
- [30] R. K. Dhandra, "Loneliness and Parasocial Interaction With Media Characters," University of California Davis, 2009.
- [31] R. Nordahl and N. C. Nilsson, *The Sound of Being There: Presence and Interactive Audio in Immersive Virtual Reality*. Oxford University Press, 2014.
- [32] D. Kayser, H. Egermann, and N. E. Barraclough, "Audience facial expressions detected by automated face analysis software reflect emotions in music," *Behav. Res. Methods*, doi: 10.3758/s13428-021-01678-3.
- [33] S. R. Livingstone, W. F. Thompson, and F. A. Russo, "Facial expressions and emotional singing: A study of perception and production with motion capture and electromyography," *Music Percept.*, vol. 26, no. 5, pp. 475–488, 2009, doi: 10.1525/mp.2009.26.5.475.
- [34] S. Dijkstra-Soudarissanane, S. Gunkel, A. Gabriel, L. Feroselle, F. ter Haar, and O. Niamut, "XR Carousel: A Visualization Tool For Volumetric Video," in *Proceedings of the 12th ACM Multimedia Systems Conference*, 2021, pp. 339–343, doi: 10.1145/3458305.3478436.
- [35] L. F. Yeung, Z. Yang, K. C. C. Cheng, D. Du, and R. K. Y. Tong, "Effects of camera viewing angles on tracking kinematic gait patterns using Azure Kinect, Kinect v2 and Orbbec Astra Pro v2," *Gait Posture*, vol. 87, pp. 19–26, 2021, doi: 10.1016/j.gaitpost.2021.04.005.
- [36] Á. Uhlár et al., "Kinect azure-based accurate measurement of dynamic valgus position of the knee—A corrigible predisposing factor of osteoarthritis," *Appl. Sci.*, vol. 11, no. 12, p. 5536, 2021, doi: 10.3390/app11125536.
- [37] T. E. Lee et al., "Camera-to-Robot Pose Estimation from a Single Image," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 9426–9432, doi: 10.1109/ICRA40945.2020.9196596.
- [38] K. E. Onderdijk, F. Acar, and E. Van Dyck, "Impact of Lock-down Measures on Joint Music Making: Playing Online and Physically Together," *Front. Psychol.*, vol. 12, 2021, doi: 10.3389/fpsyg.2021.642713.