Sttera: Software for Audio Recognition and Orchestration as Medium for Music Ritual

Luís Arandas[1] and Rui Penha[2]

[1] Faculty of Engineering, UP – INESC TEC
luis.arandas@inesctec.pt
[2] School of Music and Performing Arts, IPP – CESEM
ruipenha@esmae.ipp.pt

Abstract. Contemporary tools support new ways to mediate ritual practices such that they are collaborative, inclusive and shared via digital devices. This research presents an interaction framework that can be used for the development of music rituals mediated by digital devices, and a proof-of-concept software entitled *Sttera*. It starts from the relationship that performers have with space, the embodied activity of making music and the concept of *musicking*. For the development of the proof-of-concept, cloud computing technologies are used to establish communication between participants as well as machine learning (ML) for audio detection and orchestration. The aim of this research was to design a platform and interaction model that could mediate and enhance the establishment of musical rituals without affecting the participants' sense of presence in the performative space.

Keywords: Music ritual, audio recognition and learning, orchestration, software-based mediation.

1 Introduction

Rituals are intertwined with culture. Whatever the ritualistic practice and irrespective of format and context, it is always loaded with representative and contextual meaning. As put by Couldry (2005), the term *ritual* inevitably brings with it religious legacies, offering a certain polysemy given the plurality of interpretation. Starting from Turner's comprehensive and conceptual position (1969, p. 96), ritual can be defined as "a *moment* in and out of secular social structure, which reveals, however fleetingly, some recognition (in symbol if not always in language) of a generalized social bond that has ceased to be and has simultaneously yet to be fragmented into a multiplicity of structural ties". In ritual we encounter an inherent relationship between the subjectivity of each participant's experience and their contextual place in a group. The search for

such a moment, as well as its recognition, is made up of diverse relations such as the space in which the person is situated and how this relates to its history. Such abstract relationships can be promoted by various modalities such as dance and music, often as territorial and *site-specific* practices (Wiflihani et al. 2019).

Music has been present in ritual practices for as long as we have evidence (Trehub et al. 2015). It is often the strongest component of connection between people, and even promoted the development of instruments such as the lyra (Dawe 2007, in Bates 2012, p. 368). Existing through sound sculpted by space, music rituals – being in themselves a social practice embedded in culture, where playing is often the main goal – end up being a phenomenon embodied by those present. The present research explores this practice, the music ritual, in which humans come together as a group to perform and listen to certain musical structures that connect to a culture dear to them, representing and exploring certain cultural meanings and signs that can only be experienced through relationships (Penha 2019; Benjamin 2015).

Leveraging contemporary technological tools and functionalities, we present a framework that promotes and allows for the performance of music rituals, seeking to stimulate the sonic specificities of different instruments. As a proof-of-concept, a software that uses ML and audio recognition capable of orchestrating digital devices using the cloud is presented. Named *Sttera*, the system is able to detect, learn and play audio on various digital devices present at the ritual, as well as to define when to play them through a customised orchestration system. The system is able to learn how to detect various sound sources with different acoustic properties and it allows them to be used as trigger to control playback (e.g., by an audience). The same can be done in a concert hall with a group of people or, given the nature of the web, in several different locations around the world. The system learns to identify audio segments – the default length being two seconds – and whenever the microphone detects them it creates an impulse stream to other connected devices. The ritual members, who possess playback devices, have a passive role, and they are connected to the musicking process as listeners, spatialising the musical gesture.

Using web-based tools (Smilkov et al. 2019), the contribution of this research is that it augments and mediates the ritual practice, such that it is collaborative, inclusive and democratically shared via digital devices.

2 Music Ritual and Contemporaneity

Given the connection that ritual has with context, there is a strong need to understand it as a historical practice. Ancestral practices – for example, the ritual of tea carried by the Zen Buddhists, where they gather before the

image of Bodhi Dharma and drink out of a single bowl with the profound formality of a holy sacrament (Okakura 1906) – are inaccessible to those who do not exist in that specific place, as well as to those who do not share a certain way of seeing the world. However, many ritual practices are inclusive, even promoting the insertion of audiences in the performance itself. Ritual is a performative temporal activity; it is something that must be developed over time, even if the ritual is performed by a single musician. This is directly related to the fact that music itself can only be understood in time. In music ritual, we endow elements such as the human voice and acoustic instruments with intrinsic meaning (Staal 2007), and we adopt various forms of participation. Small (1998) contended that music rituals exist as a performative art best defined by the term *musicking*, because the act of participating in performance is such an important element of the ritual.[1]

What music ritual performance means to those who practice it is multifaceted. Although music rituals have been present since the beginning of human communication, the performance of a musical ritual is necessarily contemporary, translating the tastes and choices of both the musician and the group, and fostering a subjective understanding of each other (Keil 1966).

3 Media-based Rituals and their Anthropology

The way we interact with each other as a social group in contemporary society is significantly different from previous centuries (Bryant and Peck 2006). The way we perceive the world and relate to things is also, in itself, a changing characteristic. One of the most important mechanisms of our life as members of a group is communication; it is from this interaction that we are able to make art, and to engage in the rituals surrounding artworks (Kádár 2013). With the polyvalence of interpersonal ties that can be established on the basis of the methods of communication existing in the twenty-first century, the concept of ritual takes on new dimensions, even if historically adapted and corresponding to canons established before the appearance of digital media and the internet. It is no longer common to have ritualistic moments of deep reflection, something that may also be linked to the speed and freneticism of current technology. Nevertheless, technology often plays an active role in moments of reflection where they do exist. Often, even if we have an especially transcendent moment in our daily lives, we tend to focus only on the digital object that allows it to

[1] Small (1998, p. 9) defines that to music is "to take part, in any capacity in a musical performance, whether by performing, by listening, by rehearsing or practicing, by providing material for the performance (what is called composing), or by dancing".

exist within social-media relationships, moving our mental relationship away from the experience and corporeality of a group action in physical space (Rothenbuhler and Coman 2005).

With the term *media rituals* (Couldry 2002), we can refer to "the whole range of situations where media themselves stand in for something wider, something to do with the fundamental organizational level on which we are, or imagine ourselves to be, connected as members" (Couldry 2005, p. 3). Globalisation and the large number of news sources we have access to nowadays, empowered by the most varied methods of automation in social media, redefines our conscious experience as members of society. Online communication, and the associated exchange of symbols and experiences, may also be thought of as a form of ritual. It may redefine part of the cultural meaning of ritual and the often necessarily site-specific participation of several people, for example when playing instruments. Digital devices are not just a means of connection between people, a way to access something different while keeping us physically in the same space; they are a space of existence in their own right (Herrera 2007). They engender a second world in which we live and in which we develop specific rituals, fostered by their poetic specificities, with which we can identify and develop specific behaviors. Using digital devices individualises ritual, even if from a behavioural point of view it shows richer dynamics (Wojtkowiak 2018). It undoubtedly expands the significance that humans can give to other types of objects and their use, however it loses the physical interpersonal relationship and it attenuates the vibrancy present in a ritual around a campfire, for instance, or inside a temple with other people around us.

Taking for granted that digital technologies and the enormous flow of information we exchange between ourselves will continue, we might start thinking about how we can use such infrastructure to promote a symbiosis with the plural musical ritual and its physical format in space (Small 1995). Respecting the place that technology has in our daily lives and in our continuous relationship with history and the act of musicking – often in centuries-old formats – we come across an immense possibility for expression empowered by the digital and the need to interconnect it with innovative forms of application (Salazar et al. 2018).

4 Promoting Musical Ritual from Contemporary Media

Digital media often mediate the very meaning of the things that happen in everyday life (Lievrouw 2009). Even if the media ritual does not exist as an event in its own right, the various media we use can reshape the way we look at the world, with both positive and negative effects. By simulating musical

practice and ritual in both physical and virtual space, we can raise questions such as: how can a digital device exist as a bridge between humans, granting the organisation of interaction structures in space, allowing the development of musical rituals to happen in that same space?

The specific focus of the present research concerns how digital infrastructure can be applied to physical musical ritual, how this happens, and what the promotion of these methods actually means. A context-agnostic position of music is used, and it is from here that we propose to develop an interaction framework between digital elements and the participants, treating them as nodes in a connection graph (Wellman 2001). The number of instruments used in a given music ritual is potentially very large, and this could become a problem for even the most sophisticated computer systems (Kazakos et al. 2021). However, with the approach taken here, the digital platform is not taken to be the place where musicians will enact the ritual virtually, but rather provides the means for the ritual to happen in physical space. Mobile devices come to be treated as resonant machines, and are the physical point of connection between people sharing the ritual in another location. The present research is grounded upon Small's definition of musicking (1998) as an inclusive multimodal performance, and takes into account the different cultural positions that musicking can afford among diverse cultures, with a specific focus on digital devices as a platform for connecting participants. We consider that media rituals are important given their pervasiveness (Payne 2009), but in this work they do not supplant the classical mode of performing in physical spaces.

5 An AI-based Element Framework for the Promotion of Musical Rituals

Musical ritual occurs through sound. It is possible to take advantage of new developments in audio representation and artificial intelligence (AI) to address problems such as the reconstruction of different types of instruments used in rituals on a digital platform (Grierson et al. 2019), just as it is possible to use the speed of communication established by web channels as a vehicle for near-instant transmission (Jia et al. 2018).

As a starting point for a model of interaction in music rituals, the framework proposed here takes as its context at least two people in different physical spaces who have the goal of performing a music ritual in both synchronous and asynchronous communication (Fig. 1). We furthermore assume that each of the performers has their physical instrument as a sound source in space, and a digital device to use as a link between them. A remote server is

also used, to provide the tools necessary to transmit and manipulate data (Raschka 2015). From here, the flow of information takes place after learning, analysing and detecting audio using the microphones of each of the performers' devices. We use the capacity of digital sound representation and its transmission between machines to be able to send signals considered to be significant between participants (McCallum and Grierson 2020).

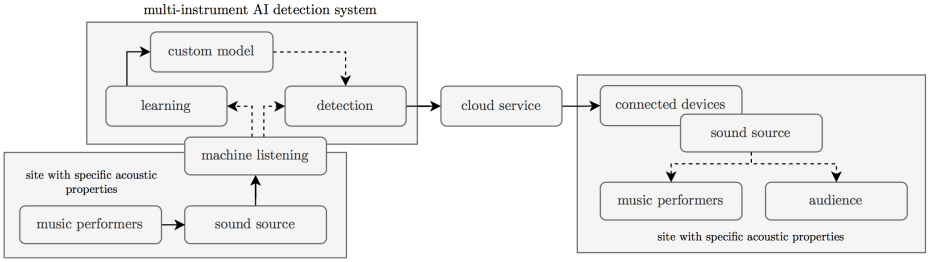

Figure 1. Proposed AI-based music ritual interaction framework.

Since there are two or more separate performers, each with their own instrument and interface, it becomes necessary to choose and define which are the important signals to transmit. These are defined relative to the performers, based on an audio stream analysis that is capable of perceiving patterns in the incoming data; in Fig. 1 this is described as machine listening. In real time, as the performers play, the devices start and maintain the information-processing chain. From here, and using AI for signal classification (Choi et al. 2017), we can detect and/or train a new audio model, so that whenever the system detects a certain sound, it sends signals to the receiver. This form of communication is bidirectional, and allows audio playback whenever an impulse is received, so that each device can not only listen, but also play (Biasutti 2015).

This approach is extensible, both for multiple performers and multiple audiences.[2] Allowing the interfaces to play audio files means that in addition to performers getting a rhythmic sense of what is happening elsewhere, they will also hear some musical material through the device. However, in the proposed framework, there is not constant audio streaming from one performer to other. Rather, what the framework tries to do is to find a robust way to learn certain musical gestures in time, and asynchronously transmit them to other members

[2] In cases where there is an audience wanting to participate in the ritual, they should be able to connect to the systems as a *receiver*, functioning only as a reproduction device in space.

of the ritual in real time, so that they can acknowledge and react to what is happening. This also allows for the transformation of interfaces into resonant objects if users so wish. Nevertheless, the main goal of a platform implementing this framework is simply to send impulses from one place to another quickly (see Robaszkiewicz and Schnell 2015).

Again, the reason for building an interface with these characteristics comes from the fact that many people who are in different spaces may want to perform a ritual together. It allows an ecological democratisation of musical practice, and above all it can connect different people in different locations (Clarke and Doffman 2017). Such a system might not make as much sense if used in the same room only by the performers, but it is possible to extend the act of musicking from one location to another if the infrastructure allows it, or even to fill a room with people mixing their devices with the performers' instruments in the acoustic space (Siddall and Waterman 2020).

6 *Sttera*: A Music Ritual Platform

Starting from the proposed framework of music rituals in Fig. 1, a platform was developed as proof-of-concept of the desired communication. Using the web, the largest communication space between humans nowadays (Sarhan and Gawdan 2018), *Sttera*, software that allows the detection, learning, analysis and orchestration of digital devices in the browser, was developed. It contains the necessary functionalities for performers to be able to communicate with each other, taking into account interaction and the specifics regarding sound and its representation. Building upon the various points listed in the proposed framework, *Sttera*: i) can automatically structure multiple connections between devices spread around the world, offering the possibility to define their interactions in different rituals and allowing them to interoperate with each other and create communication stations independently; ii) can analyse, learn, play, orchestrate and detect audio signals, given the plurality of instruments or unusual sound sources used in ritual contexts; and iii) can be used in a distributed and pervasive way given the capacity of cloud connections, promoting the integration of various participants in the ritual of both listening and playing music.

In developing *Sttera*, we encountered distinct problems that are often the subject of specific research, e.g., displaced interaction (Arandas et al. 2019) and multi-platform audio reproduction (Mason et al. 2015). This motivated our choice to use web technologies and focus on one of today's most relevant digital spaces: the browser (Taivalsaari and Mikkonen 2011). Browser-based software

already supports the application of ML in order quickly to solve problems related to the detection of instruments, and it offers the possibility of quickly training models in different acoustic contexts (Magnusson and Mendieta 2007). The several additional functionalities of *Sttera* were developed taking into account the different elements of the framework, and will be described in detail in the next sections.

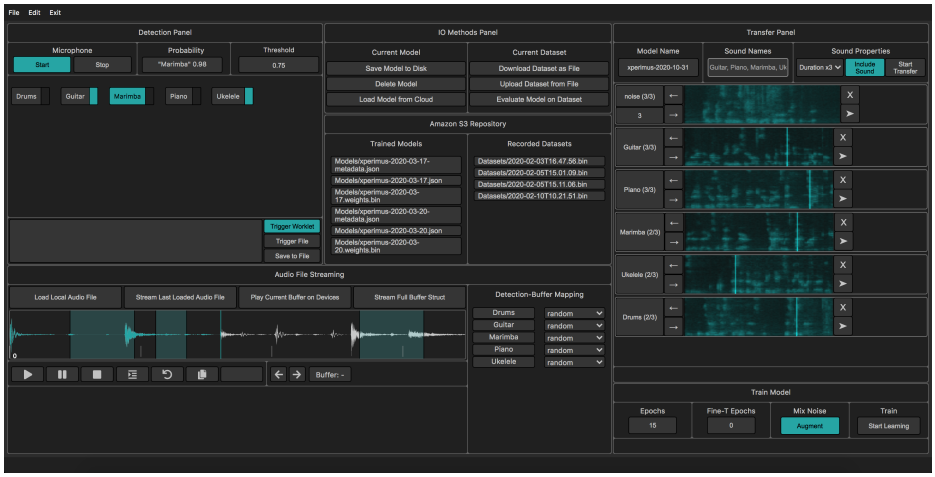

Figure 2. *Sttera*'s user interface, composed of several panels and interaction elements that allow the exploration of its features. Presented in the display are elements such as audio spectrograms, buffer segmentation and multiple buttons that relate to the ritual's setup.

6.1 General Features of the Platform

Web technologies are endowed with both obstacles and opportunities (Adedugbe 2020) that are important to consider when developing a platform that involves collaboration (Lind and McPherson 2017). The availability and accessibility of *Sttera* on mobile devices is also important given their ubiquity, quality of network technologies and multimedia playback capabilities (Clément et al. 2016). From the multiplicity of devices that ritual participants may have, browsers are the most readily available cross-platform solution. A browser-based solution melds with *Sttera*'s other main features, namely: i) a custom server hosted in the cloud, as a virtual connection centre; ii) the possibility to capture and play audio, as well as the application of ML and spectral analysis as recognition of what happens in space; iii) file and data management, allowing the recording of ML models for future reuse; iv) the orchestration of connected

devices via scheduling; and v) a multi-panel GUI mediating both server and device functionalities.

6.2 Architecture and Cloud Processing

The platform architecture developed is server-based (Arkko 2020) with a modular information stream. Established in *Heroku* (as in Muñoz-Lago et al. 2020) it is built on top of *Flask* (Grinberg 2018), a library that has been used previously in projects whose goals intersect with the present research, such as audio visualization (Pauwels and Sandler 2018), the *AcousticBrainz* project (Porter et al. 2015) and ML training systems (Yeager et al. 2015). For two-way communication the *SocketIO* library is used (as in Schlangen et al. 2018), and is bridged to Amazon AWS S3 services to download or upload trained models (Zadka 2019). In this proof-of-concept the server does not undertake any audio signal processing, being used only for the management of connections between devices. As shown in Fig. 2, the user interface has a set of graphical control elements that define the features to be used by the performer. In order to have access to all of *Sttera*'s functionality it is necessary to use the system from a device with a screen with sufficient pixels, such as laptops or tablets. The interface uses native browser technology, both javascript and HTML. The way the various elements communicate with each other is depicted in Fig. 3, where the *Heroku* server takes the place of the *cloud service* in the framework diagram and the *Sttera* instances (depicted in orange) simulate connections of the performers to the cloud. The edge devices can communicate with each other via the server and are divided as controllers and receivers to make a distinction between the devices that have full functionality and those that only play audio.

6.3 Audio Analysis and Supervised Transfer Learning

While establishing successful connections between devices, one of the most important elements of the proposed framework is robust audio recognition. Building on the immense number of spaces in which musical rituals can take place, sound analysis and recognition has to occur consistently and using the technologies described. The software must be able to capture and recognise instruments with totally different physical properties as well as in multiple reverberant spaces (see Oliveros 2007). Using browser features, the application of spectral analysis and description techniques has already been applied (Thompson et al., 2017; Jillings 2016; Rawlinson et al. 2015); there is also the opportunity to apply neural networks without a prohibitive computational cost

and with quite reliable results (Zbyszyński et al. 2017). The *tensorflowjs* library (Smilkov et al. 2019) is used and models are trained in the browser through transfer learning on top of the *SpeechCommands* model (Warden 2018). It is possible quickly to record seconds of a specific sound, assign them to a label, e.g., piano, and use them directly for instrument detection.

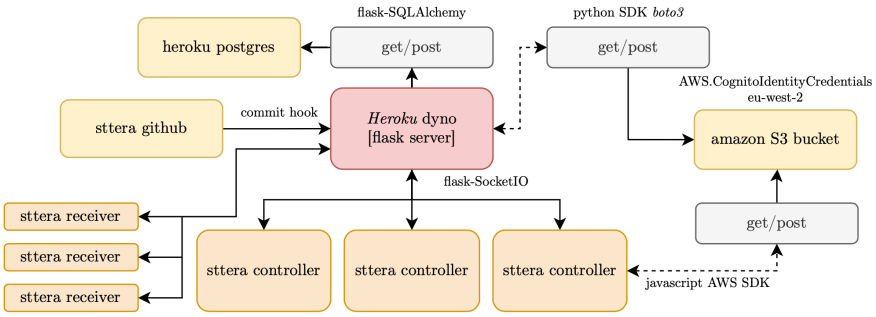

Figure 3. Diagram of communication between the platforms that are used in the proof-of-concept. The *Heroku* server, in peach, is the point of access to the cloud service; and below in orange are the various *Sttera* instances belonging to different participants in the ritual. Yellow depicts Amazon connections, a separate *Heroku* database service and the official code repository of the project through GitHub.[3]

All of this is done using the internally recognised microphone and optimised for fast usage inspired by the *Teachable Machines* project (Carney et al. 2020). Successful detection works probabilistically and by means of the type of trained model, which is built on the basis of the number of recorded examples. These functionalities are all executed without leaving the same web page, where we control the audio input, define examples to train the model, train and instantiate the model. The detection of a recognised stimulus is the starting point for triggering a transmission to another instance of *Sttera*.

6.4 Signal Transmission and Mapping

When there is a successful detection by the system, an impulse is transmitted to the remaining connected devices. This impulse is mediated by the server and is received by the other performers who are present. When conducting an experiment with more than two people, it is also possible to use

[3] https://github.com/luisArandas/xperimus.

Sttera's user-allocation mechanism for customised establishment of groups. This is done on a server-based identification, which allows the system to be scaled while maintaining some control. The transmission is lightweight, and little scalar data is sent, so it is fast and effective after detection. When the devices are used for playback, the material they will play has to be established beforehand, and the correspondence between transmitted signals and audio files to play is defined by a specific mapping.

Signal transmission allows performers to have contact with each other. It allows them to receive sets of impulses in time that can be used for a variety of multimedia playbacks. However, when a performance is collaborative, it is essential for the various participants to be able to hear each other. This does not happen directly in the proposed framework nor in *Sttera*. Rather, *Sttera* supports an extension of the musical gesture – and in turn the phenomenon of ritual – mediated by computational processes (Bishop 2018).[4] In order to give meaning to this set of impulses, and as mentioned before, it is possible to use the receiver devices as sound-reproducing machines in space. This can be done from previously transmitted audio files or from procedural generation using web audio. We can upload several files at the same time, and they will automatically be decoded to be sent by sockets. There is the possibility to map each detection label to a specific file or part of it, which specific part of the buffer is also automatically sent to the desired devices, making each impulse serve as a trigger for a specific sound. In this way impulses are not only sent from one performer to the other, but those same impulses reproduce musical material. The performer has an extension from one physical space to another, reproduced for another performer who may or may not relate directly to what is happening.

The possibility of procedurally generating audio on the various devices comes with a vast range of potential future applications when it comes to the playback of musical material using mobile phones and the browser (Correya et al. 2020). Also implemented in *Sttera* is a way to inject code into the devices' *AudioWorklet* and generate audio in a functional way using the FAUST programming language (Letz et al. 2015, 2017). This approach supports rapid prototyping and simulation of specific instruments that use the various impulses as a starting point. In the future, a code editor (as in Rietveld and Hoekstra 2013) could be embedded in the interface shown in Fig. 2; however, at the moment, a mapping from code blocks to the detection labels is not present, and demonstrates possible future work.

The interaction with the platform during the ritual happens through sound, and the machine tries to interpret the data while the performer plays.

[4] For research related to constant streaming in collaborative interfaces see Wozniewski et al. (2008).

The choice of material that the devices will play is up to each group and each ritual, and it is possible to have the devices play audio files that represent the instrument being played in the space, as well as any other kind of musical content, offering considerable freedom to the performers (Buffa and Lebrun 2017).

6.5 Device Orchestration and Spatial Arrangement

From the way the server identifies each established connection to *Sttera*, it is possible to control each one independently. This allows a possible customisation of the space and experimentation with the position of the members during the ritual. When there are more than two participants, we can not only choose what each detection plays on each device but also which one will play what. As a simple prototype, after the devices have received the material they are supposed to play, they do not all play at the same time, but in an ordered way according to the desired aesthetic experience. It is possible to make rhythmic sequences and to orchestrate what is intended to be played on the devices, as well as to disperse the participants in the space where they are located. Whether it is a performance hall or a cathedral, spatiality can eventually promote inclusion (Mitchusson and Allison 2018) and ultimately foster an experience that appropriates the device as a physical sound object (Bown et al. 2015).

6.6 Proof-of-Concept Usage Example

In order to introduce an example of the use of *Sttera*, a ritual setup is proposed to strengthen the relationship between its technical features and the underlying framework. A ritual scenario is constructed in which three performers play three different instruments – a piano, a marimba and a conga – and have a laptop with them in the space as well. Running on this laptop is an instance of *Sttera*, in which a new ML model was trained with small examples of each instrument, each one having a label with its name. The system is active, with the microphone on and making predictions on the basis of what the performers play. When a detection is successful, the system sends impulses through the cloud server, shown in Fig. 4 as *impulse matrix*, where a signal selection is made. Each impulse is sent directly to the other *Sttera* instances in *receiver* mode connected to the platform, depicted as mobile devices. In the receiving devices, each impulse is mapped into an audio file[5] which must also

[5] Only one is used in this simplified example.

be sent in advance, and which is played whenever it receives an impulse using the built-in speakers.

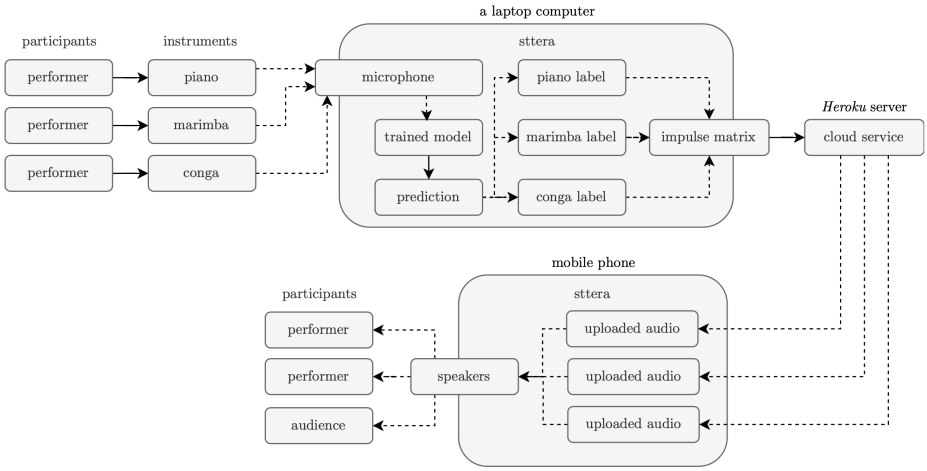

Figure 4. Proposed example of the use of *Sttera* in the context of musical ritual.

This example of interaction between performers demonstrates a possible and fast application of the software in a ritual context. It is also possible, in case any of the participants has the possibility to control the laptop, to change the audio files that are active in the mobile phones. For that, it is necessary to have some precision and control of the interface, though this need not be in the form of sending audio buffers in real time to the participants.

6.7 Possible Future Evaluations and their Approach

Following the proposal of a new framework and its proof-of-concept, it is also necessary to discuss the type of evaluation that should be employed (e.g., Wiggins et al. 1993). Given the multimodal nature of the experiences that rituals can have, such as a ritual in which dance is present or in a place where the internet is not available, we should question what in fact is our object of study to validate the application of such a system. From here we propose three types of approach. One would be to build on top of the theory of music

anthropology (see Merriam 1964),[6] where social behavior and its contextual specificities are fundamental to understanding the phenomenon of music ritual. A possible application of this approach is to bring together a specialist in a particular ritual, try to integrate the potential of the system described here with his/her approval, perform the ritual, and try to understand if it still has the same intrinsic meaning it normally has. This kind of investigation can be undertaken iteratively and may even result in knowledge fostering future investigations. We would not only be validating the performance of the system as an asset, but also seeking a heuristic validation from one or more people who consider a certain practice as their own.

A second approach would be to focus on the system and the items that are required to be captured in a given ritual, whether musical instruments or otherwise. Carrying out research that focuses on the type of system and how it can amplify the performative act may also help to understand how we are able to capture a certain characteristic. For example, the act of moving a certain object, not because of its sonic result but because of the meaning it may have in a group, is something that may not be able to be done with simple microphones. Focusing on a specific characteristic a given instrument may have, we must first identify a definition of what it is to perform a certain part of the ritual, clearly state what standards we use to evaluate how the system captures that part, and test the system in ritual against those same standards.[7] This type of study requires a longitudinal design that gathers knowledge regarding the specific AI/network system and the type of practice we are trying to capture.

The third approach is to focus on the people who experience the ritual we are considering. To try to understand fundamentally what they think is necessary for those who have never practiced it to relate minimally to its origin and value. For this, a qualitative approach will certainly be necessary, involving interactions in the specific spaces in which it takes place. This will help understand the behaviour that leads to making that specific music; what it tells us; in what way it happens in space; if it is something that does not rely on the interaction between people; if music is just one of many elements; what is the place of the passive listener; and how an intelligent system could comply with and perhaps extend such practice. It should take into account the digital literacy of the audience in our case study, since the use of digital devices can be something rather alien, and what we always want to preserve is the often

[6] In this text, the author argues that ethnomusicology cannot separate the sound-analysis of music from its cultural context (Merriam 1964, p. 8).

[7] These three steps, from identification to testing, represent a triangular relationship between events in the research process similar to the approach of the standardized procedure for evaluating creative systems (SPECS) framework (Jordanous 2012).

somewhat pre-established relationship of the subject with the given ritual (see Graber and Sumera 2020).

We do not believe that it is possible completely to assess our framework by taking into account only one study focused solely on the musical material, nor do we believe that adding more and more features to a system like ours might tackle the specificities of all possible rituals. We believe that the three proposals made here would help to create future research frameworks that are grounded in the validation of a person who is fundamentally part of the ritual. Trying to integrate technologies of a certain nature in a context where they do not exist is an elusive approach, and we can only generate knowledge of what is happening after a sincere validation by someone who treats a certain practice as their own. It might be difficult to compare it with other systems previously proposed since we will end up narrowing our object of study in the capacity that it has to execute a certain task, intelligently or not – the distancing from what it actually means to perform a ritual in context does not benefit the type of understanding discussed here. We should build our understanding of a social phenomenon by always taking into account what the agents of that same phenomenon consider about it, about themselves as they practice it, and about what defines it – from objects to concepts. Software should follow this, as well as never breaking pre-existing types of relationship. It should try to show new forms of interaction in different practices, however classical they may be, and should coexist with extant relationships in a healthy way. The type of study should always build upon this premise, and this interdisciplinary proposition has the potential to bring diversity to the study and the development of musical systems.

7 Discussion and Future Work

For a richly collaborative music performance, it is necessary that participants are clearly able to interact with each other. The felt vibration in the physical space, as well as the embodiment manifested by those present, is something that is part of the nature of performance and ritual, and no matter how many different media we may have mediating a given event, the feeling of presence remains fundamental. The identification and sharing that one person has with another is special when they share the physical space, and this is difficult to simulate in digital media. Nevertheless, digital media can promote some forms of sharing and allow kilometres of distance to be neutralised using web applications, even though one form of connection does not replace the other (Lindgren 2017). The present research starts from premises that build upon

classical views of ritual, and adopts Small's (1998) definition of the act of musicking, and also relies on his anthropological view of the phenomenon of performing, listening to and participating in musical rituals. These premises point to activities that have characteristics that are sometimes completely different from communication mediated by networks and algorithms (Frosh 2019).

In mediated interactions the relationship with space can be lost because, through the manipulation of attention, we focus only on the virtual processes that allow us to communicate. But, with the support of the proposed framework and its proof-of-concept implementation, digital communication technologies do not need to be completely excluded from physical ritual (Malloch and Wanderley 2017). Indeed, if the digital infrastructure is deployed in a noninvasive way and primarily used to establish contact between people, it can enhance the performance of the ritual. Importantly, the connection is not being approached as a virtual ritual space itself. We establish a structure of interaction by taking advantage of the cloud as a bridge, which allows for the inclusion of different people in different spaces. We use the facility of the web, as well as contemporary sound detection and analysis tools, as a starting point for embodied communication to happen between people (Akoumianakis et al. 2008).

As future work it is possible to propose an extension of the framework as well as to raise different tests of the proof-of-concept. For example, (i) we can focus on specificities of the medium and perform quantitative tests on network fidelity and data transmission between people; ii) we can correlate ML models on certain sound sources and their prediction; and iii) we can develop specific spatial languages and apply them in a given ritual. But it is also possible to stress-test the proposed framework in order to try to provide a better experience and to try to understand what kind of technologies we can adopt and apply in the context of musical rituals, which would help the physical experience of the musicians and audience members (Hagman 2005). We could further explore the way users identify with the interface, to determine whether it is relevant or whether it acts as a distraction that takes them away from the ritual. Different participants relate differently to technology and because no specific musical ritual is targeted, there is considerable openness in potential applications of *Sttera*. The functionalities of the system encourage this freedom, training models in a fast and adaptable way so that the system is not tied to a specific instrument and affording the possibility to make and work with different groups of people within the system, so that several rituals happen at the same time without affecting each other.

8 Conclusion

This article has presented research into music ritual and its performance supported by digital media, on what it means as historical action, and how it can happen in physical space mediated by contemporary devices and technologies. A framework for the development of platforms that promote the performance of musical rituals with several participants was proposed and described, along with a software system named *Sttera* that offers a proof-of-concept. By means of the necessary functionalities and the various elements of the framework, it is argued that it is possible to develop a platform that maintains the embodied act of practicing musical rituals, supported by the mediation of a cloud-based network, ML and interaction models that facilitate the orchestration of devices by performers and other participants in musical rituals.

Acknowledgements

The research leading to these results was financially supported by the Portuguese Foundation for Science and Technology (FCT), through the individual research grant 2020.07619.BD and by the project "Experimentation in music in Portuguese culture: History, contexts and practices in the 20th and 21st centuries" (POCI-01-0145- FEDER-031380), co-funded by the European Union through the Operational Program Competitiveness and Internationalization, in its ERDF component, and by national funds, through the Portuguese Foundation for Science and Technology.

References

Adedugbe, O., Benkhelifa, E., Campion, R., Al-Obeidat, F., Bani Hani, A., & Jayawickrama, U. (2020). Leveraging cloud computing for the semantic web: Review and trends. *Soft Computing*, *24*(8), 5999–6014.

Akoumianakis, D., Vellis, G., Milolidakis, I., Kotsalis, D., & Alexandraki, C. (2008). Distributed collective practices in collaborative music performance. In *Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts, Greece* (pp. 368–375). Association for Computing Machinery.

Arandas, L., Gomes, J. A., & Penha, R. (2019). Towards large-scale artistic practice with web technologies. In *Proceedings of the Web Audio Conference, Trondheim, 4–6 December 2019* (p. 87). Norwegian University of Science and Technology.

Arkko, J. (2020). The influence of internet architecture on centralised versus distributed internet services. *Journal of Cyber Policy*, *5*(1), 30–45.

Bates, E. (2012). The social life of musical instruments. *Ethnomusicology*, *56*(3), 363–395.

Benjamin, A. (2015). *Towards a relational ontology: Philosophy's other possibility*. Suny Press.

Biasutti, M. (2015). Assessing a collaborative online environment for music composition. *Journal of Educational Technology & Society*, *18*(3), 49–63.

Bishop, L. (2018). Collaborative musical creativity: How ensembles coordinate spontaneity. *Frontiers in Psychology*, *9*(1285). https://doi.org/10.3389/fpsyg.2018.01285.

Bown, O., Loke, L., Ferguson, S. J., & Reinhardt, D. (2015). Distributed interactive audio devices: Creative strategies and audience responses to novel musical interaction scenarios. In *International Symposium on Electronic Art, Vancouver*. New Forms Art Press.

Bryant, C., & Peck, D. (Eds.) (2006). *21st century sociology: A reference handbook*. Sage Publications.

Buffa, M., & Lebrun, J. (2017). Web audio guitar tube amplifier vs native simulations. In *Proceedings of the Web Audio Conference, London, 21–23 August*. Queen Mary University of London.

Carney, M., Webster, B., Alvarado, I., Phillips, K., Howell, N., Griffith, J., ... & Chen, A. (2020). Teachable machine: Approachable web-based tool for exploring machine learning classification. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–8). Association for Computing Machinery.

Choi, K., Fazekas, G., Sandler, M., & Cho, K. (2017). Convolutional recurrent neural networks for music classification. In *IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 2392–2396). IEEE.

Clarke, E. F., & Doffman, M. (Eds.) (2017). *Distributed creativity: Collaboration and improvisation in contemporary music*. Oxford University Press.

Clément, A., Ribeiro, F., Rodrigues, R., & Penha, R. (2016). Bridging the gap between performers and the audience using networked smartphones: The a.bel system. In *Proceedings of the International Conference on Live Interfaces* (pp. 96–102). Experimental Music Technologies (EMuTe) Lab.

Correya, A. A., Bogdanov, D., Joglar-Ongay, L., & Serra, X. (2020). Essentia.js: a JavaScript library for music and audio analysis on the web. In *Proceedings of the 21st International Society for Music Information Retrieval* (pp. 605–612). ISMIR.

Couldry, N. (2005). Media rituals: Beyond functionalism. In E. Rothenbuhler, & M. Coman (Eds.), *Media anthropology* (pp. 59–69). Sage Publications.

Couldry, N. (2002). *Media rituals: A critical approach*. Routledge.

Dawe, K. (2007). *Music and musicians in Crete: Performance and ethnography in a Mediterranean island society*. Scarecrow Press.

Frosh, P. (2019). *The poetics of digital media*. Polity Press.

Graber, K. J., & Sumera, M. (2020). Interpretation, resonance, embodiment: affect theory and ethnomusicology. *Ethnomusicology Forum, 29*(1), 3–20. https://doi.org/10.1080/17411912.2020.1808501.

Grierson, M., Yee-King, M., McCallum, L., Kiefer, C., & Zbyszyński, M. (2019). Contemporary machine learning for audio and music generation on the web: Current challenges and potential solutions. In *Proceedings of the International Computer Music Conference, New York, 16–23 June 2019*.

Grinberg, M. (2018). *Flask web development: developing web applications with Python* (1st digital ed.). O'Reilly Media.

Hagman, G. (2005). The musician and the creative process. *Journal of the American Academy of Psychoanalysis and Dynamic Psychiatry. 33*(1), 97–117.

Herrera, G. (2007). Cyberspace and sovereignty: Thoughts on physical space and digital space. In M. Cavelty, & V. Mauer (Eds.), *Power and security in the information age: Investigating the role of the state in cyberspace* (pp. 67-93). Ashgate Publishing.

Jia, R., Guo, H., Wang, Y., & Zhang, J. (2018). Analysis and test of sound delay on web audio under different situations. In *Proceedings of the 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)* (pp. 1515–1519). https://doi.org/10.1109/ICIEA.2018.8397949.

Jillings, N., Bullock, J., & Stables, R. (2016). Js-xtract: A realtime audio feature extraction library for the web. In *Extended Abstracts for the Late-Breaking Demo Session of the 17th International Society for Music Information Retrieval Conference*. ISMIR.

Jordanous, A. (2012). *Evaluating computational creativity: A standardised procedure for evaluating creative systems and its application* [Doctoral dissertation, University of Sussex]. Sussex Research Online. http://sro.sussex.ac.uk/44741/1/Jordanous,_Anna_Katerina.pdf.

Kádár, D. (2013). *Relational rituals and communication: Ritual interaction in groups*. Springer.

Kazakos, E., Nagrani, A., Zisserman, A., & Damen, D. (2021). Slow-fast auditory streams for audio recognition. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*) (pp. 855–859). University of Bristol.

Keil, C. (1966). Motion and feeling through music. *Journal of Aesthetics and Art Criticism*, *24*(3), 337–349.

Letz, S., Denoux, S., Orlarey, Y., & Fober, D. (2015). Faust audio DSP language in the web. In *Proceedings of the Linux Audio Conference, Mainz, Germany* (pp. 29–36). Musikinformatik & Medientechnik.

Letz, S., Orlarey, Y., & Fober, D. (2017). Compiling Faust audio DSP code to WebAssembly. In *Proceedings of the Web Audio Conference, London, 21–23 August*. Queen Mary University of London.

Lievrouw, L. (2009). New media, mediation, and communication study. *Information, Communication & Society 12*(3), 303–325. https://doi.org/10.1080/13691180802660651.

Lind, F., & MacPherson, A. (2017). Soundtrap: A collaborative music studio with Web Audio. In *Proceedings of the Web Audio Conference, London, 21–23 August*. Queen Mary University of London.

Lindgren, S. (2017). *Digital media and society*. Sage Publications.

Magnusson, T., & Mendieta, E. H. (2007). The acoustic, the digital and the body: A survey on musical instruments. In *Proceedings of the 7th International Conference on New Interfaces for Musical Expression* (pp. 94–99). Association for Computing Machinery. https://doi.org/10.1145/1279740.1279757.

Malloch, J., & Wanderley, M. (2017). Embodied cognition and digital musical instruments: Design and performance. In M. Lesaffre, P. Maes, & M. Leman (Eds.), *The Routledge companion to embodied music interaction* (pp. 438–447). Routledge.

Mason, A., Jillings, N., Ma, Z., Reiss, J. D., & Melchior, F. (2015). Adaptive audio reproduction using personalized compression. In *Audio Engineering Society Conference: 57th International Conference: The Future of Audio Entertainment Technology–Cinema, Television and the Internet* (pp. 1–4). Audio Engineering Society.

McCallum, L., & Grierson, M. (2020). Supporting interactive machine learning approaches to building musical instruments in the browser. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 271–272). Birmingham City University.

Merriam, A. (1964). *The anthropology of music*. Northwestern University Press.

Mitchusson, C., & Allison, J. (2018) Lost in space: Indoor localization for virtual environment exploration. In *Proceedings of the Web Audio Conference, Berlin, 19–21 September 2018*. Technische Universität Berlin.

Muñoz-Lago, P., Llorens, A., Parada-Cabaleiro, E., & Torrente, Á. (2020, September). SymPlot: A web-tool to visualise symbolic musical data. In *24th International Conference Information Visualisation (IV)* (pp. 537–543). https://doi.org/10.1109/IV51561.2020.00092.

Oliveros, P. (2007). Improvising with spaces. In *Proceedings of the 13th International Conference on Auditory Display, Georgia Institute of Technology*. http://hdl.handle.net/1853/50007.

Okakura, K. (1906). *The book of tea*. Duffield & Company.

Pauwels, J., & Sandler, M. (2018). pywebaudioplayer: Bridging the gap between audio processing code and attractive visualisations based on web technology. In *Proceedings of the Web Audio Conference, Berlin, 19–21 September 2018.* Technische Universität Berlin.

Payne, D. (1989). The Wizard of Oz: Therapeutic rhetoric in a contemporary media ritual. *Quarterly Journal of Speech, 75*(1), 25–39. https://doi.org/10.1080/00335638909383860.

Penha, R. (2019). On the reality clarified by art. *ÍMPAR: Online Journal for Artistic Research, 3*(2), 3–44. https://doi.org/10.34624/impar.v3i2.14149.

Porter, A., et al. (2015). Acousticbrainz: A community platform for gathering music information obtained from audio. In *16th International Society for Music Information Retrieval.* NYU Web Publishing. http://hdl.handle.net/10230/34997.

Rawlinson, H., Segal, N., & Fiala, J. (2015). Meyda: An audio feature extraction library for the web audio API. In *Proceedings of the Web Audio Conference, Paris, 26 January 2015.* IRCAM.

Raschka, S. (2015). *Python machine learning*. Packt Publishing Ltd.

Rietveld, L., & Hoekstra, R. (2013, May). YASGUI: Not just another SPARQL client. In *Extended Semantic Web Conference* (pp. 78–86). Springer.

Robaszkiewicz, S., & Schnell, N. (2015). Soundworks: A playground for artists and developers to create collaborative mobile web performances. In *Proceedings of the Web Audio Conference, Paris, 26 January 2015.* IRCAM.

Rothenbuhler, E., & Coman, M. (2005) *Media anthropology*. Sage Publications.

Salazar, S., Piepenbrink, A., & Reid, S. (2018). Developing a performance practice for mobile music technology. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 59–64). Virginia Tech. http://doi.org/10.5281/zenodo.1302679.

Sarhan, Q. I., & Gawdan, I. S. (2018). Web applications and web services: A comparative study. *Science Journal of University of Zakhoii 6*(1), 35–41.

Schlangen, D., Diekmann, T., Ilinykh, N., & Zarrieß, S. (2018). Slurk: A lightweight interaction server for dialogue experiments and data collection. In *Short Paper Proceedings of the 22nd Workshop on the Semantics and Pragmatics of Dialogue (AixDial/Semdial 2018)* (pp. 225–226). Universität Potsdam.

Siddall, G., & Waterman, E. (Eds.). (2016). *Negotiated moments: Improvisation, sound, and subjectivity*. Duke University Press.

Small, C. (1998). *Musicking: The meanings of performing and listening*. Wesleyan University Press.

Small, C. (1995). Musicking: A ritual in social space. http://jlarrystockton.com/Stocktoj/small.pdf.

Smilkov, D., Thorat, N., Assogba, Y., Yuan, A., Kreeger, N., Yu, P., ... & Wattenberg, M. (2019). Tensorflow.js: Machine learning for the web and beyond. *arXiv preprint arXiv:1901.05350*. https://arxiv.org/abs/1901.05350.

Staal, F. (2007). The search for meaning: Mathematics, music, and ritual. *The American Journal of Semiotics, 2*(4), 1–57. https://doi.org/10.5840/ajs1984241.

Taivalsaari, A., & Mikkonen, T. (2011). The web as an application platform: The saga continues. In *37th EUROMICRO Conference on Software Engineering and Advanced Applications* (pp. 170–174). IEEE. https://doi.org/10.1109/SEAA.2011.35.

Thomas, M. (2020). Digital performances: Live-streaming music and the documentation of the creative process. In E. Mazierska, L. Gillon, & T. Rigg (Eds.), *The future of live music* (pp. 83–96). Bloomsbury Academic.

Thompson, L., Cannam, C., & Sandler, M. (2017). Piper: Audio feature extraction in browser and mobile applications. In *Proceedings of the Web Audio Conference*, *London, 21–23 August*. Queen Mary University of London.

Trehub, S. E., Becker, J., & Morley, I. (2015). Cross-cultural perspectives on music and musicality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140096.

Turner, V. (1969). *The ritual process: Structure and anti-structure*. Routledge.

Warden, P. (2018). Speech commands: A dataset for limited-vocabulary speech recognition. *arXiv preprint arXiv:1804.03209*. https://arxiv.org/pdf/18 04.03209.pdf.

Wellman, B. (2001). Computer networks as social networks. *Science*, *293*(5537), 2031–2034.

Wiflihani, W., Silitonga, P. H., & Hirza, H. (2019). Digitalization of north Sumatera Malay ritual music using Cubase 5 software. *Budapest International Research and Critics Institute (BIRCI–Journal): Humanities and Social Sciences*, *2*(4), 556–566. https://doi.org/10.33258/birci.v2i4.6 57.

Wiggins, G., Miranda, E., Smaill, A., & Harris, M. (1993). A framework for the evaluation of music representation systems. *Computer Music Journal*, *17*(3), 31–42.

Wojtkowiak, J. (2018). Towards a psychology of ritual: A theoretical framework of ritual transformation in a globalising world. *Culture & Psychology*, *24*(4), 460–476.

Wozniewski, M., Bouillot, N., Settel, Z., & Cooperstock, J. R. (2008). Large-scale mobile audio environments for collaborative musical interaction. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 13–18). Università degli Studi di Genova. http://doi.org/10.5 281/zenodo.1179651.

Yeager, L., Bernauer, J., Gray, A., & Houston, M. (2015). Digits: The deep learning gpu training system. In *Proceedings of the International Conference on Machine Learning, AutoML Workshop*. Association for Computing Machinery.

Zadka, M. (2019). *DevOps in Python: Infrastructure as Python*. Apress.

Zbyszyński, M., Grierson, M., Fedden, L., & Yee-King, M. (2017). Write once run anywhere revisited: Machine learning and audio tools in the browser with C++ and emscripten. In *Proceedings of the Web Audio Conference, London, 21–23 August*. Queen Mary University of London.