LEVELS OF TEMPORAL RESOLUTION IN SONIFICATION OF NETWORK PERFORMANCE

Chris Chafe and Randal Leistikow

Center for Computer Research in Music and Acoustics
Music Department, Stanford University
Stanford, CA USA 94305
cc@ccrma.stanford.edu
randal@ccrma.stanford.edu

ABSTRACT

The standard "ping" utility provides a momentary measurement of round trip time. Sequences of ping events are used to gather longer-term statistics about jitter and packet loss in order to describe the quality of service of a network path. A more fine-grained tool is needed to evaluate paths which carry interactive media streams for collaborative environments. Natural interaction depends on obtaining consistent low-latency, low-jitter service, something which normally requires several ping "takes" to assess and even then only provides an averaged picture of quality of service. We have designed a stream-based method which directly displays the critical qualities to the ear by continuously driving a bidirectional connection to create sound waves. The network path itself becomes the acoustic medium which our probe sets into vibration. The granularity of this display better matches the time-scales of variance that are important in interactive applications (for example, bidirectional audio streams for long-distance musical collaboration or high-quality teleconference applications). The ear's acuity for pitch fluctuation and timbral constancy make this an unforgiving test.

A related sonification technique is discussed which is a sonar-like mapping of momentary ping data to musical tones. Temporal levels of musical foreground, middleground and background can be heard in the melodies derived from the data and correspond to structures that are of importance in the analysis of network performance.

1. INTRODUCTION

Our study is concerned with developing sonically-based tools for evaluating network performance. There exists a kind of "music" in the sound of performance measurements displayed to the ear, and this paper discusses temporal aspects of listening to it which, if exploited, can enhance the design of a useful evaluation tool. The overall quality of service of a network connection is a manifestation of phenomena at different temporal levels, from short-term jitter of transmission rate to long-term flux in throughput. By translating performance measures to sound with appropriate mappings, we can apply a sensory apparatus (musical hearing) that is *simultaneously* sensitive to variance and structure over a wide range of time scales.

Techniques for sonification of time-series data fall into two broad classes. *Transduction* simply plays the data as sound waves, sometimes requiring a transformation to make it perceptible. For example, seismic data can be sped up to audio rates so that earthquake events become percussion tones. *Parameterized display*, on the other hand, "plays" an analysis of the data by causing the analytic result to drive a sound-generating method. This study has explored both kinds of techniques for evaluation of network performance. As will be seen, our interest has been to create qualitative and intuitive methods which supplement rather than replace existing quantitative tools.

SoundWIRE [1] is a transduction-type method especially useful for evaluating very fine-grained jitter and packet loss. The network itself is used as a sound-producing medium as if it were a clarinet's air column or a guitar's stretched string. The resulting waveform is played to the listener in real time, and they literally can "pluck" the internet connection that is being tested.

SoundPing is a parameterized display which is applied to different kinds of analysis. In its most basic form, it is a one-for-one display of the output of the standard ping utility [1], which measures momentary *round trip time* (RTT) and packet loss. For each successful ping event (indicating that an echo was received), a musical tone is synthesized with a pitch that is inversely related to the measured RTT. Since RTT is the basis for several kinds of higher-order analyses, e.g. throughput, the method can be musically inflected to represent more than one dimension. For example, jitter is present in most RTT series. Jitter can be brought to the listener's attention by applying its running value (the windowed standard deviation of RTT) to the tone's timbral quality, such that stridency is enhanced during a strong jitter episode.

Parameterized displays which are useful for intuitive, real-time, "read-outs" of network condition can also be used to "graph" statistics from archives. Several research projects [3][4][5] have collected enormous quantities of world-wide Internet ping data, which is a bit like taking weather data over the entire networked planet. To ascertain long-term trends,"zoomed" listening is possible that reveals long-term structure. For instance, daily pings can be played back at a fast, but still musical, tempo such that one year's worth of ping data is compressed to about one minute. Spring break vacations on college campuses become evident as events.

2. QUICK PRIMER ON IP NETWORKING

Data is exchanged between network sockets. One of the tasks of a computer's operating system is to answer requests from processes

[1] Internet Control Message Protocol (ICMP) Ping [2]

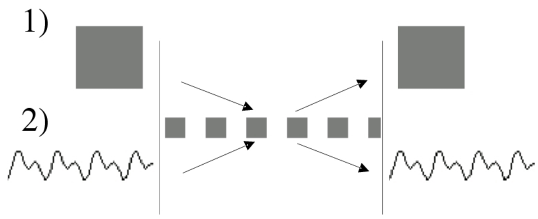

Figure 1: 1) Transfer of large blocks of data or 2) transmission of real-time signals involves packetizing the data into a series of *datagrams* no larger than the network's *MTU*.

to pipe data into and out of the network. After being granted a socket, a process passes internet protocol (IP) formatted data packets, or datagrams, as fast as possible to the receiving socket. If the amount of data is larger than fits into the network's maximum transfer unit (MTU), it will automatically be split into multiple datagrams which are sent in a sequence. In the case of real-time signals, the process instead transmits regularly-spaced packets containing media such as teleconference data. In either case, there will be a stream of datagrams sent over the network path as shown in Figure 1. The travel time that it takes for a given portion of data to be received will depend on several factors: operating system load, path, network protocol choice, buffering, network type, network topology, network difficulties, etc. – all of which interact, often in a complex manner.

Ping is a special command used to test connectivity between two computers ("hosts") and gain an idea of the travel time. The RTT it reports indicates the time it takes for a ping message (a specially-formatted datagram) to complete a trip from the host's network interface to the target interface and back. A host may be asked to ping its own network subsystem by targeting the special IP hostname,"localhost" (available on most systems). Typically, this will report a fast RTT $< 30\mu s$. External hosts will show ping times in a range of $0.1 - 1000ms$. What makes the difference is, of course, the intervening network. Its extent (number of subnets and their qualities) governs the latency and can be broken into two ranges – local area networks (LAN) which are usually much lower latency, $< 1ms$, and wide area networks (WAN) whose paths transit networks external to an organization. These administratively separate segments may differ in terms of available hardware bandwidth, policies, and congestion.

Statistics on packet loss, RTT min / max and RTT average are gathered by automatically issuing a series of pings, typically sent at one-second intervals. *Traceroute* is a related utility that reports the approximate transit time across individual segments on a given path. The relay points that are traversed are called "routers," computers that cooperate to pass data between interconnected networks.

Next-generation networks are capable of supporting a new class of bidirectional applications. Examples include teleconferencing, telemanipulation with haptic feedback, and remote musical collaboration. Experiments have been performed on streaming high-quality signals such as real-time HDTV and professional, uncompressed audio [6]. However, glitch-free transmission over WAN is still difficult when congestion occurs. Policies and protocols are being designed which will ensure proper QoS. We have tested our methods in trials for two of them.

3. FINE-GRAINED LISTENING: SOUNDWIRE

Our transduction method is a utility for fine-grained listening to packet flow on the order of $< 100\mu s$. (the granularity of an audio frame). We create "sound waves on the Internet from real-time echoes" as an easy-to-use, easy-to-understand evaluation for quality of service (QoS). Users of the most demanding interactive, real-time applications (such as high-quality audio collaboration) can only operate within a low-latency, low-jitter QoS environment that the standard ping utility is too coarse to verify. Our version implements the SoundWIRE technique with the same features as ping, with options for packetsize, "flood-rate" (i.e., audio sample rate), and start / stop timing.

SoundWIRE uses physical modeling synthesis [7]. Delay unit generators which are typical building blocks of "lumped circuit" models are replaced by the network path's delay. For example, it becomes the loop delay in the well-known Karplus-Strong algorithm [8]. As a first experiment, we have developed a plucked network-string. When the string loop (the recirculating network audio stream) is excited, the pitch of the resulting wave corresponds to the inverse of the (instantaneous) RTT. The resulting wave is directly transduced to the audio output. Because the stream is driving the network at audio rates (e.g., from 8kHz to 96,000kHz), fine-grained jitter is heard as fluctuations of the loop's pitch. With longer network delays, SoundWIRE produces actual echoes (or very long strings); with shorter delays, it sounds like a pitched instrument.

The first test involved a QoS policy that we could enable / disable while sending the audio round-trip between the SC2000 conference in Dallas and Stanford, Figure 2. In order to simulate network congestion, four routers were added to the edges of the wide area network (Abilene), two at each location (labeled SND-GW / Q-GW and Edge / Access), and artificial traffic was injected into the private segment between each pair. We used UDP, a protocol in which lost packets are not retransmitted and which relies heavily on QoS for consistent performance. A scheme to prioritize packets according to source / destination addresses was implemented, so that our audio traffic could take the "carpool lane." With the QoS policy disabled, those packets were subject to network congestion resulting in loss, delay and obvious audio glitches. With the policy enabled, the audio data received priority queueing at the routers, and the audio was as clean as if it had been streamed through an uncongested network.[6]

A second, ongoing test involves SoundWIRE and TCP, a protocol which guarantees data will arrive intact (through a scheme that re-transmits ones that are lost). TCP is "network-friendly" because it backs off its re-transmission under congestion, however, this behavior can result in intolerably late arrivals for interactive real-time applications. TCP-RTM [9] is designed to reduce this latency under network load and make the overall QoS acceptable without changes to router policies.

4. LONGER EVENTS: SOUNDPING

Parameterized display can map few or several dimensions for any of several performance measures which are of interest in monitoring networks. SoundPing is based on the standard ping and translates its output to musical parameters.

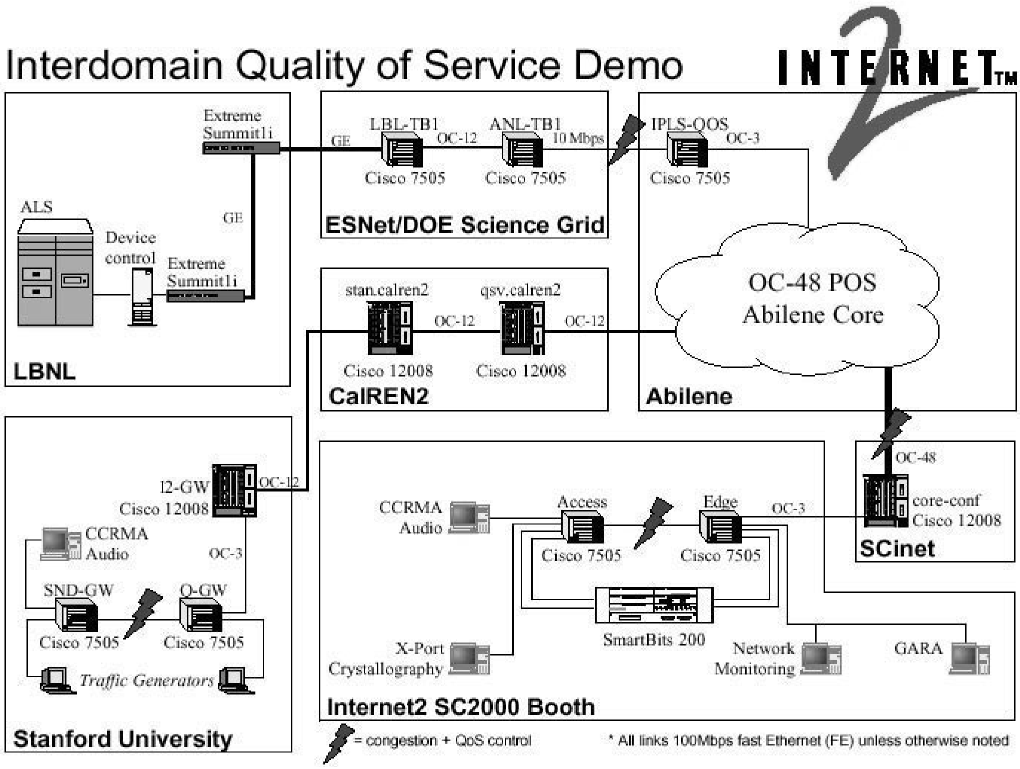

Figure 2: Schematic diagram of network path between CCRMA, Stanford and SC2000, Dallas for audio-based testing of a QoS method. Artificial traffic was injected between pairs of routers at the network's edge. With the QoS method, consistent audio delivery was possible under heavy congestion. Users of this first long-haul SoundWIRE deployment experienced ca. 150ms. RTT latency. Future versions of the utility will be closer to the actual network latency (on the order of 50ms. for the same path).

4.1. RTT, Average RTT

SoundPing synthesizes a tone for each real-time ping event or data point in an archive. Round trip time (R) is mapped inversely to pitch (f), and a scaling constant k is applied to the RTT of wide-area networks to raise the pitches to the same range produced by faster local-area networks:

$$f = \frac{1}{sR}, \quad \text{where} \quad s = \begin{cases} 1 & : \quad R \leq 5ms \\ k & : \quad R > 5ms \end{cases} \tag{1}$$

We have found that a useful value of k is 0.025.

Another parameter of interest is to accumulate an average RTT value over a time window. Useful mappings include making the ping tempo (T) at time n directly depend on the average RTT:

$$T[n+1] = T[n] + T_{offset} + cR_{avg}, \tag{2}$$

where T_{offset} is a minimum interval between pings and c is a scalar on the average.

4.2. Packet Loss

For a variety of reasons, even well-functioning wide area networks incur a certain amount of packet loss. Packets can be dropped due to errors in routing dynamics (router congestion, route changes) or possible equipment malfunction. Ping loss is expressed as the percentage of test packets sent but not echoed. Values $< 5\%$ are not uncommon and affect the choice of protocol for a given application, e.g. TCP for guaranteed delivery or UDP for data that can tolerate dropouts (but which need to avoid the added TCP overhead and latency).

The simplest sonification of loss uses the method above in which lost packets are simply musical rests (non-events).

4.3. Jitter

Variance in RTT, or jitter, is expressed as the percent standard deviation of a sequence of RTT values. Low jitter along with known RTT bounds (minimum and maximum values) have direct bearing on the specification of *playback buffers* for real-time signals. The application's buffer length must be long enough to cope with late arrivals, so that uninterrupted signals are delivered to media devices (such as the audio DAC). Timbre parameter mappings, for example, can be used to highlight level of jitter.

4.4. Other Measures – Future Work

Sonification of some further performance measures are being implemented in SoundPing's final form.

4.4.1. Asymmetry

Obtaining one-way trip time is difficult because it requires precise synchronization of both hosts. An external reference such as GPS is needed to reach millisecond-level accuracy. In the future, we hope to test a sonification method in which a plucked string tone, with its pitch sounding the RTT, chooses its pluck position according to the directional asymmetry. A perfectly symmetric path would have the timbral character of a perfectly mid-string pluck (which is missing the fundamental). The familiar changing timbre often heard on stringed instruments would provide a revealing display.

NETWORK TYPE	delay (ms.)	freq (Hz.)	PERCEPT
localhost	.025	40k	ultrapitch
LAN	.25	4k	pitch
WAN (1)	25	40	pitch
WAN (2)	50	20	infrapitch
WAN (3)	100	10	vibrato
WAN (4)	200 - 1000	5 - 1	rhythm

Table 1: RTT vs. musical rates, 1) within West Coast, USA next-generation backbones, 2) SC2000 to Stanford [6], 3) national-scale next-generation, 4) commodity Internet

4.4.2. Throughput

Maximum transfer rate (expressed in bits-per-second) can be calculated indirectly from RTT and packet loss rate:

$$Rate \leq \frac{MSS}{RTT} * \frac{1}{\sqrt{p}} \tag{3}$$

where $Rate$ is the transfer rate or throughput, MSS is the maximum segment size (fixed for each internet path, typically 1460 bytes), and p is the packet loss rate [10]. An improved, but more complex, form of the above formula which takes into account additional information about TCP, is found in [11]. Throughput, calculated as a running parameter, can be mapped in a number of ways. Pluck strength or amplitude seem good choices for our final design.

4.4.3. Traceroute

The traversing of data through multiple routers can be represented as a "ping chord." We will use traceroute and ping statistics to create multiple pitches for each event. As the segment times vary, so will the chord constituents.

5. TIME SCALES

Our ears are sensitive to structures across a wide range of time scales. Coincidentally, LAN and WAN RTT values fall directly within the gamut corresponding to pitch and rhythm. Table 1 shows the range of these temporal correspondences. Note that the fastest ping times approximate the typical audio sample rate of 44.1 kHz.

Listening to music, we are simultaneously experiencing different temporal levels, from short-term phenomena to long-term structures. A similar layering is reported in a model for mapping spatial features to different time scales of auditory display in [12]. We will show that the rough classification scheme which was reported is also relevant for the features in the present temporal data.

One can imagine a pattern of SoundPing tones in a hypothetically perfect network: the "melody" would be monotonously constant. Opposite to this, one can imagine an extremely poor network whose pings are random. From the graphs in Figure 3, we can see that actual network tests lie between these extremes. In general, things are well-behaved, but never constant and fit the class of "ambient time – perceived as always present." Occasional isolated outliers occur and fit the class of "event time – perceived as irregularly spaced singular events." Also, as in e), faster "rhythmic time – perceived as changes to events inside auditory streams."

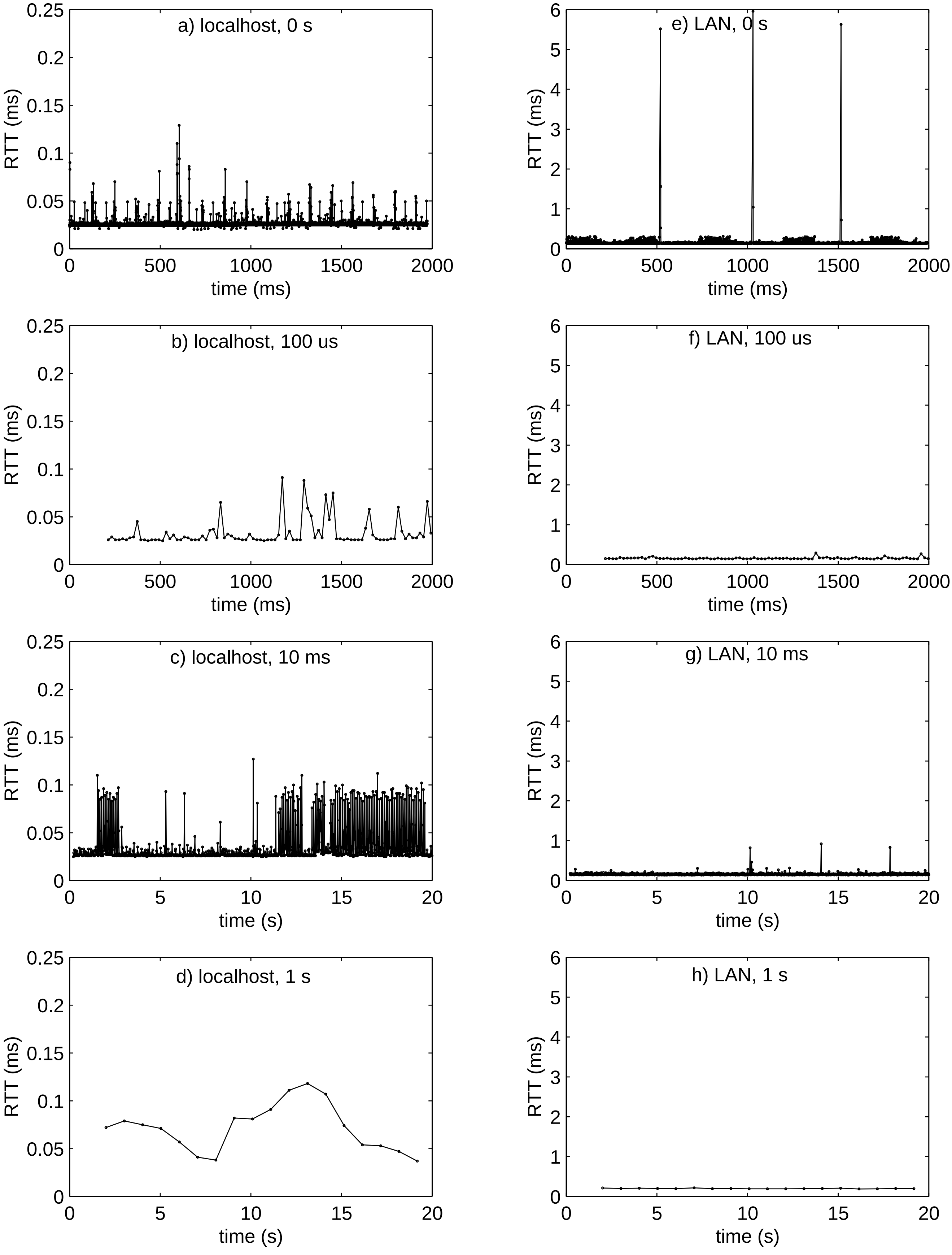

Figure 3: Measured RTT series from pinging localhost (left) and a LAN host (right) at different rates (full flood, $100\mu s$, $10ms$, $1s$). Temporal layers "zoom in" going up.

Additionally, we see the presence of very fast jitter in a) or "spectral time – perceived as variations in timbre (and/or localization)." And sometimes, broad shifts of regimes which are closer to formal events of music as shown in c) and d).

Important structures are masked from view by choosing different temporal perspectives. The large, regular, discontinuities of e) are missing from slower samplings of the same path. Through listening experiences, one learns to discern the differences between operating systems (which affect how pings are serviced) and to detect system load variation and network congestion.

As the design evolves we intend to offer a utility with a choice of granularity from fine to coarse that will be able to "encode" a combination of simultaneous temporal layers. The roster of cues underlying instrumental identity suggest possibilities in addition to those of pitch and rhythm which we're already using:

Faster than the pitch period, ultrapitch phenomena include aperiodicity, upper partials, stochastic components, and from the point of view of the listener, the arrival times of different radiation angles. These qualities belong to the instrument and acoustic chain (the playing/listening/recording environment) and are not under conscious control of the performer (though some of these are modulated in a pitch-synchronous way [13] and some are modulated as nuances of performance [14]).

Infrapitch phenomena include subharmonics, attack types, vibrato and other modulations and shadings of dynamic level. While these are still qualities that give strong cues to instrument identity, some are under the player's control as a "byproduct" of expression. They serve as strong cues to style and the qualities of an individual performer, and as with pitch and ultrapitch layers these qualities are extremely audible.

Longer still are the fixed qualities of an instrument: choice of reed, formant structure, etc. which are choices made ahead of time by the player, the instrument builder (or the sound designer).

The most elaborate experiment we have performed yet in terms of SoundPing parametric display is a music and art installation entitled "Ping" [15][16]. The gallery viewers experience a continuous monitoring of RTT from hosts which they can select or which the piece cycles through. A rich variety of melodies and rhythms result from combinations of up to four simultaneous targets, each of which is receiving and sounding two streams of pings in stereo. Shadings of instrumental quality and tonal harmony follow jitter and average RTT.

6. ACKNOWLEDGMENTS

The authors are indebted to many collaborators. Special mention is due individuals at CCRMA: Scott Wilson, Dave Chisholm, Fernando Lopez-Lezcano; SLAC: Les Cottrell, Warren Matthews; Internet2: Ben Teitelbaum, Stanislav Shalunov; Ping Art: Greg Niemeyer, Richard Humphrey. The project was funded by NSF Award 9977246.

7. REFERENCES

[1] C. Chafe, et al., "A Simplified Approach to High Quality Music and Sound Over IP," in *Proc. COST-G6 Conference on Digital Audio Effects (DAFx-00)*, Verona, 2000.

[2] D. Comer, *Internetworking with TCP/IP Vol. 1: Principles, Protocols, and Architecture*, Prentice Hall, 2000.

[3] PingER Project, Stanford Linear Accellerator Center, URL: http://www-iepm.slac.stanford.edu/pinger/

[4] Surveyor Project, IP Performance Metrics Working Group, URL: http://www.advanced.org/surveyor/

[5] T. McGregor and H.W. Braun, "Balancing Cost and Utility in Active Monitoring: The AMP Example," in *INET 2000 CD-ROM Proceedings*, Yokohama, Japan, 2000.

[6] SoundWIRE, "QoS Enabled Audio Teleportation," demonstration at SC2000, URL: http://apps.internet2.edu/html/qos-enabled-audio-demo.html.

[7] J.O. Smith, "Physical Modeling Using Digital Waveguides," *Computer Music Journal*, vol. 16, no. 4, pp. 74-91, 1992.

[8] K. Karplus and A. Strong, "Digital Synthesis of Plucked-String and Drum Timbres," in *Computer Music Journal*, vol. 7, no. 2, reprinted in C. Roads (ed.), *The Music Machine*, MIT Press, 1989.

[9] S. Liang, "TCP-RTM: Using TCP for Real Time Applications," at http://gregorio.stanford.edu/sliang/rtm.pdf.

[10] M. Mathis, J. Semke, and J. Mahdavi. "The Macroscopic Behavior of the TCP Congestion Avoidance Algorithm," *Computer Communication Review*, vol 27, no. 3, July 1997.

[11] J. Padhye, V. Firoiu,et al. "Modelling TCP throughput: A simple model and its empirical validation," in *Proc. SIGCOMM Symp. Communications Architectures and Protocols*, Aug. 1998, pp. 304-314.

[12] S. Saue, "A Model for Interaction in Exploratory Sonification Displays," in *Proc. Intl. Conf. Auditory Display*, Atlanta, 2000.

[13] C. Chafe, "Pulsed Noise in Self-Sustained Oscillations of Musical Instruments," in *Proc. IEEE Intl. Conf. Acoust. Sp. and Sig. Proc.*, Albuquerque, NM, USA, April 1990.

[14] M. Wanderley, "Non-obvious Performer Gestures in Instrumental Music," in A. Bradford et al. (eds), *Gesture-Based Communication in Human-Computer Interaction*, Springer-Verlag, Heidelberg, Germany, 1999.

[15] C. Chafe and G. Niemeyer, "Ping," in *010101: Art in Technological Times, exhibition catalog*, San Francisco Museum of Modern Art, 2001, pp. 54-55.

[16] C. Chafe and G. Niemeyer, "Ping," online streaming version, URL: http://crossfade.walkerart.org/