

Experimental Investigations and Future Possibilities in Network-Mediated Folk Music Performance



Chrisoula Alexandraki

Abstract This chapter is intended to acquaint music ethnologists with the paradigm of Networked Music Performance (NMP). NMP facilitates computer networks to allow musicians from distant geographic locations to synchronously collaborate during performance, improvisation or more generally music-making. The chapter comprises two parts. The first part is devoted to providing an overview of research approaches in NMP and elaborates on the technical and perceptual impediments restricting the wide availability of this type of technology. The second part presents an experiment involving three musicians performing folk music over the network. The experiment serves to reveal not only technical and perceptual difficulties in the communication of performers, but more importantly their attitude towards engaging in this novel practice. The chapter concludes by discussing future perspectives on the use of NMP technology in the context of ethnic and folk music.

1 Introduction

Nowadays, network communications are used for almost every form of daily activity. In music and musicology, this abundance of networking applications demands for reassessing conventional practices not only for music preservation and distribution, but also for music-making.

In line with this reasoning, the adoption of digital audio recordings in music preservation has offered significant improvements not only in terms of the quality of the recorded music, but also in terms of documenting and maintaining relevant information in music archives. Furthermore, the recent proliferation of cloud computing technologies has enhanced the visibility of music archives as well as of pertinent research instruments. Yet in more recent days, advances in semantic technologies and computational intelligence allows redefining the role of music archives from isolated

C. Alexandraki (✉)

Department of Music Technology and Acoustics Engineering, School of Applied Sciences, Technological Educational Institute of Crete, Heraklion, Greece
e-mail: chrisoula@staff.teicrete.gr

© Springer Nature Switzerland AG 2019

R. Bader (ed.), *Computational Phonogram Archiving*, Current Research in Systematic Musicology 5, https://doi.org/10.1007/978-3-030-02695-0_10

207

storage repositories to linked data resources, hence promoting novel research perspectives in massive processing of music data including the unveiling of transcultural characteristics and transdisciplinary research possibilities. A potential next step to this progress is on the actual act of music-making, including the improvisations of folk musicians and the musical interactions taking place among different cultures and ethnic groups.

This chapter focuses on NMP technologies and discusses perspectives emerging from their adoption by folk musicians. NMP can be thought of as a distinct type of teleconferencing application, which allows remotely located musicians to engage in synchronous music performance using network infrastructures and dedicated software tools. Teleconferencing and Voice over IP technologies have a history of more than thirty years now and are being broadly used for an abundance of daily collaboration activities. They are possibly the most prominent type of groupware in computer-supported cooperative work. However, as music performance constitutes a special creative activity with severe restrictions in timing synchronization as well as in motor and cognitive engagement of participants, the progress in teleconferencing has not been effectively propagated to the music domain. As a result, the potential of folk musicians using NMP technology to engage in distributed music performance remains to be elucidated.

The rest of this chapter is structured in two parts. The first part presents an overview of NMP research approaches by elaborating on the challenges faced by this line of research and the workarounds or achievements in meeting these challenges. The last section of this part elucidates that depending on the actual scenario or interaction context among musicians, NMP sessions may have a different degree of technical complexity and that there are in fact scenarios in which synchronous collaborations of musicians may be feasible through the network. The second part of the chapter is devoted on describing the experience of synchronous collaborative performance in the absence of co-presence and how it can be realized in the context of folk and traditional music. An experiment of three musicians performing two pieces of the traditional music of Crete is presented along with an evaluation concerning not only technical measurements, but also qualitative aspects delineated by interviewing performers. Finally, the chapter discusses the potential of folk musicians widely adopting NMP technology and the influence of this technology on the appearance of new music styles emanating from cross-boundary music collaborations.

2 Networked Music Performance Research

Since its inception, NMP was not intended to substitute conventional performance. In fact, professional musicians, unless acquainted with the practice of avant-guard music performance, appear sceptical of the idea of being physically separated from their peers. Unlike speech, music collaboration relies on sharing common ambience in terms of sound reverberation as well as in terms of visual communication concentrating on motor interactions and eye contact between musicians. Hence, physical

proximity of musicians and co-location in physical space are typical prerequisites of collaborative music performance. Yet, musicians are captivated by the use of network-mediated music performance not only as an experimental music practice, but also as an enabling practice when physical co-presence is not possible. NMP is considered to be an enabling technology in cases of travelling obligations of musicians or as an opportunity to reach musicians of a distant geographic region.

The idea of music performers collaborating across distance was remarkably intriguing since the early days of computer music research. Early experimental attempts on exploring the aesthetics of network music interaction date back to the 70s [1, 2]. However, in these approaches the focus seems to be placed on machine interaction rather than on the absence of co-presence, as in all of these initiatives the performers were in fact collocated and connected through Local Area Networks (LAN). Telepresence across geographical distance initially appeared in the late 1990s [3] either as control data transmission, noticeably using protocols such as the Remote Music Control Protocol (RMCP) [4] and later the OpenSound Control [5], or as one way transmissions from an orchestra to a remotely located audience [6] or to a remote recording studio [7]. True bidirectional audio interactions across geographical distance became possible around 2001 with the advent of academic network infrastructures, specifically the Internet2 in the US and later the European GEANT. These infrastructures offer highly-reliable broadband connections that are necessary for collaborative music performance.

Since then, a number of research projects have been initiated. Despite the almost two decades of a substantial number of research efforts, the main challenges faced by the implementation of NMP applications have not been defeated. Currently, NMP is only feasible on academic networks and under certain limits in geographical distance. On the other hand, widely available network infrastructures are characterized by a number of technical constraints that impede meeting the perceptual prerequisites of musicians during live performance. Hence, NMP facilities are only available to the academia.

At present, NMP research is highly interdisciplinary as it involves numerous aesthetic, technical and perceptual aspects. An extensive overview of research efforts on NMP is beyond the purposes of this chapter as there are dedicated works available in the relevant literature [8, 9]. The following sections attempt to provide an overview of basic concepts, challenges and approaches that will allow interested researchers to gain an understanding on the current issues and future perspectives of NMP research.

2.1 Technical and Perceptual Challenges of NMP Research

Undoubtedly, one of the main reasons why NMP remains an unsolved problem is related to what we know of the cognitive processes and the perceptual qualities involved in synchronous collaborative music performance. Unlike conventional music performance and as portrayed in Fig. 1, NMP fosters collaboration of musicians located in dissimilar environments, with respect to several modalities. To understand



Fig. 1 NMP fosters collaboration of musicians located in dissimilar environments, with respect to several modalities

this type of interaction, one needs to understand the cognitive processes that enable musicians to synchronise during conventional ensemble performance. A number of dedicated studies [10–12] confirm the fact that defining metrics and thresholds for successful collaboration of musicians is a poly-parametric and tremendously complex problem.

The next sections discuss the main technical impediments of NMP systems and their relation to perceptual aspects of collaborative music performance.

3 Communication Latency and Latency Tolerance

The most prevailing problem of NMP technology is the communication latency occurring among remotely located performers. This latency is due hardware and software equipment, network infrastructures, and the physical distance separating collaborating musicians. Specifically, in the simple case of performer A transmitting audio signals to a remote peer (performer B), the lifecycle of an audio segment will undergo the following processes:

- (a) *Audio acquisition*: Audio gets captured and digitized in small segments by the hardware equipment of performer A (i.e. microphone, sound card).
- (b) *Buffering*: These segments are buffered in small blocks of a predefined size.
- (c) *Packetization*: Each resulting block is integrated with additional data to form a network packet. This additional data is known as *packet header* and it contains network specific information such as the address of the destination of a network packet.
- (d) *Transmission*: Each network packet is propagated to the network through a routing path, which is determined by the instant network traffic, and therefore not known beforehand.
- (e) *De-packetization*: The packet is received at its destination and the audio block is retrieved out of the packet
- (f) *Playback*: The audio block is queued for playback by the hardware equipment (i.e. soundcard and speakers or headphones) of performer B.

Clearly, this sequence of processes is bidirectional, i.e. not only from performer A to performer B but also from performer B to performer A. Each of these processes has a different contribution to the entire communication latency. In common setups, the higher contributions of latency occur during process (b) buffering, resulting in what is called *buffering delay* and process (d) transmission, causing *network latency*.

Buffering delay refers to the time required for acquiring an audio segment from the sound card. The length of this segment corresponds to a certain number of samples and therefore a certain time interval, which depends of the sampling rate. So, for example buffering 256 samples of CD quality audio corresponds to $(256/44,100) s = 5.8$ ms, which, for the purposes of music performance, corresponds to a significant amount of delay.

Network latency refers to the time elapsed for a network packet to reach its destination. The routing path of a network packet is neither known beforehand nor can be controlled. Depending on the actual transmission path, a packet may require a long time to reach its destination, because it may be held up in long queues, or take a less direct route to avoid congestion.

Besides the value of network latency, a more important obstruction in NMP is related to the fact that the different network packets will reach their destination with different delays. Variation in the delivery time of different packets is known as *network jitter*. Network jitter may be due to queuing network packets on different network devices across the transmission path, or due to packets being driven in different routing paths. Since audio playback requires a steady pace, jitter must be eliminated. Reducing jitter in the network requires stable routes, which are generally not feasible on the Internet on an end-to-end basis.

In teleconferencing and VoIP, the total amount of communication latency is known as Mouth-to-ear latency. Speech-based human interaction is highly tolerant to latency, with a threshold of approximately 150 ms [13]. Unluckily, in music performance this threshold is lower by an order of magnitude, which is the main reason why the progress of teleconferencing has not been effortlessly propagated to the music domain.

Since the first years of NMP research, a number of studies are being performed for the purpose of effectively measuring latency tolerance during ensemble performance. For Schuett [14], this objective was defined as identifying an *Ensemble Performance Threshold (EPT)*, or “the level of delay at which effective real-time musical collaboration shifts from possible to impossible”. Schuett observed that musicians would start to slow down performance tempo when the communication delay was raised above 30 ms. This value was further confirmed by a number of subsequent studies [15, 16].

Yet, all of these studies acknowledge that the adaptability of musicians to performing with latency depends on various factors such as their music background, their skills and level of proficiency as well as their age and their familiarity with technology in general [17–19]. Besides the profile of musicians, a number studies show that adapting to latency is highly dependent on certain attributes of the music performed. Such attributes are for example the rhythmic structure [20], the tempo [21] and the timbral qualities of the instruments participating in the music ensemble [22, 23].

4 Audio Quality and Network Throughput

Bandwidth availability or *network throughput* refers to the capacity of the network to accommodate certain data rates. Due to varying load from disparate users sharing the same network resources, the bit rate that can be provided to a certain data stream may be too low for real-time audio communication if all data streams get the same scheduling priority. When the load of the network is greater than its capacity can handle, the network becomes congested. Characteristics of a congested network path are queuing delays, packet loss and sometimes the blockage of new connections.

To date, the majority of NMP architectures use uncompressed audio signals for the communication of remotely located performers. The minimum quality of these signals corresponds to the characteristics of CD quality audio, i.e. signals are sampled at a rate of 44.1 kHz with a sample resolution of 16 bits. In the case of monophonic signals, this audio quality results in a constant data rate of 689 kbps per audio channel. Considering network communication, this is less than the actual data rate required by the network, since, as previously discussed, network packets comprise not only audio data but also header information. Moreover, offering improved music collaboration experience would commonly require higher sound quality as well as multiple channels of audio.

To address this problem a number of research initiatives are focusing on compressing audio data prior to packetization and transmission, hence reducing the required network throughput. Despite the reduction of data rate, this optional step of audio encoding increases the total communication latency for two reasons. Firstly, because it introduces an algorithmic delay caused by signal encoding and secondly because, for efficient reduction of data rates, an audio segment of substantial length needs to be acquired, hence increasing the buffering delay. For instance, the Opus codec [24]

which is currently the de facto standard for real-time audio streaming over IP, recommends a buffering delay of 20 ms for full-band monophonic audio and is associated with a processing delay of 5–65 ms. This is rather prohibitive if one considers the approximate value of the EPT (i.e. 30 ms).

As a result, low latency perceptual codes are increasingly taking into account the requirements of NMP systems. Examples of such works are presented by the Soundjack application, which uses the Fraunhofer Ultra-Low Delay (ULD) Codec [25] and the integration of the WavPack codec by researchers at the Technical University of Braunschweig [26].

5 Audio Dropouts and Data Loss

Finally, *packet loss* occurs when one or more packets of data travelling across a computer network fail to reach their destination. In Wide Area Networks (WAN), packet loss is frequently observed and caused by congested network paths or data corruption by faulty networking hardware across the path. In the case of audio, losing network packets will result in audio dropouts at the receiving end. Audio dropouts correspond to signal discontinuities perceived as glitches, which, in the case of NMP, can seriously hinder the collaboration of music performers. To provide a better insight to the type of distortions caused by packet loss, Fig. 2 depicts a piano signal with severe packet loss occurring during transmission over a ADSL network.

In the domain of network technologies, there are several approaches to recovering from packet loss. For instance, for the widely used Transmission Communication Protocol (TCP), in the event of a lost packet, the receiver asks for retransmission or the sender automatically resends any packets that have not been acknowledged by the receiver. This method of error handling is known as Automatic Repeat reQuest (ARQ). Clearly, ARQ is not an appropriate error correction method for real-time multimedia communications, as in teleconferencing or NMP the packets received after retransmission will be outdated.

Alternative methods for recovering from packet loss include *error concealment* [27] and *Forward Error Correction* (FEC) [28]. Error concealment attempts to

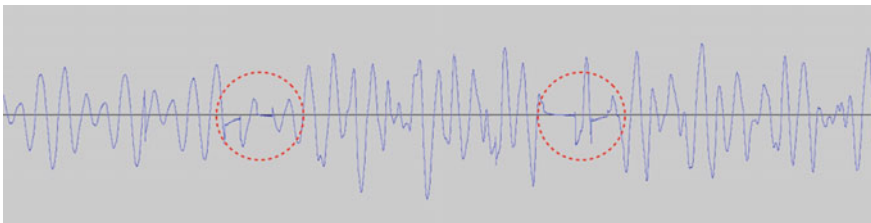


Fig. 2 Heavily distorted piano signal by packet loss during transmission, shown by dashed circles, over ADSL network

recover missing signal portions by using signal processing techniques such as interpolation, pattern repetition or silence substitution. Conversely, Forward Error Correction methods transmit redundant information in addition to actual data packets and attempt to recover losses by reading this redundant information. Information redundancy may be systematic, if it is a verbatim copy of the original data, or non-systematic, if it represents some code that can be facilitated to recover the original data. Clearly, error concealment has the drawback of adding a certain amount of algorithmic latency while FEC methods have the drawback of increasing the requirements in bandwidth availability.

6 Lack of Immersion and Relevant Interaction Affordances

Evidently, communication during music performance does not simply account to musicians listening one another. Performers rely on multiple interaction modalities to synchronise and efficiently communicate. Qualitative studies concerning the requirements of musicians in the absence of co-presence [29], have shown that in addition to sound, visual interactions are necessary for effective synchronization. For this reason, most NMP frameworks employ real-time video, in addition to audio streams, hence further increasing the requirements in network throughput. Video data rates are normally higher than audio ones. However as in music performance, vision is less critical than sound with respect to latency and quality, video communication allows for applying extensive data compression. State of the art NMP frameworks have been experimenting with the use of MJPEG [30], MPEG4 [31] and H.263 [32] video codecs. Evaluation experiments [30] of NMP systems offering video communication in addition to audio, revealed that the main problem in the video communication of performers was the range of viewing angle, rather than the performance of the video codecs. This fact suggests that motor interactions are significant during performance and therefore the positioning of cameras must be carefully considered. Moreover, the use of a data projector as an alternative to the computer monitor can further improve the comfort of musicians.

Besides effective video communication, the use of immersive audio has been considered as an enhancement to performers' feeling of immersion. Specifically, in the work of Chafe [33], a Distributed Internet Reverbarator for Audio Collaboration (DIRAC) was implemented using comb filters to provide the illusion of performers sharing common room reflections of the audio signals reproduced in different locations. The Distributed Immersive Performance (DIP) experiments (Sawchuck et al. [31] used a 10.2 channel immersive audio system so as to represent temporally and spatially distributed audio cues that were created by the interactions of each sound source with the acoustical elements of the environment. Finally, in the MusiNet project the use of spatial audio techniques was considered as an attempt to render the spatial attributes of the audio scene of each performer, thus achieving a more realistic acoustic sensation [34].

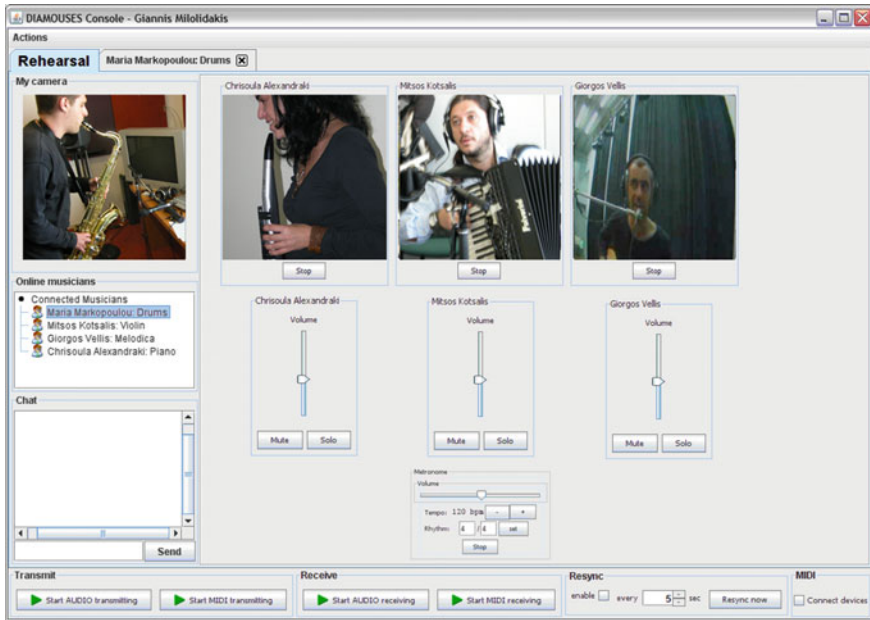


Fig. 3 A GUI developed for the purposes of a networked music rehearsal in the context of the DIAMOUSES project

Yet, a further aspect of research concerns the study of man machine interfaces that can efficiently replicate interaction practices employed by musicians during music performance. Such practices include for example the use of music notation, the use of a metronome or the presence of a conductor. The requirements for employing such concepts or artifacts depend on several defining characteristics of the music performance context. As further elaborated in the section that follows, these characteristics may for example be the kind of music being performed or the purpose of a music performance session. Clearly, different contexts of use raise different requirements in terms of the interaction practices that must be supported. For instance, in remote music-learning scenarios the research focus is on supporting appropriate pedagogical paradigms and on exploring methods for the evaluation of student progress [35]. Alternatively, in network-mediated collaborative music composition [36], a key challenge relates to representing musical events effectively and devising appropriate musical notations [37]. In the context of the DIAMOUSES framework [30], research investigations focused on accommodating diverse user requirements in music performance across different collaboration scenarios such as rehearsals, stage performances and music lessons. Figure 3 depicts an instance of the Graphical User Interface (GUI) developed for supporting distributed music rehearsals. Besides video communication, this GUI provides a chat facility, a metronome as well as the possibility to control the audio levels of individual peers.

Such interfaces integrate the audiovisual communication of musicians with shared collaborative objects permitting synchronous manipulations accessible to all participants, hence maintaining a sense of user focus and promoting a collaborative perspective. Computer Supported Collaborative Work (CSCW) is a focal point of research for several application domains, including e-gaming, e-learning and enterprise groupware, to name a few. In the domain of network-mediated music performance, this perspective has not been adequately investigated.

6.1 Challenging and Feasible NMP Setups

In agreement with the discussion of the previous section, it becomes apparent that the context of music performance severely affects not only the interaction practices employed among performers, but also their requirements in the quality of audiovisual communication. The efficiency of an NMP system to support communication and interaction throughout a networked music session, depends on various parameters that are portrayed in Fig. 4. Four independent parameters determine the feasibility, the requirements and the efficiency of a NMP setup. These are the purpose of the performance session, the characteristics of the music performed, that may be summarized as the music genre, the type of the communication network facilitated and the supported modalities in the communication of performers. Selecting different values for each of these parameters can make up more or less feasible NMP scenarios with varied degrees of complexity.

Fig. 4 Parameters determining the feasibility of a networked music performance session



Out of these parameters, the most prevailing is the purpose of the performance, as it determines the requirements in synchronisation and the required quality of audio-visual communication. Specifically, in the course of a music lesson or a masterclass, musicians will rarely perform at the same time. Usually, the instructor dictates a rhythmic or melodic pattern to be reproduced by the student or the audience of a class. Thus latency is not crucial to the outcome of the session. In contrast, rehearsal or jamming sessions, require a great amount of synchronisation and are highly intolerant to latency. Further possibilities include distributed stage performances and remote recording sessions [7]. In the case of live stage performances requirements will be different depending on whether musicians collaborate from distance, in which case low latency and high fidelity is critical or whether the performance of an collocated ensemble is transmitted to a remote audience. Although, audience feedback forms an essential stimulus for performers, this second scenario is much more feasible as the communication in the direction of the audience to the performers is neither time- nor quality-critical. Hence, this setup is highly feasible and it represents what is widely known as live streaming.

The success of a NMP session is largely dependent on intrinsic attributes of the music performed (e.g. melody rhythm, instrumentation). This was previously discussed as a determining factor to musicians' latency tolerance. Recently, there appears to be debate on whether music genre can reflect any of these attributes [38, 39]. Such considerations have emerged from the recent proliferation of novel music services and applications such as music streaming applications (e.g. Spotify, Last.fm, Pandora) as well as of computational models for automatic genre classification [40]. Consequently, the above schema is susceptible to a great amount of critique.

However, even though musical genre may not readily translate to musical structure or musical style, when considering the act of performance it does translate to the interactions taking place among musicians (e.g. use of notation, leading/accompanying performer intertwining, rhythm interlocking etc.) as well as the occasion for which performance takes place (e.g. festival, celebration etc.). Moreover, music genre may roughly describe the profile of musicians and hence their willingness or motivation to engage in network music performances. With respect to the occasion of performance and the profile of musicians, the two extremes are possibly ethnic/folk music in which there is commonly no scheduling of performance and electronic/electroacoustic performance, which, due to its experimental nature and the intrinsic use of technology, can creatively accommodate network deficiencies, as for example done by Cáceres and Renaud [41] for network delays.

A further parameter of Fig. 4, namely the modalities supported by the NMP session determine the required availability of network throughput. Audio and video streams are more demanding in terms of data rates, while gesture and control data protocols (i.e. those controlling electronic instruments like MIDI or Open SoundControl) are lighter with respect to their data rates.

Finally, the network infrastructure determines the technical efficiency that can be provided during NMP, as well as the geographic distribution of performers. Although the idea of NMP implies that performers are geographically distributed, most NMP research experiments are performed using a Local Area Network in which performers

are distributed in different rooms or buildings sharing the same LAN. This is due to the fact that LAN is more reliable in terms of latency and bandwidth and it practically exhibits zero packet loss. Hence, LAN is more appropriate for studying the attitude of performers towards being physically separated. Moreover, LAN allows for artificially inducing adjustable amounts of latency and packet loss either coded in the NMP software [15] or via using network simulation software such as GNS3 (<https://www.gns3.com/>) [32], therefore investigating the tolerance of musicians to various network conditions.

6.2 Assessing NMP Experience

The experience of participating in NMP experiments is difficult to describe. My personal involvement in NMP research began in 2005. Since then, I have participated in two funded research projects. The *DIAMOUSES* (www.teicrete.gr/diamouses) project that took place during the years 2006–2008 and the *MusiNet* project (<http://musinet.aueb.gr/>) that took place between 2012 and 2015. Both of these projects resulted in the development of technological frameworks enabling NMP. In *DIAMOUSES*, developments were based on the assumption that technology will eventually prove feasible for certain NMP contexts. Thus, the main focus of the *DIAMOUSES* framework was to support collaboration during different NMP scenarios (e.g. lesson, jamming, live performance), hence fostering the formation of virtual music communities. *MusiNet* had a different research orientation. In this framework the focus was to develop a user-friendly ‘Skype-like’ approach for music. Therefore developments were based on state of the art low-latency codecs for audio and video compression, as well as on network protocols and architectures that can provide affordances such as presence awareness (i.e. indication of the status of the user e.g. online, away, busy) and conference calls. Providing further technical details on the development of these frameworks is far beyond the purposes of the present chapter and may be found in dedicated publications [30, 32].

During these years, a number of evaluation experiments were conducted by inviting musicians of different backgrounds to experiment with the frameworks under development. These experiments spanned several combinations of the parameters depicted in Fig. 4 and they had a dual purpose. On one hand to evaluate technical performance of the frameworks under development, the so called *objective evaluation* and on the other hand to evaluate the experience of musicians using these frameworks, namely the *subjective evaluation*.

6.3 Folk Music Experiment

Of the experiments carried out during my involvement in NMP research, the most relevant to this chapter is the Folk music experiment, conducted in the context of

the MusiNet project. The setup involves three musicians performing two pieces of the traditional music of Crete. The following sections describe the pilot setup and the evaluation results, presented as objective measurements and subjective qualities. The findings are inspiring for discussions on the adoption of NMP technology by different cultures, its capacity to foster cross-cultural collaborations and its future role on the development of new music.

7 Pilot Setup

The experiment was conducted in October 2015 in the city of Rethymnon at the premises of *Department of Music Technology and Acoustics Engineering* of the *Technological Educational Institute of Crete*. Table 1 depicts the performers’ instruments and their music background. All three instruments are string instruments that are typical for the traditional music of Crete. Musicians performed two pieces, namely *Protos Syrtos* or *Chaniotikos*, which is a traditional dance piece with the rhythm of 2/4 and *Paradosiakes Kondylies*, which is an improvisation on melodic patterns of Crete performed in 4/4.

The performers were initially situated in the same room to agree upon their performance, as shown on Fig. 5. Subsequently, they were asked to move to different buildings of the Department campus, where the MusiNet client equipment had been setup. The framework was configured to allow communication through LAN using a streaming server that was developed for the purposes of the MusiNet project [32, 34]. This server was receiving the audio and video streams from each musician, which were relayed to the remaining two performers using the *Realtime Transport Protocol* (RTP), which is a protocol that is typically used for media telecommunications.

Table 1 The profile of performers that participated in the Folk music experiment

Musician	Instrument	Music background
Alexandos	Iaóúto (λαοΰτο)	He says that he comes from a ‘musical family’ and he sings since he was a little boy. He plays the piano for ten years and he has a certificate in Byzantine music, theory and harmony
Stratis	Mandolin	He holds a degree in Byzantine music, he is a self-taught musician of traditional Cretan music and he works as a professional musician
Minas	Iýra (λ΄ύρα)	He is a self-taught musician of traditional Cretan music and he works as a professional musician



Fig. 5 Musicians performing at the same site prior to the NMP experiment. The instruments in the order of left to right are Mandolin, Cretan Lyra and Laouto

During network-mediated performance, the hardware equipment of each performer comprised a microphone, a pair of headphones, a camera and a high fidelity sound card connected to a Mac OSX computer which was running the MusiNet client software. As depicted on Fig. 6, the client software provided a GUI which was displaying the video feeds from the other two performers as well as a self-view that permitted proper positioning of the local camera. Audio quality was set to 48 kHz, 32 bit, mono, compressed using the Opus codec and captured/transmitted with a buffering delay of 5 ms. The captured video had a 352×288 pixels display resolution, a frame rate of 25 fps and it was encoded using the H.263 codec. According to these quality characteristics and if no media compression would be applied, the data rates would be 1.46 Mbps for each audio stream and 2.42 Mbps for each video stream. As each MusiNet client was transmitting one audio and one video stream to the server and receiving two audio and two video streams, a minimum of 3.88 Mbps upload rate and a 7.76 Mbps download rate was required for the communication of the three performers, excluding the overhead of the headers of network packets.



Fig. 6 Stratis performing the mandolin and communicating using the MusiNet pilot setup

8 Evaluation

To evaluate the technical efficiency of the MusiNet network in supporting NMP sessions, a number of measurements were captured using a dedicated network analyser software, called *Wireshark* (www.wireshark.org). Wireshark permitted capturing the upload and download network traffic at the location of each network node (i.e. each performer as well as at the location of the MusiNet server). Analysis of this traffic permitted the estimation of average values for latency, data rate and packet loss. These averages are shown on Table 2.

In network experiments, measuring data rates and packet loss is pretty straightforward, as the software estimates the number and size of transmitted and received packets. The most challenging task is to measure network latency and this is due to the fact that computers participating in the communication are not synchronized with the required accuracy which needs to be of order of few milliseconds. One workaround to this problem is to use the *Network Time Protocol* (NTP) to synchronise different computers to the same time.

Using NTP to measure latency, the average value of latency shown on Table 2 corresponds to the total time elapsed after capturing 5 ms of audio, encoding, trans-

Table 2 Average values depicting network traffic during the folk music experiment

	Audio	Video
Latency (ms)	9.3	7.3
Outbound bandwidth (kbps)	17.92	81.06
Packet loss (packets)	0	0

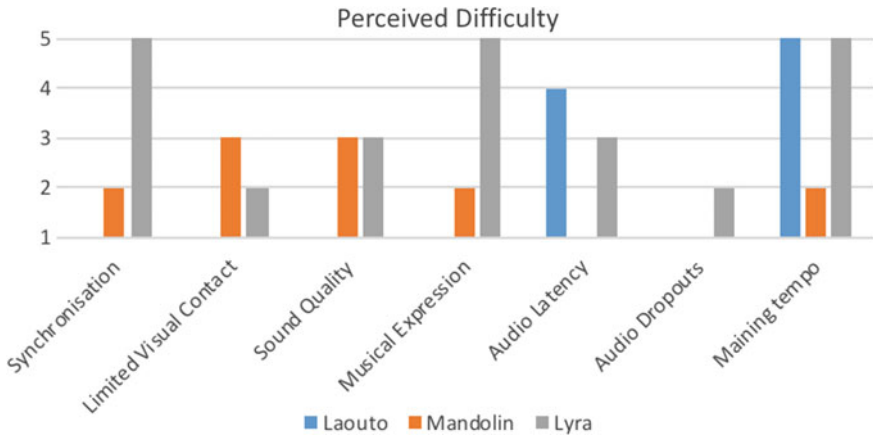


Fig. 7 The perceived difficult of musicians during network-mediated performance. Different aspects of communication were rated in a scale of 1–5, with 1 denoting no difficulty

mitting, decoding and reproducing it to the receiver. Therefore it includes all processing and network delays, apart from the buffering delay, which is 5 ms. The value of 9.3 ms is one third of the approximate EPT value of 30 ms, and should therefore be negligible by performers. As for bandwidth, Table 2 depicts the audio and video rates at the direction of transmission (i.e. the outbound network traffic). These rates correspond compression ratios of 30:1 and 83:1 for audio and video respectively compared to the raw, uncompressed rates that were mentioned earlier. Finally, no packet loss was observed, which was anticipated due to the fact that the experiment was based on a properly configured LAN.

To assess the subjective qualities of the experiment, we interviewed performers and asked them to fill in a questionnaire. Among other questions, the questionnaire asked them to rate the perceived difficulty with respect to different aspects of interaction during performance. This rating is depicted in Fig. 7. It appears that their main difficulty was on maintaining the tempo during performance and, according to the interview, this was particularly perceived when performers attempted to intentionally perform with a very slow or a very fast tempo. They felt more comfortable when performing with a moderate tempo. They reported having synchronisation problems and that they were able to perceive ‘some latency’. Limited visual communication was reported as having a moderate influence on their performance. This can be attributed to the fact that, as video and audio were not precisely synchronised led to participants ignoring the video feeds during performance. Finally, musical expression was problematic, especially for the performer that is less acquainted with audio technology, i.e. Minas performing the lyra, who reported that technology limits freedom in folk music performance. He also mentioned that the amplitude levels of his co-performers were problematic, which made him feel uncomfortable.

To conclude this experiment, it appears that even in ideal network conditions (e.g. those available on LAN) the communication of performers is not adequately

efficient to effortlessly accommodate the ‘fine-grained’ audio-visual interactions of musicians. However, despite the fact that performers were not totally comfortable with the provided setup, they reported that they are willing to adapt, especially when they are unable to meet. In fact, one of these performers, namely Alexandros performing Laouto, expressed a high interest in using this system to teach Cretan music to Greek expatriates. Although, as previously discussed, teaching is relatively more feasible to support, we informed Alexandros that using the MusiNet setup on a commonly available network, would result in high percentages of packet loss perceived as excessive signal distortion. Consequently, his desirable scenario would only be achievable, either by introducing error concealment techniques, hence further increasing the communication latency or over a highly reliable network infrastructure, which can only be provided in academic setups. In conclusion, we suggested that using the MusiNet system to teach Greek expatriates on a common network infrastructure would require a configuration that would not offer any reward to using existing teleconferencing solutions (e.g. Skype).

9 Discussion

Besides technical observations and feasibility considerations, the presented experiment provided insight on the attitude of folk musicians towards the novel practice of performing music while being physically separated. Although professional folk musicians may be familiar with audio technology, for example in course of stage performances or recording sessions, when it comes to remote collaborative improvisation they are sceptic of the efficiency of technology to support music expression. Moreover, as ethnic music typically appears to be performed in the occasion of another activity such as working, eating, drinking or in social events such as weddings, funerals, celebrations, it is difficult to envisage NMP being widely adopted by performers of this style of music. Nevertheless, traditional music should not be considered stagnant, as it does and will continue to embrace technological developments in a myriad of ways.

It is widely acknowledged that today, technological advancements are rapidly embraced by the public at large and used in our daily activities. This is also true for music and music information not only for consumers but also for music artists. Musician-blogs, social networking of musicians and indie-artist friendly streaming services have provided a new paradigm to artist promotion. With respect to ethnic music, this fact has led to the emergence of new genres as well as a constructive criticism for the validity of the existing ones [38, 39]. The genre of *world music*, although heavily criticized for its use by the music industry [42], reflects a global trend for the creation of controversial musical sounds by amalgamating music constructs originating from diverse geographical regions. In the last century, world music, fusion and hybrid genres appear to be one way to foster music controversy, possibly the other one being the development of electroacoustic music. The ever increasing appeal of music artists to these two genres, appears to evolve in parallel with technological

developments. Besides electroacoustic music, which is inherently based on technology, the growth of world music styles can largely be attributed to the fact that musicians from diverse cultures can easily access high-fidelity samples of ethnic music, sound bites and loops as well as the fact that musicians can easily reach one another by means of social networking platforms.

A reasonable extension to this trend can be realized by the advancement of NMP technology. NMP, compared to artist promotion and social networking, provides a dimension of real-time synchronous interactions, hence fostering the immediate and imminent development of new music styles. As already discussed in the introductory section, NMP falls in the category of new technological developments that have not yet been realized as a widely common practice. Indeed, this due to the various technical constraints that were discussed in the previous sections. However, as network infrastructures become increasingly efficient and as more and more research is devoted on improving NMP technology, it is reasonable to expect that at some point in the near future, NMP technologies will be adopted by music communities at large, hence inspiring cross-cultural music collaborations and promoting the fruition of new music.

With respect to improving NMP technology, a distinct line of recent research efforts concentrates on devising computationally intelligent approaches that can allow for overcoming the main technical impediment of NMP, that of latency. Audio signals traversing large geographical distances will always exhibit substantial communication latency, even with the most sophisticated future networks reaching the speed of light. Consequently, a natural workaround to this problem seems to be in predicting music performance ahead of time and rendering the predicted performance to the remote site at a steady pace and before the actual signal reaches its destination [43]. This idea is closely related to a cognitive phenomenon of music perception known as *music anticipation*. Anticipation is a fundamental characteristic of ensemble performance and it refers to the fact that when the members of an ensemble know each other's performing style very well, they know exactly when their peer will play a note in advance, so before the note is actually played. This type of intelligence emanates from different knowledge processes including the cognitive understanding of a performance plan (e.g. the music score or any alternative construct of pre-existing arrangement), the build-up of the music piece up to that time and finally the experienced gained through past rehearsals of the music ensemble. Therefore, one can develop computational models that can be trained according to these knowledge processes. This approach is employed in an alternative research domain, that of *computer accompaniment systems*. There, the objective is to develop intelligent computer agents that are able to replace any of the members of an ensemble performance [44, 45].

Possibly the first work exploiting computer accompaniment systems to enable synchronous networked performance was the development of a system called *TablaNet* [46]. *TablaNet* was a real-time online musical collaboration system for the *tabla*, a pair of North Indian hand drums. Hence, this system was in fact dealing with Indian ethnic music performed over the network. The two drums produce twelve pitched and unpitched sounds called *bols*. The system recognizes the performed bols and

the recognized bols are sent as symbols over the network. A computer at the receiving end identifies the musical structure from the incoming sequence of symbols by mapping them dynamically to known musical constructs. To cope with transmission delays, the receiver predicts the next events by analyzing previous patterns before receiving the original events. The predicted events are synthesized by triggering the playback of pre-recorded samples.

More recent initiatives on predicting sound events during performance and rendering prerecorded signal segments at remote network locations, used score following techniques for monophonic acoustic instruments [47]. Most forms of ethnic music performance employ a great amount of anticipation. For example, monophonic vocal music or percussion music of different ethnic groups may be highly suitable for making short-term predictions and can therefore provide the ground for experimental research on anticipatory models. Ultimately, ethnic music may be influenced by NMP technology as well as provide the basis for the advancement of new achievements.

10 Summary and Concluding Remarks

This chapter attempted to provide insight into current NMP technology and its potential use in ethnic and folk music performance.

The first part discussed the main technical constraints that impede meeting the perceptual prerequisites of musicians during live performance. These constraints are the focal points of NMP research and concern: (a) the minimisation of communication latency which leads to synchronisation problems among performers, (b) the high requirements in network throughput that are necessary to provide high-quality audio and video communications, (c) the elimination of network packet loss which results in signal distortions severely hindering the communication of musicians and (d) the lack of immersion in common ambience with respect to various interaction modalities.

Despite constraints, the chapter elaborates on the fact that for different NMP contexts, these restrictions may be more or less important and that there are certain interaction contexts in which NMP collaboration may be highly feasible. For instance, teleconferencing in music learning is less sensitive to communication latencies, as teacher and student rarely perform at the same time and therefore need not synchronise their performance. Equivalently, remote music recording requires high quality in one direction only, while electroacoustic music may be performed with the use of low data rate control signals. Nevertheless, realistic bidirectional music interactions of acoustic instruments, as for example rehearsals, distributed stage performances or jamming sessions in jazz, rock, classic or folk music are only feasible within short geographical distance and over highly-reliable network infrastructures. Currently, the infrastructures that appear to be more appropriate for NMP are those available to the academia. Hence, true, realistic, bidirectional NMP is restricted to academic environments.

To further inform on the actual practice of network-based music performance and how it can be realised in folk music, the second part of the chapter presented an experiment of three musicians performing pieces from the traditional music of Crete through a Local Area Network. As LANs are private networks that span short geographical distances, they are highly efficient in terms of latency, throughput and packet loss. Although they do not represent realistic remote interaction scenarios (i.e. musicians are located in near proximity), they are highly appropriate for experimental research. This is due to the fact that they allow for studying the attitude of performers towards being physically separated and that, if needed, they permit introducing adjustable amounts of latency, throughput and packet loss therefore monitoring the relevant perceptual thresholds.

The presented experiment revealed that folk musicians were not fully satisfied with the quality of their interactions, however they expressed their interest in compromising with this new technology so as to reach a wider audience, namely teach traditional music to Greek expatriates. Overall, the experience gained through the experiment inspires discussions on cross-cultural music collaborations, enabled by the use of NMP technology. As new technological advancements are rapidly diffused in our everyday lives, feasible NMP scenarios are expected to have a significant impact not only on the dissemination of indigenous music, but also on the emergence of new music resulting from broader cross-cultural collaborations.

Acknowledgements I would like to especially thank Alexandros Aggelakis, Eustratios Gounakis and Minas Sfakianakis for volunteering to participate in the folk music experiment. Part of this research has been co-financed by the European Union (European Social Fund—ESF) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF)—Research Funding Program: THALIS–MusiNet.

References

1. Barbosa Á (2003) Displaced soundscapes: a survey of network systems for music and sonic art creation. *Leonardo Music J* 13:53–59. <https://doi.org/10.1162/096112104322750791>
2. Follmer G (2005) Electronic, aesthetic and social factors in Net music. *Organis Sound* 10:185–192. <https://doi.org/10.1017/S1355771805000920>
3. Kapur A, Wang G, Cook PR, Davidson P (2005) Interactive network performance: a dream worth dreaming? *Organis Sound* 10:209–219. <https://doi.org/10.1017/S1355771805000956>
4. Goto M, Neyama R, Muraoka Y (1997) RMCP: remote music control protocol—design and interactive network performance applications. In: *Proceedings of the 1997 international computer music conference, Thessaloniki, Hellas, ICMA*, pp 446–449
5. Wright M, Freed A (1997) Open sound control: a new protocol for communicating with sound synthesizers. *Proc ICMC 1997*:101–104
6. Xu A, Woszczyk W, Settel Z, Pennycook B, Rowe R, Galanter P, Bary J, Martin G, Corey J, Cooperstock J (2000) Real time streaming of multi-channel audio data through the internet. *J Audio Eng Soc* 48(7/8):627–641
7. Cooperstock JR, Spackman SP (2001) The recording studio that spanned a continent. In: *Proceedings of 1st international conference on WEB delivering of music, WEDELMUSIC 2001*. Institute of Electrical and Electronics Engineers Inc., pp 161–167. <https://doi.org/10.1109/wdm.2001.990172>

8. Gabrielli L, Squartini S (2015) Wireless networked music performance, wireless networked music performance. <https://doi.org/10.1007/978-981-10-0335-6>
9. Rottondi C, Chafe C, Allocchio C, Sarti A (2016) An overview on networked music performance technologies. *IEEE Access* 4:8823–8843. <https://doi.org/10.1109/ACCESS.2016.2628440>
10. Goebel W, Palmer C (2009) Synchronization of timing and motion among performing musicians. *Music Percept* 26(5):427–438
11. Keller P (2007) Musical ensemble synchronisation. In: Proceedings of the international conference on music communication science, pp 80–83
12. Rasch RA (1988) Timing and synchronization in ensemble performance. In: Sloboda JA (ed) *Generative processes in music: the psychology of performance, improvisation and composition*. Clarendon Press, Oxford, pp 70–90
13. Wu X, Dhara KK, Krishnaswamy V (2007) Enhancing application-layer multicast for P2P conferencing. In: Proceedings of the 4th IEEE consumer communications and networking conference, pp 986–990
14. Schuett N (2002) The effects of latency on ensemble performance. Honors Thesis
15. Chafe C, Gurevich M, Leslie G, Tyan S (2004) Effect of time delay on ensemble accuracy. In: Proceedings of the international symposium on musical acoustics, pp 3–6
16. Driessen PF, Darcie TE, Pillay B (2011) The effects of network delay on tempo in musical performance. *Comput Music J* 35:76–89. https://doi.org/10.1162/COMJ_a_00041
17. Farmer S, Solvang A, Asbjørn S, Svensson UP (2009) Ensemble hand-clapping experiments under the influence of delay and various acoustic environments. *AES J Audio Eng Soc* 57:1028–1041
18. Bartlette C, Bocko M (2006) Effect of network latency on interactive musical performance. *Music Percept* 24:49–62. <https://doi.org/10.1525/mp.2006.24.1.49>
19. Chew E, Sawchuk A, Tanoue C, Zimmermann R (2005) Segmental tempo analysis of performances in user-centered experiments in the distributed immersive performance project. In: SMC conference, p 28
20. Rottondi C, Buccoli M, Zanoni M, Garao D, Verticale G, Sarti A (2015) Feature-based analysis of the effects of packet delay on networked musical interactions. *AES J Audio Eng Soc* 63:864–875. <https://doi.org/10.17743/jaes.2015.0074>
21. Carôt A, Werner C, Fischinger T (2009) Towards a comprehensive cognitive analysis of delay-influenced rhythmical interaction. In: Proceedings of international computer music conference. <http://hdl.handle.net/2027/spo.bbp2372.2009.107>
22. Barbosa Á, Cordeiro J (2011) The influence of perceptual attack times in networked music performance. In: Proceedings of 44th international conference: audio networking, 2011, p 10. <http://www.aes.org/e-lib/browse.cfm?elib=16133>
23. Mäki-Patola T (2005) Musical effects of latency. *Swomen Musiikintutkijoiden* 9:82–85
24. Valin, J.-M., Maxwell, G., Terriberry, T.B., Vos, K., 2013. High-Quality, Low-Delay Music Coding in the Opus Codec. 135th AES Convention 73–82
25. Kraemer U, Hirschfeld J, Schuller G, Wabnik S, Carôt A, Werner C (2007) Network music performance with ultra-low-delay audio coding under unreliable network conditions. In: Proceedings of the 123rd audio engineering society convention. New York, Curran Associates, pp 338–348
26. Kurtisi Z, Wolf L (2008) Using WavPack for real-time audio coding in interactive applications, in: 2008 IEEE International Conference on Multimedia and Expo, ICME 2008 - Proceedings, pp. 1381–1384. <https://doi.org/10.1109/icme.2008.4607701>
27. Tatlas N-A, Floros A, Zarouchas T, Mourjopoulos J (2007) Perceptually-optimized error concealment for audio over WLANs. *Mediterranean J Electron Commun* 3:77–86
28. Xiao J, Tammam T, Chunyu L, Zhao Y (2011) Real-time forward error correction for video transmission. In: 2011 visual communications and image processing (VCIP). IEEE
29. Alexandraki C, Kalantzis I (2007) Requirements and application scenarios in the context of network based music collaboration. In: Proceedings of the AXMEDIS 2007 conference. Florence: Firenze University Press, pp 39–46

30. Alexandraki C, Akoumianakis D (2010) Exploring new perspectives in network music performance: the DIAMOUSES framework. *Comput Music J* 34:66–83. <https://doi.org/10.1162/comj.2010.34.2.66>
31. Sawchuk AA, Chew E, Zimmermann R, Papadopoulos C, Kyriakakis C (2003) From remote media immersion to distributed immersive performance. In: *Proceedings of the ACM SIGMM 2003 on workshop on experiential telepresence*. New York, ACM Press, pp 110–120
32. Akoumianakis D, Alexandraki C, Alexiou V, Anagnostopoulou C, Eleftheriadis A, Lalioti V, Mastorakis Y, Modas A, Mouchtaris A, Pavlidi D, Polyzos GC, Tsakalides P, Xylomenos G, Zervas P (2016) The MusiNet project: addressing the challenges in networked music performance systems. In: *IISA 2015—6th international conference on information, intelligence, systems and applications*. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/iisa.2015.7388002>
33. Chafe C (2003) Distributed internet reverberation for audio collaboration. In: *Proceedings of the 24th AES international conference*, pp 13–19
34. Akoumianakis D, Alexandraki C, Alexiou V, Anagnostopoulou C, Eleftheriadis A, Lalioti V, Mouchtaris A, Pavlidi D, Polyzos GC, Tsakalides P, Xylomenos G, Zervas P (2014) The MusiNet project: towards unraveling the full potential of networked music performance systems. In: *IISA 2014—5th international conference on information, intelligence, systems and applications*. IEEE Computer Society. <https://doi.org/10.1109/iisa.2014.6878779>
35. Ng K, Nesi P (2008) I-Maestro framework and interactive multimedia tools for technology-enhanced learning and teaching for music. In: *Proceeding—fourth international conference on automated solutions for cross media content and multi-channel distribution*, Axmedis 2008. Florence: Firenze University Press, pp 266–269
36. Hajdu G (2005) Quintet.net: an environment for composing and performing music on the internet. *Leonardo Music J* 38(1):23–30
37. Hajdu G (2006) Automatic composition and notation in network music environments. In: *Proceedings of the 2006 sound and music computing conference*. Marseille: Centre National de Creation Musicale, pp 109–114
38. Greenberg DM (2016) Musical genres are out of date—but this new system explains why you might like both jazz and hip hop. *Economies*. <http://www.econotimes.com/Musical-genres-are-out-of-date-%E2%80%93-but-this-new-system-explains-why-you-might-like-both-jazz-and-hip-hop-244941>
39. Wong J (2011) Visualising music: the problems with genre classification. *Masters of Media*
40. Ezzaidi H, Bahoura M, Rouat J (2010) Taxonomy of musical genres. In: *Proceedings of 5th international conference on signal image technology and internet based systems, SITIS 2009*, pp. 228–231. <https://doi.org/10.1109/sitis.2009.45>
41. Cáceres JP, Renaud A (2008) Playing the network: the use of time delays as musical devices. *Proceedings of International Computer Music Conference* 244–250
42. Byrne D (1999) Crossing music's borders: 'I hate world music'. *The New York Times*. <https://archive.nytimes.com/query.nytimes.com/gst/fullpage-9901EED8163EF930A35753C1A96F958260.html>
43. Alexandraki C (2014) Real-time machine listening and segmental re-synthesis for networked music performance. PhD dissertation, University of Hamburg. <http://ediss.sub.uni-hamburg.de/volltexte/2014/7100/>
44. Dannenberg R (1984) An online algorithm for real-time accompaniment. In: *Proceedings of the 1984 international computer music conference*. Computer Music Association, pp 193–198
45. Vercoe BL (1984) The synthetic performer in the context of live performance. In: *Proceedings of the 1984 international computer music conference*, Paris, pp 199–200
46. Sarkar M, Vercoe B (200) Recognition and prediction in a network music performance system for Indian percussion. In: *Proceedings of the 7th international conference on New interfaces for musical expression NIME 07*, pp 317–320
47. Alexandraki C, Bader R (2016) Anticipatory networked communications for live musical interactions of acoustic instruments. *J New Music Res* 45:68–85. <https://doi.org/10.1080/09298215.2015.1131990>