TOWARDS RESPONSIVE SCORING TECHNIQUES FOR NETWORKED MUSIC PERFORMANCES

Rebekah Wilson
STEIM
Amsterdam, The Netherlands
rebekah@steim.nl

ABSTRACT

The latent and unstable nature of networked performances, where the delayed transmission and uncertain, unstable, and compressed reception of transferred information demands scoring conceptualizations that consider the loss of the presence information traditionally expected by musicians when performing together in a shared space and time. The focus of this study is to develop electronic network-aware responsive scoring techniques that consider the primary constraints of networked music performances: i.e., latency, uncertainty, multilocated, and digital. Using machine-learning techniques to investigate and enhance digitally mediated presence technology, scoring possibilities are discussed that promote the experience of performing together while being remote from each other—connected via a public network and subject to latency. This study also looks at compositional and technical approaches to creating responsive scores for networked music performances using analysis of transferred sound as a means to generate and control metadata and symbolic notation.

1. INTRODUCTION

The pianist reaches out and strikes a few notes, and a phrase glitters across the piano. A moment later, from somewhere that cannot be seen, an echo arrives, yet it is not a perfect echo. We can hear clearly another piano—another pianist—out of sight with a slightly different timbre, a slightly different tuning. Our pianist responds to the notes she hears; she plays her own echo back, a slight, un-avoidable transformation. Again, notes arrive from else-where that mirror the sequence—but not exactly, a respon-sive spectral apparition. The pianists iterate notes and phrases that sometimes align and sometimes do not, result-ing in unexpected harmonies and timbres. Separated by distance and connected via a digital network, the pianists engage with each other using a score that responds to their actions; i.e., a performance takes place where synchroni-zation is impossible, but music happens.

With the Internet's extended reach and the development of transmission technologies over the last two decades, music has become reproducible in unimaginable ways, leading to explorations of performing music together over the network. However networked music performance has yet to enter the common-use arena. When connecting remotely, musical participants experience the loss of significant amounts of multisensory information that they rely on when playing music together in composite space. While we can simulate and transfer certain kinds of sensory data, such as high-quality audio and video, outside of dedicated networks, this transfer is subject to latency and high compression rates, which cause the loss of infor-mation and network instability. These factors cause musi-cians to perform in unfamiliar territory where traditional forms of scoring and time-keeping fail. We may instead look towards designing new forms of scores and interac-tion that are well-suited to the technical, latent nature of the networked performance environment, leading to the development of positive conditions with the aim of wider adoption of the medium.

While technical advances in intermachine music networking protocols such as MIDI, Open Sound Control (OSC), and Dante may be now ubiquitous, creating music together over the Internet is largely limited to collabora-tions involving the non-real-time transfer of files and per-forming together is extremely limited in practice. As Nevejan states, "sharing rhythm, without having to make any extra effort, creates an initial mutual sense of trust" [1]. The latency and unreliability of the Internet affects musicians' effortless ability to trust that they are, or even can be, in rhythm with each other. This loss of trust dis-rupts the very act of making music together. However, mu-sicians have a deep capacity to adapt to new technologies. Considering how the Internet is deeply embedded within society, and the "transcendence of time and place has been a human drive for centuries" [2], musicians and audiences will accept networked music as a platform for expression given a solution for the disruption of rhythm.

In this paper, I present an overview of the primary characteristics of networked music performance that must be considered when developing network-aware responsive scoring interfaces specifically designed for performances over the public Internet. I introduce the concept of time-based "musical agents" that act as both score and represen-tation of the performance environment, where musical agents are detected, analyzed, and transformed in real-time by software. My work F Not F (2019) is presented as an ongoing experiment of this concept into practice.

2. PRIMARY CHARACTERISTICS OF NET-WORKED MUSIC PERFORMANCES

The primary characteristics of networked music perfor-mances are defined by technology and the imposition of that technology on participants. The public Internet is a packet-switched network made resilient through its decen-tralized design, where data is broken up into pieces and transmitted along many possible paths to its destination where it is put back together. This design is subject to la-tency and instability: i.e., packets may arrive late or not at all, which leaves the receiving side unable to faithfully re-produce the original. Network latency cannot be removed because it is a property of the natural laws of physics: i.e., information cannot travel faster than the speed of light and latency cannot be reliably predetermined or fixed at a cer-tain value. For example, latency may change throughout the duration of a performance. Consequently, latency and transmission uncertainty are primary unavoidable charac-teristics of networked music performances.

The second primary characteristic of networked music performances is that participants are multilocated co-crea-tors. A networked music performance requires at least two participants, each inhabiting their own distinct acoustic space and influencing a musical result that is experienced only by them and their local audience. The projection of sound from one location to the other creates new acoustic challenges that must be considered from a technical stand-point; e.g., is each musician provided with a microphone that is mixed with as little feedback as possible or are room microphones used carefully to provide a reverberant acoustic effect? How the acoustic performance details are determined often depends on the technical resources avail-able from one venue to the next. The third primary charac-teristic of networked music performances is that presence is digitally mediated: i.e., music is performed over a digital network where the signal is certain to be digitally encoded, compressed, transmitted, and decoded. Sounds, images, and performative relationships are managed purely by the means of this digital transmission and in the process of this unavoidable operation, we afforded the means to further manipulate and modify transmissions towards new timbral and aesthetic experiences.

[1] Sarah Weaver defines "composite space" as performing in the same space at the same time [3], in contrast to "networked space," where peo-ple are not in the same space, but are performing at the same time.

3. VITAL INFORMATION AND RHYTHM

Traditional music performances occur in composite spaces,1 i.e., spaces where people are together, here and now. In composite spaces, music is synchronized not only through music, but also through visual gestures and glances, a conductor's nod, or the down-movement of a violinist's bow. Musicians constantly exchange "relational information," the "tacit, the personal, experiential, ethical, aesthetic" [4], and respond to that information for the du-ration of their performance. By contrast, networked music is mediated presence: i.e., a partial presence that lacks the full-sensory relational experience of real-life interactions. The performative relationships between two or more peo-ple when communicating over a network are mediated dig-itally; i.e., they cannot sense or exchange presence infor-mation without technology. The technology of transmit-ting information over the network is full of uncertainties; it is latent and unstable. However, it is also capable of transmitting what musicians need to perform together. Nevejan suggests that "when in trouble, one needs good information and good communication; i.e., one needs 'vi-tal information'" [5], which is information that "supports survival for a specific person in a specific place at a spe-cific time" [5].

Transferring vital information in networked music per-formances is crucial in facilitating synchronization be-tween performers. In contrast to the traditional perfor-mance setting where ensemble musicians can rely on low latency audio and visual communication, when we connect remotely, we anticipate that the information that informs us about another's presence will be limited. For example, when we video chat with a friend who is on a limited net-work, we anticipate glitches and network drop-outs that in-terrupt a smooth conversation and we accommodate ac-cordingly by perhaps using text chat or rescheduling the call for a later time. These interruptions result in synchro-nization difficulties where we easily misunderstand each other's rhythmic intentions. In a musical performance, par-ticipants must consider what conditions can be controlled that limit these disruptions and still achieve a sense of shared time. Sharing time and rhythm is an important fac-tor because it creates performance intimacy between re-mote participants [6].

4. NETWORK-AWARE RESPONSIVE SCORING

A network-aware responsive scoring system is one that is in tune with both the conditions of the network and the mu-sical performance at the present moment in time. Respon-sive scores are graphical scores that in some way can be modified in real time to adapt to conditions, e.g., graphical scores that are connected to environmental input or that update a notation display as time passes. Responsive scores present information to the musician that reflect how musical relationships change over time in ways that static scores cannot, offering "ways to deal with time from a realistic standpoint" [7]. Responsive scores also afford mu-sicians to engage in the moment to change their actions and create a meaningful musical experience, where there "might be an urgency and a will to do what is required to effectuate that change" [8]. Combining responsive scores with techniques, such as telemetrics, machine learning, and statistics, responsive scores become network-aware: i.e., they can now adapt to latency and bandwidth changes as they occur, as well as responding to musical content from remote musicians as it is arriving over the network. Thus, the network-aware score can respond to and provide vital information about the performance environment to the musicians by addressing the primary constraints. A network-aware score must account for latency, it must acknowledge that no single location is the primary author or focus of a performance, and it must interface with digital network technology.

4.1 Existing network-aware interfaces

Network-aware and -responsive interfaces have been developed since networks have been in existence. For an excellent overview of historical real-time music systems for networked music, see Georg Hadju's "Automatic Composition and Notation in Network Music Environments" [9]. To date, network-aware scores have been primarily focused on transmitting and processing metadata or symbolic information or musical instructions for the performance itself due to limitations in compression technology and bandwidth. Hadju's Quintet.net transmits events between locations over the low-latency User Datagram Protocol (UDP), presenting musicians an interactive score of "certain notes or phrases to be played within time brackets" [10]. Extending this concept is Whalley's Graphic Networked Music Interactive Scoring System, where distributed clients communicate over OSC and map various musical parameters for interactive performances, with the common purpose of network-aware scores to "address practical timing problems in coordinating network-distributed participants following a score" [11]. Combining various approaches, the graphic and animated notation Decibel ScorePlayer system connects over a wide area network to transfer score instructions and permutations in real-time between remote participants [12].

With more recent developments and freely available high-quality audio transmission due to the growing ubiquity of broadband networks and concurrent software developments that offer high-quality, low-latency audio, we can now also analyze and modify the transmitted audio content itself, satisfying the network-aware requirement to consider the digitization and transmission of sound as an opportunity to seek new aesthetic approaches. Working directly with the transmitted audio, Ethan Cayko's toporhythmic research addresses "trans-chronotopic metricity" [13] by realigning decoded remote audio streams according to a telemetric-derived compensatory delay to achieve a fixed latency. This technical manipulation allows musicians to reliably perform precomposed rhythmic patterns while acknowledging the multiplicity of the distributed performance environment where each location experiences time—and therefore rhythm—differently.

4.2 Integrating network characteristics

Musicians are highly intimate with latency: i.e., the time it takes for a generated sound to reach the ears depends on the size of the performance environment and the distance from each other within that environment when performing in composite space where the response time is expected to be within some milliseconds. With the added network transmission time; however, response time is further in-creased. As Chris Chafe notes. "response time is variable depending on [the musician's] attention and what they hear, but it's way longer than the network delay" (private interview with the author, 2016). When a musician's re-sponse time is disrupted by network transmission, this la-tency extends musical opportunities with a unique dimen-sion because it affords further machine-processing latency that would otherwise be considered a disadvantage in a performance situation.

Network transmissions not only create response time de-lays and the inability to synchronize, but also cause con-gestion in networks, leading to unstable connections. Un-like in composite space, we cannot reliably repeat a musi-cal performance when connected over a network because of these unforeseen instabilities that may arise at any time. The musical score for remote participants must be de-signed for uncertainty and therefore be considered a guide to be followed rather than a score that can be reliably re-produced from one performance to the next given similar conditions.

4.3 Types of music representation

Hadju and Didkovsky state that in "current [networked music performance] environments one can typically differentiate three types of music representation being transmitted over the network" [14]. They go on to list the following:

1. Low-level audio, which can either be multichannel uncompressed or compressed (which increases latency, but allows audio transmission over consumer lines).
2. Mid-level performance data, which include note event or continuous-control messages generated by MIDI or alternate controllers.
3. High-level score data for symbolic representation of music.

I propose a fourth type of representation, i.e., the analysis and musical representation of digitally captured perfor-mances in audio format. With the aid of music infor-mation-retrieval methods and machine-learning methods, performative behavior patterns or "musical agents" can be detected in near-real time, processed locally, and transmit-ted over the network.

5. MUSICAL AGENTS

Hatten states that a "semiotic attribution of agency typically involves a sentient being that may set into action var-ious tools" [15]. Hatten introduces the idea of virtual agency, where a musical gesture or idea implies a creative force that generated that gesture or idea. The gestures then become agents of intent. When performing over a remote network where presence is mediated by technology, thesource of agency is hidden from us, yet we assume some-one, or something, created the sound we are hearing or at least triggered a process that set that sound into motion. Given we are communicating via technical means, we have data to detect and measure musical agents, and apply transformations and return them back to the remote space. By collecting and extracting these data, we can process and analyze these agents and expose new methods of operating within the constraints of the performance environment. Where natural human senses cannot detect the source of an action, technology can help us navigate this unknown territory through the transmission and generation of vital information and create new performance perspectives.

5.1 Detecting and manipulating musical agents

Musical agents serve to illuminate and respond to the primary characteristics of networked music performance and provide a channel for the transmission of vital information. This is achieved by classifying a musical gesture for each agent that can be documented, performed, modified, and transmitted. For example, the agent might be a series of pitch sequences in combination with certain loudness en-velopes and timbral fingerprints. Given a set of data, dis-tinct patterns can be stored and later recognized with the assistance of machine-learning applications. By detecting the transmission of the pattern, we know whether the agent has been transformed in some way, either by the musician or the network performance conditions.

5.1.1 Agent classification

Classifying a musical agent requires training on a certain pattern and consequently detecting that pattern using feature detection. Once the composer has decided which musical features [16] she wishes to detect, a convolutional neural network, which is "a set of filters that are trained to extract the most relevant features for detection from the received signal" [17] can be trained and deployed to ex-tract these features [18]. Where the agent is comprised of multiple components, several passes of sequence detection permits classification as a whole.

Machines are very good at classifying if they have been provided with sufficient and suitable training data, and the classification challenge matches the training data to an ad-equate degree. Deciding how to encode and train sound data so that it can be classified becomes an a~~esthetic deci-sion, as does choos~~ing the degree of accuracy and accepta-ble processing latency during a performance. Classification also requires significant attention to the data-capture method: e.g., when the input data changes, the result also changes. For music classification, this means that we must clearly define the classes and train as many different kinds of input as possible that reflect the classes under different conditions, i.e., with a variety of recording environments and instruments. As classification and pattern detection/ analyses improve, more complex decision-making tasks can take place. For now, the processing latency, classification errors, and statistical uncertainties due to insufficient data and processing time must be embraced when designing machine-learning score integrations.

[2] Quality of experience (QoS) is a set of network performance metrics (see RFC 6390, https://tools.ietf.org/html/rfc6390).

5.1.2 Agent comparison

Once a musical agent has been detected, we can obtain fur-ther detail through comparison with ideal, statistical, and historical agents. This creates a layer of comprehension: i.e., not only have we detected agents, but we also have coherence though means and variance. We might also compare the real-time transmitted model with statistical analysis of randomness generated by a computer model to generate yet more layers of data. For example, detecting vertical note alignment patterns through a loudness-based chromagram [18] permits the composer to input a series of chromagrams to be matched with more nuance than simply detecting a series of pitches. With the additional factor of loudness, a variety of variation is allowed for in the com-position.

5.2 Data collection

Networked music performances inherently contain multiple forms of information that expose its past and current states, including those of digitization, encoding, transmission, and decoding. The aim of data collection is strongly task-oriented, where data is not inferred through analysis, but retrieved from existing sources of measured information. Data sources in networked music performances in-clude telemetric data from the network itself, machine data, and metadata. Data collected from these sources do not tend to require additional processing: i.e., data in its raw format tell us the state of the system at the time we requested it.

5.2.1 Telemetric data

Telemetric data are the information automatically derived from the information generated by machine protocols during the transmission process and are designed for remote monitoring of equipment and services, which allows the calculation of statistical parameters (usually means and variance). Real-time protocol telemetry tells us detailed in-formation about the transmission, including latency or the time it takes for the data to be transmitted between loca-tions, the average rate of packet loss, the rate of change or "jitter" in latency, the number of packets discarded or re-paired, and a number of other useful measurements, such as the cumulative number of packets lost. Using this information, we can keep the participants informed of the state of the performance environment by providing quality of experience,2 and we can use the data as variables in musical algorithms and processes, such as is demonstrated in section 8. Figure 1 shows a real-time matrix of extracted data presented in a format that allows participants to absorb the present status of the system over the last 60 seconds at a glance.

Property	-60s	-50s	-40s	-30s	-20s	-10s	NOW
currentDelayMs	10	10	11	10	10	11	9
jitterBufferMs	4	5	5	4	4	5	4
jitterReceived	19	20	23	18	17	27	20
RTT	19	19	19	19	19	18	19
packetsLost	29	29	29	29	29	29	29
packetsSent	6330	6380	6429	6480	6530	6581	6682
bytesSent	162138	163388	164613	165888	167138	168434	169754
bytesReceived	232019	233381	234743	236105	237267	238829	240191
requestsSent	202	203	204	205	206	207	208
audioInputLevel	12	5	12	9	30	91	29
audioOutputLevel	80	67	57	67	9	23	25

Figure 1. Historical real-time output statistics of a networked music performance.

5.2.2 Music information retrieval

Music information retrieval primarily takes one of four forms [19]: images, such as a score; symbolic, such as MIDI; metadata, such as the instrumentation or knowledge about the performance and its environment; and digital audio. In networked music, we might classify telemetry as metadata. Music information retrieval permits us to make decisions specifically pertaining to music and performance: i.e., a score provides instructions on how and when to emit sound. MIDI and similar note-event systems tells us what electronic signals have been emitted, and information about digital audio gives us performance cues and markers. Such information is highly useful in determining context, particularly when musicians are remote from one another.

5.3 Data extraction

In contrast to data collection, data extraction is a creative, analytic process where, instead of gathering readily exposed information, meaningful information is inferred using algorithms. The extraction itself becomes part of the composer's process as she identifies and implements suitable algorithms, selects which data streams to input, and determines how to present the output to the data processors. Two immediate methods in networked music performances are those of processing the audio signal itself via digital signal processing (DSP) and extracting information about the transmission through a time-geography analysis.

5.3.1 Digital signal processing

One of the methods for extracting data from sound is DSP. As opposed to metadata collection, DSP is an algorithmic process that demands consideration of a variety of factors; therefore, it is a creative practice. DSP techniques can extract vertical and horizontal spectral and intensity information [20]. From these, we can deduce musical descriptions from the sound, such as invariance in pitch, loudness, duration, and spatial position. Using machine-learning techniques, we can extract musical features, such as beat tracking, chromagrams, and instrument detection and separation [21]. With this information at hand, the composer can implement a system that, e.g., listens to pitch or melodic content and responds accordingly, or can synchronize events according to note onsets.

5.4 Time-geography approach

Time and space are intertwined in networked music performances. Time-geography [22] is an investigation into spatiotemporal processes that tell us about "potential encounters between agents" [23] and how those agents are constrained by space and time: e.g., by capability, coupling, and authority [22], where capability is the constraints dictated by the system in which the agent exists, coupling is being together in a place and time, and authority tells us about the forces that the agent is subject to. With these topologies, we expose the interconnected nature of the performers with the musical environment including the transmission network, where time-geography's "two main tenets are that time and space are seen as resources and that the constraints which operate on human beings particularly in the physical environment, are the primary dictates of human experience" [24]. A time-geography approach detects how agents are moving through time and space, and "distinguishes between fixed and flexible activities based on their degree of pliability in space and time" [25]. By combining telemetric analysis with musical feature analysis, time-geography techniques can tell us about under what limitations the system is operating: e.g., if a pattern experiences deviation due to congestion occurring over the network, which creates ripple effects for each agent's response time and thereby influences the musicians' ability to perform.

5.5 Pattern recognition

Where situations can be described as abstract logical problems, machine-learning algorithms are better than humans at detecting patterns. Automated pattern recognition techniques can use training data to detect relationships to existing known sequences, or they use the sequence itself to describe relationships within that sequence as the time series occurs. Pattern recognition implies having a pattern to recognize. For example, in networked music performances, we would simply train a recurrent neural network, such as long short-term memory or bidirectional recurrent neural networks [17] depending on our requirements, with initial telemetric data giving a base reading of the environment before a performance begins. Once the neural network has been trained, we will output data that can be folded back into the score.

5.6 Sequence deviation

Statistical reporting of deviations from a sequence, where anomalies and detected and measured, can be applied to both high-level agent movements in time, and within the agent itself when investigation change in pitch or harmonic sequences. In networked music performances, this determines when musical agents are subject to variation, and can be applied in scoring situations to detect both the creative decisions made by musicians and evaluate transmission performance within the network.

5.7 Sequence prediction

To add yet more layers of interpretation to her application, the composer may choose to predict the next events in a given time-series. Where the objective of a predictive model is to estimate unknown variables for two-dimensional sequences, such as pitch prediction, note onsets, envelopes, and network telemetrics, statistical forecasting algorithms are most useful where, at least for short-term forecasting, machine-learning algorithms offer little benefit at the expense of complexity [26].

5.8 Composite models

Integrating multiple algorithms and processes allows the detection and generation of multiple types and layers of sequences, allowing a more-sophisticated nonisomorphic approach to working with digital interfaces. When a massive data cloud can be processed, this exposes information about music as it happens and delineates the spatiotemporal relationships between remote participants. Data analyses give the composer and musicians insight towards both creating an informed performance environment and considering future potential opportunities.

6. *F NOT F*: ANALYSIS OF A NETWORK-AWARE SCORE

F not F (2019) is a continuous research project that puts into practice the principles introduced in this study: i.e., using machine-learning and statistical techniques to develop a responsive network-aware score interface. By analyzing the performance conditions, the score reacts to time-based pattern synchronization where the intent is to react to trajectories and thresholds that must be met before making certain decisions. To achieve this, a series of musical agents are deployed, where musical patterns that are distinguishable by the machine are analyzed.

The title of *F not F* refers to Nevejan's work on presence, where participants share or do not share time, space, and action. They may be "here" or "not here," and they may be participating "now" or "not now." That is, perhaps the musicians are responding to each other in real time or they are performing alongside a recording [5]. The title *F not F* suggests the mathematical concept of a function where input relates to an output.

6.1 Version 1

The first iteration of *F not F* was a custom Python program and a fixed notated score created for a live musical experiment between two pianists located in Ghent, Belgium, and Rotterdam, The Netherlands, which was presented at the Orpheus Institute's March 2019 conference on Simulation and Computer Experimentation in Music and Sound Art. During the performance, the pianist's sound was continuously analyzed by musical feature extraction methods using the Aubio[3] and Librosa[4] libraries. The program ingested real-time data that was sent to a computer vision program built on TensorFlow[5] and Keras[6] with the aim to determine what musical phrase was currently being performed. The program had been trained prior with 16 musical phrases, recorded hundreds of times in a variety of conditions, and translated into chromagram images (Figure 2). The chromagram was selected over the spectrograph or other feature- visualization methods after the testing of all of Librosa's options determined that the chromagram was the most easily recognized image by the computer vision software.

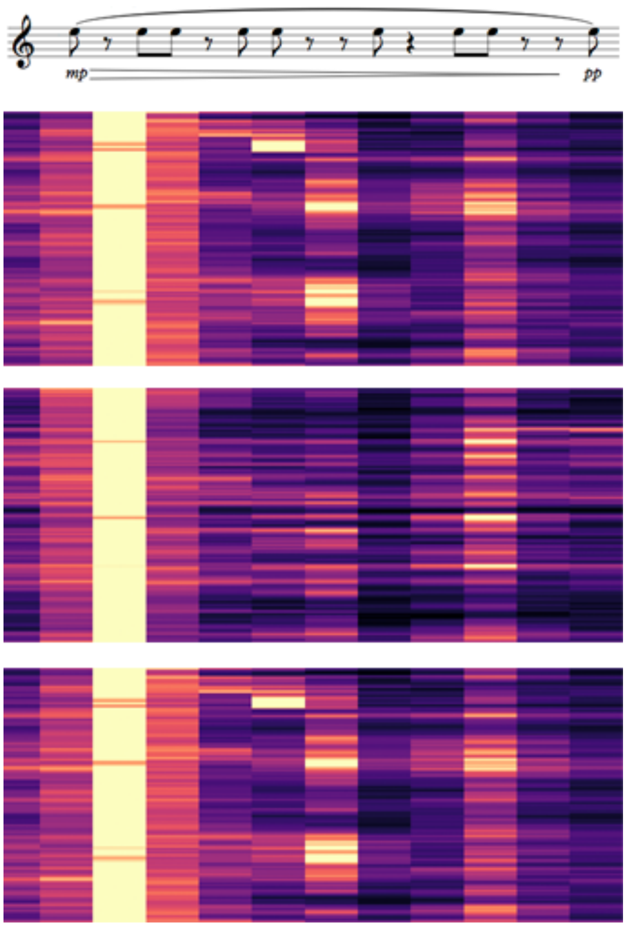

Figure 2. A musical phrase in *F not F* v. 1 and the chromagram representation of three recordings of that phrase.

These images served as training data to the program's machine-learning computer vision system. The performance consisted of the musicians playing a phrase a predetermined number of times, and then moving on to the next phrase. The discrete input of each pianist's mono-summed signal was recorded in real time by the program,

[3] Aubio is a C library and Python interface for the extraction of annotations from audio signals (see https://aubio.org).
[4] Librosa is a Python package for music and audio analysis for the purposes of music information retrieval (see https://librosa.github.io).

[5] TensorFlow is a platform for developing and training machine-learning models (see https://tensorflow.org).
[6] Keras is a high-level neural network Python API (see https://keras.io).

saved in 5-second wav files, one for each channel, and then immediately translated into chromagram images. The image was sent to the machine-learning program where the most-likely match was determined. This process took a few milliseconds. Once the program recognized a phrase, it triggered playback of samples from a bank selected according to the detected phrase. In all, a <6-second delay was experienced for the program to ingest and return the result. This delay informed the compositional structure where the musicians were instructed to repeat the musical patterns several times.

6.2 Version 2

F not F is a necessarily simplified implementation of the technical ideas discussed here; i.e., it is a constant work in development as the sophistication of tools improves. Machine learning is currently prone to error and ambiguity must be anticipated. Consequently, *F not F* is designed to embrace machine ambiguity while being structured overall for development of compositional narratives. Version 2 allows for a greater number of musicians and a variety of instruments to engage with musical agents. While version 1 was a static instructional score, a real-time score is generated for version 2, which can be viewed on a tablet connected to a central computer over a local or remote network. Like in the first version, an agent is a predefined microscore containing a distinct musical pattern where the rhythm and pitches can be easily detected by a machine. In version 2, agents become more sophisticated: i.e., a software program listens to each musician independently for the pattern and applies an interactive musical response. Table 1 lists events that are sent to the score once analysis has returned a result.

Analysis result	Action
A pattern has been partially detected	Trigger array of detected pattern string with threshold float value
A complete pattern has been detected	Trigger detected pattern string
The degree of synchronicity between any two data streams	Continuous stream of sync percentage (float array)
The degree of synchronicity between all streams (global synchronicity)	Continuous stream of sync percentage (float)
Telemetric data	Continuous stream of latency, packet loss, etc., to allow for synchronization calculations (e.g., mixed float, integer array)

Table 1. Analysis results and corresponding events

Version 2 of *F not F* is structured as a series of semi-notated precomposed instructions. For each series, musicians are presented with a series of notes. Figure 3 shows three examples of a note series that correspond to sections of a harmonic series, quantized to the semitone.

The musicians are instructed to improvise on the provided note series, playing notes in any order at a specified tempo and dynamic, but with deviating rhythms and timbres of their own choosing. For example, a resulting rhythm might sound something like the two rhythmic patterns shown in Figure 4.

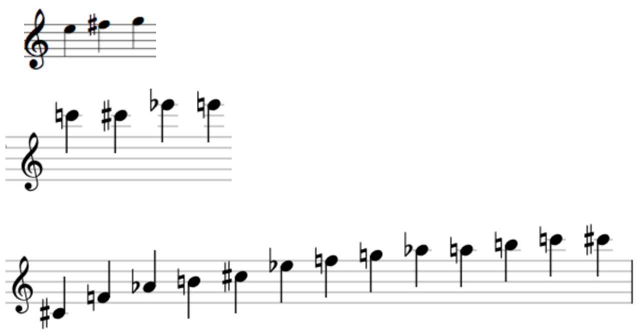

Figure 3. Three examples of a series: one for each musical agent.

Figure 4. Two musician-determined rhythmic pattern examples.

From this improvisation, all instruments in the ensemble combine to create a semi-chaotic timbral and rhythmic effect that continues for an indefinite period of time. The machine listens for a pattern it recognizes, which initially will be a partial pattern. Once this event occurs, the detected pattern becomes the focus: i.e., the "musical agent" that matches that pattern is initialized and the agent pattern begins to emerge in the score. As illustrated in Figure 5, the musicians begin to interleave the agent pattern (top sequence) with the chaos series (highlighted lower sequence). The score encourages the musicians to gradually play the agent pattern gradually more and more by flickering between the two layers, while the machine increases its reading of partial-pattern detection length.

Figure 5. Two examples of an agent pattern that would be animated and interlaced with its corresponding chaos series.

The musicians should now be starting to play the agent pattern more than not, which allows the machine to match complete instances of the agent pattern and calculate the level of horizontal rhythmic synchronicity between musicians, reading telemetric data to calculate for latency. Once synchronicity crosses a determined threshold, the next series is triggered: i.e., the previous agent pattern disappears and a new chaos pattern is shown to the musicians. This cycle can continue as long as the participants and composer wish, by supplying more or fewer patterns until the machine has completed its tasks.

An agent may also have a particular set of parameters that determine its musicality: e.g., it may respond to selected types of data streams or unique triggers specific to that agent to generate correlated effects using granular synthesis or other processing effects. In *F not F*, an agent may be responsive to the current global sync value in real time and react as outlined in Table 2.

NOT IN SYNC (<20%)	WITHIN THRESHOLD (>20% to <80%)	IN SYNC (>90%)
No effects	Clicks, noise bursts, sine tones	Agent sample triggers
	Increasing granular synthesis level	Granular synthesis immersion
		Countdown to next series

Table 2. Triggering electronic effects in *F not F*

When the musicians are not in sync, there are no effects. This decision was made so they can more easily seek synchronicity with each other. As their sync increases, so do the effects and granular synthesis levels. The intent of increasing and decreasing the intensity of electronic effects is to offer the musician the musical choice of whether she wishes to move towards or away from sync, depending on her musical intentions in accordance with the ensemble.

7. FUTURE WORK

The long-term goal of this research study is to develop advanced methods for detecting how musical agents are created and transmitted between networks. Learning how they interact and modified and transformed, whether by musician or machine, guides us towards understanding how participants experience and react to being in-time and in-rhythm with each other when performing remotely over a network. While being a disruptive factor in music, latency can be used for positive gains with machine interaction. The unavoidable delay can be well-utilized, where the transmission wait can be used to analyzed and generate performative interfaces for musical developments. Developing complex network-aware systems demands the intersection of real-time machine protocols with human-level research on mediated presence technology and theory and computer music studies. Collecting data from as many sources as possible at the time of creation and processing that data with algorithmic approaches allows us to create sophisticated sound processing applications and score creation for distance-aware composition structures while building strong musical relationships and transferring the vital information that musicians need when performing remotely. When working with sophisticated graphical interfaces, the transmission latency could be highly beneficial, e.g., when developing complex, immersive works using virtual and augmented reality tools.

Most machine-learning and statistical technologies to date are focused on genre classification or harmonic analysis. A real-time library for detecting and comparing sound features such as timbre, note onset and decay, and agent detection and comparison, e.g., would allow the composer to consider greater score and compositional complexities. As detection and analysis tools become more sophisticated, such as the real-time application of deep learning tools to recognize activities [27], composers will have the ability to model and detect musical features at fine resolution and great speed at their fingertips [28]. In addition, structural and decision-making events can be analyzed and cocreated by musicians and machines. In step with advances in analysis, innovations in digital scoring technologies increase the possibilities of representing this information with highly sophisticated responsive musical agents, which leads to more meaningful encounters between remote musicians and computing systems. As greater amounts of data become available, human perception tends towards perceiving greater meaning; i.e., we attribute a machine's complex recognizable patterns to mechanisms imbued with agency.

Ultimately, there are undiscovered depths of exploration in relation to performance and the machine, and our continued curiosity will certainly uncover many more aesthetic and technical applications. This research works towards the creation of a functional library of performance and scoring tools, while growing the portfolio of networked music performance works by this composer. She hopes to entice others to approach networked music performance as a source for novel and intriguing aesthetic opportunities.

8. REFERENCES

[1] C. Nevejan, Urban reflection: On the design of diverse engagement in the networking city of Amsterdam. [Online] Available: https://nevejan.org

[2] C. Nevejan and F. Brazier, "Granularity in Reciprocity," *AI & Society*, vol. 27, no. 1, 2012, pp. 129–147.

[3] P. Oliveros, S. Weaver, M. Dresser, J. Pitcher, J. Braasch, and C. Chafe, "Telematic Music: Six Perspectives," *Leonardo Music Journal*, vol. 19, no. 1, 2009, pp. 95–96.

[4] S. P. Gill, "Tacit Engagement: Betwixt and In-between," in *Tacit Engagement*, Springer Nature, 2015, pp. 127–160.

[5] C. Nevejan, *Presence and the Design of Trust*. Dissertation, University of Amsterdam, 2007.

[6] J. Croft, "Theses on Liveness," *Organised Sound*, vol. 12, no. 1, 2007, pp. 59–66.

[7] R.R. Smith, Composer Ryan Ross Smith lectures on his 'Study no. 47' at the New Music Gathering. [Online] Available: https://www.youtube.com/watch?v=p66Quh8zDl8

[8] S. Emmerson, M. Baalman, and O. Brandtsegg, "Instrumentality, Perception and Listening in Crossadaptive Performance," in *ICLI 2018 – International Conference on Live Interfaces (Porto)*, June 14–16, 2018, pp. 86-95.

[9] G. Hajdu, "Automatic Composition and Notation in Network Music Environments," *Proceedings of Sound and Music Computing Conference, SMC 2006 (Marseilles)*, 18–20 May, 2006.

[10] Quintet.net. [Online] Available: https://quintetnet.hfmt-hamburg.de/wiki/projects/quintetnet/Quintetnet.html

[11] I. Whalley, "GNMISS: A Scoring System for Internet2 Electroacoustic Music," *Organised Sound*, vol. 19, no. 3, 2014, pp. 244–259.

[12] S. James, B. Carey, X. Fu, and G. Hajdu, "Establishing Connectivity Between the Existing Networked Music Notation Packages Quintet.net, Decibel ScorePlayer and MaxScore," in *Proceedings of the International Conference on Technologies for Music Notation and Representation, TENOR'17 (A Coruña)*, 2017, p. 13.

[13] E. Cayko, *Rhythmic Topologies and the Manifold Nature of Network Music Performance*. Dissertation, University of Calgary, 2016.

[14] G. Hajdu and N. Didkovsky, "On the Evolution of Music Notation in Network Music Environments," *Contemporary Music Review*, vol. 28, no. 4–5, 2009, pp. 395–407.

[15] R. S. Hatten, *A Theory of Virtual Agency for Western Art Music*. Indiana University Press, 2018.

[16] G. Peeters, *A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project*. IRCAM, 2004.

[17] N. Farsad and A. Goldsmith, "Neural Network Detection of Data Sequences in Communication Systems," *IEEE Transactions on Signal Processing*, vol. 66, no. 21, 2018, pp. 5663–5678.

[18] T.L. Li, A.B. Chan, and A.H. Chun, "Automatic Musical Pattern Feature Extraction Using Convolutional Neural Network," *Genre*, vol. 10, 2010, pp. 1x1.

[19] J.A. Burgoyne, I. Fujinaga, and J. S. Downie, "Music Information Retrieval," in S. Schreibman, R. Siemens, and J. Unsworth (Eds.), *A New Companion to Digital Humanities*, John Wiley & Sons, Ltd, 2015, pp. 213–228.

[20] M. Muller, D. P. W. Ellis, A. Klapuri, and G. Richard, "Signal Processing for Music Analysis," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 6, 2011, pp. 1088–1110.

[21] B. Srebrenik, Machine learning and music classification: A content-based filtering approach. [Online] Available: https://towardsdatascience.com/machine-learning-and-music-classification-a-content-based-filtering-approach-f2c4eb13bade

[22] T. Hägerstraand, "What About People in Regional Science?," *Papers in Regional Science*, vol. 24, no. 1, 1970, pp. 7–24.

[23] S. Winter and Z.-C. Yin, "The Elements of Probabilistic Time Geography," *GeoInformatica*, vol. 15, no. 3, 2011, pp. 417–434.

[24] N. J. Thrift, "An Introduction to Time-Geography," In *Geo Abstracts*. University of East Anglia, 1977.

[25] H. J. Miller, "A Measurement Theory for Time Geography," *Geographical Analysis*, vol. 37, no. 1, 2005, pp. 17–45.

[26] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward," *PLoS One*, vol. 13, no. 3, 2018, p. e0194889.

[27] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep Learning for Sensor-based Activity Recognition: A Survey," *Pattern Recognition Letters*, vol. 119, 2019, pp. 3–11.

[28] R. Fiebrink, "Machine Learning as Meta-Instrument: Human-Machine Partnerships Shaping Expressive Instrumental Creation," in Musical Instruments in the 21st Century, T. Bovermann, A. de Campo, H. Egermann, S.-I. Hardjowirogo, and S. Weinzierl, Eds. Singapore: Springer Singapore, 2017, pp. 137–151.