Living with Net Lag

Chris Chafe[1]

[1]*CCRMA, Stanford University, Stanford, CA, 94305, USA*

Correspondence should be addressed to Chris Chafe (cc@ccrma.stanford.edu)

ABSTRACT

Internet latency can be ignored, tolerated or exploited in making music together at a distance. The premise that these are distinct ways of relating to lag is examined in case histories of recent projects. Scores of examples of split ensembles collaborating remotely in real time can be cited from the last few years. Five scenarios from this musical world have been selected to look closely at music being made across networks and the differing relationships to lag. Each involves large or multi-site ensembles. The first four represent academic/contemporary idioms (involving jacktrip software and advanced university networks) and the last is a working band which uses commodity tools to rehearse pop music covers (using jamLinks and standard residential networks).

1. INTRODUCTION

Playing music together in real time across networked locations presents a whole list of challenges and primary among them is signal latency (or lag). Different kinds of music have different relationships to lag either ignoring, tolerating or embracing it. Challenges also include difficulties in the video realm and how to best solve audio mic'ing, monitoring and mixing needs. As an acoustic performer, your sound at the other site(s) can only be as good as what comes through those remote loudspeakers.

Presented here, are some experiences with different kinds of music created over distance. These brief case histories are followed with some concluding musical and technical comparisons.

Net lag wouldn't exist as a topic without advances in network audio media. A number of latency-optimized streaming solutions have been engineered specifically for music. This class is distinct from other streaming technologies engineered for e.g., teleconference, voip or internet radio (all of which have greater lag). Some background on the development of the underlying streaming technology is useful in situating the case histories technically.

The following briefly describes the *jacktrip* software project and jamLink, a related hardware device, which were the audio streaming solutions employed in the cases cited. Both send audio peer-to-peer and require wired ethernet. Wireless will eventually achieve the essential QoS, but at present it does not provide equivalent performance. For testing future wireless solutions, the jacktrip application can play a role. *SoundWIRE* is a tool for evaluating QoS that is built into jacktrip and in fact, is the reason jacktrip was originally conceived.

1.1. SoundWIRE / StreamBD / jacktrip

SoundWIRE is a networking utility for auditory monitoring of QoS. It reveals fine-grained behavior of a roundtrip path, similar to the *ping* and *iperf* utilities but different by virtue of how it exercises the network. The tool is a network-based synthesizer (not a network-controlled synthesizer, but one in which audio flows between synthesizer components across networks). Its various components are situated remotely from one another. The synthesizer system can provide an intuitive way to assess QoS on networks capable of supporting low-latency, interactive applications like telesurgery, high-definition video conferencing and musical collaboration. The synthesizer uses audio rate feedback across a bidirectional connection to create a "plucked string" physical model synthesis ala the Karplus-Strong algorithm [1, 2] . By "plucking" or exciting the recirculating network audio signal, a tone is created which reveals latency, jitter and loss as pitch, vibrato and distortion.

The sound is exactly the same as a physical model synthesis computed in real time on a software or hardware synthesizer. The only difference is that the delay memory component is the network itself. Or rather, it's the audio packets which are flying on the network such that delay

"length" is the time it takes for a signal to make a round trip. Short network lag creates a higher pitch and very long lag ($> 50ms$ round trip) gets into very low pitches and even echoes.

An uncompressed, low-latency audio streaming engine was required to realize the SoundWIRE network synthesizer. Its first version engine which was able to stream in both directions was called *StreamBD*. In addition to supporting development of the SoundWIRE algorithm, StreamdBD found immediate use in one-sided and two-sided concerts. The next version was written to connect three sites. This new architecture incorporated the jack-audio-connection-kit [3]. Two interconnected bi-directional instances of the engine were run at each site to support the first "jack-connected triple-site" performance and was named "jacktrip" [4]. The software is an open-source command line application compiled for host computers which can support low-latency audio. Users of jacktrip must configure their soundcard, network address and ports in a manner consistent with their remote partners.

1.2. jamLink

MusicianLink, Inc. provides a system featuring core elements of jacktrip's functionality but with an emphasis on plug-and-play setup and ease of use. The jamLink hardware box was introduced at the 2010 NAMM show and is engineered for connecting up to 4 boxes at a time. These standalone devices use the company's backend service and user accounts to simplify connection creation. Users plug in their audio (instrument / talkback mic / headphones) and network (typically ethernet from their home router). A javascript control GUI on a web page provides the means to select remote partners by name and adjust audio balance. Network configuration is transparent to the user (the backend service uses the STUN protocol for port discovery and NAT traversal).

2. CASE HISTORIES

Both jacktrip and jamLink have supported a large amount of online music in the last few years. The following case histories illustrate the attraction of the medium. When distance is no object, we have the ability to make music in new ways and with new partners. Richard Scheinin of the San Jose Mercury News wrote this about the first concert discussed below (the concert pieces variously adapted to, tolerated or ignored lag):

> "But as was the case in the '60s, what happened Tuesday night was about more than the music. It raised basic questions: What does it mean to "be here," when here is there, and there is here?" [5]

2.1. Pacific Rim of Wire Concert (2008)

Cities Stanford, Beijing

Audio/Video jacktrip/DVTS

Notes IPv6

An online Stanford University / Peking University concert was held during Stanford's annual Pan-Asian Music Festival in April, 2008. The audience was located at Stanford's Dinkelspiel Auditorium. Performing at the venue were the Stanford Laptop Orchestra, several acoustic musicians from the Stanford Symphony Orchestra plus myself on celletto (an electric cello) and computer soloists. The audience was involved in singing for the last piece. The Beijing performers included a chamber ensemble and a computer performer.

Improvisation Telematica

Hong-Mei Wu is an accomplished erhu soloist whose musical interests extend beyond the tradition. The two of us planned a remote improvisation for the concert and checked out each others' playing briefly via a Skype audio connection. We sensed each others' musical direction and felt good about going for it. I received a DVD of her classical performance work to have something to practice with but that was it in terms of concert preparation. The next time we met was when we began the improvisation performance. As the recording shows [6], it began in a pentatonic sound world but quickly branched out to chart new territory. Despite the lag, but capitalizing on the ability to hear very well, we were able to trade "musical messages" and anticipate each others' "musical moves." Two computer accompanists (Caceres and Gremo) added subtle textural enhancements.

My impressions is that we were were "in the zone" together despite lag on the order of $120ms$ one-way. This was not the same zone that would have existed if we were on the same stage. Video was transmitted but unimportant in communicating. Instead, we listened with "bigger ears" for every possible nuance and musical idea coming from the other. What we explored and where we went

musically tolerated the long latency, which was simply a quality of our "stage."

When jacktrip is starved for incoming packets, it loops the last valid audio buffer. The looping cycles at a rate which is determined by the number of frames-per-period as set in jack. When this happens, it creates a wavetable synthesis tone whose pitch is $sampleRate/framesPerPacket$ (e.g.,$375Hz$ at $48kHz$ with$128fpp$). The tone's waveform depends on the most recently-arrived packet. A long network stall will create a drone and sporadically-choked incoming packets elicit a sound like a synthesizer with an intermittently-varying wavetable. Our path from China experienced both drone and intermittency effects but not so severe as to interrupt the music. These occurrences were actually embraced in the improvisation.

In C

Terry Riley's *In C* was adapted for synchronization using the lag as a source for pulse timing. This realization is described in detail elsewhere [7].

Tuning Meditation

Oliveros' *Tuning Meditation* is perfectly suited to distributed ensembles and will totally ignore lag. The score calls for listening and responding such that events / entrances unfold slowly [8]. Most important is the ability to hear well and appreciate the aggregate texture as it evolves. At one instant, jacktrip was starved for packets (as above) and sounded it's "starvation pitch." Singers keyed into the pitch and incorporated it, influencing the subsequent tonality.

Video streaming used DVTS [9] and firewire DV cameras, one-per-site. The audience watched a full-sized image above the musicians. Players in China appeared "larger-than-life" and slightly behind the audio as far as gestural synchronization.

The Chinese portion of the network required IPv6 for connectivity. We created a jacktrip version with IPv6 sockets and requested IPv6 routing from Stanford Network Services. This was the first time ever routing IPv6 at Stanford and was implemented with tunneling.

2.2. *The Thing* – TeleJazz Concert (2009)

Cities Banff, Toronto

Audio/Video jacktrip/iChat

Notes 16-channel multitrack recording

Three "TeleJazz" concerts were produced during the Spring, 2009 Banff Centre Jazz Workshop [10]. One of the organizing principles was to vary the remote sites over the three concerts to experience different amounts of distance lag. We formed bands including Banff musicians connected with players first in Calgary, then Toronto and lastly, worldwide. A second principle was to do "live-to-multitrack" recording of each concert with recording producers running the show at each site.

One of the compositions on the second concert is detailed here. Don Cherry's *The Thing* was performed with a 12-piece band split between Banff and Humber College in Toronto. It's an up-tempo piece which begins with an accelerando.

Lag between the two sites was $> 30ms$ one-way. We transmitted 16 channels in each direction and both sites made 32-track recordings (local + remote). Video was iChat.

Mic'ing of individual instruments was professional quality. The right combination of mic and position was carefully tuned for each player and room layouts (both sides) were prepared in advance to maximize track isolation (using specific mic patterns and gobo's). We also minimized air lag by using the closest possible placement of the remote players' monitor speakers to the local band. Close proximity is always better for lag but creates a tricky trade-off with the need to reduce feedback paths.

The theoretical lag limit before measurable tempo drag sets in is in the region of $30ms$ based on clapping studies [11]. Above the limit, (and this path was somewhat above it), various "coping strategies" kick in. One is a leader-follower strategy in which one side dominates the tempo and the other follows. But in a jazz context with flexible rhythmic give-and-take, such a strategy constrains the music.

Our one rehearsal of *The Thing* suffered enormously from excessive lag. Players had difficulty keeping time together and particularly the opening accelerando suffered. But mysteriously, not in the concert performance.

As can be heard on the recording, it worked. So. something changed, but what? My supposition is that player's attention shifted from "network lag difficulty" to "concert mode music making" once an audience was present. Despite the distance, the band was able to ignore the lag and play beautifully.

2.3. *Rock, Paper, Scissors* – ResoNations Concert (2009)

Cities Banff, New York, Seoul, San Diego, Belfast

Audio/Video jacktrip/Access Grid

Notes 5-way mixing hub

Sarah Weaver, Mark Dresser and the World Association of Former United Nations Interns and Fellows (WAFUNIF) have produced two annual telematic concerts for peace. The first, ResoNations 2009, involved a large ensemble distributed across five cities. I contributed a game-based composition, *Rock, Paper, Scissors*, to the program [12]. The piece exploits long video lag.

Four of the sites included a rock-paper-scissors game player and transmitted a closeup view of their playing hand to all other sites. During the piece, game players entice other game players into playing a match together and local musicians follow their local player. Section-by-section different sets of rules specify what gets played by the instrumentalists in response to the game playing gestures and game outcomes. The game is suspended during tutti sections where gestures involving all game players are used conduct the entire ensemble. Video lag is a feature of the piece. It enhances the uncertainty faced by the game players in determining their next move.

Video was Access Grid [13] which itself has a venue concept which allows for multiple cameras to stream into its "venue" from each site. Display producers for the actual concert venues were able to pick and choose from the streams reaching the access grid venue. The concert was also webcast from the Banff site and a single video camera was used (whose operator improvised the shots in real time).

The audio hub (located at Banff) involved a star configuration of bidirectional, stereo connections from the hub to each of the other four sites. At the hub, a separate mix was prepared for each outgoing signal (consisting of all other sources). A complete mix was provided for the webcast, plus separate mixes for the audience and stage monitors.

Difficulties with jitter on the path from Banff to New York (due to the UN's different network) required a very long playback buffer at the input to the UN jacktrip host. The **-q** argument to jacktrip was set to 20 buffers, fixing the problem but adding approximately 45*ms* to the lag on the path.

2.4. Chopper – CNMAT / CCRMA Exchange Concerts (2011)

Cities Berkeley, Kansas City, Belfast, Stanford / Stanford, Kansas City, Belfast

Audio jacktrip

Notes change of topology for two concerts

A second composition created with lag in mind is *Chopper* [14] for three remote saxophones. It consists of a soundfile played in the primary location where it "conducts" the local soloist and is mixed into the feed to the networked players. *Chopper* has been performed with baritone sax as lead soloist, tenor or soprano at one location and alto at the other. The combination can vary.

The piece belongs to a series in which computer-generated soundfiles provide the context for improvisation. Improvisers "learn" the soundfile. In this case, the lead sax player also conducts the remote players with one simple rule: a short staccato burst from the lead should be echoed as soon as possible by both remotes (pitches ad lib). The remotes additionally add other sounds such as occasional long tones.

Human reaction time (to a trigger like a tongue slap played by the lead) is on the order of hundreds of milliseconds and quite variable. The staccato call and response and its variable reaction time is a central device of the piece. An expectation of the effect builds up quickly as the piece gets going. Response times include both the relatively constant network lag and the much more variable reaction time. The "variable-ness" of the latter dominates and effectively masks the network lag.

The piece is, however, sensitive to the overall responsiveness of the remotes. The local soloist wants to feel they can "play" the call and response game tightly, as an extension of their own gestures. Too much lag or too little lag can interfere with this feeling. One performance

suffered simply from too much added air lag. The local soloist was hearing the remotes via monitor speakers located at the other end of a large stage (adding air lag on the order of 45*ms*). The slight added "sluggish-ness" made the response feel awkward. There is also a lower bound where it would be difficult to achieve the same effect with no lag and only reaction time, for example, if all players were closely positioned in the same room.

Two concerts in different primary locations with the same players were held back-to-back (Berkeley and Stanford). In the former case, the soundfile was streamed from Stanford to the soloist in Berkeley. The mobility of location provided by networked audio allows for different production topologies with an equivalent result – essentially, the "piece is the place" and actual place is simply a portal.

2.5. Pop Fiction (2011)

Cities Bay Area (Novato, Sacramento, Walnut Creek and either Oakland or San Jose)

Audio jamLink

Pop Fiction [15] is a Bay Area band with an ever-increasing song list. They rehearse online and don't physically play together until their next gig. Lag is tolerated and they rehearse with a metronome on the drummer's side. "Portals" to the rehearsals are studios, bedrooms and garages in which bandmembers' jamLinks are connected to residential networks. Four-way sessions can be up and running in a matter of a couple minutes. Pop Fiction is a professional band, has been using the tool for 3 years and has total dependence on it for their work. The furthest point they connect to is Los Angeles, sometimes needed because of other gigs which bring their players there.

The largest drawback has been bandwidth availability. In order to connect 3 remote sites to a given jamLink, approximately 3 Mbps is required both upstream and downstream. They use residential cable networks which are generally sufficient, but bandwidth is shared both inter-home (nearby customers) and intra-home (other computers in the house). When QoS deteriorates, the band throttles down the audio sample rate (from 44.1 to 24, 22 or even 16*kHz*). Half sample rate requires half the bandwidth.

Video is not used but is imagined for the future, especially for teaching substitute musicians.

3. CONCLUSIONS

Musical choices and technical properties of the medium are intertwined. Across the five projects some of these dependencies are common and some unique. The following wrap-ups detail the overall list of design choices and concerns.

3.1. Musical Wrap-up

Projects ignoring net lag attempt to be played equally whether over the net or in the same room. There's either a good fit in either situation or a "best effort" musical approach. An example of the former is Oliveros' Tuning Meditation and example of the latter is the Tele-Jazz project. In such cases, the music is not changed and no strategies are adopted. If the music is tightly synchronous and the lag approaches the "hairy edge," results may differ from one attempt to the next depending on the performers' state of mind, as mentioned in details about performance of *The Thing*. Musicians are amazingly adept at adapting to adverse acoustical conditions and in that case, did so somewhat unconsciously.

Examples of consciously tolerating lag or adapting the music to the condition are *Improvisation Telematica* for the former and *In C* and Pop Fiction for the latter. *Improvisation Telematica* and *In C* were performed with long ($> 100ms$) latency one-way. In the case of the improv, lag was a quality of the acoustic medium in which the music emerged. *In C*, written before the Internet, seems almost to have had the Internet in mind. The composition is based on a synchronizing pulse at the eighth-note. Parts are played in heterophony with single pulse (out-of-phase) shifts musically valid. Pop Fiction tolerates latencies of a much lower magnitude but which still require a strategy for keeping the beat synchronous. A discussion of strategies at different lags can be found in [11].

Music written with a medium's properties in mind are idiosyncratic. Truly "network music" would not work in another setting. Two examples are *Rock, Paper, Scissors* and *Chopper* which both intentionally exploit lag. *Rock, Paper, Scissors* capitalizes on video lag and *Chopper* depends on audio lag.

3.2. Technical Wrap-up

The above cases used either jacktrip or jamLinks for audio streaming. Video solutions were variously iChat, Skype, DVTS, and Access Grid. Special technical setups noted above included IPv6, audio mixing hubs, 16-channel transmission, and multiple topologies. The best

audio was in the TeleJazz project in which dedicated audio producers were included on the team. Odd effects occurred when jacktrip "packet starvation" sounds affected the music. Different audio mixes were variously required according to the project, including distinct mixes for individual outgoing sends, webcast audio, audience PA, stage monitoring, and recording tracks. Compromises that were noted due to QoS issues included very long jitter buffering and reduced sample rate. Added air lag was noted as a problem in an earlier performance of *Chopper*.

4. REFERENCES

[1] K. Karplus, A. Strong, "Digital Synthesis of Plucked-String and Drum Timbres" Computer Music J. 7(2): 43-55, 1983

[2] C. Chafe, S. Wilson, D. Walling, "Physical Model Synthesis with Application to Internet Acoustics," Proc. 2002 Intl. Conference on Acoustics, Speech and Signal Processing, Orlando, 2002

[3] Jack-audio-connection-kit http://jackaudio.org/files/docs/html/index.html

[4] J-P. Caceres, C. Chafe, "JackTrip: Under the Hood of an Engine for Network Audio" J. New Music Res. 34(3): 183-187, 2010

[5] R. Scheinin, "Laptop Concert Linking Stanford and Beijing Signals World Has Changed," (review) Mercury News, April 30, 2008

[6] Improvisation Telematica https://ccrma.stanford.edu/~cc/shtml/2008chduo.shtml

[7] J-P. Caceres, R. Hamilton, D. Iyer, C. Chafe, G. Wang, "To the Edge with China: Explorations in Network Performance," Proc. ARTECH 2008, Porto, 2008

[8] P. Oliveros, "Networked Music: Low and High Tech," Contemporary Music Review 28(4-5): 433-435, 2009

[9] DVTS http://www.internet2.edu/communities/dvts/

[10] TeleJazz https://ccrma.stanford.edu/~cc/shtml/2009telejazz.shtml

[11] C. Chafe, J-P. Caceres, M. Gurevich, "Effect of temporal separation on synchronization in rhythmic performance" Perception 39(7): 982-992, 2010

[12] Rock, Paper, Scissors https://ccrma.stanford.edu/~cc/shtml/2009rps.shtml

[13] Access Grid http://www.accessgrid.org/home

[14] Chopper https://ccrma.stanford.edu/~cc/shtml/2011chopper.shtml

[15] http://www.popfictionlive.com/