

Ear Talk Project: Participatory co-composition on YouTube and the Web

TOSHIHISA TSURUOKA¹, BRIAN ELLIS² and LEO CHANG³

¹Composer, Tokyo, Japan. Email: tt1694@nyu.edu

²Composer, New York, USA. Email: brian.ellis@utexas.edu

³Composer, New York, USA. Email: changh8@rpi.edu

In this article, we present *Ear Talk* – a co-composition and live performance project that enables remote music collaboration through technologically mediated systems. The *Ear Talk* project currently exists in two distinct implementations, one that repurposes YouTube’s live-streaming technology, and one that utilises a stand-alone website. Although *Ear Talk* was conceived prior to the 2020 COVID-19 pandemic, the necessity for remote collaboration became more apparent during the lockdown, when a vast majority of live events and music concerts were cancelled. The *Ear Talk* project enables a socially distanced form of online musical collaboration and offers a platform through which to respond to such a crisis, and has grown to be adopted and presented by many different performing groups across the world. In addition to describing the technical implementations of these two systems, we discuss issues that arise from our participatory practice: from musical quality concerns in regard to social aesthetics and artistic ingenuity, to accessibility concerns when designing technologically mediated collaborative systems. *Ear Talk* embraces continuous musical loops as well as highly asynchronous (i.e., perpetual) collaborative paradigms among remote participants, which raises a conceptual inquiry as to which part of its sonic and social experience constitutes music in the end. Finally, we evaluate performer–audience relationships (i.e., hierarchical versus horizontal interactions) and the efficacy of the *Ear Talk* systems at enabling socially engaged co-composition.

1. INTRODUCTION

The conceptualisation of *Ear Talk* was initiated by Toshihisa Tsuruoka and the members of Ensemble Consensus,¹ a New York-based performance group that investigates co-compositional practices via the facilitation of real-time, co-creative processes as performances, often involving improvisation, iteration and group discussion. Consensus calls this practice Group Listening and strives to ‘engage in different ways of facilitating social interaction in order to question our creative capacity for collaborative music-making ... [and blur] boundaries between rehearsal and performance; composition and improvisation’ (Ensemble Consensus n.d.). The idea for *Ear Talk*

¹Ensemble Consensus will be abbreviated as Consensus throughout this article.

came about as some members had to relocate to different countries. The place for collaborative music-making also needed a relocation from a physical space to an online environment. The *Ear Talk* project was conceived out of this necessity, and Consensus organised five YouTube live stream performances titled *Ear Talk: Sounds Worth Sharing*² (Figure 1) from October 2019 to February 2020, during which the YouTube-based *Ear Talk* system was premiered. Each stream picked up where the last session left off, allowing the music to iteratively transform and develop over the five public performances. These live streams were treated as an opportunity to celebrate the spontaneity heard in sound files collected from the ensemble members, which inherently reflected each member’s location and differing taste in sound.

The YouTube-based *Ear Talk* system was reprised for the Society for Electro-Acoustic Music in the United States (SEAMUS) 2020 National Conference as part of its community-engaged performances and workshops. The *Ear Talk: Online Sound Gathering with SEAMUS 2020*³ was organised in response to the pandemic and gave the conference participants an opportunity to interact with one another in spite of the distance while simultaneously creating music together. For this event, the same YouTube-based *Ear Talk* system used in *Sounds Worth Sharing* was utilised, and members of Consensus were called upon once again to lead the facilitation of the virtual collaboration. Two YouTube performances were subsequently live-streamed on 14 and 16 May 2020.⁴ For *Online Sound Gathering*, participants were asked to record and submit sounds in response to the COVID-19 lockdown. Notable sound files included the sound of a voice through a mask and a recording of Donald Trump’s COVID-19 briefing among many

²*Ear Talk: Sounds Worth Sharing* from 2019: <https://youtu.be/rPpRjJr1x28>. We will refer to this version of *Ear Talk* simply as *Sounds Worth Sharing* for the rest of this article.

³*Ear Talk: Online Sound Gathering with SEAMUS 2020* will be abbreviated as *Online Sound Gathering* for the rest of this article.

⁴*Ear Talk: Online Sound Gathering with SEAMUS 2020* <https://youtu.be/sWTFzG6r-nU> and <https://youtu.be/icB6D6vE2qM>.

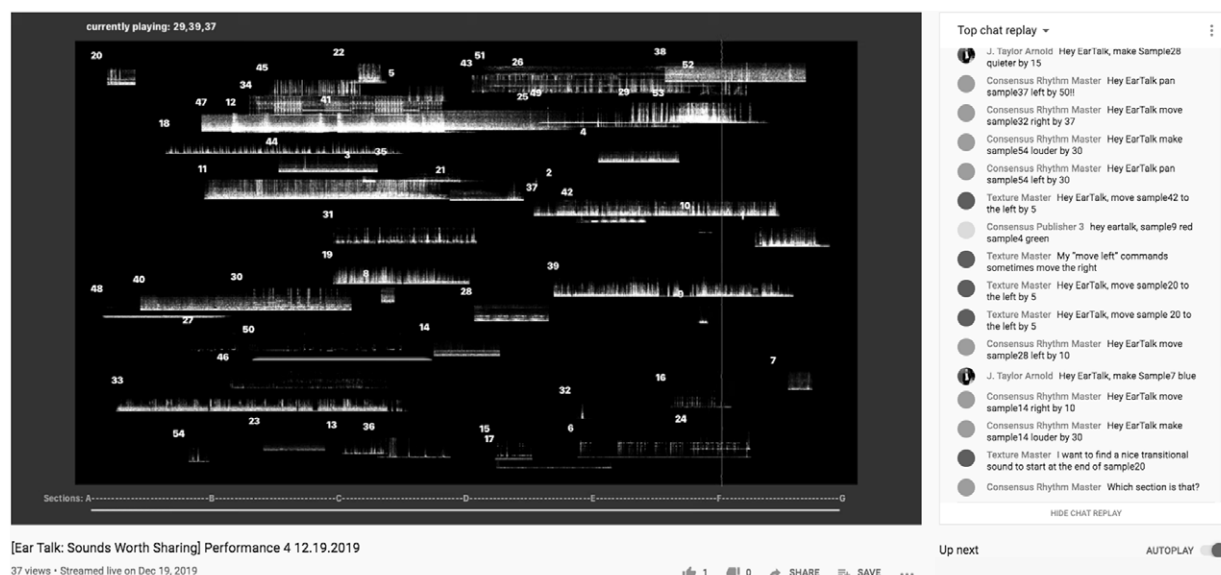


Figure 1. YouTube *Ear Talk* co-composition live stream. The chat box (on the right) is used by participants to manipulate parameters in the score (on the left).

other forms of sonic expression that reflected the time of crisis.

Ellis and Tsuruoka expanded *Ear Talk* beyond YouTube and developed a web-based version to allow for an infinitely long website installation, as well as to further the goals of the YouTube-based *Ear Talk* system by improving its accessibility. This culminated in the May 2020 premiere of the web-based system as *Ear Talk: Never Ending*, presented by Less Than 10 Music,⁵ a performing arts organisation dedicated to presenting socially distanced new music in the face of COVID-19 restrictions. It has since been presented independently in live streams by Ellis, as well as in conjunction with organisations such as zFestival⁶ and Density 512⁷ throughout 2020 and 2021.

2. YOUTUBE-BASED *EAR TALK* SYSTEM

The need for an online apparatus that could allow collaborative music-making for Consensus members quickly transformed into a desire to create a socially engaging platform that could invite anyone on the internet to participate in the music-making. Therefore, the primary impetus behind the development of the *Ear Talk* system was to be accessible for any curious participant, regardless of musical training. In order to achieve this, the system was built around free-to-use online platforms that are familiar and accessible to many internet users, namely

⁵<https://lessthan10music.com/>.

⁶<https://zfestival.wordpress.com/>.

⁷www.density512.org/.

Google Drive and YouTube, and all interactions for the sake of collaboratively making music were conducted via text-based communication over YouTube Live chat messages.

2.1. Technical Implementation⁸

The YouTube-based *Ear Talk* system enables people from remote locations to collaboratively share, shape and form music through an interactive score. The first step is to encourage people to record sounds and submit their sound files to a shared Google Drive folder. The host computer automatically downloads all submitted sounds and loads them into a Max/MSP-based system where an interactive score is generated.

The score exhibits all sounds as spectrogram images and shows the linear progress of music, similar to modern Digital Audio Workstations with a playback head and annotated sections (Figure 2). This score is live-streamed to YouTube, where participants can alter various musical parameters of the sounds remotely by commenting specific commands. The commands submitted by participants are reflected in the score on YouTube in real-time (as fast as processing and internet latency allow). Musical parameters that participants can manipulate are as follows: volume change, mute or unmute, panning, location change, colour change, reversing effect, pitch shifting effect, stretching effect and section loop. While the

⁸For more technical details and descriptions of the YouTube-based *Ear Talk* system, please refer to the ICMC paper (Tsuruoka, Chang and Hickman 2021).

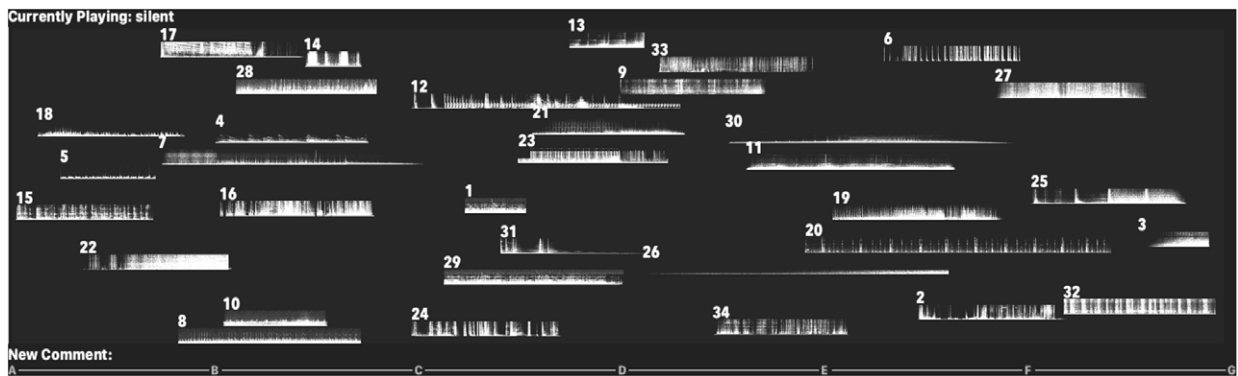


Figure 2. An example of the YouTube *Ear Talk* score. Each file is labelled with a corresponding sample number that participants can refer to in the chat box.

commenting syntax for these commands resembles conversational English and promotes an intuitive experience, participants must adhere to syntactical rules in order for the system to recognise their commands. All commands must start with the phrase ‘Hey Ear Talk’, and proceed with a specific instruction such as ‘mute sample 32’ (the *Ear Talk* system refers to sounds as samples). An example command for the volume parameter might be ‘Hey Ear Talk, make sample 21 louder by 25%.’ The duration of the entire score (usually 3–8 minutes long) is continuously looped during the performance unless a looping command for a specific section is received. Not only do the YouTube Live chat messages host these command comments from each participant but they also welcome regular conversations between participants. With this interaction, participants may influence each other to comment particular commands and thus shape the music collectively throughout the performance.

2.2. Realising the YouTube-based *Ear Talk* system

In many manifestations of the Group Listening methodology from previous in-person performances, Consensus based their performance practice around the fabrication of rules and roles that governed their collaborative strategy. The rules were often detailed in written guidelines that acted as a score, and the entire group was aware that the guidelines were always malleable and open to change. The purposefully designed roles, on the other hand, were often assigned to each member in order to facilitate their interactions. The ensemble applied many experiential facets in designing *Ear Talk*. The iterative process of the ensemble led *Ear Talk* to focus on designing roles for each member that would ensure an active and positive alteration of music during the live-streamed performance. Through the course of rulemaking and assigning roles to the ensemble members, the process of co-

composition became more interactive, and it was easier to manage the variety of collected sounds as they were shaped and reconfigured. Additionally, the role assignment was designed to balance participants’ contrary perspectives as well as to address the potential lack of attention on certain musical potentials.

For our *Sounds Worth Sharing* YouTube performances, all 15 members of Consensus were given one of seven unique roles: Rhythm Master, Pitch Master, Texture Master, Dynamic Master, Publisher, Navigator or Moderator. The Rhythm, Pitch, Texture and Dynamic Master roles were designed to help generate musical curiosity if and when participation dwindled. Members of the ensemble who were given these roles were tasked with generating prompts that would encourage participants to work together towards a certain musical goal (louder at a certain section, for instance) or draw attention to areas that had been more neglected throughout the co-compositional process. Members who were given the Publisher, Navigator and Moderator roles were tasked with maintaining a positive, troll-free environment, and with that, each role had unique authorities that no other role had. For instance, the Navigator could decide which sections to be looped while the Moderator could delete and hide comments as well as ban certain individuals from further commenting. For *Sounds Worth Sharing*, each member created a new YouTube profile with their role as their username (e.g., Rhythm Master 1), and the sounds themselves were strictly contributed and uploaded by Consensus as another precaution against trolling behaviour and to test the robustness of the system. The general public was invited to shape the sounds during the live-streamed co-composition sessions.

For our *Online Sound Gathering* performances, the rules and roles were accommodated in a different way. Sounds were contributed by members of SEAMUS, Consensus, as well as the general public. The members

of Consensus were given a new set of guidelines; the roles were simplified so that everyone acted as a 'Moderating Participant', tasked with ensuring the 'Live Chat window maintains creative and collaborative momentum' (Ensemble Consensus n.d.). The Moderating Participant's role summed up the seven distinct roles that were used during *Sounds Worth Sharing* into one, leaving all ensemble members the same general responsibility of promoting and sustaining a fulfilling experience. Consensus also decided not to use separate YouTube profiles that distinguished and anonymised members into their roles. Instead, each member participated using their personal profile (often under their own names). Though it is impossible to claim causality, these changes may have helped many more people feel comfortable participating in the co-compositional process through commenting during *Online Sound Gathering*.

Simplifying the roles illuminated aspects of the *Ear Talk* system that inherently discouraged antisocial behaviour and clarified why creating elaborate roles to address them may have been redundant. First, the inevitable latency while using API-based networked communication prevented our initial concern regarding one participant rapidly making large-scale, consequential comments that would change the overall soundscape drastically. Second, the design of one comment leading to one element of a sound being changed at a time meant that even the most drastic possible change on one of the sounds did not completely alter the overall soundscape. This limitation of incremental change built into the YouTube-based *Ear Talk* system necessitated that everyone participating embrace a collectivist, interdependent mode of creativity, which inherently mitigated drastic decisions from any one particular participant.

At the same time, moments of individual expression were still cherished. A notable progression of expression was when a participant chose to mute one of the sound samples that featured the voice of Donald Trump – a political gesture and performance. After some time had passed, another participant un-muted the sample (it is impossible to know whether or not this participant was aware that they were unmuting the sound of Donald Trump since the samples were only labelled with numbers). Another participant subsequently commented to pitch-shift the sample up to the point of it being unrecognisable as Trump's voice and then panned the sound to the right, another symbolic, politically charged expression.

2.3. Evaluating YouTube as a tool for socially engaged performances

YouTube has transformed from a simple video sharing website into a social platform where people not

only seek to entertain themselves through the viewing of content but also socially engage with other users. Khan's (2017) study argues that passive consumption of content is fuelled by the users' motive to simply relax and entertain while the social interaction motive propels active engagement in the commenting and sharing of content. The YouTube-based *Ear Talk* satisfies both desires: the live-streamed co-composition session and its archive can simply be viewed and enjoyed while the participatory elements of sharing sounds and commenting invite those who seek to interact with other participants.

Through hosting their first five YouTube performances, Consensus realised the dearth of organic reach on YouTube's live-streaming platform. Being a young experimental music ensemble, Consensus was counting on YouTube Live to have sizable organic reach and hoped for curious web users to happen upon the project, similar to pedestrians stopping by a public park. While the ensemble did not expect this project to 'go viral', the lack of *any* organic reach during hours of live-streaming came as a surprise and shed light on the purported egalitarian myth that all user-generated content is treated equally. Thus, it is important to note the distinction between hosting a participatory project in a physical public space, such as a public park with ample foot traffic, and a digital space where, theoretically, anyone *could* happen upon a participatory live stream, but what determines how many users actually listen in are algorithms heavily driven by private and monetary interests (Rosen 2008; Kim 2010). This lack of reach also meant that much of Consensus's design that accounted for moderating trolling behaviour went unused, and many of the roles became redundant. To aid in mitigating this issue, *Ear Talk* projects henceforth sought to ensure guaranteed audiences by partnering with presenting organisations (i.e., SEAMUS, Less Than 10 Music).

3. WEB-BASED EAR TALK PROJECT

Many aspects of the YouTube-based *Ear Talk* system informed the making of the web-based implementation – from its performer–audience relationship paradigms and the user interface design, to its archival and documentation processes. When Ellis joined the team, the opportunity to improve *Ear Talk*'s co-creative process through the lens of a stand-alone website presented itself. Specifically, the YouTube-based system had two areas that we believed could be re-examined. The first was an opportunity to improve how the sound manipulation was conducted: in the YouTube-based system, participants must type out very specific commands, and minor deviations or misspellings caused the commands to be unrecognised by the system. The second area was reproducibility: the

YouTube-based system relied on many different tools, APIs, services and technologies, which meant that a breaking change in any one of them would require maintenance for the system to be utilised again.

3.1. Technical implementation of www.eartalk.org

It is well established that web browsers are uniquely positioned to create accessible and powerful tools that combine synthesis, live sound production and novel interface design (Roberts, Wakefield and Wright 2013), and thus the *Ear Talk* website aims to capitalise on many of these abilities along with matured web technologies for the website's full stack implementation. The front-end of the website is built using HTML, Javascript, CSS and open source libraries for audio and visual state management, including Pizzicato.js⁹ for audio and Konva¹⁰ for visuals. The back-end server of the website is implemented in Node.JS¹¹ and hosted on an Amazon Web Services (AWS) Elastic Beanstalk¹² for instance. The usage of AWS, one of the industry standards of web hosting, gives commercial-level reliability, scaling and website deployment capabilities. The website front-end communicates to the server in near real-time, providing updates to a central server whenever a participant executes a change. These updates are then distributed back to all participants through text-based exchanges. The website renders the audio with the specified parameters live on a participant's web browser, meaning that only a light-weight text file (about 10kb) needs to be transferred over internet connections rather than live-streaming an audio and video feed from the server. The live chat feature is adopted from the YouTube-based system, allowing communication between participants during the web performance.

Although the main page for interacting with the *Ear Talk* website is relatively similar to the YouTube-based visual score, the way in which participants interact with musical parameters was changed to improve the user experience. The view consists of a graphic score display (see Figure 3) where sounds are represented as image blocks and are laid out so that their positions on the score correspond to different parameters of the sound. Instead of typing to interact with the musical parameters, a drag and drop interface was implemented such that moving blocks to different locations on the score would change the relevant parameters. The horizontal position corresponds to where sounds are played in time, and the vertical position maps to one of four parameters, determined by

the 'role' the player is assigned. The parameters available for manipulation are volume, pan, distortion and a low-pass filter termed 'clarity'. A 'role' drop-down menu allows performers to choose which parameter their vertical axis controls. For instance, when a participant selects the volume role, the sounds positioned higher in the vertical axis correspond to louder volume. The web-based *Ear Talk* system also improves the experience of uploading sound files during the course of the performance thanks to its built-in capability to record and submit sounds directly within the website. Although the YouTube-based system is capable of taking in new sound files during the performance, the web-based system's streamlined process makes it easier to contribute new sound files that are responsive to the existing musical context, along with sounds that record, remix, distort or otherwise manipulate existing content through the participant's desired technological means.

Additionally, there are two variants of the website that can be accessed via two separate links to the same performance: one with more features targeted at active performers (i.e., performer view) and one prioritising accessibility (i.e., audience view) targeted at first-time participants or curious members of the public. The main differences are that the performer view has the ability to add and delete sounds as well as change one's selected 'role'. The audience view lacks these features in favour of a less cluttered and more intuitive display. As discussed previously, the YouTube-based *Ear Talk* system was consciously designed to mitigate antisocial, trolling behaviour. Similarly, the web-based system has two main contingency vectors for dealing with potential trolling behaviour. The first is the implementation of the performer view and audience view. Although each link provides an identical layout for *Ear Talk*'s visual score, the limited feature set provided in the audience view limits potential trolling behaviour, such as uploading an inappropriate sound file in the middle of a performance. Additionally, we found during a test performance that some participants may try to quickly mute all clips, make every sound very distorted, or force all sounds to occur at the same time. While such changes may be interpreted as positive musical gestures and could be restored by other participants if necessary, we took this learning and implemented a 10-second forced pause between actions from any participant. This obligated a communal effort to make any large-scale changes and restricted any single individual from executing drastic changes.

3.2. Performance paradigms utilising the web-based *Ear Talk* system

The web-based *Ear Talk* system was designed to be used in a variety of performance paradigms. First, it

⁹<https://alemangui.github.io/pizzicato/>.

¹⁰<https://konvajs.org/>.

¹¹<https://nodejs.org/>.

¹²<https://aws.amazon.com/elasticbeanstalk/>.

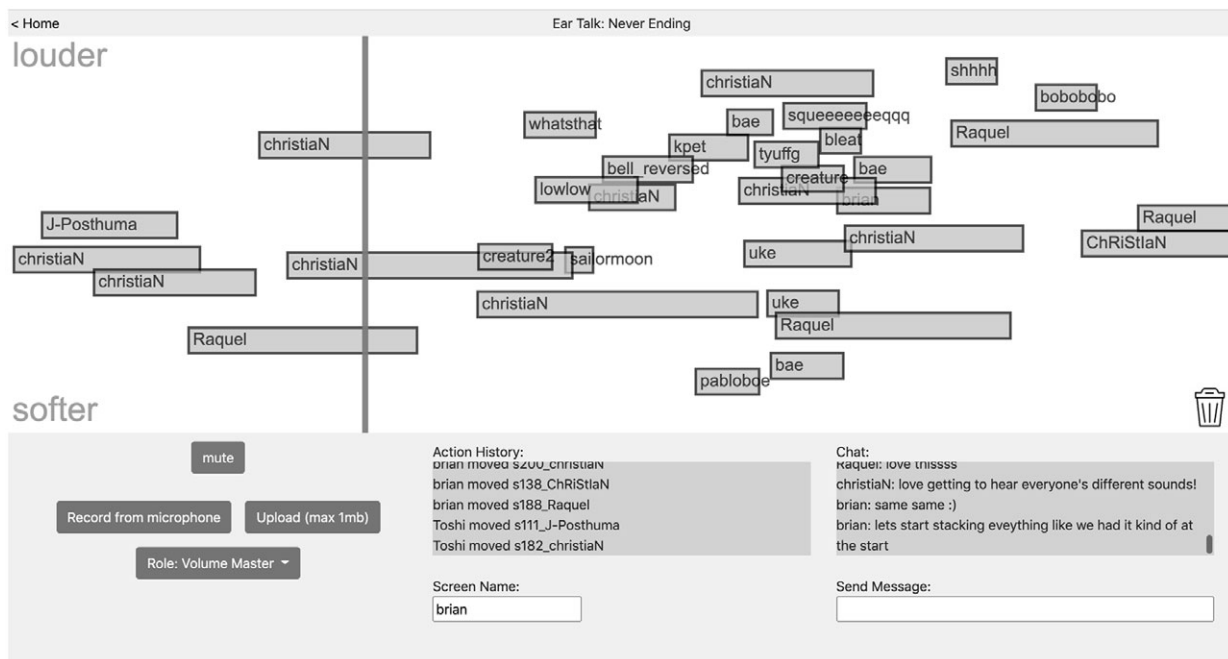


Figure 3. An example display of the performer view in the web-based *Ear Talk* implementation.

can exist purely on its own as a perpetual sound collage installation, live rendering audio into perpetuity. Under this conceptual lens, participants who interact with the website are adding their own artefacts or traces to the sound collage at an asynchronous time frame, which are then updated to all other participants (or any visitors of the website) around the world in an everlasting live performance.

The second lens through which the website is used is the context of a live performance. In this configuration, all performers and audience members gather at the *Ear Talk* website at a specified time for a specified duration. The performance takes place as participants manipulate existing sounds as well as upload and delete sounds in order to change and craft the sonic landscape to their desired aesthetic. Participants can communicate with each other via a chat feature designed for this purpose, allowing them to send text-based messages to the entire group and discuss any large-scale gestures they would like to orchestrate.

The third (and most layered) lens emerges when the *Ear Talk* website is used in conjunction with a video conference tool such as Zoom¹³ or Facebook Messenger Video Calling (Figure 4). In such a setup, all participants join a video call to see and hear one another. One of the participants then shares the *Ear Talk* website with other participants through the call, and set up thusly, they begin the performance. Every participant on the call can hear three sonic layers at

any given moment. First, they may hear performers who choose to audiate or produce sound entirely independent of the *Ear Talk* system, with audio only coming through the video conference tool. Second, participants can hear one another recording their performances into the *Ear Talk* website through the call, and thus have a ‘first hearing’ of a sound that will later be heard through the *Ear Talk* website. Third, listeners on the call can hear the *Ear Talk* website live rendering audio in accordance to the graphic score they share. The sounds in the *Ear Talk* website become an amalgamation of audio that the participants have heard before as a ‘first hearing’, as well as audio that may have already existed in the system prior to the performance. In this Zoom-aided live performance setup, the combination of different audio sources, processing mechanisms and social cues help maintain musical interest and engagement for extended periods of time. Typically, the web-based *Ear Talk* live session has occurred as a 30-minute to an hour-long performance, including *Ear Talk: Never Ending* presented by Less Than 10 Music in May 2020 zFestival in August 2020, and the *Music for Audiences*¹⁴ series presented by Density 512 in February 2021.

The ubiquity of portable electronic devices has had a significant impact on the curation of all performance paradigms discussed previously, including the

¹³<https://zoom.us/>.

¹⁴*Music for Audiences III. Ear Talk* archive <https://youtu.be/ymMW1KT9awU>.

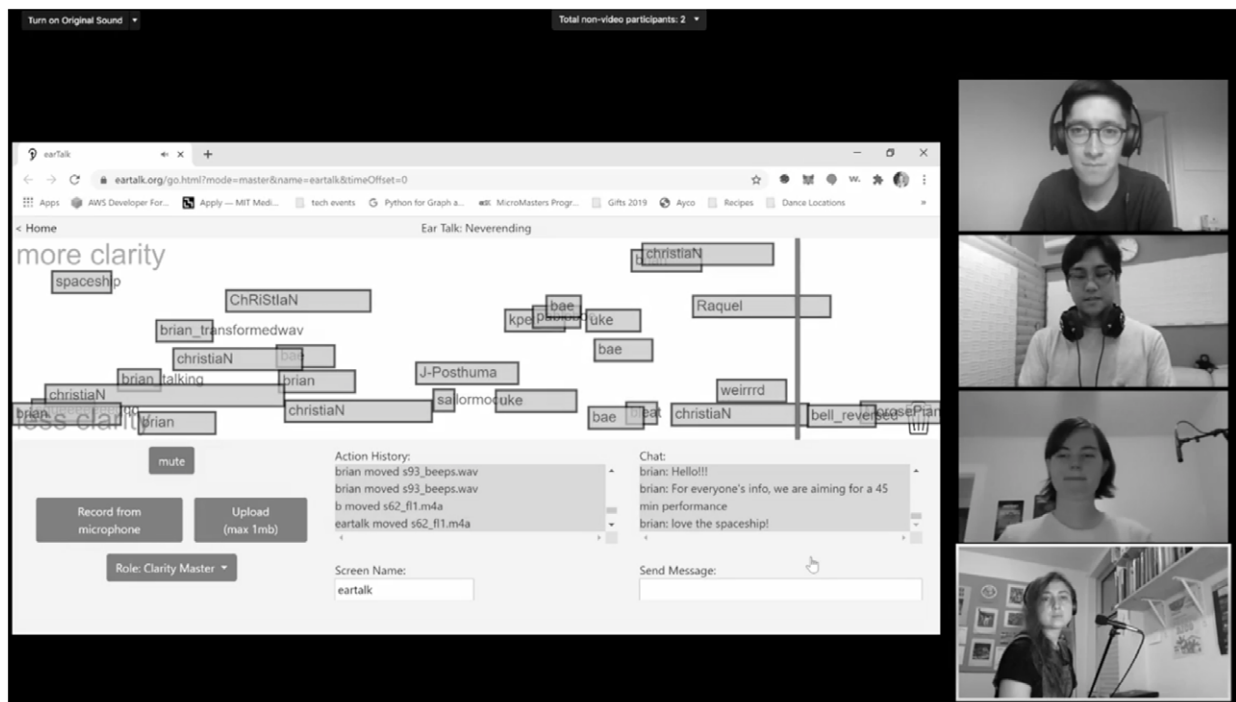


Figure 4. A web-based *Ear Talk* performance via Zoom. Four performers were featured while audience members watched via live stream and participated through the *Ear Talk* website.

YouTube-based *Ear Talk* system. As Gaye, Holmquist, Behrendt and Tanaka (2006) highlight in their paper, mobile phones pose an immense opportunity to democratise music creation, sharing and mixing. By encouraging participants to use mobile phones when recording and sharing sounds, as well as when interacting with both YouTube and web-based systems, mobile phones become a physical conduit through which to connect performers and audience members. The accessibility and portability of such electronic devices encourages participants to interact with *Ear Talk* regardless of time and place, and the fidelity of audio via modern mobile phones is sufficient to fuel creativity for our performances.

4. EVALUATING A PARTICIPATORY PRACTICE

In both the YouTube and web-based systems, *Ear Talk* invites all who are willing and able, and gives every participant the same basic controls and artistic agency. Simultaneously, there are different levels of participant engagement within *Ear Talk*. The choice of taking on the role of a performer by participating proactively (contributing sound files and freely initiating change during the co-composition process) or simply joining as a passive audience member is left to each participant.

4.1. Evaluating conditions of 'quality'

Such participatory practices often raise questions of 'quality' and 'outcome' when judged through the widely normative cultural lens of modernity.¹⁵ For instance, does the open invitation for anyone to exercise artistic agency compromise artistic quality? To what end are people participating and collaborating? As reflected in these questions, ideals of modernity establish individual artistic ingenuity (such as musical talent and training) as a prerequisite for aesthetic creation. In this context, music (the aesthetic creation) is situated as a product (often framed as 'the music') of said artistic ingenuity. Subsequently, 'quality' refers to a measure of artistic ingenuity and aesthetic clarity, with 'high quality' alluding to virtuosity via an abundance of musical talent, training and therefore aesthetic beauty while 'low quality' alludes to amateurism from a lack thereof. In contrast, embedded within the design of

¹⁵Here, modernity refers to the ideology that began during eighteenth-century European cultural thought that influences a certain normality in the way artistic creation is evaluated today: its initial influences on the arts, aesthetics and creativity is outlined in Kant's *Critique of the Power of Judgement* ([1790] 2000). Modernity has been imposed as a 'global' ideological normality via colonialist and neocolonialist strategies of Western nation states, and underlies many of the social, political and cultural aspects of 'modernised' people and places, including but certainly not limited to music and the arts (Quijano 2007; Mignolo 2011).

Ear Talk is a departure away from the modern cultural assumption that aesthetics presuppose individual artistic ingenuity. Instead, *Ear Talk* positions aesthetics as emergent from and dependent on social interaction. Therefore, a more appropriate way to discuss ‘quality’ within *Ear Talk* is to frame it as Susanne Burns (2015) did: through a ‘shared sense of quality’ fostered by the ‘conditions that make quality experiences for participants’.

Many participatory practices derive meaning from the process of engagement itself. As Turino (2008: 28) writes, ‘in participatory music-making one’s primary attention is on the activity, on *the doing* and on the other participants, rather than on an end product that results from the activity’. Conversely, Turino uses the phrase ‘presentational performance’ to indicate types of performances that imply a clear distinction between the ‘performer’ (an active contributor or the artist) and the ‘audience’ (listeners, observers or consumers). *Ear Talk* harbours both presentational and participatory elements simultaneously; our audiovisual interface allows for, but does not necessitate, interaction. *Ear Talk*’s objective is to enable people from remote locations to collaboratively gather and manipulate sounds, as well as listen to their amalgamation through interactive engagement via the World Wide Web. The act of collectively gathering and manipulating sounds by itself does not complete the *Ear Talk* project. Neither does the sole act of listening to the emergent musical ‘outcome’ of the sounds collected. The core meaning of *Ear Talk* is nurtured when the former constantly changes with the latter.

In *Ear Talk*, participants are often given guidelines for recording and sharing sounds that respond to a particular, yet broad, theme (e.g., sounds worth sharing and the 2020 coronavirus lockdown). Successful fulfilment of the guidelines is measured individually. Judgements on which sounds reflect and represent a given theme well are subjective and often very personal. For a project in this vein, individual artistic ingenuity is celebrated insofar as it helps each participant find and create sounds that reflect and represent a given theme to their liking, and any qualitative evaluation of these personal endeavours by others would be incongruous to the project. Instead, *Ear Talk* prioritises ‘quality time’ spent together by continuously changing and reorganising sounds, and this time spent results in a shared sense of contribution to the emergent, cohesive series of sounds (‘the music’), thus satisfying the project’s objective.

4.2. Evaluating accessibility

Ear Talk seeks to strategically curate enjoyable conditions for the type of audience (prospective

participants) expected. *Ear Talk*’s interface design (i.e., the ‘score’) and administrative directives (i.e., establishment of guidelines and themes) reveal our own assumptions about what we thought would be inclusive and accessible enough for a curious, English-speaking and internet-competent person. The utilisation of popular technological platforms such as YouTube, web browsers and Zoom was intended to exploit accessibility and familiarity among the prospective participants. YouTube, for instance, is a dominant platform for consumption of user-uploaded audiovisual content, and in September of 2020, it was the second most visited website on the World Wide Web (Alexa n.d.).

While some features within *Ear Talk*’s interface were developed to overcome the limitations posed by the internet-based medium (e.g., sound fidelity and immediacy) versus in-person, concertised music, our main concern when designing the interface was to establish conditions that would encourage participation and enable sonic exploration in a way that cultivates a ‘shared sense of quality’ (Burns 2015). For instance, *Ear Talk*’s interface is designed so that specialised knowledge of time and pitch-based organisation of sound is not necessary.¹⁶ The visual representation of sound (the ‘score’), for instance, aims to be accessible by utilising image blocks (see Figures 2 and 3) that do not specify time or pitch value. While participants can interact with rhythmic relationships between sounds by moving the position of image blocks or by submitting sounds that interplay with existing rhythmic elements, the lack of rhythm-managing systems (such as a tempo grid or metre) and the presence of latency inherent in online interactions means that *Ear Talk*’s interface does not allow enough control over the resolution in the time domain to enable detailed rhythmic control. The same goes for elaborating pitch-based sonic relationships; while participants can change the pitch of each sound file, coordinating an overarching pitch theory (modality or tonality) would be difficult. These conditions arise from our attempt to design a participatory system that is accessible to participants without any specialised musical training. This is not to claim that *Ear Talk* manages to remove all barriers of entry so that truly *anyone* can participate; what we assume to be intuitive for the visual layouts, musical parameters and various features built into the YouTube and web interfaces are subject to criticism and analysis.

¹⁶Specialised knowledge of time and pitch-based organisation of sound might include concepts and rules of rhythm, metre, modality, tonality and tempo among others.

4.3. Evaluating modes of engagement

Through the course of iteratively modifying and updating the *Ear Talk* systems, we observed differences in how each iteration targeted its potential performer and audience, and in turn, changed their capability to engage in social interactions. Various modes of participatory engagement made possible in our systems outlined a spectrum of possibilities between hierarchical participation, wherein participants with different levels of responsibility play towards a shared aesthetic, and horizontal participation, wherein interactivity between participants who all wield some level of artistic agency determines the emergent aesthetics. The making of the *Ear Talk* systems explicitly involved what Shelly Knotts (2015) describes as ‘the creation of the organisational structures’ that goes ‘hand in hand’ with any ‘development of new musical practices’. Knotts ties the ‘performer’s autonomy’ to the ‘relative democracy within [music] ensembles’, and in doing so, implies that the democratisation of musical processes that require ‘socially driven decision making’ offers ‘more autonomy than hierarchical structures can provide’. Drawing wisdom from improvisation studies, questions of autonomy, hierarchy and artistic agency are often complicated by social and relational dynamics that play out in real-time and are not necessarily determined solely by the design and ‘organisational structure’ of musical practices. Georgina Born’s (2017) framework of ‘social aesthetics’ acknowledges the simultaneous and continuous interplay between the multiple modalities of social and relational dynamics at play: the ‘microsocial’ (the ‘immediate, co-present and affective’ dimension), the ‘wider pre-existing social relations’ (‘class, race, ethnicity, gender, or sexuality’) and the ‘organizational, institutional, and political-economic’ (affiliations and connections to certain organizations/institutions). Thus, Born argues that relational power dynamics (hierarchy) and individual empowerment (agency or autonomy) are ever-present and ever-changing conditions of any social activity. In *Ear Talk*, we could observe how issues of hierarchy, agency and autonomy played out through the course of iteratively experimenting with different modes of participatory engagement.

As mentioned earlier, the *Ear Talk* system was initially built to facilitate an online collaborative performance for Consensus, who facilitated musical projects under the guiding principle of ‘Group Listening’. The multiple modes of engagement possible within the *Ear Talk* systems were carefully mediated through the role assignment and administrative directives. In *Sounds Worth Sharing*, only members of Consensus contributed sound files, and only Consensus members had roles that influenced their mode of engagement while participating in the

comments. By design, members of the ensemble had a greater influence on emergent aesthetics. While the open invitation for participation permitted anyone on the internet to join, those who were not part of the ensemble were to join at a different hierarchical level than the ensemble members. However, to say that *Sounds Worth Sharing* was less democratic, or fostered a more hierarchical experience because of the specialised roles, would be an oversimplification. Consensus had chosen this structural hierarchy in order to encourage ample participation and discourage antisociality concurrently. Given the ensemble members’ musical taste and background, Consensus was aware that the sounds their members would contribute might be more abstract or ‘noisy’ when compared with the type of music many prospective participants may be used to, and therefore felt that simply asking people to comment and manipulate the sounds without encouragement or guidance may lead to inaction. For instance, the Texture Master, who was tasked with cultivating *textural* richness in the music, could bring an aspect of listening to and engaging with the sounds that many participants might not have considered without encouragement. In this way, the structural hierarchy through the roles was implemented to remedy the inherently unequal power dynamics between an in-group (the ensemble members) who knew the inner workings of the system and any other participant who was not debriefed. The Consensus members’ roles encouraged them to partake in *Ear Talk* in a way that brought participants up to speed on the system’s unique possibilities and idiosyncrasies. These different levels of participation between the ensemble members and participatory audience members confound Turino’s (2008) binary of presentational versus participatory performance; *Sounds Worth Sharing* sits more comfortably within Camlin’s (2014) more nuanced idea of ‘performance-as-participation’ – a performance with a visible sense of who the main ‘performers’ are while openly inviting participation from those outside of the performing group.

For *Online Sound Gathering*, Tsuruoka and Consensus anticipated a slightly different audience because of their partnership with SEAMUS: an organisation of many self-identified musicians who would need less guidance on the type of musical collaboration that would take place. The emergence of ‘social aesthetics’ was notably distinct from *Sounds Worth Sharing*; many participants who shared similar musical backgrounds were involved, fostering inter-participant relationships more regularly. Additionally, the absence of Consensus’s specialised roles during *Online Sound Gathering* allowed all participants to move freely between the role of performer and audience member; shifting closer to

Camlin's (2014) idea of 'participation-as-performance', where the performer–audience distinction is blurred and the act of participating becomes the main focus and purpose of a performance.

The web-based system was designed to prioritise horizontal participation and further elevate conditions for accessibility.¹⁷ Incorporating features such as a drag-and-drop user interface, a built-in record function and drop-down menus for role selection introduced a level of control and accessibility that was previously impossible when limited by YouTube's proprietary interface. The web-based *Ear Talk* system also introduced highly asynchronous interactions, where only one participant may be active at a given moment during the perpetual paradigm. In this scenario, each participant manipulates sounds according to their own taste and desires. When such scenarios compound, 'the music' will have experienced multiple sequences of individualistic alterations; one participant inherits what another participant left off, each leaving their own imprint at their own time. The sense of total control or freedom to manipulate the sounds free from others' judgement during such an asynchronous collaboration could bring more satisfaction to certain participants. In a synchronous approach, such as a three-hour-long live performance, all interactions are visible to others, and the desired outcome must be socially mediated between the active participants. While each participant can simply move towards his/her/their desired musical outcome independently, such individual efforts are always intervened by group conversations that seek to refine general musical goals that all participants can take part in. Musical goals can vary infinitely and are indebted to social aesthetics, such as participants' backgrounds, interests, personalities, as well as comment-to-comment social interactions ('microsocialities') that fluctuate throughout the performance.

Networked music that utilises communication interfaces such as a live chat room has 'a unique potential to address radical democratic concerns about communication and power distribution' (Knotts 2015). In *Online Sound Gathering*, for example, comments such as 'The quiet moment after C is wonderful', which were casual and reflective, coexisted with more constructive conversations such as: 'let's see if we wanna shift some things to the very end where the guitar is. I don't think it's strong enough to stand on its own like that', 'I think we should spread out some of the sounds happening around E -> F ... those oooh's (voice recordings) should have more space to shine!' and 'I want to find a nice transitional sound to start at the end of sample20.' Comments that

expressed specific, desired musical goals often mobilised an outpouring of command comments attempting to make changes as a group in an effort to realise those goals. While momentary compositional goals might be achieved in this collective manner, there is no predetermined, ideal or 'final' outcome. Thus, an amalgamation of each participant's desires and contributions form and re-form the shape of 'the music'. Those who feel motivated, entitled and/or safe to express their desires are more likely to contribute and interact more, leaving a bigger, yet ultimately temporary, imprint than those who choose to be more tacit.

Eco (1989: 4) argues that in works that are 'open', where 'the very fact of our uncertainty is itself a positive feature', each performer becomes the 'focal point of a network ... without being influenced by an external necessity which definitively prescribes the organization of the work in hand'. This view towards a musical composition with a high degree of 'susceptibility to countless different interpretations' reflects an important element of *Ear Talk*: although there is no clear aesthetic prescribed by the organiser, a particular yet ever-evolving aesthetic experience often emerges from participant interaction and ultimately completes a musical process that is entirely unique and improvised.

4.4. Evaluating 'the music'

Ear Talk introduces a complication when defining what constitutes as 'the music' due to the fact that the *Ear Talk* project explores many different time-related paradigms; musical and social interactions span anywhere from a few milliseconds to several weeks or months. The process of collecting sounds from participants can be asynchronously organised, whereby each participant records and shares sounds at their most convenient time, or synchronously organised, whereby participants record and share sounds together with others during a specified time span. One of the most interesting aspects of the web-based *Ear Talk* system is its ability to persist in time indefinitely. Thus, asynchronous collaboration among participants is not only possible but also encouraged and inevitable when working with the system. In such a paradigm, each participant is constantly responding to and in dialogue with the existing material on the site, some of which may be recent while the rest are artefacts from performances long ago. On the other hand, the synchronous collaboration paradigms encapsulate the stretch of time during which people had come together to collectively tether seemingly unrelated sounds.

In *Ear Talk*, a single slice (or one loop duration) is only a small piece of the whole, and it raises the

¹⁷Zoom-aided live performances that involved 'featured artists' were exceptions where a clearer distinction between 'performer' and 'audience' was drawn (see Figure 4).

question of whether ‘the music’ is defined as the last loop that amalgamates all previous changes, or as the entire duration of a performance that encompasses all manifestations. In order to resolve this complication, we must look at how participants experience the piece. It is difficult to anticipate how (i.e., passively or actively) and when a participant would interact with the piece. *Ear Talk* inherently anticipates waves of varying participants to come and go throughout the performance, and those who participated early on may be surprised by the evolution of the piece when revisiting the stream towards the end of the session. One participant might even experience a very different piece upon revisiting the ‘same’ piece. In *Sounds Worth Sharing*, one participant commented ‘I literally go away for an hour and I feel like I’m listening to an entirely different piece.’ This reality makes us consider that ‘the music’ is defined as the span of time during which a participant experiences the music at any given moment. To borrow Bourriaud’s (2002: 16) words from his discourse on *relational aesthetics*, works involving social participation are ‘presented as a period of time to be lived through, like an opening to unlimited discussion ... [where] there is the possibility of an immediate discussion ... I see and perceive, I comment, and I evolve in a unique space and time’. Evidently, the resulting piece and the experience of it differs from participant to participant. The final evaluation of the piece, therefore, is in the hands of each participant, measured by their reflection of their personal interaction with *Ear Talk*.

Despite the complication in defining which part of the whole *Ear Talk* performance constitutes as ‘the music’, our participants showed positive feedback about the music created when nearing the end of a three-hour-long live stream. Some comments from *Online Sound Gathering* included: ‘Wow! The piece has changed a lot since I was here an hour ago’, ‘This was very cool. It is interesting to note that at the start many people were making lots of changes. Now as a piece has emerged from the material, the changes are less frequent’ and ‘The score has come to life :-).’ All these comments allude to and affirm a sense of completion, suggesting that while the sounds can indefinitely and continuously transform, a feeling of finality that concludes the group’s effort can be achieved.

5. CONCLUSION

In this article, two different implementations of *Ear Talk* – a technologically mediated performance system for participatory co-composition – are discussed. The YouTube-based *Ear Talk* system has enabled performers on opposite sides of the world to collaborate through socially engaged performances,

and the web-based implementation has made performances more accessible and introduced additional performance paradigms. From a real-time, co-creative performance to a perpetual installation paradigm, *Ear Talk* presents different modes of participatory engagement ranging between hierarchical and horizontal interactions. The open invitation to participate necessitated that the interface design be intuitive to use, the performance objectives be straightforward and the access to the performance be available for all who are curious, including music novices.

Such a participatory practice distinctly differs from presentational (or ‘traditional’) music-making in that the desirable musical outcome is constantly subject to change and is defined within each participant, influenced by other participants and refined as a group through a collaborative experience. Owing to the malleable nature of *Ear Talk*’s collaboration paradigms pertaining to time – where there is no one discrete span of time during which all who join will share one experience – there exists no one final piece to be interpreted, and any slice of the whole experience has a chance to become ‘the music’. To borrow Eco’s (1989) words, ‘every reception of a work of art’ (in our case, every loop or slice of time) ‘is both an interpretation and a performance of it, because in every reception the work takes on a fresh perspective for itself’.

The COVID-19 social distancing protocols, though not the impetus for this work, further highlight the value of online platforms such as *Ear Talk* that enable distributed forms of collaboration among musicians and audience members. While the future of traditional concert music is very much in flux, projects such as *Ear Talk* serve as avenues to maintain existing connections, foster new friendships and allow all who are curious to orchestrate enjoyable musical experiences.

Acknowledgements

Special thanks to Oliver Hickman for facilitating the technical infrastructure of the YouTube *Ear Talk* system and proofreading this paper, Ashley Muniz for editing and proofreading this paper, everyone from Ensemble Consensus who helped facilitate the YouTube *Ear Talk* system, and everyone involved in the administration and performances of *Ear Talk*.

REFERENCES

- Alexa Internet Inc. n.d. Alexa – Top Sites. www.alexa.com/topsites (accessed 12 September 2020).
- Born, G. 2017. After relational aesthetics: Improvised musics, the social, and (re)theorizing the aesthetic. In *Improvisation and Social Aesthetics*. London: Duke University Press, 33–58.

- Bourriaud, N. 2002. *Relational Aesthetics*. Paris: Les Presses du reel.
- Burns, S. 2015. ArtWorks: Reflections on Developing Practice in Participatory Settings. *Paul Hamlyn Foundation*. www.phf.org.uk/publications/artworks-reflections-on-developing-practice-in-participatory-settings/ (accessed: 12 September 2020).
- Camlin, D. A. 2014. Whose Quality Is It Anyway? *Journal of Arts and Communities* 6(2–3), 99–118.
- Eco, U. 1989. *Poetics of the Open Work. The Open Work*. Cambridge, MA: Harvard University Press.
- Ensemble Consensus. n.d. About – Ensemble Consensus. www.ensembleconsensus.org/about (accessed 12 September 2020).
- Gaye, L., Holmquist, E. L., Behrendt, F. and Tanaka, A. 2006. Mobile Music Technology: Report on an Emerging Community. *Proceedings of the 2006 International Conference on New Interfaces for Musical Expression*. Paris: NIME.
- Kant, I. [1790] 2000. *Critique of the Power of Judgement*. Cambridge: Cambridge University Press.
- Khan, M. L. 2017. Social Media Engagement: What Motivates User Participation and Consumption on YouTube?. *Computers in Human Behavior* 66, 236–47.
- Kim, J. 2010. User-Generated Content (UGC) Revolution?: Critique of the Promise of YouTube. PhD dissertation, University of Iowa.
- Knotts, S. 2015. Changing Music's Constitution: Network Music and Radical Democratization. *Leonardo Music Journal* 25, 47–52.
- Mignolo, W. D. 2011. *The Darker Side of Western Modernity: Global Futures, Decolonial Options*. Durham, NC, and London: Duke University Press.
- Quijano, A. 2007. Coloniality and Modernity/Rationality. *Cultural Studies* 21(2–3), 168–78.
- Roberts, C., Wakefield, G. and Wright, M. 2013. The Web Browser as Synthesizer and Interface. *A NIME Reader: Current Research in Systematic Musicology* 3, 433–50.
- Rosen, J. 2008. Google's Gatekeepers. *The New York Times*. www.nytimes.com/2008/11/30/magazine/30google-t.html (accessed 12 September 2020).
- Tsuruoka, T., Chang, L. and Hickman O. 2021. Ear Talk Project: Repurposing YouTube Live for Online Co-composition and Performance. *Proceedings of the 2021 International Computer Music Conference*. Santiago, Chile: ICMA.
- Turino, T. 2008. *Music as Social Life: The Politics of Participation*. Chicago: University of Chicago Press.