OPEN ACCESS Freely available online

H. Lindetorp1, and K. Falkenberg,
"Music Collaboration for the Masses,"
J. Audio Eng. Soc., vol. 73, no. 12, pp. 827–836 (2025 Dec.).
https://doi.org/10.17743/jaes.2022.0246.

Music Collaboration for the Masses

HANS LINDETORP,[1,*] AND KJETIL FALKENBERG[2]
(hans.lindetorp@kmh.se) (kjetil@kth.se)

[1] *The Royal College of Music (KMH), Stockholm, Sweden*
[2] *KTH Royal Institute of Technology, Stockholm, Sweden*

This study evaluates a system designed for large-scale collective music sessions, aiming to answer how it performs and how it is perceived by the musicians using it. A music system was developed and an experiment was conducted with 96 participants playing together, using their smartphones as musical instruments. A musical director controlled the pitches for all instruments using a MIDI keyboard. A set of mixed methods was employed to contribute knowledge to improve the design of such systems and to better understand how musicians experience being part of a music session like this. The session was recorded and analyzed, and data from an evaluation form were examined using descriptive statistics and correlation analysis. The result confirms that web technologies are a contributing factor to rapid onboarding, and the musicians were very positive when describing the session. The result shows significant correlations between several sentiments reported by the participants. The authors also conclude that achieving a pleasing musical result is possible even when conditions are not always ideal and that several aspects of how the music session is composed and prepared contribute to an engaging experience for the crowd.

0 INTRODUCTION

Music is a fundamental human expression found in all known cultures worldwide, and most people engage in musical activities in one way or another. A term often used to include all these aspects of engaging with music, either passively or actively, is "musicking" [1]. Musical practices such as listening, singing, and dancing are accessible to most people, whereas playing a musical instrument tends to come with higher barriers, including financial costs and the time required for practice. Digital musical instruments (DMIs) have the potential to lower these barriers and engage more people actively in musicking—in playing and making music.

A recent definition qualifies a DMI to be an electroacoustic instrument if the system includes an input action, a control engine, and sound generation with audio output [2]. It has already been argued for a long time that a laptop can be an expressive instrument [3] and that smartphones take the potential further with built-in accelerometers, gyroscopes, cameras, and multitouch pads [4].

The increase of smartphones in the world, with more than 7 billion mobile network subscriptions,[1] combined with the widely supported Web Audio technology[2] contributes to a new level of accessibility for DMIs. More than 70% of the world's population is thus already carrying a potential musical instrument in their pocket, and new opportunities emerge for musical collaboration with DMIs designed for active and inclusive involvement. This participation in musicking is central to embodied music appreciation and, in turn, closely connected to any experience of playing an instrument [5].

The rapid advancement of technology has created new possibilities for building collective DMIs; however, there are still limitations that can negatively affect the user experience, such as network latency, poor sound quality, and slow touch responsiveness. Earlier studies by the authors [6–10] have explored and exemplified the practical needs of designers when navigating the tension between technical possibilities and limitations. This study builds on those outcomes, adopting a pragmatic approach to what can be achieved with the available technologies rather than focusing on their improvement. The purpose of this project is to explore technical, musical, and collaborative possibilities by describing the design and evaluation of a large-scale collective music system, assessing its features and impact using both quantitative and qualitative methods. The questions in focus are:

*To whom correspondence should be addressed, email: hans.lindetorp@kmh.se

[1] https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/

[2] https://www.w3.org/TR/webaudio/

- How does the system perform technically and musically?
- Which are the experiences of actively performing and being a part of a mass mobile orchestra?

The study contributes knowledge that can guide designers in building systems for large-scale collective music sessions that engage participants in playing music together using DMIs.

1 BACKGROUND

A broad range of academic work, both from a technical and social perspective, has been done in the field of focus for this study. This section presents some of the most essential technologies, projects, and experiences.

1.1 Musicking and Embodiment of Music

"Music" is, in daily speech, something that can be both listened to and performed, and for many, these two perspectives will coincide in amateur music activities. Chanan wrote in *Musica Practica* [11] about the social practices of music-making that are governed by technological development and criticized institutional control of high art forms.

Small suggested a denominalization of the term—*to music*, or musicking [1]—to draw attention to the participational aspects of musical activities. He was, like Chanan, criticizing the traditions tied to Western classical music and in particular the structural focus on composers. Musicking also provides meaning to music that emerges through musical interactions, regardless of one's level of expertise.

Another perspective of music experience is embodied music cognition, which describes the relationship between playing, listening, and mediating music through, typically, an instrument [5]. This mediation is inherently multimodal, which enhances understanding. Embodied music cognition necessitates a discussion of what constitutes a musical instrument as one operates along a continuum from tight connection to the body (such as the voice or a wind instrument) to a loose or even disconnected interaction (such as a musical box or a virtual instrument).

Refsum Jensenius defines, based on the action-sound separation continuum, an electroacoustic instrument to consist of a controller, an engine, and a speaker [2]. He also questions people call an instrument (without providing an answer): Is an orchestra the conductor's instrument or a collection of independent instruments?

1.2 Collaborative Music Making

In contrast to individual musical performances that focus on virtuosity, collaborative music-making enables inclusive practices that emphasize communication and relationships between the players. Studies indicate a positive impact on social, musical, and personal skills [12] and on physiological effects, such as heart rate, when singing or playing music together [13, 14].

Experiments involving large audiences have been conducted over the years, utilizing digital technology to facilitate participation. This includes systems where the audience interacts with the musicians and impacts the musical performance using their smartphones [15], which has demonstrated significant levels of audience creative engagement [16]. Systems where a musician on stage performs with the audience indicate a sustained level of activity among the participants [17] but also highlight essential aspects when designing systems. One aspect is the accessibility of the tool for the audience, and another is the importance of restricting the freedom of the tools to retain artistic control and make it possible to realize an artistic intent [18]. In addition to systems for collaborative music making in a concert setting, similar systems have been built for other purposes, such as augmenting the practice and experience of choir singing [19] and enabling amateur musicians in geographically remote locations to collaborate over the Internet [20].

1.3 Web-Based Smartphone Instruments

Many applications have been developed that turn smartphones into musical instruments, both for commercial and academic purposes. Many are designed for novices rather than professional musicians, with the primary goal of creating instruments that are easy to learn. The tradeoff to achieve this happens at the expense of providing an upward path to virtuosity and is one of the main concerns for the designer [21]. One key to making the instrument easy to learn is to restrict the musical control, which places instruments like those used in this study somewhere along a continuum between a music box and a traditional musical instrument [2]. Another key to providing an effective interface for a DMI is to consider the complex layer of mapping an interaction to an audio parameter [22]—an aspect partly addressed by the Web Audio framework chosen for this study [23].

Web Audio API was first introduced in 2010 and became a *Web Recommendation* in 2021. It has caused a lot of activities resulting in commercial products, such as Soundtrap[3] and Amped Studio,[4] and research projects, such as Nü Soundworks [24], Playsound.space [25], and Orchestra [26]. The advantages of web technologies are many, including aspects such as *omnipresence*, *immediacy*, *discoverability*, and *freedom and openness* [27], and the platforms have proved to be excellent for learning, accessibility and distribution [28].

To simplify the development process of Web Audio applications, numerous frameworks have been introduced. Some build on existing frameworks and provide methods for bringing patches from Cycling '74 Max,[5] Pure Data,[6] or CSound [29] to Web Audio. Other frameworks are JavaScript abstractions or extensions of Web Audio API that simplify the coding, such as WAAX [30], Tone.js [31], and DrSax.js [32]. There are also declarative formats, such as Flocking [33] and WebAudioXML (WAXML) [34], that

[3]https://ampedstudio.com
[4]https://www.soundtrap.com
[5]https://rnbo.cycling74.com
[6]https://github.com/sebpiq/WebPd

separate the descriptive layer from the functional when constructing the audio graph.

This study uses a collection of technologies shared with some of the projects mentioned above and aims to make all parts as accessible, transparent, and replaceable as possible. This led to the system described in detail in SEC. 2 with Node.js,[7] Socket.IO,[8] and WAXML.[9] In this sense, this study does not present a new technology; instead, it focuses on how to utilize a combination of existing components to evaluate the aspects addressed in the research questions.

1.4 Learning From Accessible DMIs

Many smartphone instruments can be seen, in one way or another, as accessible DMIs (ADMIs) [35]. Even if smartphones themselves have shown limitations in terms of accessibility [36], the quest for an instrument that is easy to use when designing for a large-scale collaboration invites ideas from the discipline of ADMIs. Additionally, the general public targeted by mass collaboration systems may not have formal music education, including score reading or instrument proficiency, which introduces another type of accessibility challenge that is closely related to but typically not covered by the ADMI literature. Thus, it is argued here that one should learn from studies that examine improvisation with ADMIs [37] and from studies conducted in special educational needs schools [38, 10] to gain insight into designing systems for inclusive musical instruments more effectively.

It has been pointed out that ADMIs often suffer from high costs, limited availability, and sustainability issues [39]. The choice of technology used in this study aims to address this by utilizing open-source technology and hardware that the participants already possess.

2 THE SETUP

The experiment used for data collection in this study is based on a musical system comprising multiple smartphones with the *musicians' application*, one computer with a MIDI keyboard and the *musical director's application*, and a *server* that connects them all. In this system, a *musical director* can press one or several keys on a MIDI keyboard. The notes are distributed to an *ensemble* of *musicians* with smartphone instruments. The ensemble can consist of an unlimited number of users because there is no technical restriction on the maximum size. The director can also activate a *solo mode* for a randomly selected participant. When a new soloist is chosen, the previous solo instrument returns to its ensemble mode. See Fig. 1 for a schematic overview of the system.

The musical director controls the pitches and the overall dynamics of the ensemble from the MIDI keyboard. The technical system (which can be downloaded from https://github.com/hanslindetorp/Music-collaboration-for-the-masses) and the music are described in the following sections.

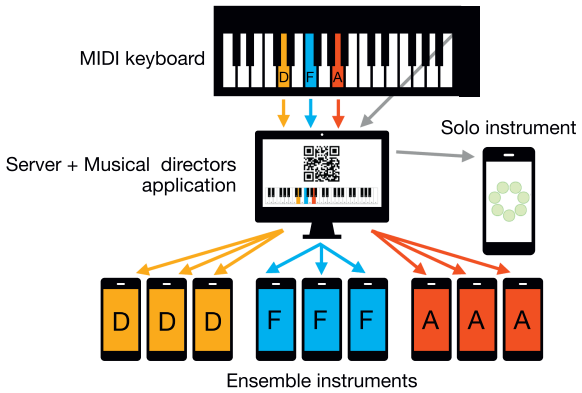

Fig. 1. The system built up with a MIDI keyboard, a server application, and a musical director's application in the computer, a solo instrument, and multiple ensemble instruments illustrating how three notes played on the MIDI keyboard are distributed evenly to the ensemble instruments.

2.1 The Musicians' Application

The participants use their smartphones to run a web application that turns them into musical instruments. The audio is generated using Web Audio and is built with WAXML to construct two different digital instruments (ensemble and solo) utilizing subtractive sound synthesis based on native Web Audio API nodes. The parameters, such as pitch, gain, and cutoff frequency, are controlled by MIDI messages from the system and by user touch events.

2.1.1 Structure

The application consists of five pages. The first two pages are used for the start-up process and have instructions on how to turn off silent mode and adjust the volume to maximum. They have a red background color to indicate that the user is not yet ready to proceed. The following two pages represent the two instruments (described below), and the last page contains a link to an evaluation form. It can be made visible by the musical director when the music session is complete. The application can be accessed from the server (see SEC. 2.3) by scanning a quick response (QR) code generated by the musical director's application (see SEC. 2.2).

2.1.2 Ensemble Instrument

The main instrument used by the ensemble is a *synth-string* sound built as a subtractive synthesis model based on a sawtooth waveform. The interface features a green background and two arrows on the left-hand side, indicating the direction in which the musician is expected to slide their finger to produce a sound. The speed of the sliding finger controls a low-pass filter and the gain and mimics the behavior of playing a violin with a very short bow. Each musician's instrument is monophonic, with its pitch controlled from the server (see SEC. 2.3). This allows each monophonic instrument to contribute to a collective, poly-

[7] https://nodejs.org/
[8] https://socket.io/
[9] https://waxml.org/

phonic ensemble, all of which are controlled by one musical director.

2.1.3 Solo Instrument

The musician's application can be set to solo mode by the musical director, which changes the sound to a brighter synth model using a square wave and a dynamic low-pass filter that follows the root frequency of the oscillator. The interface has a white background and seven green touch buttons with notes picked from a D minor scale (A4, C#4, D5, E5, F5, G5, and A5). The instrument's timbre is chosen to allow the solo instrument to be heard above all the ensemble instruments.

2.2 Musical Director's Application

The musical director's application runs on a computer, preferably one that is the same as the server described in SEC. 2.3. It has two main features: a graphical interface that displays a QR code and a MIDI receiver that forwards messages from a connected MIDI keyboard to the server application. The application is built using Hypertext Markup Language (HTML), Cascading Style Sheets (CSS), and vanilla JavaScript.

2.2.1 Graphical Interface

The graphical interface features a QR code that all participants initially scan to access the application and initiate the session. The interface also has a graphical user interface keyboard. It is mainly used for testing purposes, but it also serves as visual feedback of the notes distributed to the participants' instruments. SEC. 2.3 describes in more detail how the note distribution works.

2.2.2 MIDI

The MIDI messages are received using Web MIDI API through an implementation in WAXML. In this setup, all incoming notes on MIDI channel 1 are distributed to all the musicians' instruments. The MIDI channel 10 has two sets of dedicated notes: one set contains a few notes that are assigned to control the mode of the instruments, described in SEC. 2.1, and notes in the other set trigger different performance test programs, described in SEC. 3.1. One slider on the MIDI keyboard is mapped to control the amplitude of all ensemble instruments, allowing the musical director to balance the volume to the soloist.

2.3 Server

The server is built with Node.js and Socket.io and runs on a computer, preferably the same as the musical director's application described in SEC. 2.2. Its main features are described in the following sections.

2.3.1 Connection

The server can run with various internet connections, ranging from local networks to mobile internet connections. Any connection will make participation possible, but it affects performance (see SEC. 3.1). In the current study, the server and clients were connected via separate mobile internet connections. To simplify access to the server from various connections, Cloudflare Tunnel[10] was used to securely connect a public domain to the server without requiring an external Internet Protocol address. The server stored the userAgent string provided by the clients' `window.navigator` object on each client connection to collect the operating system version.

2.3.2 Controlling Instruments

As described in SEC. 2.2, the server lets the musical directors' application control all the musicians' applications. This is achieved by transmitting incoming MIDI messages to all connected instruments via Socket.io. Different MIDI messages are configured to control the instruments' pitches, control the master volume, activate solo mode, or trigger a performance test.

2.3.3 Note Distribution

The core musical feature of the server is to split layered incoming MIDI notes (i.e., chords) into monophonic MIDI streams. This distributes notes played on the director's MIDI keyboard to all connected instruments. Fig. 1 illustrates how three MIDI notes (colored yellow, blue, and red) are received from the MIDI keyboard and distributed to three instruments per note. The number of instruments playing the same pitch is always the total number of instruments divided by the number of MIDI notes played. Therefore, there must be at least one instrument per MIDI note played for the system to function correctly.

2.4 The Music

The composition used in this study is a mixture of improvisation and composition. The musical director is free to play any notes on the MIDI keyboard. When the solo mode is activated for a musician, that instrument can only play one of the seven notes described in SEC. 2.1, and the director is expected to play matching chords in the key of D minor. The soloist can play notes in any order or rhythm (see Fig. 2). There does not have to be a specified tempo or rhythm because the sound follows the user's interactions and the pitches follow the musical director's MIDI keyboard. The network latency affects the pitches transmitted via the server but not the sound itself because it is generated locally for each instrument.

3 METHOD

This study uses a mixed-methods approach to answer the research questions, collecting and analyzing log data, audio recordings, and participant evaluations. The data were collected during and after a music session. The session was recorded using a Zoom H3-VR audio recorder to conduct a performance test and capture the musical result. After the session, participant feedback was collected using an online form.

[10]https://www.cloudflare.com/

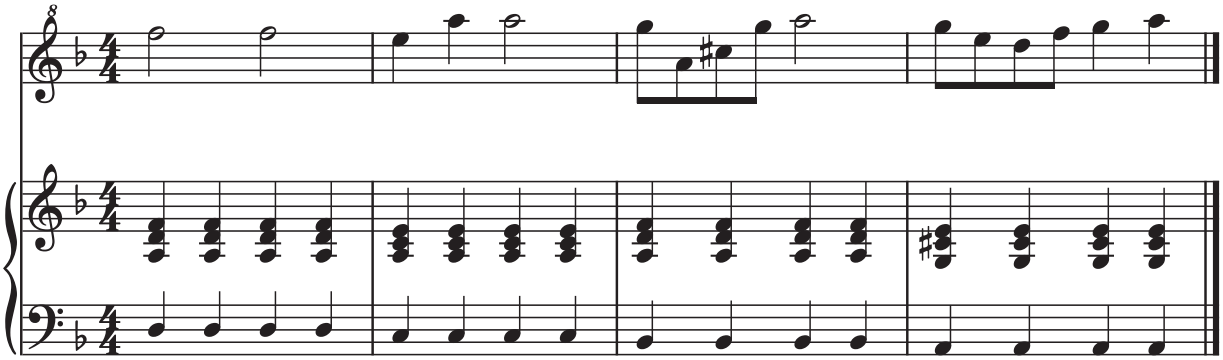

Fig. 2. Four bar excerpt from the music session, illustrating the improvised melody played by the soloist, the chords played by the musical director, and the rhythm played by the participants.

The experiment was performed as the last part of a choir rehearsal. The choir is based in Sweden, outside Stockholm, and consists of 170 predominantly Swedish, female amateur singers aged 13–85 (approximately 10% are male). The majority are between 40 and 50 years old, and the repertoire includes popular songs from musicals, top-lists, and traditional Swedish folk music. Every week, approximately 100 singers rehearse for 90 min. During the rehearsal when this study was conducted, they were informed that the final 15 min would involve voluntary participation in a research study on music and technology. The first author is also a leader and pianist in the choir and was organizing the session.

The experiment began with a brief introduction, during which participants followed on-screen instructions and had a chance to try both instruments. The introduction was followed by the performance test, which was generated and sent out from the server.

Finally, the music session began, with instructions given to the musicians to follow a shared tempo (indicated by the choir's conductor). Initially, they were to slide their fingers across the screen once per beat, and later, they were to slide forth and back on each beat, creating an eighth-note pattern. After 4 min of collaborative playing where five soloists were randomly selected, the director made a fade-out using the volume slider on the MIDI keyboard.

3.1 Performance Tests

One performance test was run at the beginning of the experiment. It played a sequence of short notes, one note at a time, on each instrument. The notes were triggered from the server, separated by 600 ms. The audio was recorded in front of all participants, and the recording was analyzed using Reaper[11] (a digital audio workstation) to identify the onsets for the notes from the different instruments.

A complementary performance test was conducted with three instruments running the same operating system connected via Wi-Fi as a reference. This test was performed in a studio setting without participants and was recorded and evaluated similar to the first test.

[11] https://www.reaper.fm/

In the survey data, two questions addressed system performance, which could not be assessed from the recorded sound. The participants reported whether they received a sound from their mobile and whether they were randomly selected to play solo.

3.2 Online Form: User Evaluation

Following the music session, participants were invited to contribute to the study voluntarily by completing an evaluation form with 11 questions. They were informed that, by filling out the form, they consented to let the authors use the data anonymously for research purposes. The participants completed the online evaluation form using their smartphones.

Two questions were used to validate the system's performance and asked if their phone emitted sound and if they were chosen to be soloists. One question was, "How many minutes would you spend participating in a session like this without getting bored?" Then, the participants were asked to rate agreement to six statements using a five-point Likert scale from "not at all" to "completely." The first three statements covered their experience participating in the session, and the last three covered purposes and situations in which they thought this type of session would be suitable.

The evaluation form had two free-text sections. One asked, "Write three words that best describe your experience of participating and playing." It was intended to collect the primary data for text analysis. The other asked "Do you have any additional comments?" to collect data treated as anecdotal reports to support or explain other findings.

3.3 Survey Analysis

The survey consisted of four parts to be analyzed: the session duration, the statements, the free-text evaluation, and the open-ended additional comments. The authors particularly wanted to investigate the extent to which the participants' experiences related to active musicking, feeling part of a collective and the embodiment of music, and whether such experiences correlated with positive or negative sentiments.

A hybrid approach was used, where a large language model performed translation and text analysis of free-text

responses, and the results were subsequently verified by manual review to ensure precision and reliability. First, the text was translated from Swedish to English. Then, each word (sentiment) was attributed to the variables valence and arousal. Valence was either negative (−1), neutral (0), or positive (1), whereas arousal was low (−1), neutral (0), or high (1). The sentiments were also annotated with several nonexclusive keywords, namely, "musical," "community," "engagement," and "embodiment." The incidences for each variable or keyword were calculated for each participant. Finally, the resulting table was examined using descriptive statistics and correlation analysis in the jamovi statistical software.[12]

4 RESULT

Ninety-six choir members participated in the musical activity, and 84 complete responses were collected from the online form. This section presents the results from the performance test and the user evaluation form.

4.1 Performance Test

It took 12 s from when the participants were asked to scan the QR code until the first smartphone emitted sound, and after 1 min, most participants were up and running. The musicians were asked to turn the volume to maximum manually, but it was never verified whether they did so. Three-quarters of the participants were using iOS, and the others used Android. Most smartphones used an updated version of their operating system, but a few used older versions, such as iOS 16 and Android 7.

The results from the performance test indicate that 89 out of 96 connected smartphones produced a sound. It is not known why the others did not work. The server log reports 112 connections before the test, indicating that there were sixteen reloads of the page during the initial phase of the experiment.

The result indicates a high variance in latency between the connected instruments (see Fig. 3), where 41 responded with a sound less than 250 ms after the server sent the triggering message. This corresponds to a less than a sixteenth note in the tempo (60 beats per minute) used for most of the music session. The musical director typically changed chords once every bar, which caused almost all instruments to change notes within an acceptable timeframe.

A reference test was also conducted in a controlled environment, where all system components were connected to a local Wi-Fi network that was connected to the internet via a fiber-optic cable. In this test, the instruments responded with a latency between 79 and 110 ms.

4.2 User Evaluation

From the 83 respondents, 246 sentiments were collected from the question, "Write three words that best describe your experience of participating and playing." Of these, 90 words were unique. The sentiments were classified into

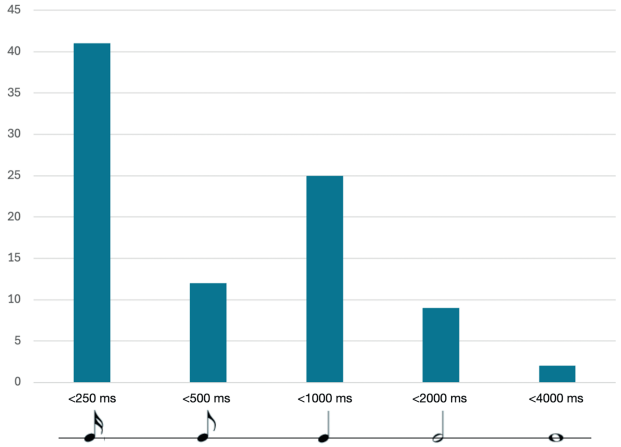

Fig. 3. Latency performance test indicating number of instruments responding faster than a given note value in tempo 60 beats per minute.

Table 1. Reported words to describe the experience, divided into five nonexclusive categories.

		Valence			Arousal		
Category	Total	−1	0	+1	−1	0	+1
Musical	35	3	4	28	12	6	12
Community	22	0	0	22	2	13	7
Musicking	46	3	4	39	12	15	14
Engagement	215	20	4	167	23	28	137
Embodiment	77	0	2	75	8	7	57

Note: Valence is rated from negative (−1) to positive (+1) and arousal is rated from low (−1) to high (+1).

five categories (see Table 1). Each word was attributed a value for valence and arousal. For valence, 182 were attributed a positive valence, 24 a negative valence, and 16 were neutral. Forty-one respondents expressed three positive sentiments, and another seven respondents expressed only positive sentiments. For the sentiments that included negative values, 13 came as the first reported word and five as the only one. Similarly, far more sentiments were classified as high arousal than low (143 high, 42 low). There was a positive correlation between valence and arousal; all correlations are listed in Table 2.

The authors also identified words related to "musical" (35) and "community" (22), such as "monophonic" and "togetherness." These were combined into one category: "musicking." In total, 46 words describe the experience in terms of musicking, whereas sentiments of embodiment (of music) were identified in 77 instances. Musicking was only correlated with embodiment and community, suggesting that these words were not attributed with specific valence or arousal and that they did not co-occur with other sentiments that carried those attributions. Embodiment also correlated with valence and arousal, indicating that sentiments of embodiment are positive, active, and co-occur with music-related words.

Sentiments of engagement correlated positively with both arousal and valence. Still, they did not correlate with

[12]https://www.jamovi.org/.

Table 2. Correlation table including all statistical tests.

		Valence	Arousal	Engagement	Musicking	Embodiment
Arousal	Pearson's r	0.665***	···	···	···	···
	p value	<0.001	···	···	···	···
Engagement	Pearson's r	0.811***	0.688***	···	···	···
	p value	<0.001	<0.001	···	···	···
Musicking	Pearson's r	0.210	−0.178	−0.053	···	···
	p value	0.058	0.110	0.634	···	···
Embodiment	Pearson's r	0.483***	0.320**	0.412***	0.397***	···
	p value	<0.001	0.003	<0.001	<0.001	···
Community	Pearson's r	0.230*	−0.141	0.017	0.717***	0.230*
	p value	0.037	0.207	0.879	<0.001	0.038

Note: degrees of freedom (df) = 80.
*$p < 0.05$
**$p < 0.01$
***$p < 0.001$

musicking or community, which suggests that reported engagement centered mainly on the situation rather than the activity. The related question, "How many minutes would you spend participating in a session like this without getting bored?" had the alternatives 1, 3, 5, 10, or 20 min, and the results show an average of around 10.5 min.

Two (out of six) statements were considered for this study, namely, "I felt I was playing along" and "It was easy to play the instrument." Both were rated 1–5 according to agreement from "Not at all" to "Fully agree." The first, about playing along, got a mean rating of 4.1, and the second, about the ease of playing, got a mean rating of 4.3. None rated the questions as "Not at all."

There were 28 open-ended responses to the question that collected further comments, approximately half of which were statements of positive valence like "it was fun," Of the other half, four reported technical issues, two wanted to play more (and be soloists), one noted "More fun with real instruments," one suggested a different approach to synchronization, and the rest were of little relevance to this evaluation.

5 DISCUSSION

Some aspects of this study have been promising—most participants were up and running within a minute, the evaluation feedback was overall very positive, and the fact that the system worked without a dedicated network or specified selection of smartphones is very encouraging and confirms that web technologies in general and Web Audio in particular are valuable technologies for making accessible interactive music systems for large crowds. With that being said, the performance test indicates that the latency in the test case was far from ideal and would have been challenging to use to trigger sound simultaneously on all instruments. One approach used in this study to overcome this was to prepare a music session that worked with the conditions rather than suffered from them. The most important aspect of this is that the messages from the server did not trigger the sound on the instruments but changed their pitch. The actual sound was generated by the musicians' finger movements on the touchscreen, so they did not experience any latency when playing, even though the pitch had not yet been updated to match the one played by the director.

The sentiment analysis and open-ended responses suggest that people were overly optimistic about the experience. The few negative comments tended to be the first reaction, and often the only one. Most of the negative ones related to not understanding the purpose and, infrequently, to network connection issues. There was no relation between negative valence and high arousal (for instance, words like "scary" or "nervous"); instead, high arousal correlated positively with high valence. This may mean that even when participants used negatively charged words to describe the experience, these typically were of low activity.

The missing correlation between musicking and valence could be interpreted as the participants were already in a musical context and used words that do not necessarily describe affect musically. Also, there was no correlation between sentiments of engagement or arousal and musicking. The experimental setting alludes to musicking being more than a passive process (like only listening). There is also a signaling effect of actively musicking because each participant holds a loudspeaker; these findings might not support Small's initial definition of the term [1], but considering theories of embodied music cognition [5], they could arguably be expected. The reason for attending the rehearsal and participating in the study can also contribute to the missing correlation between musicking and valence—participants were already in a positive, active, musicking "mode," and sentiments describing music could delve more into technical detail than being affective.

A factor that introduces bias and might affect the described experience is that the first author, who led the experiment, is also one of the two leaders of the choir. Furthermore, it should be noted that the test was a voluntary activity at the end of the choir rehearsal, which is likely to have a positive impact on their expectations of partic-

ipating. Another factor that may have contributed to the experience is that the chords for the ensemble instruments and the scale for the solo instrument were similar to those in songs the choir enjoys singing.

In the experimental setting, an embodiment of music was already implied. The participants were all choir singers who could relate to music through carrying their instrument with them always [2]. Engaging with a handheld device emitting sound was likely to elicit a strong sense of embodiment. From the results, it is clear that this was also the case: sentiments of embodiment correlated with those of musicking. Participants agreed almost entirely with the statement about having a feeling of playing along, which confirms that the intended purpose of the experience was realized.

Interestingly, embodiment positively correlates with valence and somewhat less with high arousal. The instrument's low-exertion characteristics could explain this. The fact that the instrument was easy to play was also evident from the high agreement with that statement. However, even if the instrument may be described as not requiring a lot of effort to play, participants still suggested that they could spend, on average, at least 10 min playing it before it becomes boring.

The sentiment analysis faces challenges with attribute classification because a single respondent may use contradictory words or words that can have conflicting meanings in the particular context (such as "monotonous" and "monophonic," which are similar but of opposite valence). The attribution process allows for broad interpretation, with overlapping categories, ambiguous terms, and risk of translation issues. Valence and arousal are also limited to the values −1, 0, or 1, raising questions about what "neutral" represents.

There are numerous opportunities for enhancing a music session, as presented in this study. The system has the potential for greater variation, which could be realized by strengthening the conductor's role. It would be easy to add a short "rehearsal" at the beginning, where the musicians exercise playing with different intensities and rhythms and possibly even follow a simplified "musical score." There is also an opportunity to enhance the collaborative aspect by introducing the concept of call-and-response between different groups of musicians and singers or between the musicians and a solo performer on stage. All these aspects could contribute to the experience and are good candidates for future studies.

With all these ideas being mentioned, it is essential to remember that a barrier to developing such a system is that it is challenging to predict the final result without testing it in real-life scenarios. Therefore, it is always a good idea to add real-time controllers to adjust for unpredictable aspects of the sound.

6 CONCLUSION

An interactive music system designed for large crowds has been developed and evaluated utilizing smartphones as musical instruments. By planning the music session to accommodate technical limitations, such as network latencies, the system performs well and contributes to a pleasant music experience. The system encourages musicking, reported as a predominantly positive experience, where people also want to play for extended periods of more than 10 min. The embodied nature of the smartphone instruments enables tactile interaction through touch and rubbing, creating a tight connection between action and sound. Many aspects remain to be explored, and the authors believe that, if designed to leverage technical limitations as advantages rather than problems and if the music session is composed and prepared according to the conditions, a collective music system like this could attract large crowds to more engaging music-making experiences.

7 ACKNOWLEDGMENT

The authors thank Elias Lang for setting up Cloudflare Tunnel. The authors also extend their gratitude to the members of the SOULFUL choir, who participated in the study.

8 REFERENCES

[1] C. Small, *Musicking: The Meanings of Performing and Listening* (Wesleyan University Press, Middletown, CT, 1998).

[2] A. R. Jensenius, *Sound Actions: Conceptualizing Musical Instruments* (MIT Press, Cambridge, MA, 2022).

[3] R. Fiebrink, G. Wang, and P. R. Cook, "Don't Forget the Laptop: Using Native Input Capabilities for Expressive Musical Control," in *Proceedings of the 7th International Conference on New Interfaces for Musical Expression (NIME)*, pp. 164–167 (New York, NY) (2007 Jun.).

[4] G. Essl and M. Rohs, "Interactivity for Mobile Music-Making," *Organ. Sound*, vol. 14, no. 2, pp. 197–207 (2009 Jun.).

[5] M. Leman, *Embodied Music Cognition and Mediation Technology* (MIT Press, Cambridge, MA, 2007).

[6] K. F. Hansen, C. Dravins, and R. Bresin, "Active Listening and Expressive Communication for Children With Hearing Loss Using Getatable Environments for Creativity," *J. New Music Res.*, vol. 41, no. 4, pp. 365–375 (2012 Dec.). https://doi.org/10.1080/09298215.2012.739626.

[7] L. Elblaus, K. F. Hansen, and C. Unander-Scharin, "Artistically Directed Prototyping in Development and in Practice," *J. New Music Res.*, vol. 41, no. 4, pp. 377–387 (2012 Dec.). https://doi.org/10.1080/09298215.2012.738233.

[8] L. Elblaus, K. F. Hansen, and R. Bresin, "NIME Design and Contemporary Musical Practice: Benefits and Challenges," in *Proceedings of the Practiced-Based Research Workshop at NIME (New Interfaces for Musical Expression)* (London, UK) (2014 Jun.-Jul.).

[9] E. Frid, H. Lindetorp, K. F. Hansen, L. Elblaus, and R. Bresin, "Sound Forest – Evaluation of an Accessible Multisensory Music Installation," in *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Glasgow, UK) (2019 May), paper 677. https://doi.org/10.1145/3290605.3300907.

[10] H. Lindetorp, M. Svahn, J. Hölling, K. F. Hansen, and E. Frid, "Collaborative Music-Making: Special Educational Needs School Assistants as Facilitators in Performances With Accessible Digital Musical Instruments," *Front. Comput. Sci.*, vol. 5, p. 1165442 (2023 Jul.). https://doi.org/10.3389/fcomp.2023.1165442.

[11] M. Chanan, *Musica Practica - The Social Practise of Western Music From Gregorian Chant to Postmodernism* (Verso, London, UK, 1994).

[12] D. Kokotsaki and S. Hallam, "The Perceived Benefits of Participative Music Making for Non-Music University Students: A Comparison With Music Students," *Music Ed. Res.*, vol. 13, no. 2, pp. 149–172 (2011 Jun.). https://doi.org/10.1080/14613808.2011.577768.

[13] E. Morgan, H. Güneş, and N. Bryan-Kinns, "Using Affective and Behavioural Sensors to Explore Aspects of Collaborative Music Making," *Int. J. Hum. Comput. Stud.*, vol. 82, pp. 31–47 (2015 Oct.). https//doi.org/10.1016/j.ijhcs.2015.05.002.

[14] B. Vickhoff, H. Malmgren, R. Åström, et al., "Music Structure Determines Heart Rate Variability of Singers," *Front. Psychol.*, vol. 4 p. 334 (2013 Jul.). https://doi.org/10.3389/fpsyg.2013.00334.

[15] N. Weitzner, J. Freeman, Y. Chen, and S. Garrett, "massMobile: Towards a Flexible Framework for Large-Scale Participatory Collaborations in Live Performances," *Organ. Sound*, vol. 18, no. 1, pp. 30–42 (2013 Mar.). https://doi.org/10.1017/s1355771812000222.

[16] Y. Wu, L. Zhang, N. Bryan-Kinns, and M. Barthet, "Open Symphony: Creative Participation for Audiences of Live Music Performances," *IEEE Multimed.*, vol. 24, no. 1, pp. 48–62 (2017 Jan.–Mar.). https://doi.org/10.1109/MMUL.2017.19.

[17] S. W. Lee, A. Willette, D. Koutra, and W. S. Lasecki, "The Effect of Social Interaction on Facilitating Audience Participation in a Live Music Performance," in *Proceedings of the Conference on Creativity and Cognition*, pp. 108–120 (San Diego, CA) (2019 Jun.). https://doi.org/10.1145/3325480.3325509.

[18] 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 Feb.). https://doi.org/10.1080/09298215.2020.1722181.

[19] L. Turchet and M. De, "A Web-Based Distributed System for Integrating Mobile Music in Choral Performance," *Pers. Ubiquitous Comput.*, vol. 27, no. 5, pp. 1829–1842 (2023 Oct.). https://doi.org/10.1007/s00779-023-01709-0.

[20] D. Prior, P. Reeder, M. Rofe, I. Biscoe, and S. Murray, "Designing a System for Online Orchestra: Peripheral Equipment," *J. Music Technol. Educ.*, vol. 10, no. 2-3, pp. 197–212 (2017 Sep.). https://doi.org/10.1386/jmte.10.2-3.197_1.

[21] T. Blaine and S. Fels, "Collaborative Musical Experiences for Novices," *J. New Music Res.*, vol. 32, no. 4, pp. 411–428 (2003 Dec.). https://doi.org/10.1076/jnmr.32.4.411.18850.

[22] A. Hunt, M. Wanderley, and M. Paradis, "The Importance of Parameter Mapping in Electronic Instrument Design," *J. New Music Res.*, vol. 32, no. 4, pp. 429–440 (2003 Aug.). https://doi.org/10.1076/jnmr.32.4.429.18853.

[23] H. Lindetorp and K. Falkenberg, "Audio Parameter Mapping Made Explicit Using WebAudioXML," in *Proceedings of the 18th Sound and Music Computing Conference* (Virtual) (2021 Jun.-Jul.). https://doi.org/10.5281/zenodo.5038686.

[24] S. Robaszkiewicz and N. Schnell, "Soundworks - A Playground for Artists and Developers to Create Collaborative Mobile Web Performances," in *Proceedings of the 1st International Web Audio Conference (WAC)* (Paris, France) (2015 Jan.).

[25] A. Stolfi, A. Milo, and M. Barthet, "Playsound.space: Improvising in the Browser With Semantic Sound Objects," *J. New Music Res.*, vol. 48, no. 4, pp. 366–384 (2019 Sep.). https://doi.org/10.1080/09298215.2019.1649433.

[26] D. Dziwis, H. von Coler, and C. Pörschmann, "Orchestra: A Toolbox for Live Music Performances in a Web-Based Metaverse," *J. Audio Eng. Soc.*, vol. 71, no. 11, pp. 802–812 (2023 Nov.). https://doi.org/10.17743/jaes.2022.0096.

[27] H. Choi, *Collaborative Musicking on the Web*, Ph.D. Thesis, Stanford University, Stanford, CA (2016 May).

[28] H. Lindetorp and K. Falkenberg, "Evaluating Web Audio for Learning, Accessibility, and Distribution," *J. Audio Eng. Soc.*, vol. 70, no. 11, pp. 951–961 (2022 Nov.). https://doi.org/10.17743/jaes.2022.0031.

[29] V. Lazzarini, E. Costello, S. Yi, and J. Ffitch, "Extending Csound to the Web," in *Proceedings of the International Web Audio Conference (WAC)* (Paris, France) (2015 Jan.).

[30] H. Choi and J. Berger, "WAAX: Web Audio API eXtension," in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, pp. 499–502 (Daejeon, Korea) (2013 May). https://doi.org/10.5281/zenodo.1178494.

[31] Y. Mann, "Interactive Music with Tone.js," in *Proceedings of the 1st International Web Audio Conference (WAC)* (Paris, France) (2015 Jan.).

[32] E. Hong and J. Kim, "Interactive Web Audio Networking System and Method With DrSax.js," in *Proceedings of the 3rd International Conference on Communication and Information Processing (ICCIP)*, pp. 411–414 (Tokyo, Japan) (2017 Nov.). https://doi.org/10.1145/3162957.3162968.

[33] C. B. Clark and A. Tindale, "Flocking: A Framework for Declarative Music-Making on the Web," in *Proceedings of the Sound and Music Computing Conference*, pp. 1550–1557 (Athens, Greece) (2014 Sep.).

[34] H. Lindetorp and K. Falkenberg, "WebAudioXML: Proposing a New Standard for Structuring Web Audio," in *Proceedings of the 17th Sound and Music Computing Conference*, pp. 25–31 (Torino, Italy) (2020 Jun.). https://doi.org/10.5281/zenodo.3898655.

[35] E. Frid, "Accessible Digital Musical Instruments—A Review of Musical Interfaces in Inclusive Music Prac-

tice," *Multimodal Technol. Interact.*, vol. 3, no. 3, p. 57 (2019 Jun.). https://doi.org/10.3390/mti3030057.

[36] M. Bowman, J. Robinson, E. Buehler, and S. Kane, ""I Just Thought It Was Me": How Smartphones Fail Users With Mild-to-Moderate Dexterity Differences," in *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, p. 4 (New York, NY) (2023 Oct.).

[37] K. Samuels and F. Schroeder, "Performance Without Barriers: Improvising With Inclusive and Accessible Digital Musical Instruments," *Contemp. Music Rev.*, vol. 38, no. 5, pp. 476–489 (2019 Nov.). https://doi.org/10.1080/07494467.2019.1684061.

[38] A. Förster, "Accessible Digital Musical Instruments in Special Educational Needs Schools – Design Considerations Based on 16 Qualitative Interviews with Music Teachers," *Int. J. Hum.-Comput. Int.*, vol. 39, no. 4, pp. 863–873 (2022 Apr.). https://doi.org/10.1080/10447318.2022.2049141.

[39] D. Moreno and A. Maia, "Accessible Music for Everyone: Discovering Resources," in A. Mesquita, A. Abreu, and J. V. Carvalho (Eds.) *Smart Innovation, Systems and Technologies*, pp. 883–893 (Springer, Singapore, Singapore 2022). https://doi.org/10.1007/978-981-16-5063-5_73.

THE AUTHORS



Hans Lindetorp



Kjetil Falkenberg

Hans Lindetorp studied music pedagogy at The Royal College of Music (KMH) in Stockholm, where he has been teaching music production since 1996. He earned a doctor of technology degree from the KTH Royal Institute of Technology, Department of Media and Interactive Design, with his thesis titled "Interactive Sound and Music Technology for Everyone." The work aims at building bridges for creators with little or no technical skills to contribute artistic ideas and content in applications that are traditionally limited to programmers and developers. As a platform for his studies, Hans uses web technologies and has proposed WebAudioXML (WAXML) as a potential standard format for implementing interactive audio in web applications.

•

Kjetil Falkenberg studied pedagogy and musicology at Norwegian University of Science and Technology (NTNU), Trondheim. He earned a doctor of technology degree in speech and music communication from the KTH Royal Institute of Technology, Department of Speech, Music, and Hearing, with a thesis on modeling disc jockey (DJ) scratching. He is a docent in media technology and associate professor in sound and music computing at KTH. His key research interests are music interaction, sound design, sonification, and accessibility. Kjetil is a member of the Sound and Music Computing Network, the Musical Care Network, the Aural Diversity Network, and the Nordic Network on Disability Research.