NETWORKED MUSIC PERFORMANCE : STATE OF THE ART

ALAIN B. RENAUD[1], ALEXANDER CARÔT[2], PEDRO REBELO[3]

[1] *Sonic Arts Research Centre (SARC), Queen's University Belfast, Northern Ireland, UK*
`alain.renaud@qub.ac.uk`
[2]*International School of New Media (ISNM), University of Luebeck, Germany*
`Alexander.Carot@isnm.de`
[3] *Sonic Arts Research Centre (SARC), Queen's University Belfast, Northern Ireland, UK*
`p.rebelo@qub.ac.uk`

The fast paced development of broadband internet infrastructure makes high quality real time networked music performances possible. In recent years, the implementation of over-provisioned network backbones has led to the development of several initiatives that allow real-time or time-shifted interactions of geographically displaced musicians. The development of such networks is also leading to the creation of new forms of art and communications that use the network as a core.

INTRODUCTION

This paper provides an overview of the various initiatives that are shaping the emerging network performance landscape today. This relatively new area of research examines music production using existing and emerging internet technologies and attempts to demonstrate that the network is not limited to being a platform for the unilateral distribution of digital content; it can also act as a medium for real time high quality bi-directional musical interactions. This includes real time music performances between disparate locations as well as time shifted music production processes that use the internet as a core for interactions. The subject is increasingly being examined as a combination of technological issues such as network latency and other technology-related problems inherent to the network as well as sociological issues. This includes research on the usefulness and the cultural implications of providing network performance frameworks to the general public and the music production community.

1 HISTORY OF NETWORKED MUSIC PERFORMANCE

Musicians and composers, especially in the area of new music, have always been intrigued by the possibility of remote musical collaborations even before the development of the internet. John Cage's "Imaginary Landscape No. 4 for twelve radios" is considered to be one of the earliest networked music performance experiment. The piece, "used radio transistors as a musical instrument. The transistors were interconnected

thus influencing each other" [1]. Although the levels of interactivity were limited to the dialling of radio-stations, gain and tone-colour, the desire to investigate the possibilities of cross-influence in networked instruments is evident in the piece.

It is really the development of computers that led to more realistic networked interactions as computers provided an easy way to transport binary data from one point to another. One of the first groups to experiment network practice with computers was the The League of Automatic Music Composers in the late seventies. The group was originally composed of Jim Horton, Tim Perkis, and John Bischoff started using networked computers to exchange messaging data between each other with the goal of influencing their playing. The group, later renamed as The Hub [2], started experimenting remote collaborations between the west and the east coasts of the US. Due to the limited bandwidth available at the time, the group exchanged messages and not pure audio signals.

Even though there were several initiatives to exchange audio data from one site to another in real-time, it is only the development of high-speed over provisioned internet backbones such as Internet2 [3] in the US that made real time high quality interactive audio links possible. The first initiative to take advantage of the implementation of Internet2 was The SoundWire project, led by Chris Chafe at Stanford's CCRMA [4]. The project developed a set of tools for high quality uncompressed bi-directional audio streaming.

SoundWire was one the first known initiative to take this approach.

Several projects were run in parallel or after SoundWire including the McGill Ultra Video Conferencing research group [5], which introduced low latency uncompressed video along with high quality audio.

2 THE CURRENT SITUATION

Most projects undertaken in the field of real time high quality networked music performance took place around and after 2000 and went very quiet afterwards. In the last year or so, a revival of interest for the subject has emerged not only on the technological side but also on the cultural side, where researchers are seeking to understand the cultural implications of providing such facilities to musicians and producers as well as seeking ways to increase the level of interactions between musicians collaborating over network connections.

2.1 Technology

From the point of view of latency, playing live music as if in the same location requires peers to be separated by a delay less than 25 ms [6][7]. This value is called the EPT (ensemble performance threshold), which is the maximum delay that allows for musicians to play in synchronisation.

Latency is the largest perceptible auditory effect in a networked performance and is caused by several factors. Firstly, the soundcard introduces a blocking delay, which appears on the capturing and the receiving end depending on the actual settings as shown in the table below [8].

Framesize/samples	Rate/Hz	Blocking / ms
2048	48k	42.6
1024	48k	21.3
512	48k	10.6
256	48k	5.3
128	48k	2.6
64	48k	1.3

Fig 1: soundcard blocking delay

Secondly, the speed of fiber (~ c * 0.7) , non direct paths between two locations and the internet itself all introduce delays which allow only a certain EPT coverage if we are in a situation where latency is critical. This is illustrated in Fig.2 with the European example.
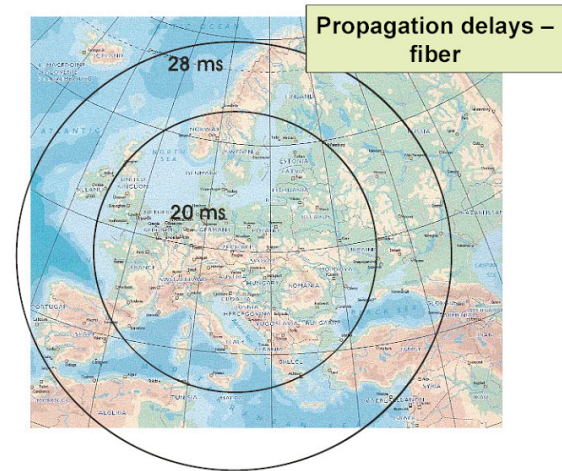

Fig 2: Average propagation delay diameter [9]

Due to various reasons, internet packets are typically not delivered with a constant delay unless special QoS-agreements have been made. The generated network jitter appears with a certain improbability, which can't be foreseen, leading to undesired audio dropouts. Hence a certain amount of packets have to be buffered on the receiving end in order to provide a solid playout of the audio stream. Depending on the number of buffered packets this generates another delay on the top of blocking and propagation delay.

As a result, due to the technical limitations it is often challenging to remain in a situation where the EPT is of 25ms or less.

Based on the latter technical challenges posed by the issues introduced by audio hardware and network problems three approaches for designing networked music performance tools have so far been pursued and classified accordingly:

2.1.1 Realistic Jam Approach (RJA)

This approach is considered when real-time live music interactions are crucial and when the goal is to get as close as having geographically displaced musicians feel like they are playing in the same space. It involves sending and receiving relevant data as quickly as possible as well as ensuring that the network quality between peers is as good as it possibly can be. The following projects are based on this approach:

2.1.1.1 Soundjack:

Soundjack [10] is a standanlone software application developed by Alexander Carôt at the ISNM in Lübeck, Germany and inspired by the SoundWIRE project. The software allows direct access to the soundcard buffer in order to send it to the desired destination via UDP/IP

without any central server in a peer-to-peer (p2p) manner. The quality of the stream depends on the soundcard settings and the network quality. In order to reduce jitter dropouts the audio buffer can be increased manually while GUI elements give information about the significant values such as delays, buffersizes and audio over/underruns.

With standard soundcard settings of 48 kHz, 16 Bit one audiostream requires 768 kbps which prevents it from using DSL lines. For this purpose the Fraunhofer ULD [8] codec has been implemented recently. The codec is able to compress the payload down to a bitrate of 64 kBit/s.
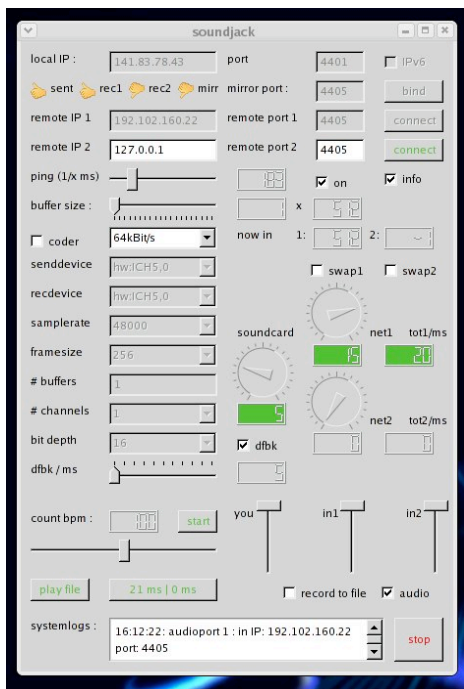

Fig 3: Soundjack Interface

2.1.1.2   eJamming:

In principle, eJamming and Soundjack follow the same idea except that eJamming [11] uses MIDI data which reduces the required bandwidth to a minimum. Since MIDI data is only transmitted when an event is triggered by the user, the system doesn't need to process a constant stream of data which makes it less vulnerable to jitter. In case jitter appears, the sent MIDI packet is simply played out later. For packages arriving far too late the user can adjust a threshold over which the signal is not played out anymore. This parameter is called the "Late-Note-Tolerance".

In order to make sessions, exceeding the EPT of 25ms, playable, eJamming introduces the idea of delayed feedback [12] which delays one's own instrument sound

for an amount of time in order to get it closer to the incoming sound of the external peer. This parameter is called the "Instrument-Feel" and can be set to tight, regular or loose depending on the short, medium or long distance scenario.

Apart from the technical background and basic principle eJamming has an advanced GUI with instrument rack, player information and other useful features.

2.1.2   Latency Accepting Approach (LAA)

This approach considers the internet as a decentralized and space independent medium and thus connecting globally, network delays of more than 200ms are common and perfectly acceptable.

Accepting these delays beyond the EPT and finding new ways of delayed musical interaction is an alternative approach to the realistic jam approach. The following projects are based on this principle:

2.1.2.1   Ninjam:

According to its website [13], the Novel Intervallic Jamming Architecture (Ninjam), puts emphasis on musical experimentation and expression and tries to establish a jamming environment under the assumption that network latency prevents true realtime synchronization of the participating musicians. To circumvent this assumption, Ninjam's creators introduce a communication principle in which latency is increased in such a way that participating performers receive each other's output with the delay of at least one measure. In fact musicians play asynchronously to the music their colleagues have played at least one measure before, which made Ninjam's creators call this principle "faketime".

In comparison to Soundjack and eJamming, Ninjam uses a central server for the session control and the delaying mechanisms. The basic idea of delaying audio streams by measures requires Ninjam to use a metronome in order to have a central control time instance.

2.1.2.2   Quintet.net:

Quintet.net is an interactive network performance environment invented and developed by composer and computer musician Georg Hajdu [14]. It enables performers at up to five locations to play music over the internet under the control of a "conductor." The environment, which was programmed with the graphical

programming language Max/MSP consists of five components:

• Server
• Client
• Conductor
• Listener
• Viewer add-on

The players interact over the internet by sending control streams to the server either using a pitch-tracker, MIDI or simply the computer keyboard. On the server, the streams get copied, processed, and sent back to the clients as well as to the listeners. In addition, a conductor can log onto the server and control the musical outcome by changing settings remotely and sending streams of parameter values as well as short commands to the players.

Quintet.net uses a sampler or MIDI for instrumental playback. It also features granular synthesis as well as vst~ plug-ins for sound processing and playback, and has additional video and graphical properties, which permit better interaction and control on a symbolical level: The performers along with the audience see the music which the participants produce on screen in "space" notation on five grand staves. In addition video clips and/or live video can be displayed by the viewer add-on and mixed with real-time music notation for an enhanced viewing experience. The conductor can also read musical scores and send parts to the performers, which will be displayed along with the notes produced by the musicians.

The music performed with Quintet.net is typically a combination of composed and improvised elements. The lack of real synchronicity due to the usual delays on the net, necessitates the adaptation of a genuine "in-ternet" performance style for which John Cage's number pieces could be considered a model: These pieces require certain notes or phrases to be played within "time brackets." [14]

2.1.3 Remote Recording Approach (RRA)

This approach involves producing music by using the internet as a medium for remote recording sessions. In this case the audio signal sent to is "time stamped" which makes it possible to ignore latency issues as there is no real human-to-human interactions.

2.1.3.1 DML:

The Digital Musician Link (DML) [15] is a plugin for sequencers compatible with the VST 2.0 standard (Steinberg VST). DML is the successor of the former ResRocket Network.

To collaborate, two logged in DML users agree on a session, one acting as the playback/recorder (A), the other acting as the performer (B). A assigns B a track in his production and starts the recording process. B receives a stream of the mix that A prepared and plays his specific track to the session, as if situated in a recording booth in A's studio. B's DML instance assigns each recorded sample a time stamp and sends it back to A who's DML instance sorts the received data and puts it into the assigned track. The playback of A's mix doesn't start for A until B's data is received, so for both participants, an environment similar to that in a recording studio and booth is maintained. The possibility to connect a webcam and a talk back channel rounds up the simulation. Before being able to connect, users need to authorize with a username and password. Depending on what payment model the user has chosen, a connection with low, medium or high quality can be established between two users. Fig.4 shows a screenshot of a Nuendo session with the DML launched.
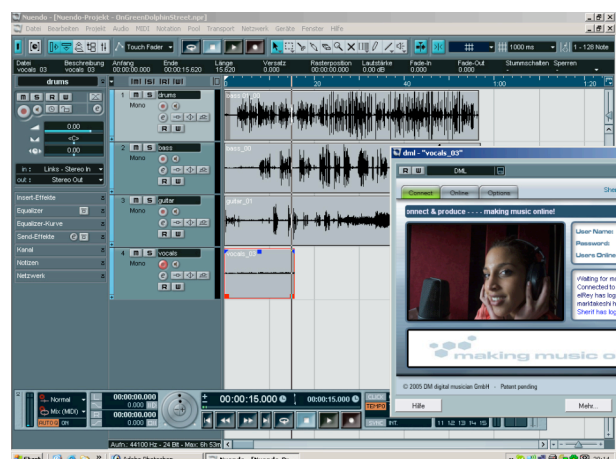

Fig 4: Nuendo session with DML-Plugin

2.1.3.2 VSTunnel:

The VSTunnel Plug-In [16] is used like an insert effect in a VST compatible sequencer's master out channel. When the plug-in is launched, the user can start a session or to join another session. When users create their own session, they decide whether to make it public or private. A list of all public sessions is offered to the user. It can be sorted by genre, speeding up the search for music of taste. By clicking on a session, it can be pre-listened. By a further click, a session can be joined.

During a session the music played is analyzed by the VSTunnel Plug-in. Local changes are recognized, compressed and transmitted to the other session

participants. In this instance the audio is mixed and sampled accurately into the audio out. Therefore every participant hears a mix of their own work and the work of others. This mix can be adjusted by each participant. The volume of each "remote line" can be set (ie. muted, etc). Text communication is achieved through the use of an integrated chat system [16].

2.2 Network Culture

The development of internet technology at the consumer level such as blogging sites where users can share opinions online as well as new types of "open source" licensing schemes for music such as Creative Commons [17] have recently led to the advancement of new collaborative cultures which use the network as a medium for exchanging creative materials in an electronic form. Several sites, which use the Creative Commons licensing schemes are allowing musicians to share, exchange and remix musical material, which remains in the public domain for free and generally forever.

Such initiatives are leading to the expansion of social networks, while compressing both time and space through the use of the internet as a medium for dynamically exchanging digital information. The logical step to take forward is to allow online communities to collaborate in real time and this is where networked music performance systems have a place. For example, recent developments in the online game environment Second Life [18] demonstrate a significant interest in the use of sound and music as a way of defining avatar identity. Second Life's "residents" stream audio to their virtual spaces and host events such as live concerts.

The cultural implications of network performance systems are potentially significant for the way music is created as they could lead to the creation of "virtual bands" which can collaborate in real time from various locations on actions such as composing, rehearsing, and recording.

However, using the network as a medium for interaction doesn't stop at the level of attempting to replicate situations that exist if the physical world as the network can also be used as an acoustic and social medium for which music is composed for. As Castells points out, "the network links up specific places, with well-defined social, cultural, physical, and functional characteristics" [19]. This specificity becomes of central importance when thinking of activities such as music. In the same way that the history of music is intrinsically linked to places and societies, the network provides a cultural condition of unprecedented complexity.

2.3 Interaction

Interaction is a real issue in network performance systems as natural visual or sensory cues, such as breathing and gesture, are completely removed from context. There has been several attempts to add video to real time bi-directional audio streams, however even though some experiments have been relatively successful, some basic interaction components such as the potential for common pulse, the synchronisation of events that require negotiation and agreement from two or more performers, the identification of performative roles and the up-beats or the ability to anticipate various perfomative actions are very difficult to re-create in a networked environment.

Research into how these types of musical interaction can be rendered in electronic music has been pursued through the development of a network tool The Frequencyliator [20], which is seeking to re-create those elementary interactions by imposing a basic structure and dynamic mechanisms to a distributed ensemble of computer musicians. The implementation of the system has proved successful in promoting exchange and negotiation of musical events in the context of live computer music. This is in part due to the hybridisation between the network tool (in this case the Frequencyliator) and the computer instruments themselves.

A purely acoustic networked ensemble presents a far more challenging situation. If we take the example of a string quartet playing pre-composed music, we will quickly notice that even though the notation is a common point of reference for the ensemble, human interactions that don't rely on a centralised score are numerous and almost impossible to render in a networked environment. Therefore, a significant amount of research needs to be conducted in order to identify what types of basic interactions are essential for acoustic ensembles to reach a performative state in a geographically displaced situation. Once the performance cues are identified, a strategy on how to communicate interactions across the ensemble will need to be defined.

Although it is clear that in a standard performance situation the types of interaction between musicians are inherently multi-modal (aural, visual, proprioceptive) it is likely that in a networked condition sound becomes the most critical source for shared performance cues due to its immediacy, specificity and for being after all the mode that musicians are trained to develop.

3 FUTURE PLANS

In terms of technological challenges, the Realistic Jam Approach (RJA) is the most challenging kind of networked performance tools to design. Although technologies such as VoIP has some specific requirements in terms of QoS and realtime network capabilities, RJA has much higher requirements to minimise delay, jitter and packet loss. Therefore the technical aspect of research is going towards implementing systems with smaller data packets, reduced delay time, large overheads and extremely low jitter tolerance. Implementing the latter will lead to reducing the latency as well as making sure that the quality of transmission remains stable throughout a performance.

Existing projects based on the LAA and RRA approaches are already attracting online communities and often work well as they are less dependant on real-time communication and optimal network conditions. Many of the existing LAA and RRA systems are likely to focus on increasing usability and user friendliness in order to attract a wider community.

Finally, in order to further understand interaction modes between musicians in performance situations, a study is being carried out at the Sonic Arts Research Centre (SARC), which attempts at simulating network conditions in order to observe their impact on performance. An understanding of rehearsing and performance at a professional level is essential for the design of interactive strategies, which can become useful in a networked environment.

4 CONCLUSIONS

When looking at the amount of projects, research ideas and publications in the field, topped up with new initiatives launched by the audio industry, it is clear that networked music performance has a future. However, as demonstrated in this paper, there are several technical and cultural barriers that need to be overcome in order to bring networked music performance in the mainstream but obviously a general acceptance can be observed.

The classification of various types of networked performance frameworks (Realistic Jam Approach (RJA), Latency Accepting Approach (LAA) and Remote Recording Approach (RRA)) is a good start as it allows the developments of different scenarios depending on the performers' needs.

A specific community of researchers and artists is being developed around those frameworks, which makes it

encouraging and certainly gives a future to networked music performance.

It is also clear that network performance is a prime example in the use of a technology for purposes to which it was not originally intended. The appropriation of network infrastructure for a new distributed and multi-spatial type of music-making provides opportunities for a significant change in the way we relate to music. With such aims, our approach requires an understanding of technical, social, cultural and musical issues, which pertain to this unprecedented type of performance. This paper presents the beginnings of such work.

REFERENCES

[1] Pritchett, J. The Music Of John Cage. Cambridge University Press, Cambridge, UK, 1993.

[2] Brown, C and Bischoff, J, Hub Origins http://crossfade.walkerart.org/brownbischoff/. Accessed on October 20, 2006.

[3] Internet2 homepage http://www.internet2.edu/. Accessed on October 19, 2006.

[4] Chafe, C, Wilson, S, Leistikow, R, Chisholm, D and Scavone, G, A simplified approach to high quality music and sound over IP. In Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00), (Verona, Italy, December 7-9,2000).

[5] McGill Utravideo Conferencing page. http://www.canarie.mcgill.ca/. Accessed on October 22, 2006.

[6] Carôt, A. "Diplomathesis: Livemusic on the Internet," FH Lübeck, May 2004

[7] Schuett, N. "The Effect of Latency on Ensemble Performance," CCRMA Department of Music, Stanford University, May 2002

[8] Carôt, A, Krämer, U, Schuller, G. "Network Music Performance (NMP) in Narrow Band Networks" presented at the 120th AES convention, Paris, France, 2006 May 20-23

[9] Rønningen, L. A., "Adaptive Video Resolution and Traffic Control in Packet Networks," Guir, UC Berkeley, March 2004

[10] Carôt, A, Renaud, A and Verbrugghe, B.,

Network Music Performance (NMP) with Soundjack. Networked Performance Workshop Paper, Presented at the 6th NIME Conference June 8, 2006, Paris, France

[11]   eJamming website at www.ejamming.com, accessed on October 24, 2006

[12]   Barbosa, A M, Phd Thesis "Displaced Soundscapes", MTG Barcelona, April 2006

[13]   Ninjam Website at www.ninjam.com, accessed on October 24, 2006

[14]   Hajdu, G. "Quintet.net – A Quintet on the Internet," Proceedings of the International Computer Music Conference, Singapore, 2003

[15]   Digitalmusician.net website at www.digitalmusician.net, accessed on October, 24th 2006

[16]   VSTTunnel Description at www.vstunnel.com, accessed on October 24, 2006

[17]   Creative Commons http://creativecommons.org, accessed on October 22, 2006

[18]   secondlife.com/. Accessed on October 24, 2006

[19]   Castells, M. (2000) The Rise of the Network Society. Vol. 1 of The Information Age: Economy, Society and Culture, Mass. Blackwell

[20]   Rebelo, P. Renaud, A. (2006) The Frequencyliator – Distributing Structures for Networked Laptop Improvisation, Nime 2006