Peter F. Driessen, Thomas E. Darcie, and Bipin Pillay
Department of Electrical and Computer Engineering
University of Victoria
Victoria, BC Canada V8W 3P6
peter@ece.uvic.ca
{tdarcie, bipin}@uvic.ca

The Effects of Network Delay on Tempo in Musical Performance

Internet-based collaborative networking applications, such as instant messaging, voice-over-IP telephony, and social networking, have displaced traditional communication services and redefined social interaction. The Internet has also transformed the music industry, revolutionizing the way music is distributed and marketed. Yet despite these two powerful trends, the intersection—where collaboration and music meet in online music-making—has remained merely a curiosity. Why? Artistically pleasing online audio collaboration requires network delay less than that encountered typically in the Internet. The bandwidth required for high-quality audio exceeds the bandwidth that is generally available on consumer-oriented broadband access (cable and digital subscriber line [DSL]) systems.

The emergence of Web 2.0, broadly defined as Web-based communities such as social-networking sites that facilitate sharing of ideas among Web users, has been significant for many existing online communities. One such community, made up of real-time Web-based music collaborators using systems for networked musical performance (NMP), is in its infancy. An online NMP application lets musicians from across the globe play together over the Internet, as if they were together in a studio. With online music-making (either improvisatory or strictly notated), musicians can create ensembles without location bounds, searching for musicians around the world. The quality of the user's experience is critical to the success of this Web application. However, because performing artists are highly sensitive to delay, network latency affects the quality of the user experience of online music-making.

To achieve a good user experience the latency over the network has to be within reasonable bounds. If the delay is excessive, then the musicians will not be able to maintain a consistent tempo. We seek to find out how the tempo of two musicians performing together via a network varies as a function of fixed network delay. Future work may consider the more general case of more than two musicians and/or variable network delay.

Two musicians making music online will independently generate rhythmic patterns. *Entrainment* refers to an interaction between autonomous rhythmic processes such that they adjust to a common tempo or related tempi. Two oscillators, like two rhythmic processes, may synchronize, but other phase relationships are also possible. Entrainment and synchronization arise in many different contexts, where there is interaction (or coupling) between oscillators, where the oscillator may be designed for a particular purpose (e. g., electronic oscillator), or occur naturally (e. g., neural oscillators). Mathematical models of electronic and neural oscillators have been developed and are used to predict behavior by analysis or simulation. We are particularly interested in models for coupled oscillators with a time delay between them.

One such model was developed for geographically separated oscillators with delay compensation (anticipation). An equivalent model was developed for mutual entrainment of two limit-cycle oscillators with time-delayed coupling. We will show that both models make the same prediction: For delays that are a small fraction of the tempo period, the mean tempo in beats per minute (BPM), or beats per second, decreases by approximately half the tempo times the delay in seconds. This result is also relevant when musicians are far apart on a stage (e. g., the opposite ends of an opera stage), as each meter of separation adds about 3 msec of delay.

This article is organized as follows. We begin with a review of previous work on network musical performance systems and musical collaboration at a distance with delay. We also review previous work on entrainment and coupled oscillators. We develop

Computer Music Journal, 35:1, pp. 76–89, Spring 2011

an analysis to predict the tempo variations of two musicians performing with fixed network delay for impulsive (clapping) music. We then describe the experimental methodology, followed by results and conclusions.

Literature Review

We review literature in three areas: NMP systems, musical collaboration at a distance, and entrainment.

Systems for Networked Musical Performance

Interconnected musical networks are defined and classified by Weinberg (2005). An early networked musical collaboration was reported by Gang et al. (1997) using MIDI, not audio. Bargar et al. (1998) summarized the state of the art in networked audio and suggests directions for future research. Topics investigated include better protocols, lower packet processing latency, and musicians' tolerance to delay.

Cooperstock and Spackman (2001) organized a live performance of a jazz band in Montreal while recording engineers mixed the twelve channels of uncompressed audio in Los Angeles. Because all the musicians were in the same room there were no ensemble delay issues. This was the first live performance over the Internet using audio, not MIDI.

Lazzaro and Wawrzynek (2001) presented a case for NMP as a practical multimedia application. Their system combined MPEG-4 for sound synthesis, RTP and SIP for networking, and MIDI for musical control. Their network spanned only from UC Berkeley to Stanford University and Caltech, with delays ranging from 6 msec to 33.5 msec. The nominal latency was readily apparent at approximately 30 msec, causing depressed keys on the electronic piano not to sound, and released keys to sometimes keep sounding for a short time period. Their experiments used hosts that were connected directly to enterprise routers and therefore were able to use low-latency routes. They concluded that last-mile technologies dominate the end-to-end delay between two hosts and can result in a total latency too high for a useable NMP. This conclusion is less valid in 2010, with some ISPs providing short ping times to accommodate gamers.

Yoshida, Ob, and Yonekura (2005) proposed a new protocol called Mutual Anticipated Session (MAS) to compensate for network delays. In the MAS protocol, the keyboard player cannot sound a MIDI musical note until he or she receives the other player's MIDI note data. Using this protocol, the researchers were capable of achieving satisfactory comfort levels from the players.

Kramer et al. (2007) minimized delay using "Soundjack," an optimized audio framework on the network layer, along with an ultra-low-delay audio coder.

Barbosa (2006) performed an extensive survey of all computer-supported collaborative music applications. Some leading companies providing live online music-making experiences using audio are JamNow, eJamming, and Ninjam.

JamNow bases its technology on a client–server model and minimizes the delay from all sources by using low-delay audio compression and network transmission.

eJamming attempts to minimize the latency using three approaches (Greene 2007). First, the application uses audio compression to reduce the amount of data to be sent on the network. Second, it uses a peer-to-peer configuration in which, they claim, the musicians are connected directly to each other. Third, the application time-stamps the packets to synchronize the music. In reality, these methods increase the latency within the application by increasing the processing time to synchronize, compress, and decompress the data. Also, the peer-to-peer connections do not reduce delay in most instances involving traffic routed between different network service providers.

Ninjam uses the clientserver model and sends compressed audio (Barbosa 2006). Ninjam actually increases the delay to a musically relevant quantity (a full measure) using the Remote Music Control Protocol (Goto, Neyama, and Muraoka 1997). The extra delay is fixed with very small deviation (Sarkar and Vercoe 2007). Effectively, this means that the performer must synchronize to audio that was generated by other artists in the previous measure. This requirement makes it particularly challenging to perform naturally, as transitions must be anticipated.

Musical Collaboration at a Distance

Maki-Patola (2005) reviewed the musical effects of latency. Chafe et al. (2004) measured the rhythmic accuracy for a clapping session between two musicians over various fixed network delays, ranging from 0 msec to 77 msec. Their study reveals that delays longer than 11.5 msec cause the tempo to decelerate, wheras delays shorter than 11.5 msec cause the tempo to accelerate. Chew et al. (2005) described an experiment with two award-winning pianists who, although able to see each other in the same room, could only hear each other via headphones through the delayed network. The experiment shows that fixed delays up to 50 msec (65 msec with practice) can be tolerated. Bartlette et al. (2006) measured the effect of network latency on the tempo of interactive musical performance of clarinet duets and violin duets.

Barbosa, Cardoso, and Geiger (2005) showed that there is an inverse relationship between network delay tolerance and tempo. They showed that more network delay can be tolerated for slower tempi. This also reveals that, depending upon the tempo chosen for the sessions, the results can vary between research studies.

Entrainment and Coupled Oscillators

Clayton, Will, and Sager (2004) introduced basic concepts of entrainment, including interpersonal synchrony in musical performance. Large (2008) introduced a resonance theory of musical rhythm based on neural resonance, and reviewed various oscillator models of pulse and meter. This theory asserts that some neural oscillations entrain to musical rhythms.

Neural oscillator models represent the spiking dynamics of a real neuron, which gradually accrues voltage until it reaches a threshold; upon reaching that threshold it fires a spike and quickly releases the energy. Izhikevich (1999) considered pulse-coupled neural networks in which each isolated neuron fires periodically and the neurons are weakly connected, and shows that such networks can be transformed into a phase model (see Figure 1). The input from other neurons delays or advances each firing, thereby introducing a phase shift.

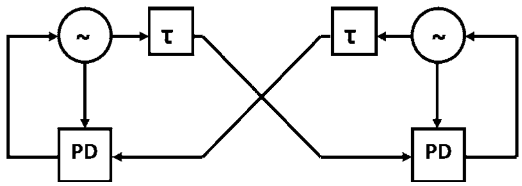
Thaut (2005) found that the entrainment process can be modeled well via resonant network functions and coupled oscillator phase models. Oprisan and Boutan (2008) predicted entrainment using a phase resetting curve method. Earlier studies with different oscillator models include Large and Jones (1999), Large and Palmer (2002), and Large and Kolen (1994). Interestingly, Large (2008) pointed out that many studies in the literature, such as Repp (2008), focus on tapping with recorded music, as opposed to real-time musical interactions.

The Kuramoto model (Acebron et al. 2005) is a well-known phase model with a sine interaction function motivated in a biological context. Eck (2002) reviewed a number of oscillator types and illustrates how nonlinear response to perturbations can lead to an oscillator naturally beating along with driving signals having compatible frequencies. A framework for characterizing different oscillator models is given by Buchli, Righetti, and Ijspeert (2009).

The earliest results for oscillator coupling with delay were found in a Bell Labs paper on mutual synchronization of geographically separated oscillators (Gersho and Karafin 1966). Motivated by biological oscillators many years later, Schuster and Wagner (1989) studied the mutual entrainment of two limit cycle oscillators with time-delayed coupling. The same model can arise from analysis of pulse-coupled systems of neural oscillators with coupling delay (Izhikevich 1999). The Kuramoto model was extended to include delays in Yeung and Strogatz (1999). Ermentrout and Ko (2009) showed that phase models with stronger coupling (with or without delays) have richer dynamics but are generally similar to the weakly coupled case.

Coupled phase oscillator models without delays were introduced in the context of mutual

Figure 1. Coupled oscillators with delay. PD = phase detector, circle = oscillator, τ = delay.



entrainment in human musical performance by Nagata, Kobayashi, and Miyake (2002). Kobayashi and Miyake (2003) presented experimental data for a network ensemble between humans with time lag; no mathematical model or solution was presented, however.

Analysis

Consider two musicians performing together via a network with symmetric delay $\tau_{12} = \tau_{21} = \tau$. Each musician attempts to play a rhythm (clapping) at a predetermined tempo or frequency $f_0 = \omega_0/2\pi$. Musicians normally express tempo M in BPM, so that $f_0 = M/60$ and 60 BPM corresponds to $\omega_0 = 2\pi$ radians per second. In general, we may assume that each musician has slightly different free-running tempi (i.e., the tempo when there is no external reference tempo to adjust to) ω_1, ω_2 which are close to ω_0.

We begin by modeling this as a system of two coupled oscillators with a time delay between them (see Figure 1). We can use the theory of the synchronization of geographically separated oscillators connected by a communications link (Gersho and Karafin 1966). This theory was originally developed in the 1960 s for the synchronization of communications networks, where there is a time delay between network nodes. The oscillator at each node has a free running frequency which, in general, will be slightly different at each node and varies with time. The frequency of the oscillator at each node is continuously adjusted to maintain synchronization using a control signal from the other oscillators at other nodes. For a network with two nodes, the signal from the local and remote oscillator are combined in a phase detector, which measures the phase difference and sends a control signal to each oscillator to adjust its frequency so as to minimize their phase difference. This work was further elaborated by Lindsay et al. (1985). These results can be extended using the theory of mutual entrainment of two coupled limit-cycle oscillators with time delay (Schuster and Wagner 1989), which potentially leads to richer dynamics, including stable synchronization at more than one frequency.

To begin, we model two performing musicians each as an oscillator with a control input and a phase detector, in a network with two nodes separated by a delay. Each musician has a free running frequency when the control input is zero, and has the ability to receive a signal from the other, detect the phase difference, and use this information (control signal) to adjust his or her tempo (clock frequency).

We carry out the analysis using the model from Schuster and Wagner (1989) for mutual entrainment of two limit cycle oscillators with time-delayed coupling

$$\begin{aligned} d\phi_1(t)/dt &= \omega_1 + K \sin[\phi_1(t) - \phi_2(t-\tau)] \\ d\phi_2(t)/dt &= \omega_2 + K \sin[\phi_1(t-\tau) - \phi_2(t)] \end{aligned} \tag{1}$$

ϕ_1, ϕ_2 are the phases of the two oscillators. The rate of change of these phases with time t are the free-running angular frequencies ω_1, ω_2 of each oscillator, respectively, plus an interaction term depending on the phase difference between the oscillators multiplied by a coupling constant K. The frequencies in Hz are the angular frequencies (in radians per second) divided by 2π. τ is the net time delay, which is the difference between the actual network delay and an anticipation factor τ^*, which is the delay as estimated and compensated for by the musicians. The difference in free-running frequencies of the two oscillators is $\Delta\omega = \omega_1 - \omega_2$ and the average is $\overline{\omega} = (\omega_1 + \omega_2)/2$. We look for the most general synchronized solution to Equation 1 with both oscillators perfectly synchronized at some angular frequency Ω not necessarily equal to $\overline{\omega}$. Given that such synchronized solutions must be in the form $\phi_1(t) = \Omega t$ and $\phi_2(t) = \Omega t + \alpha$, where α is a constant phase offset, Ω as a function of K and τ is given by Schuster and Wagner (1989):

$$\Omega = \overline{\omega} - K \tan(\Omega\tau)\sqrt{\cos(\Omega\tau) - \Delta\omega/K^2} \tag{2}$$

which is solved numerically for Ω. The steady state phase shift is given by

$$\alpha = \arcsin[\Delta\omega/(2K\cos(\Omega\tau))] \tag{3}$$

If $\Delta\omega = 0$ so that $\omega_1 = \omega_2 = \overline{\omega}$ then

$$\Omega = \omega - K\sin(\Omega\tau) \tag{4}$$

and the steady state phase shift $\alpha = 0$.

For $\tau \neq 0$ and $\Delta\omega \neq 0$ there is more than one stable synchronization frequency, and the number of stable frequencies increases with K and τ. A linear stability analysis shows that the condition for stability is $\cos(\Omega\tau) > 0$.

For $\tau = 0$, the only solution is $\Omega = \overline{\omega}$. Also, for small $\Omega\tau$, Equation 4 may be written as

$$\Omega \cong \overline{\omega} - K\Omega\tau \tag{5}$$

for which the only stable frequency is

$$\Omega = \frac{\overline{\omega}}{(1 + K\tau)} \tag{6}$$

This is the same as the solution found for geographically separated oscillators (Gersho and Karafim 1966; Lindsay et al. 1985). This same solution is also found in the work of Yeung and Strogatz (1999) and Ermentrout and Ko (2009).

For small $\Omega\tau$, the steady-state phase shift (3) becomes

$$\alpha = \Delta\omega/(2K) \tag{7}$$

corresponding to time error $\varepsilon = \Delta\omega/(2K\overline{\omega})$, which again is the same solution found by Lindsay et al. (1985) with symmetric delays $\tau_{12} = \tau_{21} = \tau$.

The steady state frequency $\Omega/2\pi \stackrel{\Delta}{=} f_3$ is reduced below $\overline{\omega}/2\pi \stackrel{\Delta}{=} f_0$ arising from the network delay τ. An average phase error $-\overline{\omega}\tau$ of one oscillator, as observed at the other oscillator, causes the frequency to reduce and the system to never recover. This explains the common observation that tempo slows down with increasing delay (Barbosa, Cardoso, and Geiger 2005).

This analysis may also be considered as a special case of general pulse-coupled neural networks (Izhikevich 1999):

$$\begin{aligned} dx_1/dt = {} & f_1(x_1) + \varepsilon g_{11}(x_1)\delta(t - t_1^* - \eta_{11}) \\ & + \varepsilon s g_{12}(x_1)\delta(t - t_1^* - \eta_{12}) \\ dx_2/dt = {} & f_2(x_2) + \varepsilon g_{21}(x_2)\delta(t - t_2^* - \eta_{21}) \\ & + \varepsilon s g_{22}(x_2)\delta(t - t_2^* - \eta_{22}) \end{aligned} \tag{8}$$

where x_1, x_2 are the membrane potentials of two coupled neurons, and the functions f_1, f_2 describe their dynamics, typically an integrate and fire model $f_i(x_i) = a_i + b_i x_i$. When x_i reaches 1 at time t_i^* the neuron fires a spike and x_i is reset to zero. The function $g_{ij} = \sin$ describes the coupling. When the jth neuron fires, the ith neuron is incremented by $\varepsilon g_{ij}(x_i)$ after some time delay η_{ij}. When each neuron can fire periodically and independently of the other neurons, Equation 8 can be transformed to a phase model equivalent to Equation 1.

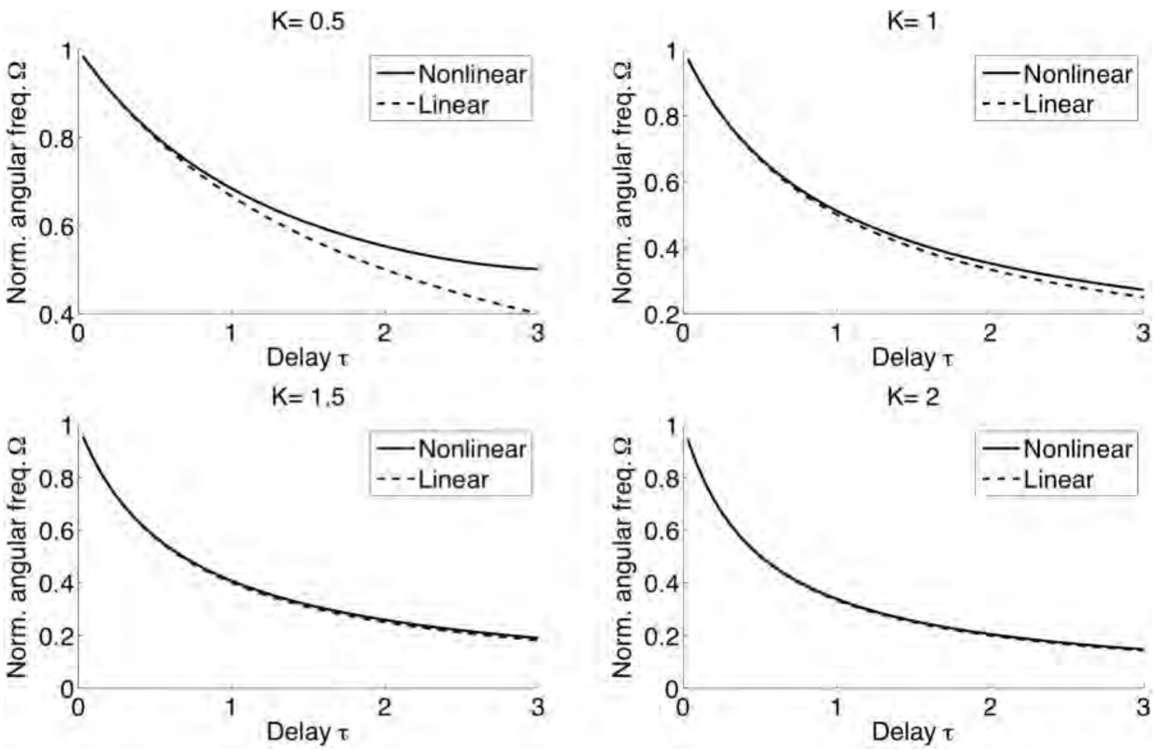
The Appendix shows some details of the analysis by Lindsay et al. (1985) which explicitly includes the anticipation factors and provides for the possibility of asymmetric delays.

Numerical Results from Theory

Figure 2 shows numerical results for angular frequency Ω versus delay τ for selected values of the coupling constant K between 0.5 and 2.0 and for τ from 0 to 3, with the nominal angular frequency $\overline{\omega}$ normalized to 1 (or $\overline{f} = \overline{\omega}/2\pi = 0.159$ Hz) with corresponding period $T = 1/\overline{f} = 2\pi/\overline{\omega} = 6.28$ sec (so $\tau = 3$ sec is a delay of about $1/2$ of an oscillator cycle at $\overline{\omega}$). The "non-linear" curve is a numerical solution of Equation 4, obtained by finding the roots of $\overline{\omega} - \Omega - K\sin(\Omega\tau)$ using the Matlab function *fzero*. The "linear" curve is from Equation 6. The two curves are very close, as the values of τ are such that $\sin(\Omega\tau) \approx \Omega\tau$.

To interpret these curves in terms of practical musical tempi in BPM, the y-axis values of Figure 2 can be scaled (unnormalized) by a factor so that $\Omega = \overline{\omega} = 1$ at $\tau = 0$ corresponds to a desired reference tempo. For example, scaling by a factor of 10, the y-axis scales up from $\overline{\omega} = 1$ ($\overline{f} = \overline{\omega}/2\pi = 0.159$) to $\overline{\omega} = 10$ ($\overline{f} = \overline{\omega}/2\pi = 1.59$ Hz $= 95.4$ BPM) and the x-axis scales down from the range 0 to 3 to a range

Figure 2. Normalized angular frequency (radians per second) versus delay (seconds). Ω is normalized with respect to $\overline{\omega}$.



of 0.0 to 0.3 sec. The y axis is in units of 1/sec, and the x axis is in units of seconds, thus if y scales up, then x scales down. This scaling also shows that more absolute network delay can be tolerated for slower tempi, as also observed by Cooperstock and Spackman (2001). As an example, using this scaling for K = 0.5, and given a tempo of 95.4 BPM with a zero delay (point $[x,y] = [0,1]$ on the curve), the results predict a tempo of (0.7)(95.4) = 67 BPM, given a delay of 100 msec (point $[x, y] = [1, 0.7]$ on the curve). A linear approximation of this curve for small delay, which permits easy calculation of tempo versus delay, is given in the Experimental Results section.

Experimental Setup

Our methodology for testing musicians employed a commercial, beta-release online collaboration application, an audio editing and analysis software package, and a music network that included a delay emulator. Two musicians performed a clapping session together with a selected delay. The musicians were asked to follow a click track (metronome) that was set at 90 BPM and to clap in rhythm for at least 60 seconds. The clapping audio files were then analyzed using Matlab to detect the claps, and then to calculate the time differences between claps, both from the same player and between players.

The recorded audio waveform was analyzed for both tempo and ensemble consistency. The musicians also answered a subjective questionnaire about their experience after each session.

Hardware and Software

The online collaboration application, JamNow, is a client–server system provided by Lightspeed Audio Labs. The music server in this system is designed to support multiple clients and separate simultaneous sessions over the public Internet. In order to control network delays, a local network

Figure 3. The music network system used for the experiments.

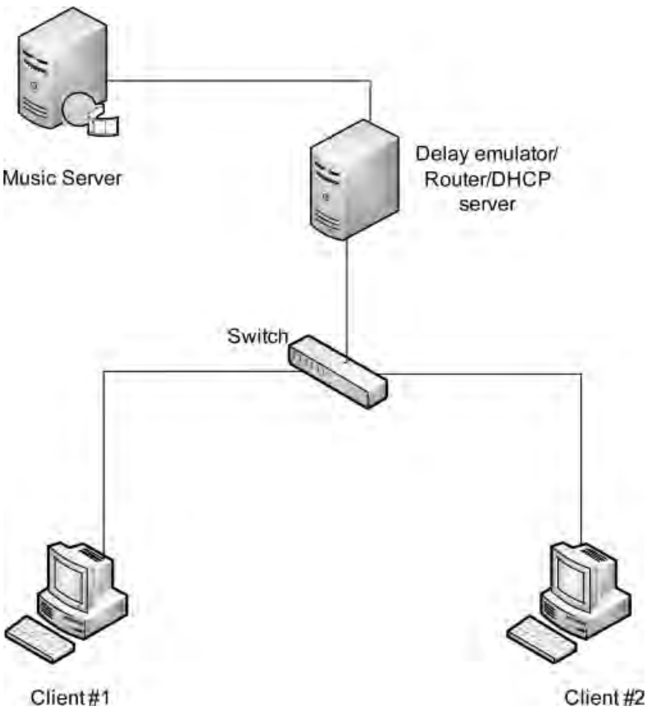


Table 1. Jam – Online Jamming Application Delay Budget

Location	*Module*	*Delay*	*Comment*
Client	Audio I/O	<1 msec	Headphones
Client	Audio Driver	<6 msec	Time to fill two 256-sample buffers at 48 kHz, plus additional delay
Client	Audio Encoder	<50 μsec	Encoding is 100 × real time for a 256-sample window
Client	Receive FIFO	5 msec	Jitter Buffer
Client	Audio receive thread timing and decode	3 msec	Average wait time for next ping pong occurrence
Server	Input FIFO	5 msec	Jitter buffer
Server	Audio decode, error mitigation, mix, and audio encode	<100 μsec	Decoding is 100 × real time
TOTAL		**20 msec**	

with a delay emulator was used in place of the Internet. In Figure 3 we show the test setup with server, delay emulator, and two clients.

The components contributing to delay, which constitute the application delay budget, are shown in Table 1, not including any delay added by the network delay emulator. This delay budget represents the minimum possible delay with zero network delay.

The delay emulator is based on NetEm, a component of the Linux kernel. It intercepts packets and can then add a fixed delay, a variable delay, and/or a delay value from a mathematical distribution. Because of the limitation of the Linux timer resolution, the timer in NetEm has a resolution of 1 msec. The emulator is also configured as a router to route packets to the appropriate clients and the music server, and further, as a DHCP server to dynamically assign an IP address to the clients (not strictly needed here, but useful for testing on the public Internet).

The minimum total delay is 30 msec, which includes 20 msec from the application delay budget and 10 msec from the delay emulator. This minimum delay of 10 msec represents a best case over the commodity Internet, but a more typical round-trip ping time from cable or DSL to the first level of network aggregation is on the order of 20 msec or more, not counting application delay.

The network-level audio flow is illustrated in Figure 4. Two rooms, separated by a concrete wall and thus acoustically isolated from each other, are used. Audio travels from a microphone via a mixer to Client # 1, then through the delay emulator to the music server, and then on to Client # 2. A similar path in the reverse direction takes audio from Client # 1 to Client # 2.

The low-level audio path is shown in Figure 5. When there is input on the microphone on Client # 1, it is received on the audio card and then read by the jamming application. The jamming application converts this audio input into 128-byte packets and sends it over the Ethernet network to the music

Figure 4. Network-level audio flow.

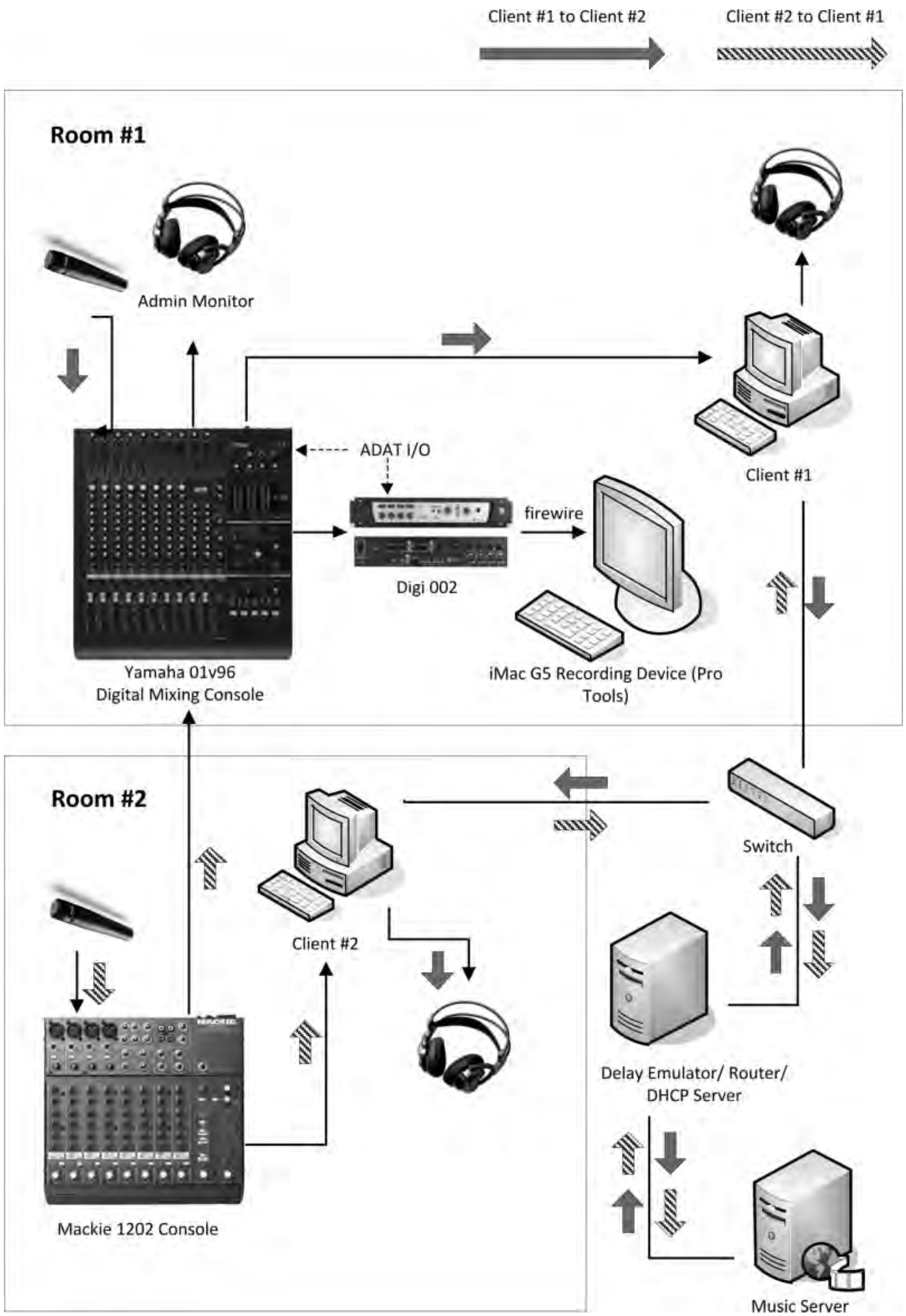
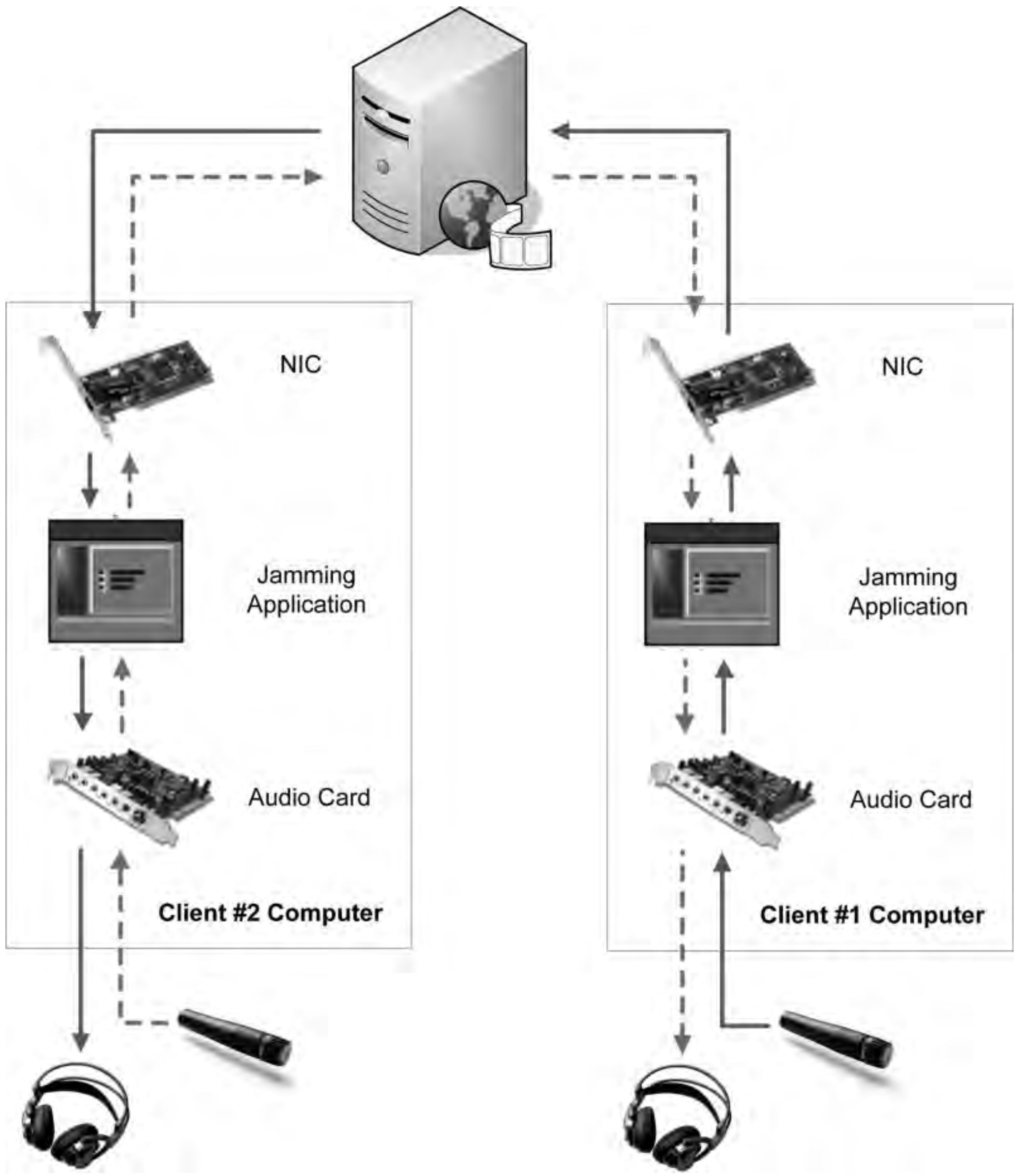


Figure 5. Low-level audio path. NIC = network interface card.



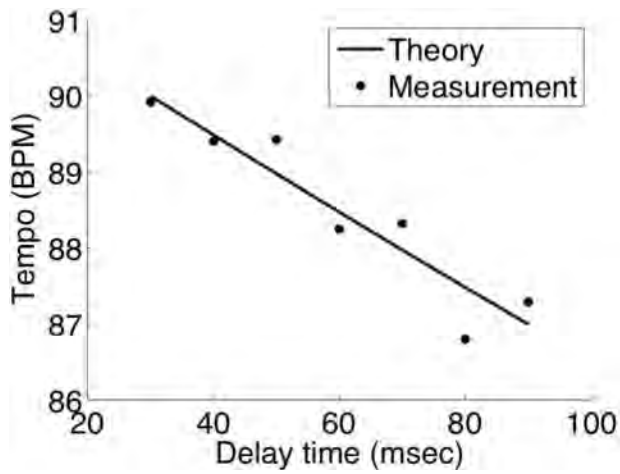
server. The server then decides which client or clients should receive this packet; in this scenario it is sent to Client # 2. It is received by Client # 2's network interface card and then read by the jamming application. The jamming application outputs packets to the audio card to be heard by the musician. This process is reversed for audio generated by Client # 2.

Testing Methodology

The following experiment illustrates a test of ensemble accuracy with total delays (network plus application) ranging from 30 msec to 90 msec. Eleven non-overlapping pairs of subjects were recorded for duo-clapping. A simple interlocking rhythm was chosen for the duo-clapping (see Figure 6). Ensemble

Figure 6. Notation of duo-clapping pattern, as shown to participants in the experiments.

Figure 7. Mean tempo versus total delay, as predicted by theory and as measured experimentally.



accuracy was measured by analyzing both the tempo of each player individually and also the time difference between the players. These two measured quantities are termed *pacing* and *coordination* by Bartlette et al. (2006). Answers to subjective questions (described in the next section) were also collected. Each session had two musicians clapping together. The session had seven trials with total delays, in 10 msec increments, between 30 msec and 90 msec in random order. The musicians were asked to clap in rhythm for at least 60 seconds for each trial. All sessions used a tempo of 90 BPM.

Experimental Results

Experimental results were obtained for tempo variation with delay, to be compared with the predictions of Equation 6. Experimental results were also obtained for the time error between musicians, to be compared with the predictions of Equation 7. The audio files were analyzed for tempo, as a function of delay, and ensemble time (phase) differences between the players, also as a function of delay. The clapping audio files were analyzed using Matlab to detect the claps and calculate the time differences between the players.

Tempo Variation with Delay

The points indicated as "measurement" in Figure 7 show the mean tempo in BPM, given an initial tempo of 90 BPM. The steady-state tempo slows down as the network delay is increased.

To obtain the theoretical curve shown in Figure 7, we plot Equation 6 using values of K and f_0 that give the best fit to the experimental data. A simple approach is to use the first and last data points $f_4 = 90/60 = 1.50$ Hz, $\tau_4 = 0.03$ sec and $f_3 = 87/60 = 1.45$ Hz, $\tau_3 = 0.09$sec. Here we do not use the subscripts 1 and 2, because we used them in Equation 1. From Equation 6, we write

$$\overline{\omega} = \Omega(1 + K\tau) \tag{9}$$

We observe from Figure 7 that the intercept frequency $f_0 = \overline{\omega}/2\pi$ (for zero delay) is somewhat higher than 1.5 Hz (90 BPM), and solve $f_0 = f_j(1 + K\tau_j)$ with $j = 3, 4$ to obtain $K = 0.5848$ and $f_0 = \overline{\omega}/2\pi = 1.5263$ (91.5 BPM).

The theoretical curve of Equation 9, expressed as $\overline{\omega}/\Omega$ (in BPM) for these values of K and f_0, is shown in Figure 7. A least-squares linear-curve fit, based on the measured data points, yields an

Figure 8. Mean time error between clients versus total delay, ten trials.

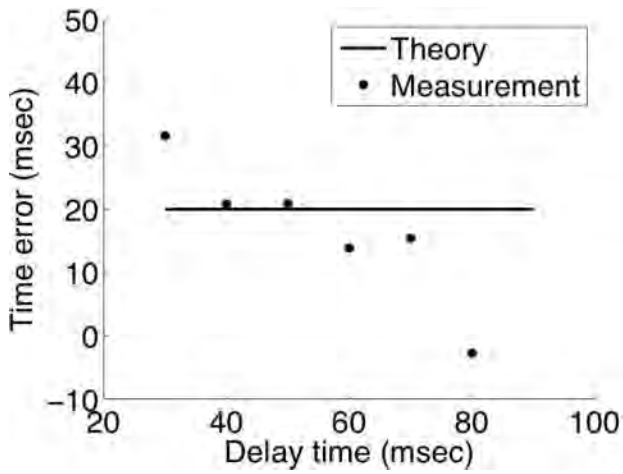
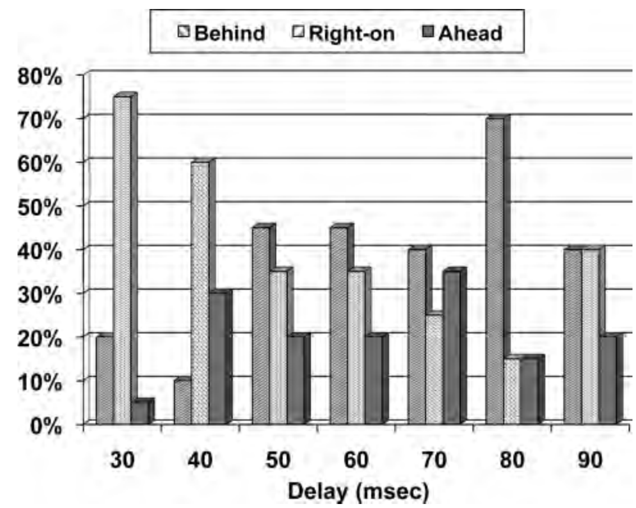


Figure 9. Subjective response of the musicians to the question "Did you perceive that you were behind the beat, right on, or ahead of the beat?" as a function of network delay.



indistinguishable curve and thus identical values of K and f_0.

A simple approximate formula for tempo versus delay, valid for delays that are small compared to the tempo period, may be obtained by writing Equation 9 as

$$\frac{\Omega}{\overline{\omega}} = \frac{f_j}{f_0} = \frac{f_0 - \lambda}{f_0} = 1 - \frac{\lambda}{f_0} = \frac{1}{(1 + K\tau)} \cong 1 - K\tau \tag{10}$$

Here we have introduced the variable $\lambda = f_0 - f_j$ to represent the difference between the zero-delay frequency f_0 and the frequency f_j with delay τ_j. From Equation 10 we can see that $\lambda \cong f_0 K\tau$. Thus for $f_0 = 91.5$ BPM and $\tau_3 = 0.09$ sec, $\lambda \cong 0.58(91.5)(0.09) = 4.7$ BPM and the steady state tempo becomes $f_j = f_0 - \lambda \cong f_0 - f_0 K\tau = 91.5 - 4.7 = 86.8$ BPM, reasonably close to the measured value of 87.4 BPM.

Time Error (Ensemble Consistency)

Figure 8 shows the average (mean) of the time (ensemble) differences between the musicians for duo-clapping as a function of delay.

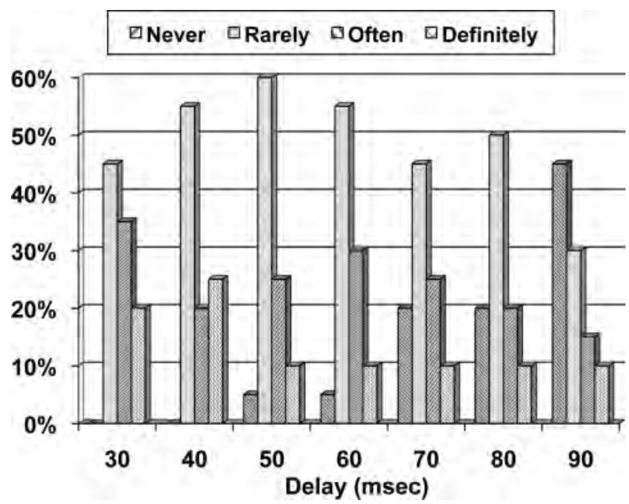
From Equation 7 the time difference $\varepsilon = \Delta\omega/(2K\overline{\omega})$ depends on the natural free-running frequencies of the oscillators but does not vary as a function of delay. The data show a roughly constant time difference of about 20 msec in the range of delays between 30 and 80 msec. Thus $\varepsilon = \frac{(f_2 - f_1)}{2 f_0 K} = 0.02$ and $f_2 - f_1 = 0.04K f_0$. This shows that the difference between the natural free running frequencies of the two musicians $f_2 - f_1 = (0.04)(0.58)(1.5)(60) = 2.06$ BPM. We believe this to be a reasonable result, because it is similar to the amount of error a musician might make when attempting to play alone at a specified tempo.

Musicians' Subjective Response

After every trial, the musicians were asked "Did you perceive that you were behind the beat, right on, or ahead of the beat?" The survey results (see Figure 9) show that the musicians felt that they were right on the beat at least 41% of the time. In most cases the musicians could not distinguish between being ahead or behind, but they knew if they were not on the beat. For total delays of 50 msec and greater (tempo of 89 BPM or less in Figure 7) the musicians started to perceive that they were behind the beat.

In response to the subjective questionnaire "Would you agree to jamming online?" 55% expressed a willingness (response of "often" or "definitely") to music-making online at 30 msec total delay, but this number dwindled down to approximately 30% as the network delay increased, as shown in Figure 10. These subjective results appear to be consistent with the experimental data.

Figure 10. Subjective response of musicians to the question "Would you agree to jamming online?" as function of network delay.



Conclusions

We measured the variations in tempo of two musicians performing together via a network as a function of fixed network delay. We used a tempo of 90 BPM and total end-to-end delays from 30 to 90 msec.

The tempo of two musicians performing together at a distance, with network delay and without any external tempo reference, will slow down as the delay is increased (at least with the tempo and range of delays used in our experiment). The amount of slowing may be predicted using theories of coupled oscillators with delay. These theories arise in the contexts of geographically separated oscillators with delay compensation (anticipation), limit cycle oscillators with time delayed coupling, and weakly pulse-coupled oscillators. The amount by which the tempo decreases is approximated by just over half (0.58) of the tempo times the delay in seconds, so that a tempo of 90 BPM with a delay of 60 msec will slow down to $90 - 0.58(90)(0.06) = 90 - 3.1 =$ about 87 BPM. As suggested earlier, we might hypothesize that this result is also applicable when musicians are far apart on a stage (e. g., opposite sides of an opera stage), as each meter of separation adds about 3 msec of delay. We have not, however, done experiments to support that hypothesis.

Acknowledgments

The authors thank Kirk McNally for providing audio engineering support. This work was supported by the Natural Sciences and Engineering Research Council of Canada.

References

Acebron, J. A., et al. 2005. "The Kuramoto Model: A Simple Paradigm for Synchronization Phenomena." *Reviews of Modern Physics* 77(1):137–185.

Barbosa, A., J. Cardoso, and G. Geiger. 2005. "Network Latency Adaptive Tempo in the Public Sound Objects System." In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME).* Vancouver, Canada: University of British Columbia, pp. 184–187.

Barbosa, A. 2006. "Displaced Soundscapes." PhD Thesis, Pompeu Fabra University, Barcelona, Spain.

Bargar, R., et al. 1998. "Networking Audio and Music using Internet2 and Next Generation Internet Capabilities." New York: Audio Engineering Society. Audio Engineering Society White Paper, Tech. Report TC-NAS98/1. Available online at http://www.aes.org/technical/documents/i2.pdf.

Bartlette, C., et al. 2006. "Effect of Network Latency on Interactive Musical Performance." *Music Perception* 24(1):49–62.

Buchli, J., L. Righetti, and A. J. Ijspeert. 2009. "Engineering Entrainment and Adaptation in Limit Cycle Systems." *Biological Cybernetics* 95(6):645–664.

Chafe, C., et al. 2004. "Effect of Time Delay on Ensemble Accuracy." In *Proceedings of the International Symposium on Musical Acoustics*. Nara, Japan: International Symposium on Music Acoustics (pages unnumbered).

Chew, E., et al. 2005. "A Second Report on the User Experiments in the Distributed Immersive Performance Project." In *Proceedings of the 5th Open Workshop of MUSICNETWORK: Integration of Music in Multimedia Applications*. Vienna, Austria: Universität für Musik und darstellende Kunst Wien. pp. 1–7.

Clayton, M., U. Will, and R. Sager. 2004. "In Time With The Music: The Concept Of Entrainment And Its Significance For Ethnomusicology." *ESEM CounterPoint* 1:1–81.

Cooperstock, J., and S. Spackman. 2001. "The Recording Studio that Spanned a Continent." In *Proceedings of the International Conference on Web Delivering of Music*

(WEDELMUSIC 2001)*. Florence, Italy: Wedelmusic, pp. 161–167.

Eck, D. 2002. "Finding Downbeats with a Relaxation Oscillator." *Psychological Research* 66(1):18–25.

Ermentrout, B., and T. W. Ko. 2009. "Delays and Weakly Coupled Neuronal Oscillators." *Philosophical Transactions* 367(1891):1097–1115.

Gang, D., et al. 1997. "TransMIDI: A System for MIDI Sessions Over the Network Using Transis." In *Proceedings of the International Computer Music Conference.* San Francisco, California: International Computer Music Association,pp. 283–286.

Gersho, A., and B. J. Karafin. 1966. "Mutual Synchronization of Geographically Separated Oscillators." *Bell System Technical Journal* 45:1689–1704.

Greene, K. 2007. "Jam Online in Real Time." *MIT Technology Review*. Available online at www.technologyreview.com/web/18783/?a=f. Accessed 17 September 2010.

Goto, M., R. Neyama, and Y. Muraoka. 1997. "Remote Music Control Protocol—Design and Applications." In *Proceedings of the International Computer Music Conference.* San Francisco, California: International Computer Music Association, pp. 446–449.

Izhikevich, E. M. 1999. "Weakly Pulse-Coupled Oscillators, FM Interactions Synchronization, And Oscillatory Associative Memory." *IEEE Transactions on Neural Networks* 16(3):508–526.

Kobayashi, Y., and Y. Miyake. 2003. "Analysis Of Network Ensemble Between Humans With Time Lag." *SICE 2003 Annual Conference* 1:1069–1074.

Kramer, U., et al. 2007 "Network Music Performance with Ultra-Low-Delay Audio Coding under Unreliable Network Conditions." *Audio Engineering Society 123rd Convention.* Paper no. 7214 (pages unnumbered).

Large, E. W., and J. F. Kolen. 1994. "Resonance And The Perception Of Musical Meter." *Connection Science* 6(2):177–208.

Large, E. W., and M. Riess Jones. 1999. "The Dynamics of Attending: How People Track Time-Varying Events." *Psychological Review* 106(1):119–159.

Large, E. W., and C. Palmer. 2002. "Perceiving Temporal Regularity in Music." *Cognitive Science* 26(1):1–37.

Large, E. W. 2008. "Resonating to Musical Rhythm: Theory and Experiment." In S. Grondin, ed. *The Psychology of Time*. Amsterdam: Elsevier Science, pp. 189–232.

Lazzaro, J., and J. Wawrzynek. 2001. "A Case for Network Musical Performance." In *Proceedings of the 11th International Workshop on Network and Operating Systems Support for Digital Audio and Video.* New York: Association for Computing Machinery, pp. 157–166.

Lindsay, W. C., et al. 1985. "Network Synchronization." *Proceedings of the IEEE* 73(10):1445–1467.

Maki-Patola, T. 2005. "Musical Effects Of Latency." *Swomen Musiikintutkijoiden* 9:82–85.

Nagata, Y., Y. Kobayashi, and Y. Miyake. 2002. "Development of New Ensemble System Based on Coupled Phase Oscillator." In *Proceedings of the 11th IEEE International Workshop on Robot and Human Interactive Communication*, pp. 77–80.

Oprisan, S. A., and C. Boutan. 2008. "Prediction of Entrainment and 1:1 Phase-Locked Modes in Two-Neuron Networks Based on the Phase Resetting Curve Method." *International Journal of Neuroscience* 118(6):867–890.

Repp, B. H. 2008. "Metrical Subdivision Results in Subjective Slowing of the Beat." *Music Perception* 26(1):19–39.

Schuster, H. G., and P. Wagner. 1989. "Mutual Entrainment of Two Limit Cycle Oscillators with Time Delayed Coupling." *Progress of Theoretical Physics* 81(5):939–945.

Sarkar, M., and B. Vercoe. 2007. "Recognition and Prediction in a Network Music Performance System for Indian Percussion." In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME).* New York: New York University, pp. 317–320.

Thaut, M. H. 2005. "Embodied Communication. Rhythm, Human Temporality, and Brain Function." In D. Miell, R. MacDonald, and D. J. Hargreaves, eds. *Musical Communication*. New York: Oxford University Press, pp. 171–192.

Weinberg, G. 2005. "Interconnected Musical Networks: Toward a Theoretical Framework." *Computer Music Journal* 29(2): 23–39.

Yeung, M. K. S., and S. H. Strogatz. 1999. "Time Delay in the Kuramoto Model of Coupled Oscillators." *Physical Review Letters* 82(3):648–651.

Yoshida, M., Y. Ob, and T. Yonekura. 2005. "A Protocol for Real-Time Remote Musical Session." *IEICE Transactions on Information & Systems* E88-D(5):919-929.

Appendix

This Appendix details the analysis by Lindsay et al. (1985). This analysis explicitly included the anticipation factors and provided for the possibility of asymmetric delays.

From Lindsay et al. (1985), an oscillator time scale or time process is obtained by dividing the clock phase by the nominal free-running frequency of the oscillator (frequency is phase divided by time, so time is phase divided by frequency). The work of Lindsay et al. was not specifically about music, but for a musical interpretation we may think of the time process as the actual time of occurrence $T_1(t)$ of an event (say a beat) time-stamped on a recorded track that is supposed to (intended to) occur at time t according to the musical score. Ideally, the time process $T_1(t) = t$ assuming that the offset $T_1(t = 0) = 0$ and that there are no time drifts or instability.

Lindsay et al. (1985) define the time scale or time process of clock 1 as

$$T_1(t) = T_1(t = 0) + t + D(t) + \psi(t) \tag{11}$$

where $D(t)$ is a time drift modeled as a polynomial in t, and $\psi(t)$ is an instability modeled as a zero-mean stationary random process. A similar equation applies for clock 2. Thus a plot of $T_1(t)$ versus t is a line with nominal slope of 1. The slope will increase or decrease slightly with time according to the linear term in $D(t)$ and will contain additional variations from $\psi(t)$.

In the model of synchronous networks from Lindsay et al. (1985), the equations that govern the network of two musicians in terms of the time scale may be written

$$\begin{aligned} dT_1(t)/dt &= W_1(t) + K_1 g[T_2(t - \tau_{12}) - T_1(t - \tau^*_{12})] \\ dT_2(t)/dt &= W_2(t) + K_2 g[T_1(t - \tau_{21}) - T_2(t - \tau^*_{21})] \end{aligned} \tag{12}$$

where $W_1(t), W_2(t)$ are the normalized instantaneous frequencies (close to 1), $g[]$ is the characteristic of the time (phase) detector, K_1, K_2 are loop gain constants (units of 1/sec), $\tau_{12} = \tau_{21} = \tau$ is the network delay (assumed symmetric), and τ^*_{12}, τ^*_{21} are the anticipation factors, i.e., the delay as estimated and compensated for by the musicians based on what they hear. We assume $\tau^*_{12} = \tau^*_{21} = \tau^*$.

In the Analysis section herein, the anticipation factor is contained within τ.

We can also write these equations in terms of the clock phase rather than time process. The clock phase is obtained by multiplying the time process by the nominal free-running angular frequency $\omega_1 = 2\pi f_1$ of oscillator 1, which in general is not the same as ω_2 for oscillator 2.

$$\begin{aligned} \phi_1(t) &= \omega_1 T_1(t) = \omega_1 T_1(t = 0) + \omega_1[t + D(t) + \psi(t)] \\ &= \phi_1(t = 0) + \omega_1(t)t \end{aligned} \tag{13}$$

where we define the time varying frequency $\omega_1(t) = \omega_1[1 + D(t)/t + \psi(t)/t]$. Thus, in terms of clock phase, we write

$$\begin{aligned} d\varphi_1(t)/dt &= \omega_1(t) + K_1 \sin[\varphi_2(t - \tau_{12}) - \varphi_1(t - \tau^*_{12})] \\ d\varphi_2(t)/dt &= \omega_2(t) + K_2 \sin[\varphi_1(t - \tau_{21}) - \varphi_2(t - \tau^*_{21})] \end{aligned} \tag{14}$$

In Equation 14, we have also made the simplifying assumptions that the phase detector $g[]$ that detects the difference in the two phases has a sinusoidal characteristic.

The form of Equation 14 is the same as Equation 1, with the same solutions for Equations 6 and 7 for symmetric delay $\tau_{12} = \tau_{21} = \tau, \tau^*_{12} = \tau^*_{21} = \tau^*$, with the change of variable $\tau - \tau^*$ in Equation 14 to τ in Equation 1.