Network Audio Performance and Installation

Atau Tanaka
NetFIVE Ltd. Tokyo, Japan
atau@ccrma.stanford.edu

Abstract
This paper describes a series of performance works in the area of network music. The work presented spans a period of time and a range of different infrastructures and technologies. As network technologies continue to evolve, we can draw from these experiences insight into musical issues surrounding the practice of music on networks.

1. Introduction

The work presented represents a period of time of five years, 1994-99. The network infrastructures used include ISDN telephony and Internet. Transmitted data includes control data, audio, video, and graphics. Despite this range of different approaches, musically we have held to a specific model: concert performance. The projects were live performances connecting remote locations where the network connection arrived onstage. Although network topologies afford the possibility of a broader conceptual re-thinking of musical presentation [1], a deliberate choice was made to preserve the established model of concert performance and to investigate the implications of applying network technology in this context.

Three types of connection methods are discussed there – point-to-point audio/video, IP based control data, and IP based streaming audio.

2. ISDN Videoconferencing

The first body of work consists of a series of concerts using videoconferencing technology over ISDN telephone lines. Data bandwidth was typically 128kbps, over which the H.320 protocol [2] was employed. Off the shelf videoconferencing hardware and software were used, typically to create point to point connections. Concerts realized in this manner include connections between Paris-New York, Madrid-Tokyo, Tokyo-Paris, Tokyo-New York, Tokyo-Dijon, Barcelona-Amsterdam, and Barcelona-Montreal.

These events used bi-directional transmission of compressed audio and video signals between two points (*fig. 1*). In several cases, there were exceptions: Twice there were three point connections established (Japan-France-Australia, and Quebec-Quebec-Holland), and on two other occasions there was the use of an additional MIDI data line (Tokyo-Brooklyn, Barcelona-Montreal).

Although it can be argued that a two point connection does not constitute a network, important lessons in remote musical performance were gained in these experiences. They include issues of delay, synchronization, and musical congruence. It was observed that two types of data delay exist: codec processing delay and network transmission delay. Codec processing delay is constant, and dependent on processor speed. Transmission delay on a dedicated line connection such as ISDN telephony is quite consistent once the connection is established but varies with each connection. This changes considerably with a migration towards IP networks.

One feature of of the integrated videoconferencing systems used in these concerts were the synchronizing of image and audio transmission. In other areas, it became quite clear that this off the shelf equipment was not designed for musical applications. Careful adjustment of certain encoding and transmission parameters, however, maximized the systems' musicality.

3. W Client/Server

Another series of work was realized using the Max environment [3] and the W Protocol [4]. In this series of performances, called **LaserOscillators**, there was no audio nor video data transmitted. Instead, Max control messages were transmitted over a TCP/IP connection to multiple sites. Recent concerts have included Montreal-Utrecht-Stockholm and Lisbon-Utrecht-Tokyo. The W Protocol implements a client/server system akin to a chat system. The client runs as a Max object in a patch at each concert site, with the server program running under Unix. Each client connects to the server and joins a channel (chat room). Data from each client is transmitted to the server, which then relays the data to all connected clients.

Time delay is different for each client-server connection, depending on the network

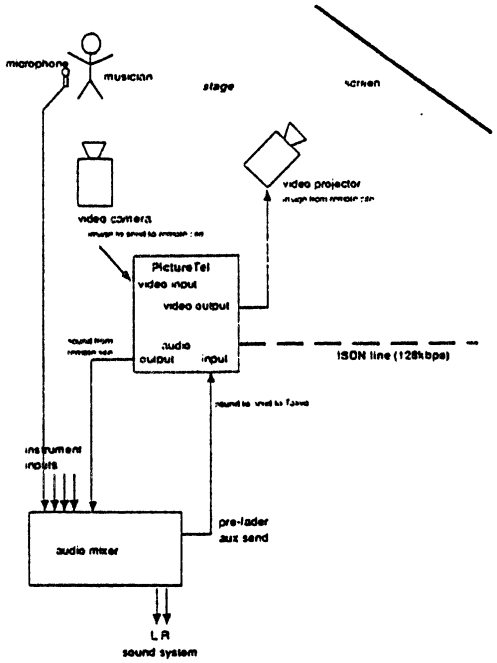

fig. 1. One site in an ISDN concert

"distance" and congestion from each client to the server. TCP data transmission packetizes data, and was observed as data bursts and clustering of musical control data.

As all transmitted data was control information, sound synthesis is done locally. Thus a remote player's actions are represented via control signals, and synthesized local with respect to his partner (*fig. 2*). Within the Max program, a text typing interface was created to allow verbal communications among the performers. This verbal data was interleaved into the musical control stream.

4. MP3 Streaming

The third type of network concert realized made use of live audio streaming using MPEG I Layer 3 format. A concert was realized connecting Paris-Hamburg-Vienna-Tokyo. At each site were installed both an MP3 streaming encoder/server [5] and multiple MP3 clients. The servers stood by, ready to serve streams upon receipt of requests from clients. Once a stream was initiated, the music from that stage was served to the remote site. Each site received streams from each of the other sites.

Time delay and synchronization in this case becomes a multifaceted problem. In addition to the encoder latency and network transmission time was the added layer of server latency. The encoder formats incoming audio to MP3 format, and redirects the data to the server. In most cases this is simply another process running on the same machine, but the encoder and server could be two separate machines in two separate network locations. The server relays the stream received by the encoder. At each request from a client, it serves a new stream. It is unclear what the time offset is between different parallel streams are. On the client end is another layer of time latency – the client typically buffers up to 28KB of data before it starts playing.

In this concert, separate channels were established for audio, visual and verbal communication. Webcams provided visual indication of what was happening at each site, and Internet Relay Chat [6] provided a means for written communication between stages. The fact that these were separate processes happening on independent streams had interesting consequences. The most apparent was that the visual images were not synchronized to the audio stream.

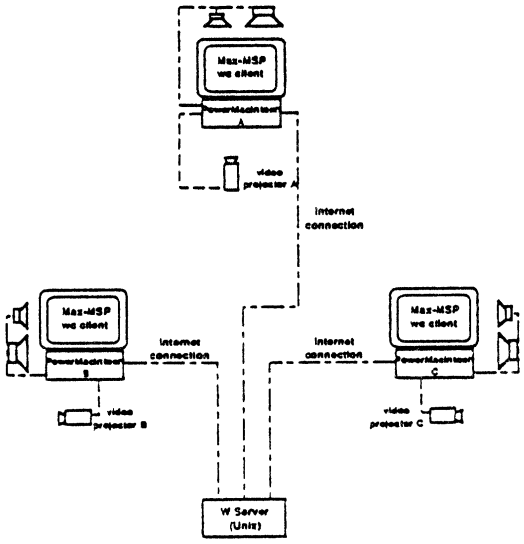

fig. 2. Sensorband's LaserOscillators configuration

5. Discussion

Performing music on these three different systems yielded vastly differing musical results – much akin to playing music on different musical instruments. The most radical case is with LaserOscillators, the performance using the W client/server. As the transmitted information was reduced to just data, the choice of sound materials also became reduced, to pure sine waves. The control signal passed via the W server were frequency controls for the single oscillator representing each site, detunings of which caused beating patterns creating acoustic resonances at each concert site. The choice was not due to technical limitations, but was rather a musical decision based on the nature of the medium.

Despite the differences among these three approaches, some common elements emerge. The most compelling characteristic of network

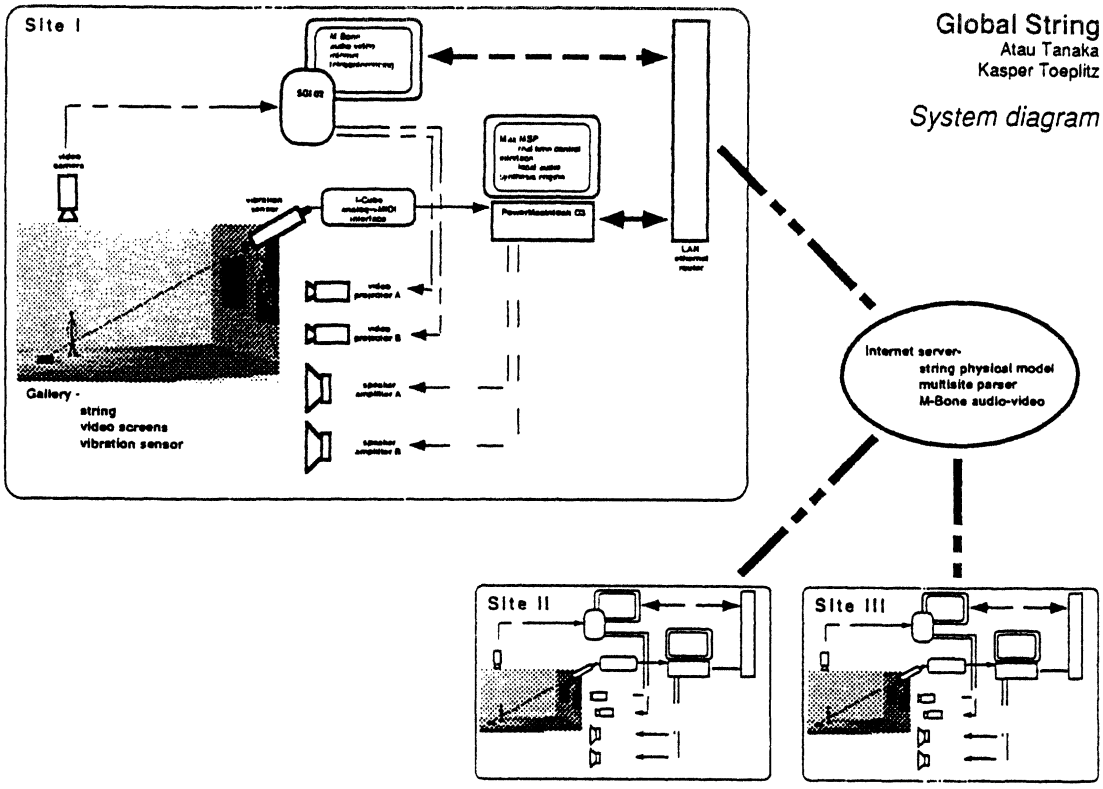

fig. 3 Global String (Kasper Toeplitz, Atau Tanaka)

based performance is time delay. The instinctive reaction of a musician is to try to improve the system by minimizing this delay. One can argue, however, that this is a misplaced motivation. The time delay can instead be regarded as the "acoustic" of the network, much in the way the reverberation time in a cathedral differs from that of a jazz club. Musically accomodating this delay may lead to the creation of a music that is idiomatic to the medium [7].

The fact that delays are often different along each leg of the connection means that relative synchronization of musical elements differ at each site, particularly when dealing with multi-point connections. This leads to an interesting phenomenon of non-congruous simultaneity. Whereas the real time nature of the performance implies one music, the result is heard quite differently at each site, creating the interesting musical dynamic of a single music with simultaneous multiple interpretations.

The different types of connection contributed insight into the role of visual communication among the performers. With regard to image quality, fluidity took primacy over image resolution. The image component's contribution was effectively nullified unless the image was synchronized with the audio. These are concerns dealing with the simple case of direct visual communication extending the notion of eye-eye contact in performance. Other, more abstract applications of visual imagery include a performance using a live network score in which this author took part [8], as well as visual abstractions of musical gesture [9].

Combinations of the different techniques above yielded interesting results. To realize a three point ISDN connection, the hub architecture native to the videoconferencing system was not musically appropriate – rather than mixing the audio from the multiple sites, the system implemented an algorithm that favored the site who was broadcasting the loudest. Instead, in the three-way ISDN concerts, we assigned one of the sites to be a hub. At this site were two independent ISDN systems, making a separate connection with each of the other sites. The local audio and video was then mixed with the incoming audio/video from one site to be transmitted to the other. This added an extra layer of complication to the time delay of reception at the endpoints.

We also experimented with the use of both control data (in the form of MIDI) alongside ISDN audio/video. As the data was sent under separate independent channels, there was no guarantee of synchronization. However, the tangible nature of seeing remote MIDI events manifested locally (especially on an instrument

like a Diskclavier MIDI piano) while seeing that event be articulated over video gave the listener a real perception of distance. Hearing that same gesture over the audio connection created interesting visual/aural/midi echoes.

Communication between the performers outside the performance itself is essential. In the videoconferencing connections, a significant amount of connection time is spent at the beginning and end in greetings and confirmations. In the W or MP3 connections, this is replaced by the typing/chat interface. In the case of the W connection, the patch used interleaved the typing in the musical control stream. This meant that if the connection went down, that it was dropped both for verbal as well as musical communications. In the MP3 connection, since the audio streamer was independent of the chat software, troubleshooting was facilitated so long as the net connection itself stayed up. In all cases, the ultimate safeguard was to have a standard telephone connection reserved for verbal communication alongside the musical connection.

6. Current work

Current work is branching out into several directions: higher bandwidths, new streaming software, and non-performative applications. The Audio Engineering Society (AES) has published a whitepaper [10] outlining future possibilities for audio transmission on Internet2. Custom MP3 encoder/decoders are being programmed [11], including streaming audio objects for MSP. UDP offers interesting advantages in musical data transmission that decrease packetized data clumping.

New projects include applications in interactive installations and distributed performance environments. **Global String** (*fig. 3*) is a network based musical instrument with physical components based at each end point [12]. Server based audio synthesis is affected by visitor data from sensors, and is streamed back to each site. A separate channel implements IP based videoconferencing and data visualization. **Netradio Feedback** is a shared performance space, organized like a distributed radio studio. Audio harvested from radio broadcasts and webcrawlers are remixed and re-transmit between sites as well as to the webspace at large. Lessons culled from the previous concerts are being applied in these new projects.

References

1. **GNUsic: An Open Studio on the Network for Electronic Musicians.** http://www.gnusic.net
2. **Recommendation H.320 (07/97) - Narrow-band visual telephone systems and terminal equipment.** http://www.itu.int/itudoc/itu-t/rec/h/h320.html
3. Miller Puckette, David Zicarelli. **MAX – An Interactive Graphic Programming Environment.** Opcode Systems. 1990.
4. David Zicarelli. **The W Protocol: A System for Collaborative MAX Patches on the Internet.** 1996.
5. Shoutcast Streaming Server http://www.shoutcast.com/info.html
6. Internet Relay Chat http://www.newircusers.com
7. Atau Tanaka. **Netmusic – a Perspective.** Catalogue, Festival du Web, Webbar, Paris. 1999.
8. **Finding Time** http://www.turbulence.org/Works/ftime/
9. Vibeke Sorensen, Miller Puckette, Rand Steiger. **Global Visual-Music.** http://www.visualmusic.org/gvm/gvmjs.htm
10. AES Technology Report TC-NAS 98/1: **Networking Audio and Music Using Internet2 and Next-Generation Internet Capabilities.** 1998.
11. Atau Tanaka. **Network Music Laboratory Research Project Report.** Keio University internal document. 1999.
12. Atau Tanaka, Kasper Toeplitz. Global String http://po.ntticc.or.jp/~sband/atau/string

Acknowledgements

These concerts were realized with the sponsorship of: PictureTel, Fujita Sanken, France Telecom
And the participation of:
Japan - Keio University SFC, IAMAS, Cyberia, CyberOz, Milk
Europe – Webbar, Atheneum, IECA Festival Aye Aye, TransEurope Halles, V2, Sonar, Art Futura
U.S.A. – HERE, Dumbo
Canada - Centre Méduse, Studio 303, FCMM

Participating artists: Les Virtualistes, Pita, Farmers Manual, Sensorband, Camel Zekri, Sheldon Steiger, Koji Asano, Gil K, Kasper Toeplitz, Starbait, Masami Sakaide, Koichi Makigami, Zack Settel, BuyoBuyo Igor, Tom Mays, Erik Montgomery, Eve Beglarian, Danny Blume, Takahiko Suzuki, Kenji Yasaka, Shunichiro Okada, Tito.