# AUDIENCE AS PERFORMER: A SMARTPHONE WEB-APPLICATION FOR REAL-TIME SOUND CONTRIBUTION

**Davor B. VINCZE** (vincze@hkbu.edu.hk) (0009-0006-3499-8931) [1] and **Maurice OESER** (maurice.oeser@swr.de) [2]

[1] **Hong Kong Baptist University**, Kowloon Tong, Hong Kong
[2] **SWR Experimentalstudio**, Freiburg, Germany

## ABSTRACT

This paper presents a participatory performance system that enables audience members to contribute directly to the sonic material of a live electronic music performance using their smartphones. The system builds on earlier iterations developed for projects *Freedom Collective* and *Exo Signals*, where smartphones functioned as a distributed speaker array and interactive control interface. The current iteration introduces a new module, allowing participants to record short sound bytes that are integrated into the performance in near real time. This development represents a shift from audience interaction through control to audience participation through sound generation. The system was evaluated in three performances at the Australian National University (Canberra), Emory University (Atlanta), and CCRMA (Stanford), followed by a post-concert survey examining usability, enjoyment, and perceived agency. The findings suggest that enabling audience members to contribute their own recorded material increases perceived engagement and sense of participation in the performance.

## 1. INTRODUCTION

In recent decades, the role of the audience in concert and performance settings has gradually shifted from that of a passive listener and observer toward more active forms of participation, agency, and, in some cases, co-creation [1]. While in traditional concert settings the audience's interaction was largely limited to social feedback such as approval (applause) or disapproval (booing), contemporary music practices increasingly explore what Majumder describes as "relational music" [2], in which the audience becomes an active component of the performance system. Recent research has described this shift in terms of different audience roles or "personae" [3], reflecting a broader move toward the "activation of the audience" in contrast to what Bishop describes as "its mythic counterpart, passive spectatorial consumption" [4].

With the development of digital and networked technologies, this shift has led to a growing field commonly described as Technology-Mediated Audience Participation (TMAP), in which audiences interact with live performances through technological interfaces that allow them to influence musical processes, trigger events, or shape musical structure in real time [5]. Within this field, personal mobile devices such as smartphones and tablets have become particularly important, as they allow large audiences to participate simultaneously through web-based interfaces and networked audio systems [6].




Figure 1. Davor Vincze performing *We're in this together* at Emory University (© Davor Vincze).

From pioneering works such as Max Neuhaus' *Public Supply* [7], in which audience members contributed sound via telephone lines, to later projects such as *Dialtones: A Telesymphony* [8], artists have long explored the use of personal communication devices as part of the musical performance system. While some works focused on the use of audience mobile phones as a distributed sound system [9], or interfaces for participatory performance and networked music-making [10], or simply to have a feedback loop [11] all of which are valuable strategies in making the audience shape the musical system rather than just observe it [12].

Along these lines, our initial design of a smartphone-based web application for the opera *Freedom Collective* was guided by a similar principle, where the audience was treated as a distributed array of speakers, creating a spatialized and immersive sonic environment [13]. A second iteration of the system, developed for the project *Exo Signals*, expanded this model by allowing audience members not only to listen, but also to trigger and manipulate sounds in real time through their smartphones, aligning with a broader range of participatory music systems in which mobile devices function as control interfaces [14].

In 2025, during the NIME conference in Canberra, Australia, we presented a new performance titled *We're in This*

*Copyright: © 2026. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*

*Together* [15], which introduced an additional functionality that we call the **RECORD** module. This module enabled audience members, at specific moments during the performance, to record a short sound using their smartphones and upload it to a central server. After the recording window closed, these audience-generated sounds were assembled and reintroduced into the performance through granular and concatenative synthesis techniques. In this way, the audience not only controlled or influenced the music, but also directly contributed sonic material that became part of the musical texture.

The performance was later repeated in 2026 at Emory University (see Figure. 1) and at CCRMA, Stanford University. On each occasion, audience members were invited to complete a short post-concert survey that addressed their experience of participation, engagement, and perceived agency. This paper presents the conceptual framework behind this participatory model, the technical implementation of the system, and an analysis of audience responses collected across the three performances.

## 2. RELATED AND PREVIOUS WORK

One of the early examples of using mobile phones in a musical performance context is Golan Levin's *Dialtones: A Telesymphony*, which utilized the audience's mobile phone ringtones as a distributed sound system, effectively transforming the audience into a grid of loudspeakers [16]. Over the following two decades, as mobile phones evolved into smartphones with increased computational power and network connectivity, a wide range of participatory performance models emerged. These include systems in which audiences and/or performers influence the outcome of a performance through networked interaction, such as massMobile [17]; systems that use smartphones as control interfaces for musical parameters, such as *Soundworks* [11] and *Imaginary Berlin* [18]; mobile phone orchestras and classroom ensembles, as developed by Essl and Wang [19]; and projects in which smartphones function either as musical instruments, such as *echobo* [20], or as playback devices within distributed sound environments, effectively acting as a network of loudspeakers [21]. Some projects relied on dedicated mobile applications, such as *Poème Numérique* [22], while others used web browser–based systems, such as the distributed music projects by Turchet and De Cet [23]. Other works have explored audience participation through social and networked interaction, such as *TweetDreams* [24] and Corporate Retreat [25], where audience activity influences the progression or structure of a performance rather than directly generating sound. Together, these works demonstrate how mobile devices have enabled new forms of audience participation, ranging from distributed sound diffusion and musical control to social interaction and networked performance environments.

Recent Web Audio Conference practice includes Dennis Scheiba's *Network Latency Feedback*, which incorporates audience-generated sounds into a live feedback loop [26]. By contrast, our system collects contributions during designated interaction windows and subsequently transforms and recombines them within a staged compositional structure.

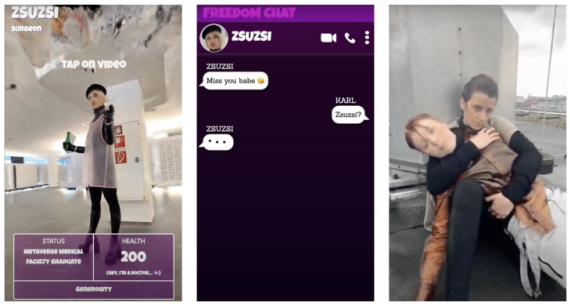



Figure 2. Smartphone interface from the first iteration of the web application developed for *Freedom Collective* (© Rosa Wernecke).

Following these developments, our initial web-application design was created for the audiovisual musical theatre project *Freedom Collective* (2024), where audience interaction was used to complement moments of low stage activity. Similar to approaches described in studies on the audience as a distributed speaker array [27], we used the audience's smartphones as a diffused sound system. In this version of the system, audience members could switch between different video scenes from the opera on their phones, while multiple pre-composed electronic voices were distributed across the audience devices, creating a spatialized sonic environment. As *Freedom Collective* was conceived as an audiovisual performance with a dedicated visual design team, the web interface was developed in close relation to the visual identity of the opera (see Figure. 2).

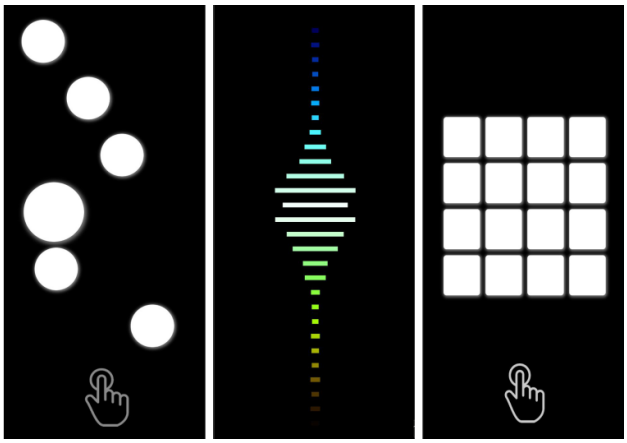


Figure 3. Smartphone interface from the second iteration of the web application developed for *Exo Signals* (© Maurice Oeser).

In the second iteration of the system, developed for *Exo Signals* and presented at the Ubimus conference in Macao in 2024, the visual components were reduced and the focus shifted more strongly toward musical interaction. While the distributed speaker array model was retained, additional interactive modules were introduced, allowing audi-

ence members to trigger sounds and control audio effects in real time through buttons and sliders on their smartphones (see Figure. 3), similar to other participatory systems that use mobile devices as control interfaces for live electronic music [17-23]. However, in both of these versions, as in many of the previously mentioned examples, the sonic material remained entirely pre-composed by the author, and thus the overall authorship and sonic outcome remained largely under the control of the composer.

## 3. SYSTEM DESIGN OF RECORD MODULE

### 3.1 Interaction Design

In most existing systems, the audience interacts with or controls pre-existing sonic material, while the generation of new sonic material by the audience themselves remains comparatively underexplored. To address this, we developed the **RECORD** module. On two occasions during the performance, audience members were given a one-minute time window to record a short sound of up to three seconds using their smartphone by pressing a microphone icon in the web interface (see Figure. 4). After confirming their selection, the recording was uploaded to the server. All submitted sound files were then converted to WAV format and concatenated into a single audio file, which was downloaded within seconds to the performer's laptop, where it could be used for further processing and integration into the performance.

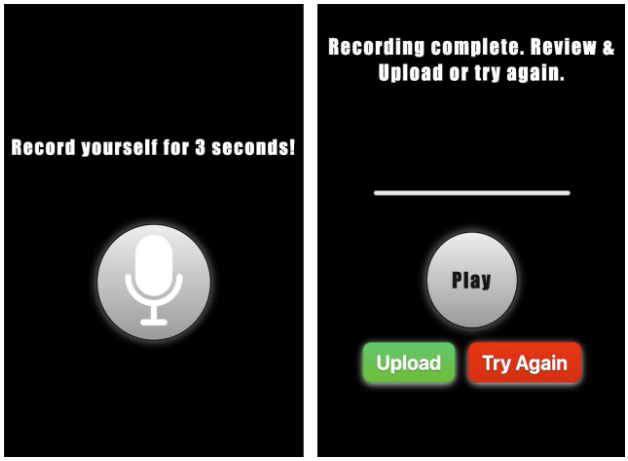



Figure 4. Smartphone interface from the third iteration of the web application developed for *We're in This Together* (© Maurice Oeser)

### 3.2 Technical Implementation

As in previous iterations of the system, the client–server architecture is based on a bidirectional WebSocket connection, allowing fast and reliable communication between the server and the audience's smartphones. This connection is used to transmit control messages, synchronize interaction windows, and trigger events across all connected devices. The main novelty and technical challenge of the RECORD module is that, unlike previous versions, it involves the recording and transfer of audio data, which is significantly

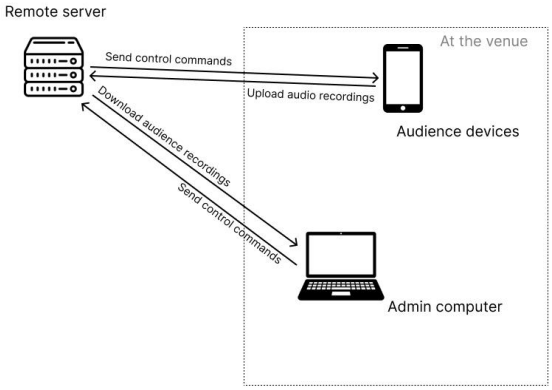



Figure 5. System architecture of the web application showing audio upload from audience devices and control communication between the server and the performance computer (© Maurice Oeser).

larger in size than the control messages used in earlier interaction modules.

To implement audio recording directly in the browser, we used the native MediaRecorder Web API, which is supported by all major web browsers [28]. This API handles implementation details such as audio encoding and user microphone permission management, allowing us to integrate recording functionality without developing a custom recording solution. Because each browser implements the MediaRecorder API slightly differently, the resulting audio files are typically encoded using compressed formats such as OPUS, which are optimized for web-based audio streaming and data transfer, where file size and transfer speed are often more critical than high audio fidelity [29] (see Figure. 5).

However, our audio processing environment is based on Max/MSP, which primarily operates with uncompressed audio formats such as WAV. Therefore, the server—implemented in Node.js—was responsible for receiving the uploaded audio recordings, converting them to WAV format, storing them on disk, and notifying the Max/MSP environment that a new recording was available for processing [30]. Since Max 8 includes integrated Node.js support, the Node.js server could be run directly inside the Max environment, significantly simplifying communication between the web server and the audio engine. Once the recordings were converted and stored, further processing and playback were carried out in Max using custom-built granular synthesis [31] and concatenative synthesis [32].

### 3.3 Musical Integration

The approximately thirteen-minute work is organized into four blocks of roughly three minutes, each foregrounding a different relationship between audience action and vocal material. In the first block, audience recordings become source material for granular synthesis. The second introduces short, button-triggered vocal clicks whose pointillistic character echoes the preceding granular texture. In the

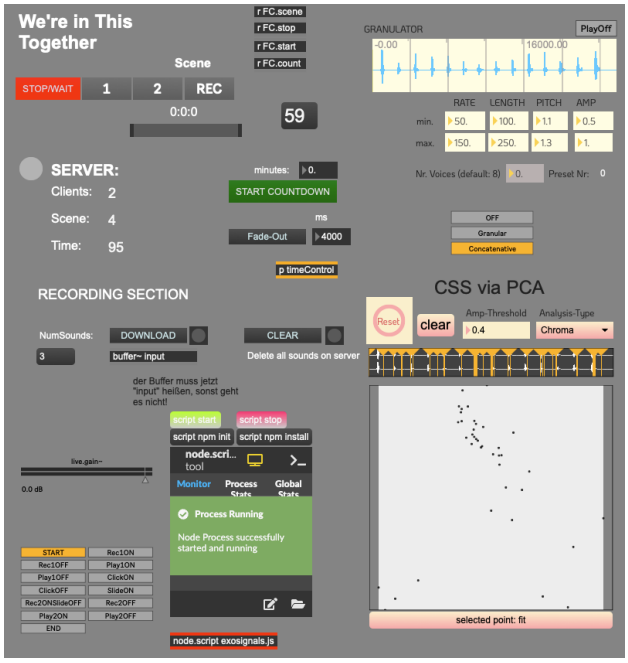



Figure 6. Max/MSP performance interface used to control the system, download audience recordings from the server, and process them using granular and concatenative synthesis during the performance (© Davor Vincze & Maurice Oeser).

third, participants use a slider-controlled filter to expose different registers within a dense, layered texture of vocal hissing. The final block returns to audience recording, but reorganizes the resulting material through concatenative synthesis. This progression gives the interactions distinct formal roles while linking them through increasingly varied transformations of vocal sound.

To highlight the contribution of the audience, in addition to the quadraphonic speaker setup used for the fixed media electronics, two additional frontal loudspeakers were dedicated specifically to the playback of audience-generated sound material. During the first instance of playback, the granular synthesizer allowed the performer to select and emphasize individual audience recordings. After the second recording session, the sound files were segmented into short fragments and organized according to their MFCC (Mel-Frequency Cepstral Coefficients) descriptors and energy levels into a two-dimensional control space [33], which the performer could navigate in real time using a mouse interface. This allowed the performer to musically match and blend the fixed media electronics with the newly generated audience material (see Max patch in Figure. 6).

The title *We're in This Together* reflects the work's central artistic premise: collective participation through a sonic resource shared by performers and audience members—the human voice. The fixed media and interactive materials derive from four broad vocal categories: whispering, singing, speech, and nonverbal sounds such as tongue clicks, lip smacks, and hissing. The button-triggered sounds and filterable textures were recordings of composer's voice, while audience members generally responded to the recording prompts with their own vocal contributions. Consequently, the audience material enters an already established vocal sound world rather than functioning as an arbitrary external input. The interactions create forms of sonic mimicry between composer and audience, framing participation as a collective and embodied compositional process.

## 4. AUDIENCE EVALUATION

### 4.1 Methodology

The system was evaluated across three performances, following common practice in research on technology-mediated audience participation, where post-performance questionnaires are used to assess user experience, engagement, and perceived agency [34]. The first performance took place during the NIME 2025 conference in Canberra, Australia, with an audience of approximately 100 people. The subsequent performances at the Performing Arts Center at Emory University and the CCRMA Stage at Stanford University had smaller audiences of approximately 15–20 people each. In total, 42 completed surveys were collected within approximately ten minutes of the performances.

In addition to basic demographic questions such as age and gender, the survey consisted of seven questions addressing the audience's experience with the three interaction modules (tapping, sliding, and recording), the perceived intuitiveness and comfort of the interface design, whether participants recognized their own contributions in the performance, their perceived sense of agency, and their perception of the overall musical result of the audience contribution. Most questions were presented using a seven-point Likert scale [35], allowing participants to rate their level of agreement with each statement.

### 4.2 Survey Results

The recording module was preferred by 52.4% of participants, followed by the tapping module (35.7%) and the sliding module (9.5%). This preference is consistent with 23 participants describing the recording module as more engaging than the other two. Although the survey did not ask participants to explain the reasons for this preference, the module's direct use of personally contributed sonic material offers one possible interpretation.

Participants reported that the application was generally intuitive to use, with an average rating of 6.29 on a 7-point Likert scale. The sense of agency or co-creation was rated moderately high, with an average score of 4.24 out of 7. Participants also reported feeling comfortable using their smartphones during the performance, with an average comfort rating corresponding approximately to "comfortable" on the provided scale (mean = 3.19 on a 4-point scale).

When asked whether they recognized their own voice or contribution during playback over the loudspeakers, 11.9% of participants reported that they clearly recognized their contribution, while 28.6% responded "maybe." A larger portion (42.9%) reported that they did not recognize their contribution, and 16.7% indicated that they were not aware

that their voice had been used. These responses indicate that the legibility of individual contributions was limited.

An exploratory Spearman rank analysis found only a weak, non-significant association between recognition of an individual contribution and perceived agency ($\rho = .15$, $p = .34$, $n = 42$). Perceived engagement with the recording module was also not clearly associated with agency ($\rho = .20$, $p = .23$, $n = 38$). These results should be interpreted cautiously given the small sample and ordinal response categories.

Regarding the musical result of the audience contributions, most participants described the result as "interesting" (57.1%), followed by "meaningful" (21.4%) and "powerful" (7.1%), while 11.9% reported that they did not notice the audience contribution (see pie charts in Figure. 7). Overall, these results suggest that the integration of audience-generated recordings was perceived positively and contributed to the overall musical experience.

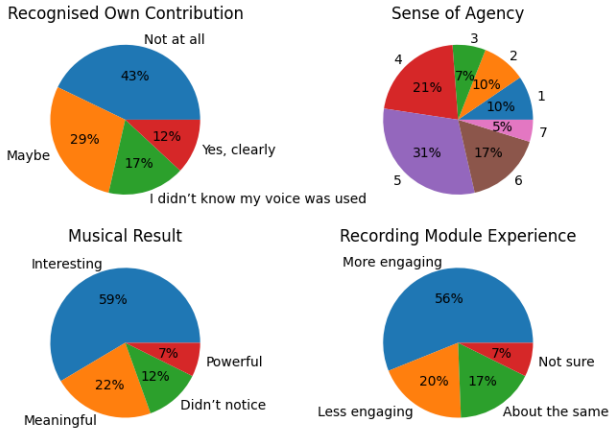



Figure 7. Survey results regarding questions of audience contribution, agency, engagement and perceived musical result (© Davor Vincze).

## 5. DISCUSSION

The performances attracted participants with varying levels of technical familiarity. Although less experienced participants required more assistance connecting to the system, engagement appeared similar once the performances began; the small venue-specific samples, however, preclude reliable demographic comparisons.

During the recording phases, audience members often interacted in playful ways, whispering, laughing, and experimenting with vocal sounds. This suggests that this form of participation introduces elements of play and social interaction that diverge from the traditional concert experience.

A potential concern when allowing audience recordings is the possibility of inappropriate verbal content. Research on participatory and interactive performance has shown that when audiences are given agency, they often engage in playful, mischievous, or subversive behavior rather than purely goal-oriented interaction [36]. To mitigate this, recordings were limited to short durations and processed electronically, effectively transforming speech into abstract sonic material and minimizing this risk.

The musical design accounted for smartphone speakers' limited low-frequency response and sound pressure level [37]: phone playback emphasized mid- and high-frequency material, while granulated and concatenated sounds were presented through high-fidelity loudspeakers. The overall electronic dynamics remained relatively soft, allowing the cumulative output of multiple devices to remain audible.

Audience size therefore presented a trade-off: more devices produced a denser aggregate texture, whereas smaller groups made individual actions easier to distinguish. A similar tension was reported in *Hermès v2*, where 15–20 simultaneous participants obscured individual feedback, leading a later version to limit active participation to four [38]. Our observation that the larger performance produced a richer collective result thus concerns acoustic and musical density, not greater participation legibility.

Recognition was not clearly associated with perceived agency in our survey, suggesting that co-creation may also arise from contributing to a collective process rather than solely from identifying one's own sound. This result does not establish that feedback is unimportant. Future versions could provide immediate visual acknowledgement when a contribution is received and project the interface or an accompanying visualization showing how audience material is captured, mapped, and transformed.

To limit divided attention [39], each interaction was followed by a longer listening phase, during which a countdown timer indicated when the next interaction would begin and allowed audience members to return their attention to the performance.

## 6. CONCLUSION

Smartphones are powerful and omnipresent devices that include microphones, cameras, speakers, and various sensors, and have increasingly been used as musical instruments [37], particularly in the context of technology-mediated audience participation [5]. While many existing systems—including our own previous iterations—have explored audience interaction through control interfaces and distributed playback of pre-composed material, the RECORD module introduces a different approach by allowing audience members to contribute their own recorded sound material to the performance.

The survey results indicate that the recording module was the preferred interaction method among participants and that, despite differences in technological familiarity and varying onboarding times, audience engagement remained high once the performance began. Participants reported that the system was intuitive and comfortable to use, and the results suggest a moderately high level of perceived agency. The compositional design of *We're in This Together*, based primarily on vocal sound material, provided a sonic context in which audience contributions could be naturally integrated into the overall musical texture.

Audience interaction was intentionally time-structured within the larger musical framework. Future versions

could make individual contributions more recognizable, incorporate video, or explore computational transformation. Holly Herndon and Jlin's *Godmother*, in which an AI trained on a human voice acts as a creative collaborator, offers one model of technologically mediated co-creation that could inform smartphone-based participatory systems [40].

The results of this project suggest that enabling audiences to contribute their own sound material, rather than only controlling pre-existing material, increases the sense of participatory agency and moves the audience closer to the role of performer within the musical system, without requiring specialised knowledge or disrupting the performance experience.

**Acknowledgments**

The initial development of the web application was supported by the SWR Experimentalstudio. Attendance at the NIME 2025 conference was supported by Hong Kong Baptist University. The performances at Emory University and Stanford University were supported by the respective institutions. The authors would also like to thank Prof. Dr. Zdenko Sonicki for his help with survey data interpretation, as well as the audiences at NIME, Emory, and CCRMA for their participation and feedback. This article was created during a Postdoctoral Fellowship supported by the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. PDFS2324-2H04).

## 7. REFERENCES

[1] C. Heim, *Audience as performer: The changing role of theatre audiences in the twenty-first century*. Routledge, 2015.

[2] T. K. Majumder, "Relational Music: A Participatory Approach Towards Live Experimental Electronic Music," PhD Thesis, University of California, Irvine, 2023. [Online]. Available: https://escholarship.org/uc/item/3w9072ps

[3] C. Sdraulig and L. d'Heudières, "Attending to attending: performing audience personae in contemporary music," *Tempo*, vol. 76, no. 300, pp. 18–32, 2022.

[4] C. Bishop, *Artificial hells: Participatory art and the politics of spectatorship*. Verso books, 2023.

[5] O. Hödl and L. Turchet, "The TMAP Framework: describing the design space of Technology-Mediated Audience Participation in live music," *IEEE Access*, vol. 13, pp. 38 593–38 608, 2025.

[6] 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, Oct. 2020. [Online]. Available: https://hal.science/hal-03033143

[7] M. Neuhaus, "The broadcast works and audium," in *Zeitgleich Symposium. Hall/Tirol*, 1994.

[8] C. Hope, "From early soundings to locative listening in mobile media art," in *The Routledge Companion to Mobile Media Art*. Routledge, 2020, pp. 46–56.

[9] B. Taylor, "A history of the audience as a speaker array," in *Proceedings of the International Conference on New Interfaces for Musical Expression*. Copenhagen, Denmark: Aalborg University Copenhagen, 2017, pp. 481–486.

[10] M. K. Koszolko and N. Wilson, "Idiosyncratic approaches to electronic music-making through mobile technologies," *Media Practice and Education*, vol. 27, no. 1, pp. 40–54, 2026.

[11] T. Hopkins, E. Doherty, N. Ofer, S. C.-C. Weng, P. Gyory, C. Tobin, L. Hirshfield, and E. Y.-L. Do, "Stringesthesia: Dynamically shifting musical agency between audience and performer based on trust in an interactive and improvised performance," in *Proceedings of the 18th International Audio Mostly Conference*, 2023, pp. 9–16.

[12] J. Oh and G. Wang, "Audience-participation techniques based on social mobile computing," in *Proceedings of the International Computer Music Conference*. Huddersfield, UK: University of Huddersfield, 2011, pp. 665–672.

[13] D. Vincze and M. Oeser, "Freedom Collective: Interactive Mobile Web-App For Immersive Audience Participation," in *Symposium 2024 (UbiMus 2024)*, 2024, p. 14.

[14] ——, "Sonic immersion with interactive mobile web-application," *Arts & Communication*, p. 025210043, 2026.

[15] ——, "We're in this together: Expanding Interactive Sonic Spaces," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2025, pp. 98–102.

[16] G. Weinberg, "Interconnected musical networks: Toward a theoretical framework," *Computer Music Journal*, vol. 29, no. 2, pp. 23–39, 2005.

[17] N. Weitzner, J. Freeman, Y.-L. Chen, and S. Garrett, "massMobile: towards a flexible framework for large-scale participatory collaborations in live performances," *Organised Sound*, vol. 18, no. 1, pp. 30–42, 2013.

[18] A. Xambó and G. Roma, "Performing audiences: Composition strategies for network music using mobile phones," in *Proceedings of the International Conference on New Interfaces for Musical Expression*. Birmingham, UK: Birmingham City University, 2020, pp. 55–60.

[19] G. Wang, G. Essl, and H. Penttinen, "The Mobile Phone Orchestra," in *The Oxford Handbook of Mobile Music Studies*, S. Gopinath and J. Stanyek, Eds. Oxford University Press, 2014, vol. 2, pp. 453–469.

[20] S. W. Lee and J. Freeman, "echobo: A mobile music instrument designed for audience to play," *Ann Arbor*, vol. 1001, no. 48109-2121, pp. 10–44, 2013.

[21] T. Shaw, S. Piquemal, and J. Bowers, "Fields: An exploration into the use of mobile devices as a medium for sound diffusion," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2015, pp. 281–284.

[22] F. Kayali, C. Bartmann, O. Hödl, R. Mateus-Berr, and M. Pichlmair, "Poème Numérique: Technology-Mediated Audience Participation (TMAP) Using Smartphones and High-Frequency Sound IDs," in *International Conference on Intelligent Technologies for Interactive Entertainment*. Springer, 2016, pp. 254–258.

[23] L. Turchet and M. De Cet, "A web-based distributed system for integrating mobile music in choral performance," *Personal and Ubiquitous Computing*, vol. 27, no. 5, pp. 1829–1842, 2023.

[24] 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, pp. 272–275.

[25] Le Vivier, "Corporate Retreat | Le Vivier," Sep. 2023. [Online]. Available: https://levivier.ca/en/concert/2023-24-season/corporate-retreat

[26] D. Scheiba, "Network latency feedback," Web Audio Conference 2025, IRCAM, Espace de projection, Paris, 2025. [Online]. Available: https://ressources.ircam.fr/fr/media/e25c61016509d3ff6a1a

[27] A. Forgette, B. Manaris, M. Gillikin, and S. Ramdsen, ""Speakers, More Speakers!!! – Developing Interactive, Distributed, Smartphone-Based, Immersive Experiences for Music and Art"," *Proceedings of the International Symposium on Electronic Art*, 2022. [Online]. Available: https://par.nsf.gov/biblio/10374099

[28] MDN Web Docs, "MediaRecorder - Web APIs | MDN," Jul. 2024. [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/API/MediaRecorder

[29] J.-M. Valin, G. Maxwell, T. B. Terriberry, and K. Vos, "High-quality, low-delay music coding in the opus codec," in *Proceedings of the 135th Audio Engineering Society Convention*. New York, NY, USA: Audio Engineering Society, 2013, convention Paper 8941.

[30] J. Bell, "Distributed Notation in the Browser, an Overview," in *Proceedings of the International Conference on Technologies for Music Notation and Representation—TENOR '20/21*. Hamburg, Germany: Hamburg University for Music and Theater, 2021, pp. 251–259.

[31] C. Roads, *Microsound*. The MIT Press, 2004.

[32] D. Schwarz, "Corpus-based concatenative synthesis," *IEEE signal processing magazine*, vol. 24, no. 2, pp. 92–104, 2007.

[33] P. A. Tremblay, G. Roma, and O. Green, "Enabling programmatic data mining as musicking: The fluid corpus manipulation toolkit," *Computer Music Journal*, vol. 45, no. 2, pp. 9–23, 2021.

[34] O. Hödl, C. Bartmann, F. Kayali, C. Löw, and P. Purgathofer, "Large-scale audience participation in live music using smartphones," *Journal of New Music Research*, vol. 49, no. 2, pp. 192–207, 2020.

[35] A. Joshi, S. Kale, S. Chandel, and D. K. Pal, "Likert scale: Explored and explained," *British journal of applied science & technology*, vol. 7, no. 4, pp. 396–403, 2015.

[36] S. Benford, C. Greenhalgh, G. Giannachi, B. Walker, J. Marshall, and T. Rodden, "Uncomfortable interactions," in *Proceedings of the sigchi conference on human factors in computing systems*, 2012, pp. 2005–2014.

[37] G. Essl and S. W. Lee, "Mobile Devices as Musical Instruments - State of the Art and Future Prospects," in *Music Technology with Swing*, M. Aramaki, M. E. P. Davies, R. Kronland-Martinet, and S. Ystad, Eds. Cham: Springer International Publishing, 2018, pp. 525–539.

[38] J. Svidzinski, R. Montferme, and J. Larralde, "Hermès v2 : création musicale en réseau," in *Journées d'Informatique Musicale 2023 (JIM 2023)*. Saint-Denis, France: Centre de Recherche Informatique et Création Musicale, May 2023, hAL Id: hal-04484035. [Online]. Available: https://hal.science/hal-04484035

[39] A. Chang, "An Investigation of Usability Factors in Designing a Companion System for an Immersive Environment," PhD Thesis, Rensselaer Polytechnic Institute, 2024.

[40] K. Lillvis, "Music video distortion and posthuman technogenesis," *More Than Illustrated Music: Aesthetics of Hybrid Media between Pop, Art and Video*, p. 153, 2023.