# Count-Me-In: A Collaborative Step Sequencer for Audience Participation

**Juan Pablo Carrascal**
Microsoft
jpcarrascal@acm.org

## ABSTRACT

Modern musical events create a boundary between performers and the audience, with only the former playing and active role. We introduce Count-Me-In, a collaborative music sequencer that uses a distributed Web architecture to promote audience participation in music performances, installations, and other related contexts. Count-Me-In uses a step-sequencer design to reduce the negative effects of network latency on the participants' experience.

## 1. INTRODUCTION

In primitive societies, music has served as cohesive aspect of social activities such as religious rituals [1]. Music events are highly participatory, with all attendees playing an active role. In modern western society, where individual roles have become more and more specialized, participants of musical events are most often split into *performers* and *audience* [1] , with only the former playing an active role and producing music onstage.

Filipino musician and ethnomusicologist José Maceda criticized the commercial proclivity of western music [3], while studying indigenous music and advocating "*for a return to the experience of music as a social and communal event*" [4]. It is thus not surprising that some of his experimental works focused on using technology to bring the music performance back to a collective, participatory experience. In his work *Ugnayan*, he enabled large audiences to become performers by means of the *distributed speaker array* approach [5]: all of Manila radio stations broadcast the piece, while inhabitants were encouraged to go out with their portable radios to create a massive city soundscape. Similar participatory speaker arrays have become popular and used by several composers ever since. Please refer to Taylor's work [5] for a story of this approach.

A more contemporary take on participatory music makes use of modern computing devices like mobile phones to provide interaction capabilities to the audience [6–8]. An important aspect of this approach is to provide the audience with a clear understanding of the cause and effect relation between their actions and the resulting musical output. This aspect, also known as agency, is critical for the quality of musical interaction [9]. It has been observed that the perception of agency when observing a digital music instrument performance can be unclear [10]. Furthermore, providing agency to the *audience* while preserving overall musical structure, becomes a balancing act. Because of this trade-off, some systems are more suitable for ambient or experimental music. Solutions like voting systems [11, 12] have been used to achieve more structured musical results at the expense of agency [12].

[1] Note that Small [2] considers many others, besides musicians and audience, to be participants of the "musicking" activity.



Figure 1. Spontaneous Count-Me-In session with children.

In recent years, Web technologies have evolved to incorporate features that make it possible to develop real-time music applications. Schnell et al [13] make use of several modern Web standards to create a collaborative sound and visual installation. Matuszewski [14] created a framework for the development of distributed multimedia applications based on Web standards. Using Web technologies on mobile devices makes it extremely easy to deploy collaborative applications, as it relieves the need for installing dedicated applications, lowering the barrier for participation. Furthermore, the introduction of the Web MIDI API allows Web applications to communicate with external musical equipment, opening new possibilities for collaborative music performance.

With Count-Me-In, we aimed at creating a participatory and playful music experience that provides participants with agency while producing structured musical output. The system was designed around a collaborative step sequencer and leverages networked computing devices and modern Web technologies. Count-Me-In can easily fit spontaneous, 2-people sessions (e.g. Fig. 1), scenarios involving a larger number of people, or even combined with other electronic or acoustic instruments.

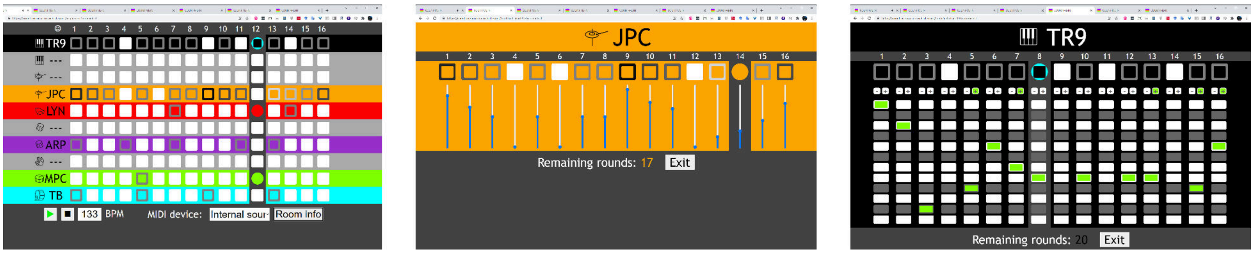


Figure 2. Count-Me-In Web apps. Left to right: the Sequencer, the Drum Track and the Synth Track.

## 2. DESIGN RATIONALE

Count-Me-In allows a group of people to become active performers and to create music together in a playful manner. We designed it with these principles in mind:

*Collaborative*. It engages audience members by making them active and conscious participants of the performance.

*Social*. It should promote social interaction. Focus should not be on each individual's interaction exclusively, but in the collaborative contribution. It should be a fun experience for all participants.

*Interactive*. In contrast to passive speaker arrays, members of the audience can control musical parameters. Causal relationship between their action and sound output should be clear (agency). Interactivity and agency allow members of the audience to effectively become performers.

*Democratic*. It should be very easy to use, regardless of age, technical knowledge, or musical proficiency. It should be platform independent, and it should not require a complicated installation or authentication.

*Scalable*. It should support a variety of scenarios with diverse numbers of people. It should work self-contained, but it should also be possible to use it along other electronic or acoustic instruments.

In alignement with these principles, we aimed for simplicity in sound and graphic design. We did not enable timbre editing or other complex functions and controls in the current iteration of Count-Me-In, as we wanted participants to focus on the collaborative experience. Some design decisions (e.g. visual cues like track color coding and labels) were aimed at helping participants to recognize their own and others' contribution to the musical activity, hence helping with agency and social awareness [15]. The effectiveness of these visual cues were evaluated in a preliminary study included in section 5.

## 3. SYSTEM DESIGN

### 3.1 Architecture

Figure 3 shows the system architecture. It comprises a back-end server running Node.js, a *Sequencer* Web application, and multiple mobile-friendly Web clients, the *Tracks*. The Web applications communicate in near-real-time with the back-end using WebSockets. Participant interaction data are logged to disk for analysis purposes. The Sequencer app runs on an orchestration computer, whose video output is meant to be shared with the audience via

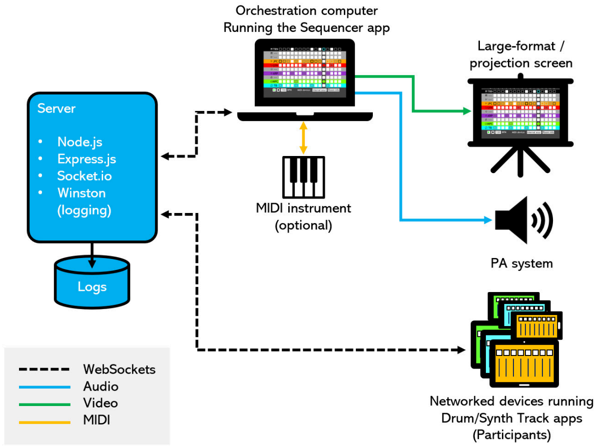


Figure 3. System architecture.

a large display or a projection screen. All system audio is rendered on the orchestration computer, hence, its audio output should be amplified.

The system is designed around collaborative step sequencer. In a step sequencer, user actions (button presses) *schedule* notes to be played—as opposed to play them immediately. The sequencer then triggers audio at the right times, according to scheduled notes (steps). As user actions are not meant to trigger sounds in real time, the negative effects of network latency in participant engagement [16] are mitigated.

#### *3.1.1 Sequencer and Track Apps*

The Sequencer features 10 tracks and 16 steps and it is depicted in Figure 2 (left). It is worth noting that the number of tracks is not hard-limited to 10. We used a number that allowed all tracks to be comfortable shown on a typical 16:9 display. However, more tracks could be fit using a large vertical display. Adding new tracks only requires trivial code adjustments. Sequencer controls include Play and Stop buttons, Tempo value in beats per minute, a selector to switch between internal or external sounds, and a button to display the current *session* information. As the tempo is determined by the Sequencer app, master tempo adjustments should be done in the orchestration computer. We plan to implement a feature to overcome this limitation in a future iteration of the Count-Me-In.

The Drum and Synth Track applications are shown in Figure 2 (middle and right, respectively). The Drum track features 16 switches to turn individual steps on or off, and

16 slider controls to adjust the velocity of each step. The Synth Track features a 1-octave keyboard per step centered at the C2-B2 octave (C2 = 65.41Hz), with switches to transpose an octave up or down.

#### *3.1.2 Sound Generation*

We implemented a simple square-wave-based monophonic synthesizer and a 8-track drum sample player using the WebAudio API [17]. The synth was based on a simple square wave, with 30 milliseconds of portamento. Drum samples were a subset of the popular Roland TR-808 [2] sounds (Kick Drum, Snare Drum, Clap, Hi Tom, Mid Tom, Cowbell, Closed Hi-Hat, and Open Hi-Hat). Two instances of the synthesizer are assigned to tracks 9 and 10 respectively, while different drum sounds are assigned to tracks 1 to 8. Further sound design for the system is possible but out of the scope of this paper. However, the Sequencer output can also be sent to an external MIDI device if external sounds are preferred. The use of MIDI also makes it possible to synchronize Count-Me-In with other electronic instruments. MIDI implementation was done using the WebMIDI API [18].

### 3.2 Session Workflow

Figure 4 depicts the workflow of a Count-Me-In *session*, as explained next:

❶ The Orchestration computer should first access the Count-Me-In server [3] and provide a name for the session. ❷ The computer will display the Sequencer app. Its tracks will all be empty and greyed out. A modal window (❸) containing session information will be initially overlaid over the Sequencer. It includes the name of the session, a URL for participants to join the session, and a QR code. At that moment, the session has started. ❹ Once the session information modal is closed, playback will begin. Playback can also be manually started and stopped, and tempo can be adjusted with the relevant controls (Figure 2, left).

❺ Multiple participants can join the session by scanning the QR code or entering the URL in a Web browser. ❻ They are asked for their initials, and they can click a button (labelled "Go!") to enter the session. When a participant enters a session(❼), they are randomly assigned one of the sequencer tracks. ❽ The Track app (either Drum or Synth) will be displayed on their device. Besides the track controls, Track apps display an icon that represents the instrument controlled by the participant and the participant's initials. Each Track has a distinct color, depending on the assigned instrument (Figure 2, center and right). When a participant joins, the Sequencer app will highlight the corresponding app with a color that matches that of the participant's Track, and will display the participant's initials at the left side of the Sequencer track (note the matching colors and initials in Figure 2). We hypothesized that with this the combination of visual cues (color-coded tracks, initials, icons), each participant can easily notice what Sequencer track they are controlling, and hence understand how their interactions result in changes in the sound output.

[2] https://en.wikipedia.org/wiki/Roland_TR-808

[3] Currently deployed at https://count-me-in.azurewebsites.net

❾ As soon as a participant starts interacting with the Track app (e.g. by clicking a step button), 20 loop iterations will be counted. After the 20 iterations have passed, the participant will lose control of the track and will be dropped in the initial screen (their initials are remembered by means of a client cookie). The track is then released for use by other participants, but the changes previously made to the Track remain so as not to disrupt the current composition being played. This turn-taking mechanism prevents participants to take exclusive controls of a track for a long period of time. This also allows a number people higher than the number of Sequencer tracks to participate in the session. We thought it made more sense to set this timeout in musical terms, hence we chose a number of iterations instead of a period of time. 20 loop iterations (40 seconds at 120bpm) seemed adequate in our tests, but further evaluation with participants might be necessary to find the sweet spot. Also, there is no reason for this to be a set number, it could be adjusted depending on the Sequencer tempo or the number of participants.

## 4. PRELIMINARY EVALUATION

We conducted a simple evaluation study with Count-Me-In. This was not meant to be a thorough evaluation of the system, but a way to obtain preliminary feedback to validate some of our design decisions. Specifically, we wanted to determine whether the combination of visual cues (track color, initials, icons), and simple interface controls used in the design of the Sequencer and Track apps, allowed participants to understand the cause and effect relationship between their individual actions and the resulting collaborative output.

We invited 5 people with ages ranging from 8 to 46 years (mean 34.4), 4 of them female, to participate in the evaluation. None of the participants had a background in music, although the youngest (an 8-year old child) has started his musical training. The participants were attending a small social gathering. Given the social aspect of Count-Me-In, we consider this was an adequate context for evaluation.

We set up a computer with a projector and a speaker, started a Count-Me-In session and invited participants to scan the QR code, enter their initials and join the session. We were intentionally brief and vague when providing instructions to participants. We simply asked them to try to use the Web app on their mobiles (we provided an 11-inch tablet to the child) and to observe what happened in the projected app. We then observed and took note of participants' reactions and behaviors.

To avoid interrupting the social activity, we did not ask participants any follow-up questions immediately after the Count-Me-In session took place. Instead, we asked participants what they liked and what they disliked about the session the following day, via text messaging. When needed, we also asked some clarifying follow-up questions, including some regarding the Drum and Synth Track user interfaces (UI). It remains part of future work to conduct a more in-depth study that might include a post-session focus group or individual interviews.

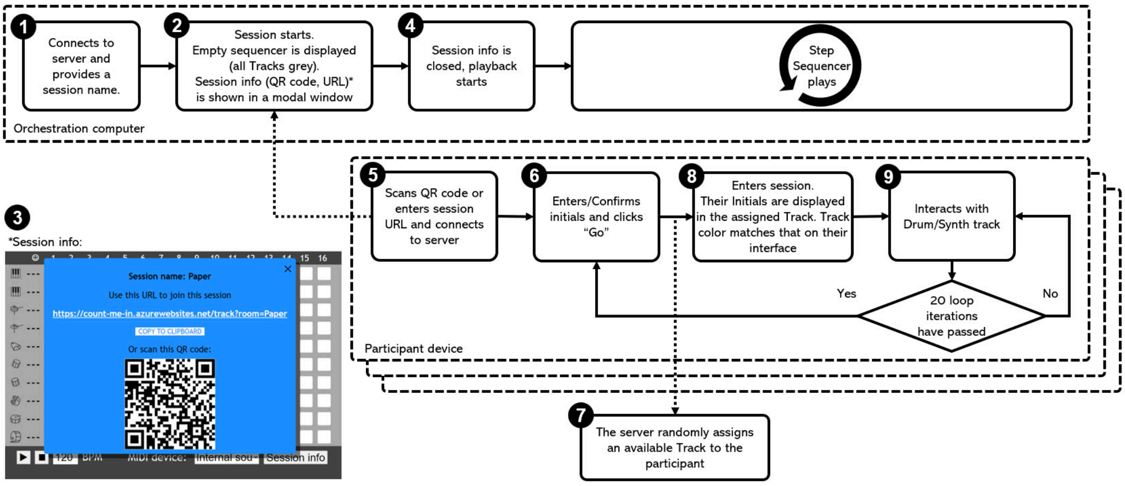


Figure 4. Session workflow and session information window (lower left corner).

Figure 5. Evaluation study setting.

## 5. RESULTS

We logged activity information in the server during the almost 17 minutes that the session lasted. Our logs were time-stamped and included connection and disconnection of each participant as well as interaction events. Interactions could be button clicks (either on the Drum Track or the Synth Track) or slider movements (only available in the Drum Track). During the session, the 5 participants produced a total of 2018 interaction events (min: 237, max: 620, mean: 403.6), for an average of 23.7 events per minute per participant. While there is no baseline to compare with, this is evidence that participants were engaged in the activity.

Overall, the reception to the Count-Me-In session was positive. Based on the way people interacted with the system and with each other, we observed two phases in the activity. In the first phase, as soon as the activity started, participants experimented with the interface, clicking buttons and tweaking controls to see what happened. Phase two came after a few minutes of experimentation, when participants started to have a more strategic approach, teaming up to have structured results. Post-hoc discussion with P2 corroborated the existence of these 2 phases. She explained: *"What I liked most about the instrument was the social element and the way we could explore and make sense of it together, both individually in our own devices, and as a group, on the wall projection and in our discussion. At one time, someone said, let's work together, and I tried coordinating more with the others and creating some sort of visual harmony that I hoped would translate into musical harmony"*. P4 also appreciated the collaborative possibilities of the system and observed that *"[it is] an activity that is shared among several people to form a team together"*. The experience of "making music" with Count-Me-In was generally considered fun and playful, and seemed to elicit engagement and social connection between participants: *"It was fun to be able to interact with a melody and with others"* (P5). A most concise piece of feedback came from P3, an 8-year child, who simply said *"I liked it all"*. We tried to obtain more critical opinions from him, to no avail.

### 5.1 Effectiveness of Visual Cues

Our observations during the session and subsequent responses to post-hoc questions indicated that the visual cues provided participants a sense of agency. After the initial instructions were given and they had joined the session, participants did not ask for additional guidance from the researchers. During the activity, we asked participants if they understood which Sequencer Track was theirs, and the answer was affirmative in all cases.

Visual cues, specially color coding, seemed to be effective in guiding participants through their interaction. Colors were also considered a pleasant aspect of the interface. As explained by P1 *"The interface on the screen was pretty, and the little colors were easy to understand"*. When we asked P4 what helped her the most to identify what sound she was in control of, she stated *"with the button pushing and the assigned icon"*, indicating icons were indeed a helpful visual cue for participants. However, P2 also observed that the icon used in the Synth Track (a generic 1/2 octave keyboard image) misled her to expect the sound of an actual piano.

Participants' answers to post-hoc questions did not reference the use of the initials as Track identifiers, but we observed a lot of lively conversations about who was in control of each of the active Tracks. Initials were therefore used to recognize what Tracks *others* were in charge of, as opposed to which one was their own Track. Understanding the contributions of peers by reading each other's initials helped developing the observed collaborative strategies after the initial exploration phase.

### 5.2 Sound Recognition by Participants

While the visual cues seem to be helpful for participants to identify their Tracks, recognizability of sounds varied between instruments. P1 indicated that *"Some instruments could be heard or were more obvious then others, which made it difficult to recognize what one was doing"*, In P2's words, *"Some of the most difficult to understand and feel like I was indeed controlling them, were the pianos and some drums. I couldn't tell what kind of drums they were, so it was difficult to hear myself, but others, like the clapping, were easy to identify."* P4 even suggested to *"use a single earphone per person to better listen each one's instrument"*. A more extreme case, P5, felt she could not never find her sounds, and explained that *"I could find myself clearly in the projection [but] I would have liked to be able to find [the sound of] my instruments"*. We hypothesize the difficulty to recognize some of the timbres had to do with the sub-optimal quality of the sound amplification system used during the session (the speakers in the projector), as we observed sounds like the kick drum were particularly difficult to hear.

### 5.3 Feedback on the Drum and Synth Track UI

Two out of 5 participants considered the Synth Track interface difficult to use. In P1's words, *"I didn't understand the more complex interface of the keyboard"*. We used a vertical button strip with black and white keys color coding (see Figure 2, left). Users with musical training would probably notice the black and white pattern and vertical orientation commonly used in piano rolls but, admittedly, it does not look like a real keyboard. In contrast, we observed that participants used the Drum Track interface without noticeable issues or complaints.

We did not hear any criticism regarding the Drum Track during the activity, and we observed that all participants seemed to quickly understand how to use it. However, most of the interaction events were button presses (1379, 68%), with less than 1/3 being slider adjustments. To understand why, we asked participants what they thought the sliders were for, and we discovered that they were not well understood. P4 said *"I have no idea. I presume for turning up the intensity"*, while P1 stated *"I think it was the volume"*. The remaining participants had other assumptions, but nobody seemed to clearly understand the function of the sliders.

Finally, there was some confusion among participants regarding the notes that were already set in the Track when they joined or re-joined the session after their turn expired (see ❾ in section 3.1.2). They thought they were placed there as an example, or simply did not know where they came from. This was not problematic, but it was confusing, nonetheless.

In the next section we elaborate on the results and propose potential solutions for some of the issues we identified.

## 6. DISCUSSION

The results of the preliminary evaluation indicate that Count-Me-In has potential for creating playful musical engagement in social contexts. Visual output, specially color coding and icons, effectively helped participants to understand their contributions to the music activity. Initials identifiers were more frequently used by identifying the contributions of *other* participants, which was an unexpected use of the feature. We think this obeyed the need for social awareness, which is a relevant aspect of communal music activities. As explained by Professor Iain Morley, when discussing what constitutes participation in musical and ritual performances, *"to consider oneself a participant [...] in a performance (ritual or musical) relies upon a strong sense of the perception of you by others engaged in the activity, that is, it relies on well-developed theory of mind and social awareness"* [15]. Being able to identify each other's contributions encouraged participants to develop collaborative strategies during the music making activity.

The aesthetics of the Sequencer interface were appreciated by participants, suggesting that the session provided a rich sensory experience: *"I tried coordinating more with the others and creating some sort of visual harmony that I hoped would translate into musical harmony"* (P2).

There is still work to be done regarding the Track interfaces. The vertical piano roll in the Synth Track was difficult to understand. We will consider other approaches. One could be to redesign it to be more piano keyboard-like. Another approach might be to use a slider to control pitch. Unfortunately, sliders were not easily understood by participants in the Drum Track, where they were used to adjust the velocity of each step. Participants indeed explored the sliders, but presumably did not understand the connection between the slider position and the velocity of the drum sounds. We will need to further explore whether this lack of clarity remains when the sliders control pitch, as opposed to velocity. As for the Drum Track, possible improvements include adding labels to the sliders or replacing the sliders altogether with radio buttons with different levels.

We did not receive any critical feedback that could be related to the responsiveness of the system. Further technical evaluation is necessary, but our preliminary study indicates that the step-sequencer design successfully made the effects of network latency and jitter irrelevant to the participant engagement.

We conclude that the combination of visual cues and a step-sequencer design generally helped providing agency to participants. However, this result was not consistent in all circumstances. Participants found it difficult o recognize some of the sounds used by Count-Me-In. Low-frequency sounds, like the kick drum and the center octave used by the Synth Track (C2 = 65.41Hz) were harder to hear. The latter case was aggravated by using an interface that was hard to understand by our non-musically-oriented participants. We think that the use of a low-quality amplification system had an important influence in this result, as low frequencies were not reproduced in an adequate manner. Further evaluation with a proper, full-range amplification system will help clarifying this. Using a higher center octave for the Synth Track would also help users to recognize its sound.

### 6.1 Future Work

The initial results of the Count-Me-In are promising, but there are areas of improvement. In future work we will include the evaluation of improvements mentioned in the precious section; understanding the influence of sound quality in the perception of the different sounds; and a study with focused on verifying that the turn-taking system allows larger number of users to engage with the system.

We also envision different scenarios in which Count-Me-In can play an important role in engaging attendees. First, we are interested to see how current results vary with expert subjects, thus we plan to evaluate the system with users with musical training. We hypothesize that there will be less issues with the interface and sound recognition, and we think the input of musicians might be valuable on other aspects. Additionally, we want to evaluate the value of using the system along with live performances of other instruments. In that scenario, we think Count-Me-In will help creating a cohesive, engaging and fun experience involving both performers and the audience.

## 7. CONCLUSION

We introduced Count-Me-In, a collaborative music making and performance system. We conducted a preliminary evaluation and found evidence that it helps creating a fun, playful and engaging experience for social contexts, satisfying most of our design principles. We identified shortcomings of the interface and observed that the experience of individual users might be affected by the timber of the instrument they are controlling. We also observed that the sound quality of the amplification system is important for an optimal experience. We plan to refine the system based on the feedback received in the preliminary study, and to evaluate it in different contexts, including larger user groups and sessions with musicians.

## 8. REFERENCES


[1] B. Nettl, *Music in primitive culture*. Harvard University Press, 2013.

[2] C. Small, *Musicking: The meanings of performing and listening*. Wesleyan University Press, 1998.

[3] J. Maceda, "A primitive and a modern technology in music," *Unpublished typescript. University of the Philippines*, 1978.

[4] E. Bentcheva, "From ephemeral experiences to lasting legacies: Discourses on experimental art in the philippines during the 1960s and 1970s," *Tate Papers*, vol. 32, 2019.

[5] B. Taylor, "A history of the audience as a speaker array." in *NIME 2017*, 2017, pp. 481–486.

[6] N. Weitzner, J. Freeman, S. Garrett, and Y.-L. Chen, "massmobile-an audience participation framework." in *NIME 2012*, vol. 12, 2012, pp. 21–23.

[7] A. Hindle, "Swarmed: Captive portals, mobile devices, and audience participation in multi-user music performance." in *NIME 2013*, 2013, pp. 174–179.

[8] J. J. Arango and D. M. Giraldo, "The smartphone ensemble. exploring mobile computer mediation in collaborative musical performance." in *NIME 2016*, 2016, pp. 61–64.

[9] O. Bown, A. Eldridge, and J. McCormack, "Understanding interaction in contemporary digital music: From instruments to behavioural objects," *Organised Sound*, vol. 14, no. 2, pp. 188–196, 2009.

[10] F. Berthaut, D. Coyle, J. W. Moore, and H. Limerick, "Liveness through the lens of agency and causality," in *The 15th International Conference on New Interfaces for Musical Expression, 31 May-3 June, 2015, Louisiana State University, USA*, 2015.

[11] M. Gimenes, P.-E. Largeron, and E. R. Miranda, "Frontiers: Expanding musical imagination with audience participation," in *NIME 2016*. Brisbrane, Australia, 2016.

[12] L. Zhang, Y. Wu, M. Barthet *et al.*, "A web application for audience participation in live music performance: The open symphony use case," in *NIME 2016*, 2016.

[13] N. Schnell, B. Matuszewski, J.-P. Lambert, S. Robaszkiewicz, O. Mubarak, D. Cunin, S. Bianchini, X. Boissarie, and G. Cieslik, "Collective loops: multimodal interactions through co-located mobile devices and synchronized audiovisual rendering based on web standards," in *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction*, 2017, pp. 217–224.

[14] B. Matuszewski, "A web-based framework for distributed music system research and creation," *Journal of the Audio Engineering Society*, vol. 68, no. 10, pp. 717–726, 2020.

[15] I. Morley, "Music and ritual: Parallels and practice, and the upper palaeolithic," 2009.

[16] J. Lazzaro and J. Wawrzynek, "A case for network musical performance," in *Proceedings of the 11th international workshop on Network and operating systems support for digital audio and video*, 2001, pp. 157–166.

[17] "Web audio api," https://www.w3.org/TR/webaudio/, (Accessed on 01/27/2022).

[18] "Web midi api," https://www.w3.org/TR/webmidi/, (Accessed on 01/27/2022).