The Technophobe and the Madman: An Internet2 Distributed Musical

Robert Rowe* and Neil Rolnick†

*New York University
robert.rowe@nyu.edu
†Rensselaer Polytechnic Institute
neil@neilrolnick.com

Abstract

This paper reports on the composition, technology, and performance of The Technophobe and the Madman, a collaborative work commissioned by the New York State Council on the Arts and performed simultaneously at Rensselaer Polytechnic Institute and New York University over an Internet2 connection. The report is framed by a discussion of some issues surrounding distributed performance and concludes with documentation of the work and a consideration of its lessons for interactive multimedia using Internet2.

1 Introduction

The Technophobe & the Madman was a collaborative work specifically developed for simultaneous performance at two sites connected by an Internet2 communications link, making it the first Internet2 distributed musical. Funded by a New York State Council on the Arts (NYSCA) grant to New York University, Rensselaer Polytechnic Institute, and Harvestworks, The Technophobe & the Madman was conceived and realized through a collaboration of three composers (Nick Didkovsky, Neil Rolnick, and Robert Rowe), a video artist (Don Ritter), two writers (Tyrone Henderson and Quimetta Perle), and a theatrical director (Valeria Vasilevski). The performing forces consisted of two singer/actors and four musicians, with one singer/actor and two musicians located at each of the two venues.

Technically, Technophobe required continuous two-way streaming of six channels of video and twelve channels of audio between New York City and Troy, New York. That is, three video streams were sent from Troy to New York and three others were sent the other way. Similarly, six channels of audio (stereo tracks accompanying each video channel) were sent in each direction.

Figure 1 shows the technical setup in the house for the New York site of the performance. Five computers are in evidence: flanking the mixer are computers manned by Nick Didkovsky (left) and Robert Rowe (right), both running real-time compositional software; immediately behind the mixer to the left are two computers that monitored the quality of the Internet2 connection between NYU and RPI; and behind those are the video switches and computer for Don Ritter’s generation, mixing, and manipulation of the video signals being routed to the three projections onstage (figure 3).

Figure 1. Technical setup for New York site

The Technophobe project required guaranteed, uninterrupted connectivity of approximately 20 Mbps (megabits per second) between Troy, New York and New York City for several rehearsals and the concert. In keeping with the general manner of operation of Internet2, the delivery of that connectivity was arranged by direct contacts with systems administrators along the entire data path.

The data sent between the two sites consisted of three channels of video information in each direction, and six channels of audio each way, for a total of 6 video and 12 audio channels transmitted within the 20 Mbps connection. The realization of these connections was handled by a commercially available codec, the Vbrick 3000.

The Vbrick 3000 codec uses MPEG-1 compression to send 30 field per second, 1/4 frame video as well as two channels of audio over an internet connection. There must be a separate device at each node of a multisite performance, and a connection with no dropouts requires bandwidth of 1.5 to 3 Mbps each way. Consequently, Technophobe required six Vbricks, three at each site. A

Proceedings ICMC 2004
more recent device, the Vbrick 6000, uses MPEG-2 compression to send full video and stereo audio over a 6–10 Mbps connection (www.vbrick.com).

2 Latency in distributed performance

Interactive multimedia requires three stages of processing: in the first, sensors register and transmit some representation of the performance of one or more human and/or machine player(s). Second, a computer program analyzes the output of the sensors and prepares an algorithmic response based on that information and its own internal state. Finally, instructions emanating from the analysis/composition stage are used to control the generation and playback of audio, video, and other signals.

All of these stages take time. The longer they take, the greater the delay between the human player’s actions and the computer system’s response. Beyond the processing necessary for each stage, the time needed to transmit information between groups of machines performing each task can become significant as well.

Figure 2. Video image created by Don Ritter for The Technophobe & The Madman. The vision of Alma, avatar of the technophobe, is a manipulation of a photograph of Valeria Vasilevski, theatrical director of the production.

2.1 Latency tolerance

In a recent paper, Nathan Schuett attempts to identify the Ensemble Performance Threshold, or EPT, which is defined to be “the level of delay at which effective real-time musical collaboration shifts from possible to impossible” (Schuett 2002). Musical collaboration here is used in the traditional sense of the term, according to which performers play together with the audible and musical sense that what they are doing is happening at the same time. Schuett’s methodology was to observe the performance of two musicians trying to perform rhythmic patterns in synchrony, e.g. by clapping simple two-part rhythms together. The musicians were separated either in time (by introducing an electronic delay into the transmission of sound from one player to the other) or in space. “The direction of the tempo was a very useful indicator of whether a performance was being hindered by the effects of latency. If the delay was greater than 30 msec, the tempo would begin to slow down. This gives a solid indication that EPT for impulsive, rhythmic music lies between 20–30 msec” (Schuett 2002).

3 Processing

Interactive multimedia processing latencies include those due to the operating system, and those arising from the nature of the processing performed. Brandt & Dannenberg (1998) documented end-to-end audio latencies on various operating systems running from a minimum of 2 ms (Windows 98, cpu load 60%) to a maximum of 821 ms (Windows 95, cpu load 90%). Beyond the limitations of the OS, computation of audio can introduce significant delays as well. Users of software synthesizers and programming environments such as Max/MSP (Winkler 1998), a standard for interactive multimedia applications, can set a buffer size for successive blocks of audio processing. This size is often set to 512 samples when computing at a sampling rate of 44100 samples per second, introducing a latency of approximately 11.6 milliseconds to the system (1000 ms / 44100 samples per second * 512 samples in the buffer).

3.1 Distributed performance

Transmission rates over the Internet range from a 28.8k bps telephone dialup line to the 100 Mbps or more available on high-speed broadband connections. Clearly, a 28.8k bps telephone connection is too slow to keep up with the 705.6k bps needed for even one channel of CD-quality digital audio. Even when a transmission channel with sufficient theoretical bandwidth is used, signals going into and coming out of the link must be buffered to compensate for network congestion between the two machines. Depending on the nature of the signals being sent and the quality of the transmission channel, these buffers may range anywhere from 15 to 3000 milliseconds or more.

Given all of these potential sources of delay, it becomes apparent that current Internet technology is unlikely to deliver the 20 ms throughput required to meet Schuett’s EPT. Accordingly, artists using the Internet for distributed performance have found models other than traditional ensemble coordination to realize their aesthetic goals. Jörg Stelkens has developed a peer-to-peer model for networked performance that makes use of latency as a parameter of synthesis (Stelkens 2003). Users logging in to Stelkens’s application peerSynth are presented with graphic interfaces representing each connected node. A high-resolution ping process constantly measures the communications latency between a user’s machine and all other connected nodes. The measured latency is then used as a control value for synthesis algorithms running on the user’s computer.
“The sound of a fixed synthesis unit is modified through further signal processing in such a way that a sound with relatively high latency values is heard as ‘further away’ and with low values as ‘closer by’. This corresponds to the experience of acoustic musicians who play spatially separated in a large room or outside (‘alpenhorn effect’). Sound propagation occurs with a corresponding delay and each musician perceives the music that is being played together differently” (Stelkens 2003). Here the audible character of the sound is changed by the speed of the connection between any two participating nodes. Rather than minimize or camouflage the time delays between networked computers, that very element becomes a governing factor in the production of the output.

Stelkens’s work is an example of adapting compositional strategies to the nature of distributed performance, rather than trying to find ways to realize traditional ensemble playing across a latency-filled network. Atau Tanaka argues for a similar aesthetic position: “Transmission delays will be considered a hindrance as long as we try to superimpose old musical forms onto the network. Rather, a new musical language can be conceived, respecting notable qualities of the medium, such as packetized transmission and geography independent topology. These features can be said to define the ‘acoustic’ of the network, a sonic space that challenges existing musical notions of event, authorship, and time” (Tanaka 2000).

4 Internet2

Internet2 is a research consortium of universities, industry, and government agencies. The purpose of the consortium is to develop protocols and tools for guaranteed, high-bandwidth transmission of data across the Internet. Such guaranteed quality of service has the effect of reducing or eliminating the buffers needed to compensate for network congestion as described above.

One obvious use of fast, reliable connections is the real-time transmission of multimedia content, and several recent projects have explored just that possibility. A demonstration at the Audio Engineering Society conference in 1999 sent surround sound audio from a stage at McGill University to a movie theater at New York University (Xu et al. 2000). Mara Helmuth's SoundMesh application was used to mix audio files in a live Internet2 improvisation (Helmuth 2000). The SoundWIRE project at Stanford's Center for Computer Research in Music and Acoustics (CCRMA) is concerned with developing techniques for streaming high-quality audio across Internet2 and assessing its reliability (Chafe, Wilson, & Walling 2002). In 2002 this led to a networked jam session between musicians at Stanford and McGill.

The purpose of the NYSCA commission of *The Technophobe and The Madman* was to explore the artistic consequences of composing for the medium of Internet2. One expression of this exploration was an extensive rehearsal period that spanned several months of trials across a connection between the two sites. Repeated encounters with the realities of distributed performance had a profound effect on the piece as it was being developed.

4.1 Lessons learned

The latency of signals sent from one site to the other in our application was roughly 250 milliseconds. The largest part of this delay was due to the MPEG-1 video compression/decompression performed by the Vbrick 3000 interfaces. A more important measure is round-trip latency. Consider the following scenario: a musician plays in New York, and that sound arrives in Troy 250 ms later. Another musician in Troy hears that and plays some sound in response. The audience in Troy hears both together, but the audience (and performers) in New York hear the Troy contribution 500 ms after the first sound was played in New York. This has some immediate artistic consequences: first, the audiences in Troy and New York are not hearing the same thing; and second, it is virtually impossible in such circumstances for musicians at the two sites to play rhythmic material in synchrony.

![Figure 3. Stage configuration of New York site during performance of *The Technophobe & The Madman*](image)

A related issue is feedback: site A has live microphones that are being sent to site B. Site B is playing the feed from site A over its local sound system. Site B also has live microphones that are being sent to site A. The live microphones at site B pick up the transmission from site A coming over the local loudspeakers and send it back, and vice versa. The result is a persistent echo effect that is actually feedback delayed by the system latency. In our application this was on the order of 500 ms, producing a repeating echo that sounded like a decaying delay line.

Such echoes are a well-known problem in video conferencing and similar applications. Solutions developed there, however, are not appropriate to distributed performance. In particular, the video conferencing technique of dynamically muting all inputs but one works well when speakers are taking turns but cannot be applied when several
musicians are all playing simultaneously. Similarly, feedback suppressors work to notch out particular resonant frequencies, not the broadband echo effects suffered by Internet performances (Galanter 2002).

The feedback issue was successfully addressed by virtuosic live mixing. An extensive period of mixer normalization between the two sites was required before each rehearsal and the final performance, to ensure that signals would not begin to distort at the remote site when levels were adjusted during performance at the local site. Further, mixing engineers at each location followed the score to mute every microphone that was not in immediate use during the show to minimize the occurrence of feedback echoes.

The composition of the piece similarly reflected the realities of a distributed realization. One technique was the use of "floating progressions," a continuously varying trade-off of leadership in which one site would initiate a change of harmony (for example), followed by the other site when that change became locally audible. At times these changes of leadership were scored on a measure-to-measure basis, so that neither site appeared to its own audience to be simply following along, karaoke-style, with an external source.

At other points the performance differed between the two sites. Particularly when the material at one location or the other was scored to be based on a pulse, the following strategy was used: strongly pulsed material was performed at site A, and sent to site B. At site B, the local performers could perform in perfect synchrony with the remote beats, even though this was happening after a delay relative to site A. To prevent interference with the pulse at site A, the return from site B was simply muted. This is an extreme case of two different performances happening simultaneously: the audiences at the two locations heard dramatically different contents at that moment in the composition. A critical extension, then, was the full video documentation of the performances at both sites. These records are used as the basis of web and DVD renditions of the piece, in which viewers can choose their own path through the material — focusing on the New York rendition of the pulsed section, or the Troy version, or creating some other sequence combining the two.

The totality of the piece is not represented fully to the audience in either venue: they see different versions of the piece, and later viewers who tune in to view it on the web get a third, different view. A DVD document of the piece shows still another perspective. Using the camera-selection feature of DVD video, users can select their own path through the material by remote control. In some ways, the work is like a sculpture: it is multi-dimensional, though at any one time you can only have a single view from a single perspective. Yet, to fully experience the piece, you have to view it from a variety of perspectives, like walking around a sculpture.

The experience at either of the live sites was much like a traditional concert. The sound and video quality of signals from the local and remote sites was virtually indistinguishable, and the extended rehearsal period led to a performance in which the two forces meshed musically and dramatically. It is in the consideration of the work as a whole, encompassing the variants presented at each site and its continued existence in multifaceted web and DVD versions, that the unique qualities of a distributed performance are best realized.

5 Acknowledgments

The Technophobe and the Madman was created by a commission from the New York State Council on the Arts to Harvestworks, Inc., Rensselaer Polytechnic Institute, and New York University. Thanks to Carol Parkinson of Harvestworks for her work on the production. Photographs of the performance were taken by Chia-nan Yen. Philip Galanter, Jeff Bary, Igor Broos, Keith Engel, Tom Beyer, Jamie Forrest, Stephan Moore, Tom Doczi, Rachelle Menshikov. Don Neumuller and many more lent invaluable technical assistance.

References


