To the Edge with China: Explorations in Network Performance

Juan-Pablo Cáceres, Robert Hamilton, Deepak Iyer, Chris Chafe, and Ge Wang
{jcaceres, rob, ideepak, cc, ge}@ccrma.stanford.edu

Center for Computer Research in Music and Acoustics (CCRMA), Stanford University
The Knoll, 660 Lomita Dr., Stanford, CA 94305, U.S.A.

***Abstract*—The complex nature of distributed network-based musical performance served as the starting point for the Stanford University SoundWIRE group's 2008 collaboration with Peking University in Beijing, China. In planning and executing the multi-ensemble networked concert entitled "Pacific Rim of Wire" at Stanford on April 29, 2008, musicians and engineers from Stanford and Beijing undertook issues—technical and musical—ranging from the use of incompatible networking address protocols to the synchronization of performers, human and computer, across a 6000 mile span of network. This paper outlines the technical and musical strategies employed to support the production's demands, as well as specific methodologies employed for the realization of Terry Riley's *In C*.**

I. INTRODUCTION

Musical and technical strategies in local performances need to be re-factored when musicians are separated by long physical and acoustical distances. We address some of these issues in a large scale concert between two venues separated by 6000 miles: Stanford, California and Beijing, China. The SoundWIRE research group at Stanford [1] has been organizing and implementing technologies for real-time distributed performance for the last several years. The "Pacific Rim of Wire" concert highlights a number of key musical and technical challenges that still loom large above current attempts to perform using communication technologies.

The technical and musical demands of the "Pacific Rim of Wire" collaboration required the initiation of new types of network connectivity, the development of software to deal with next-generation Internet backbones, the implementation of musical strategies to deal with network-induced acoustical delays, and the organization of network-based metronomic systems with which a laptop orchestra synchronizes its performance with an ensemble of acoustic instruments.

Performing with ensembles based in other countries affords musicians and researchers the opportunity to explore edges of musical and technological strategies. The "Pacific Rim of Wire" performances included an ensemble of traditional Chinese instruments performing with the Stanford New Ensemble, using traditional Western instruments, all mixed together with the newly minted Stanford Laptop Orchestra (SLOrk) [2]. Such a collaboration also offers a unique opportunity to explore the manner in which traditional pieces can be performed in this medium. As part of this effort, a performance of Terry Rilley's *In C* is showcased.

II. COLLABORATION WITH CHINA

In April of 2008, the annual Stanford Pan-Asian Music Festival [3] turned its focus towards China, featuring its music and musicians from that country. Seeking a forward-looking production that would bridge the geographical distance between American and Chinese cultures, maestro Jindong Cai and CCRMA faculty member and SLOrk director Ge Wang conceived the "Pacific Rim of Wire", a networked musical collaboration between musicians at Stanford and in Beijing, performing together across fast Internet connections. Based on the ongoing research of CCRMA's SoundWIRE group, directed by Chris Chafe, the "Pacific Rim of Wire" concert would make use of SoundWIRE's JackTrip software [4] running on Linux, allowing the use of low-latency uncompressed bi-directional multi-channel audio streams.

Working with Kenneth Fields of the Peking University and China's Central Conservatory of Music, network testing between CCRMA and the Computer Science/Networking department of Peking University initially showed encouraging results, but network traffic patterns and insufficient stability left a number of problems with no immediate resolution. With the goal of better understanding the specific issues at hand, members of the SoundWIRE group traveled to Beijing to meet with their team and to run additional tests.

Although Peking University serves as a key hub in China's CERNET2 next-generation education and research network, testing conducted from a multimedia conference room within the University grounds showed that the Stanford-Beijing connection was in fact not utilizing the CERNET2 connection, running instead on China's standard first-generation Internet. As CERNET2 required the use of the IPv6 protocol, and as JackTrip only supported IPv4, some software changes were required.

A. Reaching China: a path to CERNET2 and IPv6

Connecting to Beijing with the requirements for a high-quality low-latency musical experience requires the use of the fastest and most stable high-bandwidth networks and backbones available. CERNET2 [5] is the most well-provisioned Internet backbone in China, with speeds of 2.5∼10 Gbps (Giga bit per second). CERNET2 is a native IPv6 [6] backbone; to connect to it, hosts are required to use IPv6 protocols.

Fig. 1. Simplified network path between Stanford and Peking University during the "Pacific Rim of Wire" concert

The development of IPv6 started with the need to address the scaling problems caused by Internet growth, and hence the need for more IP numbers than the ones presently provided by IPv4, the most widespread protocol currently in use. IPv4 uses a 32 bit address space, while IPv6 was designed with 128 bits, with a potential (assuming 100% efficiency) of addressing $3.4 \cdot 10^{38}$ nodes. Based on even the most pessimistic estimates, IPv6 may provide over 1500 addresses per square foot of the earth's surface [6].

To communicate with CERNET2, Stanford University peers via Internet2 [7], the U.S. research and education network. IPv6 static routing was set up, allowing direct communication with the IPv6 Internet from Internet2's Abilene backbone.[1]

Stanford University's network was still exclusively IPv4 and required that Stanford hosts use tunneling to connect to the IPv6 router. The portion of the connection that runs inside the university, i.e. from the CCRMA computer to the IPv6 router, runs within a *tunnel*: over the piece of network that only understands IPv4, the IPv6 packets "travel inside" IPv4 ones. Packets are encapsulated and decapsulated at each end of the tunnel—the *host* and the *router*. Across the rest of the network packets travel as normal IPv6 ones. In testing, the overhead of the encapsulation/decapsulation was found to be insignificant, without any important processing spikes on the IPv6 Stanford router, translating into no additional latency.

Figure 1 shows a simplified version of the network path between Stanford Campus and Peking University during the concert. The packets first travel from Stanford to Los Angeles, then cross the Pacific Ocean, passing through Korea before finally reaching China's CERNET2. The connection was symmetric—packets in both directions follow the same network path—and highly stable. The round trip time (RTT), measured with `ping6`, was ∼220 milliseconds.

The "Pacific Rim of Wire" performance made use of full-duplex, uncompressed audio (thus avoiding additional latency and audio artifacts of perceptual-audio compression), at 16bits and a 44.1KHz sampling rate, equivalent to Compact Disc quality. Two channels of audio were sent to the concert hall stereo PA speakers with one extra channel used for synchronization in the performance of Terry Rilley's ***In C*** (see Sec. III). The software used was an IPv6 version of JackTrip[2] [10], [4], a system for multi-channel uncompressed audio streaming.

Video streaming was done using the free open-source and cross platform software VLC [11], that supports various video codecs and streaming protocols. The input stream from the digital video camera was set to high-definition quality (720x480 pixel resolution). The deinterlaced video was transcoded using the MPEG4 codec and streamed as UDP packets, both operations provided by VLC. As there is an inherent trade-off between compression and bandwidth, the encoder settings were selected to minimize processing and thereby avoid

---

[1] The static router was specifically connected with the CENIC HPR (high-performance research network) 10Gbps backbone [8].

[2] The machine used in the performance was running Fedora distribution with Planet CCRMA [9].

Fig. 2. "Pacific Rim of Wire" concert: performance of Terry Riley's *In C*. Onstage: Stanford Laptop Orchestra (SLOrk) and Stanford New Ensemble. Onscreen: real-time video of musicians at Peking University.

latency, making use of available bandwidth as needed.

Camera movements in concerts are typically not very aggressive (low-motion). Video encoding exploits this fact effectively with a peak bandwidth utilization that did not exceed 8 Mbps (Mega bit per second) with camera movement and less than 1 Mbps without.

The total video latency was on the order of one second, composed of image capture delay in the camera, network delay and encoding/decoding time. The audio and video were not synchronized as both were streamed separately and had significantly different latencies. Although this might seem disruptive for the performers, our previous experience in network performance shows that musicians usually don't look at the video when they perform; it serves primarily the purpose of providing an experience for the audience—while also adding additional reassurance and comfort to the musicians during setup, discussion and other communication needs. Until video can match audio in terms of latency, the trade-offs for synchronizing video and audio are a significantly higher bandwidth utilization (for uncompressed video) vs. a correspondingly longer latency for audio (to match video codec lags).

With audio, any small dropouts or artifacts can be very noticeable and potentially annoying, focusing attention on glitches and sound quality [12]. In comparison, dropouts and latency in video delivery seem more tolerable. While uncompressed video is preferable, its enormous bandwidth demands and the difficulty of obtaining video cameras with fast capture motivated us to employ video encoding this time around. Uncompressed should provide a much better solution for the future.

B. *Nested Rims of Wire and Laptop Orchestra*

Within our wide-area, ocean-spanning network connecting Stanford and Beijing, an onstage *local area network* at Stanford University kept the computers in the laptop orchestra tightly synchronized. Each of the 20 hosts was connected via wireless Ethernet to an 802.11n switch, and used Open Sound Control [13], [14] to transmit low-latency control messages across the ensemble. Equipped with a custom hemispherical speaker array and paired with a human performer, each laptop station represented a single, localized meta-instrument with its own sonic presence and identity. With 20 such stations, the Stanford Laptop Orchestra leveraged its capability to project an ocean of sound, while fusing it with that of acoustic instruments playing on the same stage (Fig. 2). In the "Pacific Rim of Wire", these two aspects of synchronization and sound projection were fully explored in our networked performance of Terry Riley's *In C*. Here the laptop orchestra contributed a point of local synchronization as well as a centralized, dynamic sonic "anchor" for musicians at Stanford and Beijing. In the next section, we present and discuss our computer-mediated, wide-area, and yet rather traditional realization of *In C*, combining SoundWIRE, laptop orchestra, and acoustic musicians.

III. A BI-LOCATED TERRY RILEY'S *In C*

To best showcase the trans-continental collaboration between Stanford and Peking University, the decision was made to perform an ensemble musical work featuring performers located in both locations. A performance of Terry Riley's *In C*, led by Michael Bussiere, was performed by participants of the 2008 ANET II (High Quality Audio over Networks) Summit at the Banff Centre for the Arts [15]. This experience suggested that Riley's work might prove a good choice for the Beijing collaboration.

For the performance between Stanford and Beijing, instrumental performers at Stanford playing a variety of traditionally Western instruments joined instrumental performers at Beijing, playing a variety of traditional Chinese instruments, and the Stanford Laptop Orchestra. The choice to perform *In C* was in hindsight even more fitting than previously intended as Terry Riley himself brought the work to Beijing in 1989 where he performed and recorded the work with Chinese musicians of the Shanghai Film Orchestra performing on traditional Chinese instruments [16].

A. *Performance Details*

Composed and premiered in 1964, Riley's *In C* consists of 53 melodic patterns—or cells—each composed with a loose tonal center based around the pitch-class C. Instrumentation for *In C* is not set by the composer and can be performed by virtually any instrument capable of producing diatonic pitches. The instructions for the score require that beginning with the first musical cell, each pattern must be played in sequence by each performer, moving through the sequence of cells at their own discretion. Performers may choose to repeat cells as many times as they wish and may also pause between performance of different cells. The work ends after all performers have arrived at the final cell of the composition.

Riley's instructions in the written score [17] include the following:

> Each pattern can be played in unison or canonically in any alignment with itself or with its neighboring patterns. One of the joys of *In C* is the interaction of the players in polyrhythmic combinations that spontaneously arise between patterns. Some quite fantastic shapes will arise and disintegrate as the group moves through the piece when it is properly played. [...] The ensemble can be aided by the means of an eighth note pulse played on the high c's of the piano or on a mallet instrument. [...] All performers must play strictly in rhythm and it is essential that everyone play each pattern carefully. It is advised to rehearse patterns in unison before attempting to play the piece, to determine that everyone is playing correctly.

In C presents several challenges for a distributed network performance context, one of the most important being that it requires tight synchronization between musicians. It is a well known phenomenon that rhythmic synchronization is problematic when the acoustic delay between musicians becomes too long, with significant problems occurring at delay thresholds of just 20 milliseconds [18]. Faced with a single-direction base delay path between Beijing and Stanford of approximately 110 milliseconds—a delay already significantly greater than this 20 millisecond threshold—it was clear that the goal of a rhythmically-synchronized distributed ensemble performing with a signal path greater than 6000 miles would require a different solution.

B. *Distributing the Pulse*

As Riley's instructions indicate, one of *In C*'s most striking features is the work's ability to create complex polyrhythms through the repetition and alignment of each musical phrase into patterns of tight rhythmic synchronization. The use of an audible metronomic pulse allows performers to concentrate on the phrasing and alignment of tonal and rhythmic patterns between their own individual performance and the performances of each member of the ensemble, safe in the knowledge that each musician is locked in step with the same pulse. However, while a metronomic pulse will clearly aid performers sitting in the same performance space, the introduction of a significant signal path latency and potentially a dynamic latency effectively renders a static metronomic pulse useless: this leaves the two ensembles out of step.

The solution put forth to provide a stable pulse for the entire distributed ensemble for this performance of *In C*—the technique of *feedback locking* [19]—relies on a metronomic pulse transmitted to both performance locations, with its rate based on the current dynamic signal path delay between Stanford and Beijing. The tempo of the audible eighth note pulse, set, performed and adjusted by a musician listening to a hidden audio channel (i.e. not broadcast to the ensemble and audience) is based on the round-trip network feedback. In this manner, a tight rhythmic alignment between both locations can be maintained. The RTT in the "Pacific Rim of Wire" concert was $\sim$220 milliseconds. A simple calculation serves to obtain the tempo for the performance in beats per minute (BPM):

$$\text{Tempo} = \frac{60 \text{ (seconds in a minute)}}{0.220 \text{ (seconds)}} = 272.73 \text{ (BPM)}$$

This result is used for the tempo of the eighth note (♪) pulse. The piece was performed at approximately 270 BPM.

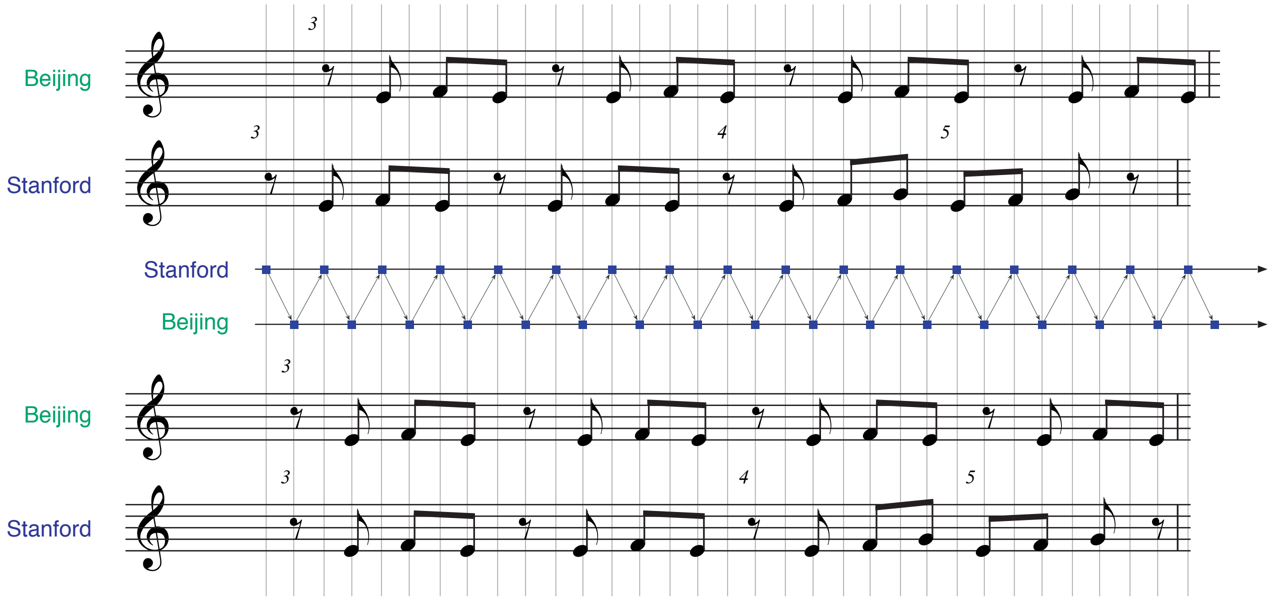

Fig. 3. Feedback locking *In C*. The center of the figure shows in blue squares the pulse as heard in both locations. A performer at Stanford locks with its own feedback. The top part shows the musical cells as performed at Stanford and heard from Beijing. The bottom shows the musical cells performed in Beijing and the ones heard from Stanford.

Figure 3 shows the feedback locking approach, with a metronomic pulse originating at Stanford sent to performers in Beijing. The blue square at the center of the figure represents the pulse. The horizontal time axis illustrates the arrival of the pulse to each location. Three musical cells performed by one performer at each location are shown as well as an example of an interesting extension of Riley's desire for a variable interlocking composition: at any given time, the performance will sound significantly different in each location. This extension of the composer's intent is happily furthered by the act of network distribution, simultaneously introducing two unique variations on the piece into the world during any given performance. Figure 4 shows the compound musical figure heard at one point in performance, where a performer in Beijing loops musical *Cell 3* four times against a performer in Stanford performing two loops of *Cell 3*, immediately followed by *Cell 4* and *Cell 5*.
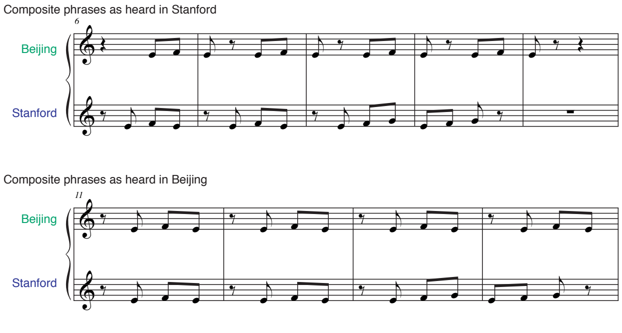
Fig. 4. Composite phrases resulting at Stanford (top staff) and Beijing (bottom staff)

C. *ChucK Implementation*

To create both a sonified metronomic pulse which could not only be easily tuned by the network engineer but could also regulate timing for the 20 laptop performers, as well as a performable version of *In C* for Laptop Orchestra, a custom client-server implementation was written in the ChucK language [20]. A slider on the screen of a laptop in front of the network engineer simultaneously regulated both the audible metronomic pulse and an inaudible Open Sound Control data pulse, projected to each of the 20 laptop performers over a wireless 802.11n network. Performers on each laptop would select an instrument at the start of the performance from a selection of digital physical models in the Synthesis ToolKit (STK) [21]. As seen in Figure 5, each laptop performer was presented a small GUI window with controls to start and stop musical cells numbering 1-53, as well as a toggle switch to loop the selected cell. The timing for each individual note was clocked to the Open Sound Control pulse and subsequently was always perfectly in sync with the network-synchronized audible metronome. In this way, the laptop instruments served as an audible reinforcement to the metronomic pulse.

IV. CONCLUSIONS

The application of recent low-latency audio transmission technologies to a real musical scenario served as a good example of the challenges facing musicians and engineers alike in the realization of real-time networked-based musical performance. The showcasing of these techniques resulting in a concert between Stanford University and China's Peking University brought forth a series of network-based obstacles which required novel

Fig. 5. *In C*'s ChucK client interface for laptop performer. The picture shows the interface as used by one of the 20 SLOrk performers.

solutions to produce satisfying engineering and musical results.

The successful "Pacific Rim of Wire" concert has shown the SoundWIRE group that with the addition of networking and distributed performance practice, it is possible to "enhance" the experience of existing musical repertoire. The presentation of Terry Riley's *In C*, expanded Riley's own concept of a loose-but-synchronous ensemble to include dual related but significantly different performances in each distributed location. The application of musical strategies that use network-based time-delay to synchronize and to distribute musical patterns was successfully applied in this performance, paving the way for future distributed performances of rhythmically-strict works. The successes outlined above have shown us that through the use of these techniques, we can successfully synchronize musicians through a consistent distribution of the musical pulse.

Network performance has proven to be also a good opportunity to experiment with non-traditional instrumental combinations. A laptop orchestra performing in real-time with a traditional Chinese erhu, combined with western orchestral instruments may not be the most standard instrumental ensemble, but through the use of distributed performance practices, such a grouping, even performed in a small performance space, is made possible. In this manner, the use of powerful networking technologies has shown itself as an effective paradigm for musical performance, well worthy of future technological and musical efforts.

ACKNOWLEDGMENTS

Ma Hao, Kenneth Fields and Haku Wang from Peking University, Lea Roberts of the Stanford's Networking Systems group, Carr Wilkerson, Fernando Lopez-Lezcano and Chryssie Nanou from CCRMA, Scott Gresham-Lancaster and maestro Jindong Cai.

REFERENCES

[1] (2008) SoundWIRE research group at CCRMA, Stanford University. [Online]. Available: http://ccrma.stanford.edu/groups/soundwire/

[2] G. Wang. (2008) Stanford Laptop Orchestra (SLOrk). [Online]. Available: http://slork.stanford.edu/

[3] (2008) Stanford Pan-Asian Music Festival. [Online]. Available: http://panasianmusicfestival.stanford.edu/

[4] J.-P. Cáceres. (2008) Jacktrip: Multimachine jam sessions over the Internet2. [Online]. Available: http://ccrma.stanford.edu/groups/soundwire/software/jacktrip/

[5] (2008) CERNET2. [Online]. Available: http://www.edu.cn/cernet%202_1382/

[6] L. L. Peterson and B. S. Davie, *Computer Networks: A Systems Approach, 3rd Edition*, 3rd ed. Morgan Kaufmann, May 2003.

[7] (2008) Internet2. [Online]. Available: http://www.internet2.edu/

[8] (2008) Corporation for Education Network Initiatives in California, CENIC. [Online]. Available: http://www.cenic.org/

[9] F. Lopez-Lezcano. (2008) Planet CCRMA. [Online]. Available: http://ccrma.stanford.edu/planetccrma/software/

[10] C. Chafe, S. Wilson, R. Leistikow, D. Chisholm, and G. Scavone, "A simplified approach to high quality music and sound over IP," in *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*, Dec. 2000.

[11] (2008) VideoLAN (VLC). [Online]. Available: http://www.videolan.org/

[12] S. Gulliver and G. Ghinea, "The perceptual and attentive impact of delay and jitter in multimedia delivery," *Broadcasting, IEEE Transactions on*, vol. 53, pp. 449–458, 2007.

[13] M. Wright, A. Freed, and A. Momeni, "OpenSound Control: State of the art 2003," in *NIME '03: Proceedings of the 3th international conference on New Interfaces for Musical Expression*, Montreal, Canada, 2003, pp. 153–159. [Online]. Available: http://cnmat.berkeley.edu/publications/open_sound_control_state_art_2003

[14] M. Wright. (2002) Open sound control 1.0 specification. [Online]. Available: http://opensoundcontrol.org/spec-1_0

[15] (2008) The Banff Centre Programs—ANET II: High Quality Audio over Networks Summit. [Online]. Available: http://www.banffcentre.ca/programs/program.aspx?id=721

[16] D. M. Liang, T. Riley, and Shanghai Film Orchestra, "In C," Audio CD, Nov. 1992.

[17] T. Riley, "In C," Musical score, 1964.

[18] C. Chafe and M. Gurevich, "Network time delay and ensemble accuracy: Effects of latency, asymmetry," in *Proceedings of the AES 117th Convention*, 2004.

[19] J.-P. Cáceres and A. B. Renaud, "Playing the network: the use of time delays as musical devices," in *Proceedings of International Computer Music Conference*, Belfast, Northern Ireland, 2008, pp. 244–250.

[20] G. Wang and P. R. Cook, "ChucK: A concurrent, on-the-fly, audio programming language," in *Proceedings of International Computer Music Conference*, Singapore, 2003.

[21] P. R. Cook and G. P. Scavone. (2007) The Synthesis Toolkit in C++ (STK). [Online]. Available: http://ccrma.stanford.edu/software/stk/