REGION ONE: THE INTERNET FOR ENSEMBLE PERFORMANCE?

Panel hosted by ROBERT CUTIETTA; organized by CHRISTOPHER SAMPSON
University of Southern California

DISTRIBUTED IMMERSIVE PERFORMANCE

ELAINE CHEW AND ALEXANDER SAWCHUK

ROGER ZIMMERMANN; THE TOSHEFF PIANO DUO (VELY STOYANOVA and ILIA TOSHEFF);
CHRISTOS KYRIAKAKIS; CHRISTOS PAPADOPOULOS; ALEXANDRE FRANÇOIS; and ANJA VOLK

University of Southern California

Synopsis

The goal of Distributed Immersive Performance (DIP) is to allow musicians to collaborate synchronously over distance. Remote collaboration over the Internet poses many challenges, such as delayed auditory and visual feedback to the musicians and a reduced sense of presence of the other musicians. We are systematically studying the effects of performing under remote conditions so as to guide the development of systems that will best enable remote musical collaboration.

First, we present a narrative of our evolving distributed performance experiments leading up to our current framework for the capture, recording, and replay of high-resolution video, audio, and MIDI streams in an interactive collaborative performance environment. Next, we discuss the results of user-based experiments for determining the effects of, and a partial solution to, latency in auditory feedback on performers' satisfaction with the ease of creating a tight ensemble, a musical interpretation and adaptation to the conditions.

Overview

The DIP project explores one of the most challenging goals of networked media technology: creating a seamless environment for remote and synchronous musical collaboration. Participants in the performance are situated at remote locations, and the interaction occurs synchronously, as in ensemble playing rather than a master class scenario. One might ask:

- WHY create and study remote synchronous music collaboration environments? (Are we crazy?)
- WHO else has tried this? (Related work)
- WHAT have we done? (Recent experiments)
- WHAT have we found? (Latest results)
- HOW is this of relevance? (Impact for musicians)

This paper was presented by Elaine Chew and Alexander Sawchuk at the Region One meeting. E-mail {echew, sawchuk}@usc.edu

Is Synchronous Collaboration over the Internet Plausible?

We argue that synchronous collaboration over the Internet is indeed possible in many cases. Consider a trio distributed over distance on the North American continent as shown in figure 1(a). In the best of circumstances, when there is no network congestion and direct paths exist between all locations, the travel time (at the speed of light) between the different locations are on the order of tens of milliseconds (ms), as shown in figure 1(a). Consider the musicians in a large orchestra as shown in figure 1(b). Sound travels at a considerably slower speed than light—330 meters per second. Figure 1(b) shows some typical time delays between the time a musician makes a sound and the time his/her colleague hears the sound in a different section of the orchestra. Note that this delay is also in the order of tens of milliseconds. There is one main difference between the scenarios depicted in figures 1(a) and (b). In the remote ensemble in figure 1(a), the visual cues from the conductor are delayed, while in the orchestral situation in figure 1(b), there is negligible visual delay between the conductor and the musicians.

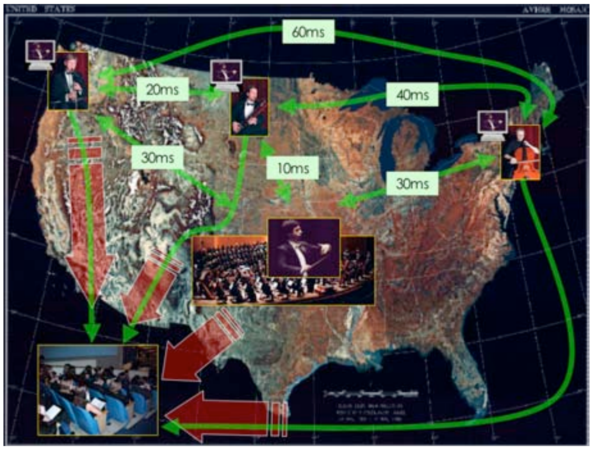


Figure 1(a). Musicians connected by a network.

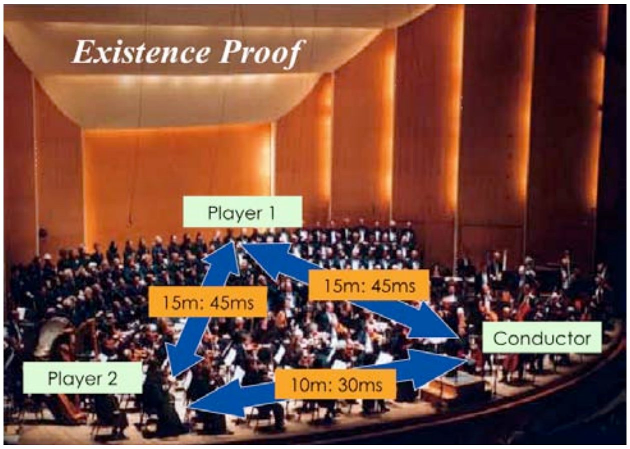


Figure 1(b). Musicians on stage.

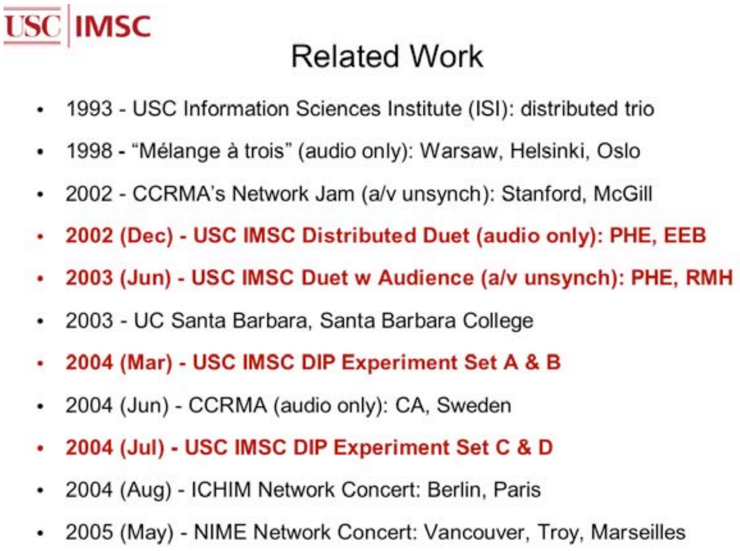
A viable remote collaboration environment for musical ensembles must minimize the audio and video signal latency among the musicians. Traffic on the Internet does not always flow at a constant rate. Hence, such a system must also ensure constant delay between the players.

Related Work

Many other groups have proposed and implemented systems for remote musical ensembles. One of the earliest attempts took place in 1993 at the University of Southern California's Information Sciences Institute in the form of a distributed trio (see descriptions by Schooler, 2001). In 1998, a performance titled "Mélange à trois" for three musicians was connected by audio signal only between Warsaw, Helsinki, and Oslo (see Kanki, 1998). More recently, several experiments have originated from Stanford's Center for Computer Research in Music and Acoustics, including a Network Jam (with unsynchronized audio and video) between Stanford and McGill Universities (2002), and an ensemble performance (audio only) between California and Stockholm (see Chafe, 2004). In 2003, a remote performance took place between University of California, Santa Barbara, and Santa Barbara College, in 2004, a network concert took place between Berlin and Paris at the International Culture Heritage Informatics Meeting, and in 2005, Gresham-Lancaster presented a three-way network concert between Vancouver-Troy-Marseilles at the International Conference on New Interfaces for Musical Expression.

The Distributed Immersive Performance Experiments

The Distributed Immersive Performance experiments at the Integrated Media Systems Center have been taking place since late 2002. Figure 2(a) shows the list of experiments in the context of the related work mentioned in the previous paragraph. Figure 2(b) shows further details of the experiments in the context of related work on media streaming at the University of Southern California (USC). Each experiment will be described in greater detail below.

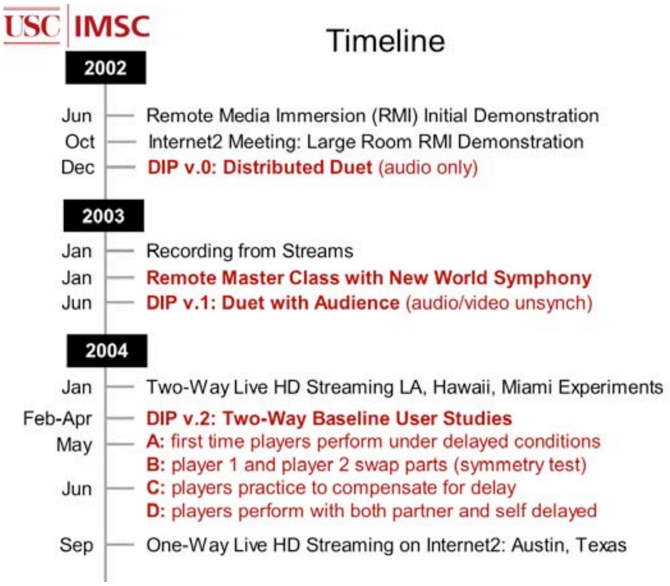


USC | IMSC

Related Work

- 1993 - USC Information Sciences Institute (ISI): distributed trio
- 1998 - “Mélange à trois” (audio only): Warsaw, Helsinki, Oslo
- 2002 - CCRMA’s Network Jam (a/v unsynch): Stanford, McGill
- **2002 (Dec) - USC IMSC Distributed Duet (audio only): PHE, EEB**
- **2003 (Jun) - USC IMSC Duet w Audience (a/v unsynch): PHE, RMH**
- 2003 - UC Santa Barbara, Santa Barbara College
- **2004 (Mar) - USC IMSC DIP Experiment Set A & B**
- 2004 (Jun) - CCRMA (audio only): CA, Sweden
- **2004 (Jul) - USC IMSC DIP Experiment Set C & D**
- 2004 (Aug) - ICHIM Network Concert: Berlin, Paris
- 2005 (May) - NIME Network Concert: Vancouver, Troy, Marseilles

Figure 2(a). USC DIP experiments and related work.



USC | IMSC

Timeline

2002

Jun — Remote Media Immersion (RMI) Initial Demonstration
Oct — Internet2 Meeting: Large Room RMI Demonstration
Dec — **DIP v.0: Distributed Duet** (audio only)

2003

Jan — Recording from Streams
Jan — **Remote Master Class with New World Symphony**
Jun — **DIP v.1: Duet with Audience** (audio/video unsynch)

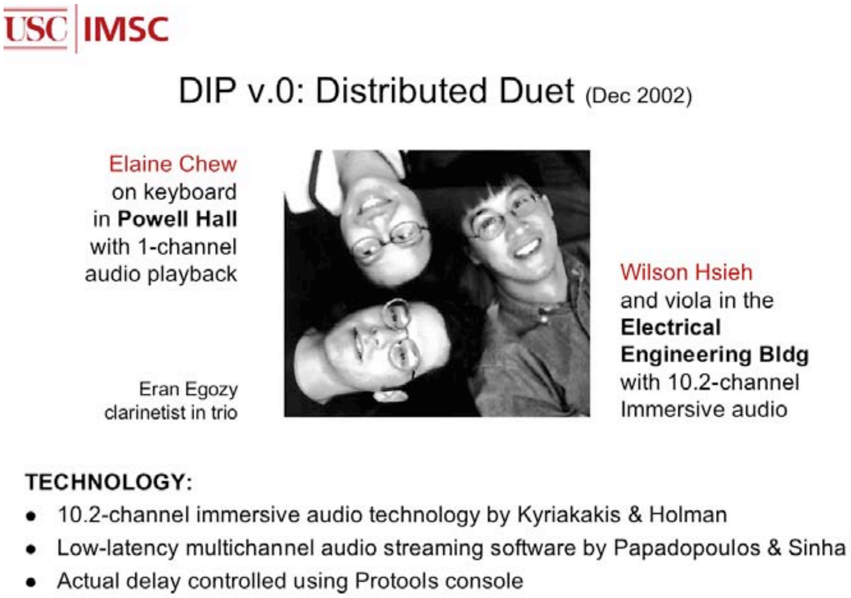
2004

Jan — Two-Way Live HD Streaming LA, Hawaii, Miami Experiments
Feb-Apr — **DIP v.2: Two-Way Baseline User Studies**
May — **A:** first time players perform under delayed conditions
B: player 1 and player 2 swap parts (symmetry test)
Jun — **C:** players practice to compensate for delay
D: players perform with both partner and self delayed
Sep — One-Way Live HD Streaming on Internet2: Austin, Texas

Figure 2(b). DIP timeline of experiments.

DIP v.1: Distributed Duet (December 2002)

Our first remote duet experiment took place on the USC campus between two buildings, Powell Hall (PHE) and the Electrical Engineering Building (EEB). The players were Elaine Chew in PHE on a piano keyboard with one-channel audio playback, and Wilson Hsieh in EEB playing the viola (see Figure 3) with 10.2-channel immersive audio technology developed by Christos Kyriakakis and Tomlinson Holman (see Kyriakakis et al, 1999, and Mouchtaris et al, 2000, 2003, 2005). The two locations were linked by low-latency multichannel audio streaming software created by Christos Papadopoulos and Rishi Sinha, and the actual audio delay between the two sites were controlled using a Protools console. The musicians played selections from Hindemith's *Sonata No.4* and Piazzolla's *Le Grand Tango*; the controlled audio delay ranged from close to 0ms to over 300ms.



USC | IMSC

DIP v.0: Distributed Duet (Dec 2002)

Elaine Chew
on keyboard
in **Powell Hall**
with 1-channel
audio playback

Eran Egozy
clarinetist in trio

Wilson Hsieh
and viola in the
Electrical
Engineering Bldg
with 10.2-channel
Immersive audio

TECHNOLOGY:

- 10.2-channel immersive audio technology by Kyriakakis & Holman
- Low-latency multichannel audio streaming software by Papadopoulos & Sinha
- Actual delay controlled using Protools console

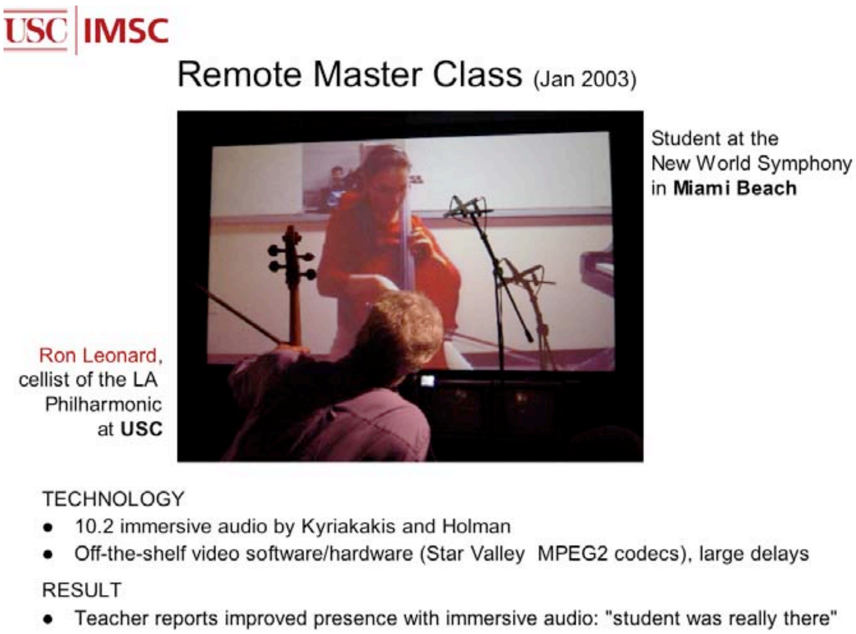
Figure 3. Members of the Aurelius Trio and conditions of first remote duet.

What we learned from these initial sets of experiments was that the musicians' latency tolerance was dependent on (1) the tempo of, and types of onset synchronization required in the piece; and, (2) the timbre of the instrument. For example, latency tolerance was higher for the languid first movement of the Hindemith Sonata No. 4 than for the final movement, which contains sharp and sudden attacks. For *Le Grand Tango*, the latency tolerance increased from 25ms to 100ms when the keyboardist switched from the accordion to the piano sound.

After some calibration of the 10.2-channel audio at EEB to make the acoustics sound more "natural," as they would in a concert hall, the violist felt more at ease. Finally, there was a distinct difference in the perspective of the performance at the two sites. To the violist, the pianist was almost always late, and to the pianist the violist was mostly late; this is because by the length of time it takes an audio signal to travel from one site to the other, its arrival is later than intended. This perspective difference would require that future experiments record the experience at both sites.

Remote Masterclass (January 2003)

In January of 2003, a remote masterclass took place between Powell Hall at USC and the New World Symphony as documented in Figure 4. This marked the first experiment combining audio and video streaming. The audio technology was Christos Kyriakakis and Tomlinson Holman's 10.2-channel immersive audio. We used off-the-shelf video software and hardware by Star Valley (MPEG2 codecs), which had large delays. The teacher, Los Angeles Philharmonic cellist Ron Leonard, remarked that he felt that the 10.2-channel immersive audio helped him feel that the "student was really there." The life-sized image was also important in improving the sense of a shared space. At one point, when the projector's bulb was overheating and a small monitor took its place, the teacher asked if the audio volume had been turned down.



Remote Master Class (Jan 2003)

Student at the
New World Symphony
in **Miami Beach**

Ron Leonard,
cellist of the LA
Philharmonic
at **USC**

TECHNOLOGY

- 10.2 immersive audio by Kyriakakis and Holman
- Off-the-shelf video software/hardware (Star Valley MPEG2 codecs), large delays

RESULT

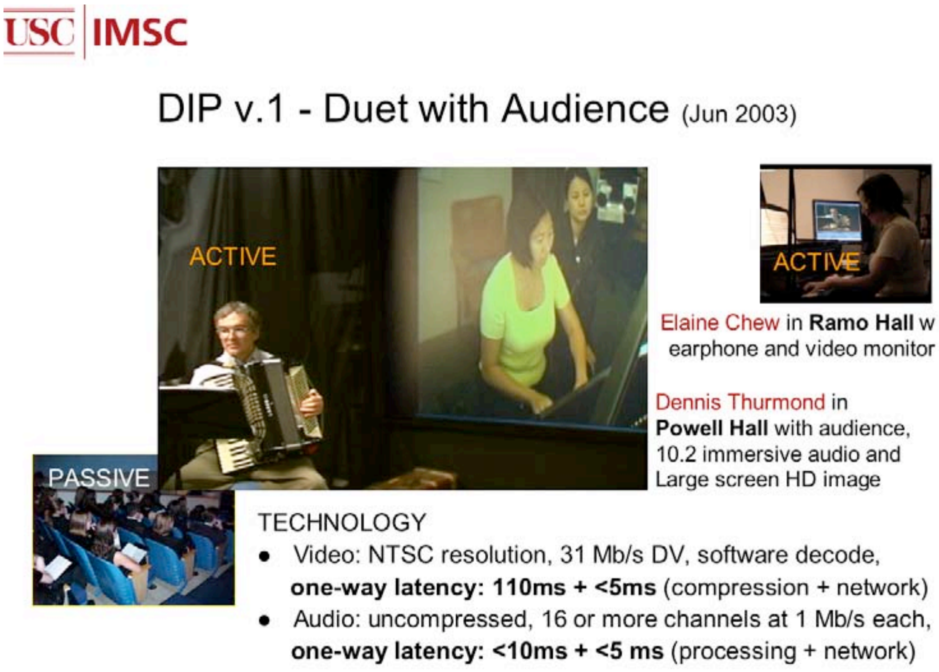
- Teacher reports improved presence with immersive audio: "student was really there"

Figure 4. Ron Leonard and New World Symphony student in remote masterclass.

DIP v.1 – Duet with Audience (June 2003)

Our first distributed ensemble experiment with audio and video links took place in June 2003 at the Integrated Media System's National Science Foundation site visit. The two musicians were located in Ramo Hall and in Powell Hall. Elaine Chew, on piano in Ramo Hall, had an earphone and video monitor as shown in the top right of Figure 5. Dennis Thurmond on accordion in Powell Hall was co-located with the audience with 10.2-channel immersive audio and large screen NTSC resolution (TV resolution) image.

The video latency was on the order of 115ms one way, and the audio latency approximately 15ms one way. Note that one has to consider the round-trip delay because the time from the moment a note is sounded until the time the musician hears the response to that note is essentially the roundtrip delay. The musicians performed Piazzolla's *Le Grand Tango*, which had an overall tempo of 120 beats per minute. The granularity of the events was mostly at the 16th-note level, meaning that the inter-onset-interval was around 125ms. At this rate, even a roundtrip delay of 60ms could be debilitating.



DIP v.1 - Duet with Audience (Jun 2003)

Elaine Chew in **Ramo Hall** w earphone and video monitor

Dennis Thurmond in **Powell Hall** with audience, 10.2 immersive audio and Large screen HD image

TECHNOLOGY

- Video: NTSC resolution, 31 Mb/s DV, software decode, **one-way latency: 110ms + <5ms** (compression + network)
- Audio: uncompressed, 16 or more channels at 1 Mb/s each, **one-way latency: <10ms + <5 ms** (processing + network)

Figure 5. Distributed duet with Dennis Thurmond and Elaine Chew.

We learnt that the large video delay (230ms roundtrip) made it unusable as a source of cues for synchronization. The musicians relied only on the audio signal, which had a roundtrip delay of under 50ms, for ensemble cues. The musicians compensated for the delay by anticipating each other's actions and scaling back on spontaneity to present a low-risk performance. Some artistic license was exercised to "make ends meet." Furthermore, co-location of the audience with one musician caused an imbalance in the ensemble dynamics. No matter what happened, the performer at the audience site, the accordionist at Powell Hall, had to make the final performance "work" and was thus at the mercy of the pianist at Ramo Hall.

DIP v.2 – Two-way Base-line User Studies (2004)

The objective of our next set of experiments is to measure and document qualitatively and quantitatively the effects of delay and other variables on immersion, usability, and quality in the DIP scenario. For these experiments, we enlisted the help of the Tosheff Piano Duo (*www.tosheffpianoduo.com*), Vely Stoyanova and Ilia Tosheff (see Figure 6). Founded in 1997, the duo has won prizes at international competitions in Tokyo, Bulgaria, Italy, Spain, and the United States. They are the first pair of pianists to be admitted to the Thornton School as a duo and are pioneers in the school's Protégé Program.

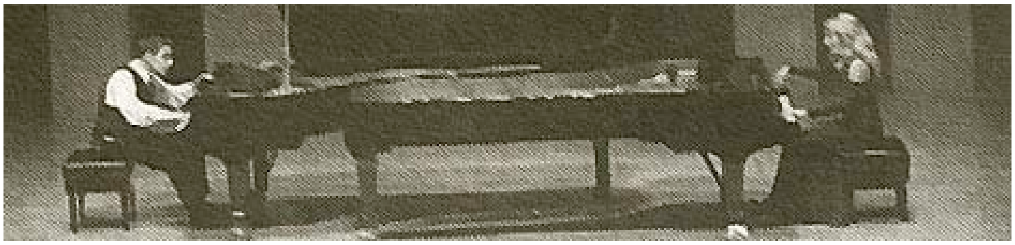


Figure 6. The Tosheff Piano Duo in concert (picture from *www.tosheffpianoduo.com*).

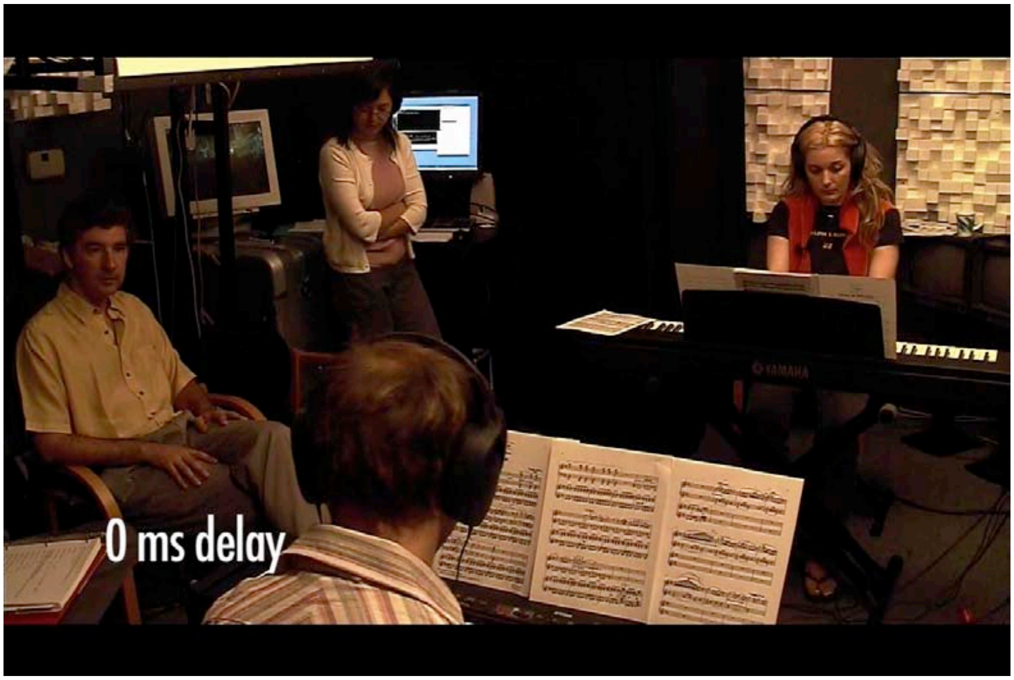


Figure 7. The Tosheff Piano Duo in face-to-face keyboard setup common to all experiments.

In our two-way baseline user studies, the two pianists were seated facing each other in the same room as shown in Figure 7. The audio and MIDI output from each keyboard and video from three high-definition (HD) cameras were streamed to the HYDRA database developed by Roger Zimmermann et al. (2003, 2004, 2005). Low-latency multi-channel audio streaming was made possible by Papadopoulos and Sinha. Audio delay was controlled from a Protools console. Figure 8(a) shows the equipment associated with each player, the database server, and a hypothetical remote audience. Figure 8(b) shows the data stream connections in the experiment setup.

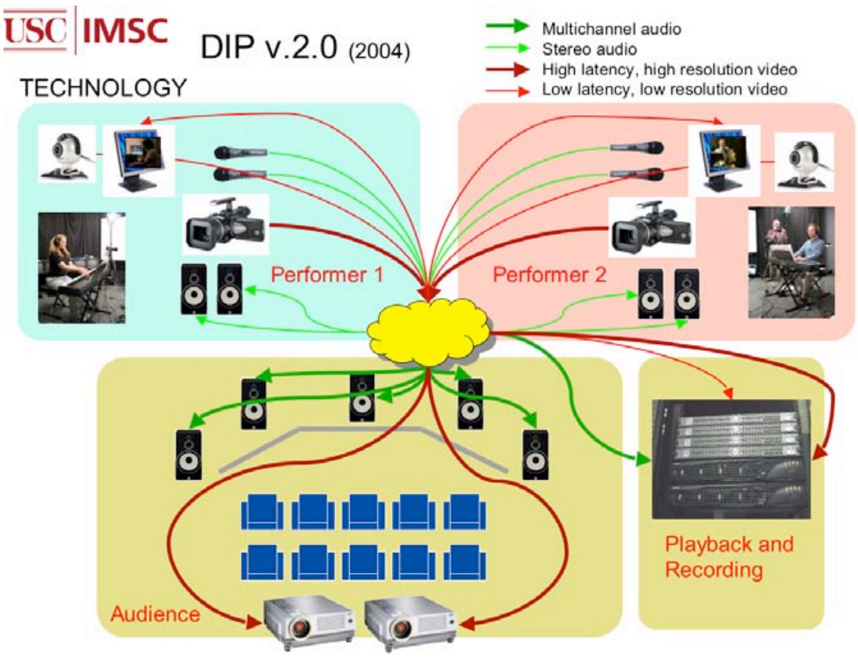


Figure 8(a). DIP v.2 equipment specifications.

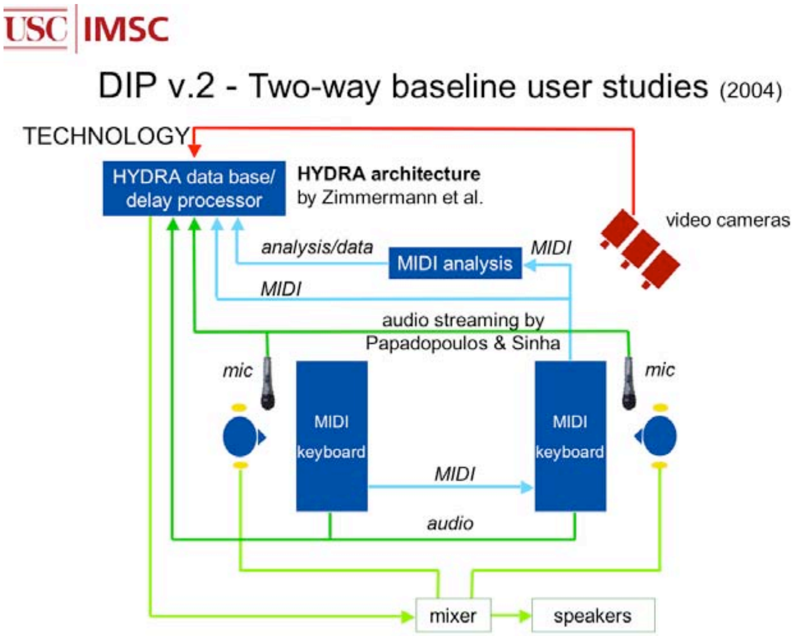

USC | IMSC
DIP v.2 - Two-way baseline user studies (2004)
TECHNOLOGY
HYDRA data base/ delay processor
HYDRA architecture by Zimmermann et al.
video cameras
analysis/data
MIDI analysis
MIDI
MIDI
audio streaming by Papadopoulos & Sinha
mic
mic
MIDI keyboard
MIDI keyboard
MIDI
audio
mixer
speakers

Figure 8(b). DIP v.2 data stream connections.

The Tosheff Duo was asked to play Poulenc's Sonata for Piano Four-Hands on two keyboards. The three movements of the sonata are the *Prelude* (tempo = 132bpm), the *Rustique* (tempo = 46bpm) and the *Finale* (tempo = 160bpm). At the end of each performance of each movement, the two pianists are asked the following questions:

- *How would you rate the ease of ensemble playing?*
- *How would you rate the ease of creating a musical interpretation?*
- *How would you rate the ease of adapting to this condition?*

Each rating was made on a scale of one to seven, with one being the easiest and seven being the hardest. Members of the duo are then debriefed, and their observations recorded. Chew et al (2004, 2005) are currently developing quantitative methods for measuring musical synchronization. We summarize here the players' responses to the questions for the following experiments:

A: first time players perform under delayed conditions
B: player 1 and player 2 swap parts (symmetry test)
C: players practice to compensate for delay
D: players perform with both partner and self delayed

In experiment set A, the players perform under delayed conditions for the first time. To eliminate any possible player-based bias in the data, we also conducted experiment set B, where the players swap parts. In each experiment, members of the duo sat facing each other so that the visual delay was essentially 0ms, and the audio delay was a randomly chosen number from the set {0ms, 10ms, 20ms, 30ms, 40ms, 50ms, 75ms, 100ms, 150ms}.

The overall result, shown in Figure 9, demonstrated that delays 50ms and under were generally considered to be tolerable. At 50ms, the musicians were conscious of the delay but were often able to compensate. Delay conditions at 75ms, 100ms, and 150ms were increasingly difficult, with 100ms being extremely difficult and 150ms almost impossible.

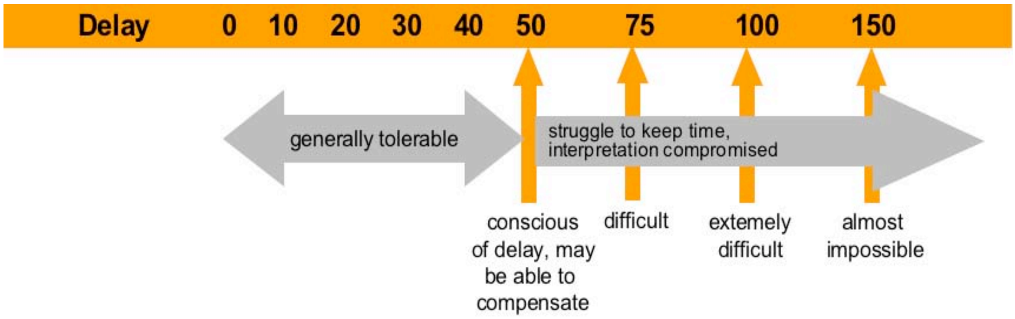

Delay
0
10
20
30
40
50
75
100
150
generally tolerable
struggle to keep time, interpretation compromised
conscious of delay, may be able to compensate
difficult
extemely difficult
almost impossible

Figure 9. Audio latency tolerance in experiment sets A and B.

Because the delay tolerance threshold appeared to be around 50ms, our next two sets of experiments focused on the region around 50ms. In experiment set C, the duo was asked to practice and strategize to compensate for the delay. The players were generally frustrated with the outcomes and with each other's perceived inability to stay together. At one point, they had the opportunity to put on the other person's headphones to better understand the different delay situations at both ends. After this experience, they asked to hear what it is the audience hears, which meant that the audio signal from their own keyboard would be delayed in transmission to their own headphones as well. This request resulted in experiment set D, where each player heard the audience's perspective, that is, both their own *and* their partner's playing delayed. Scenario D is shown in Figure 10, a composite from the video streams captured during the experiment.

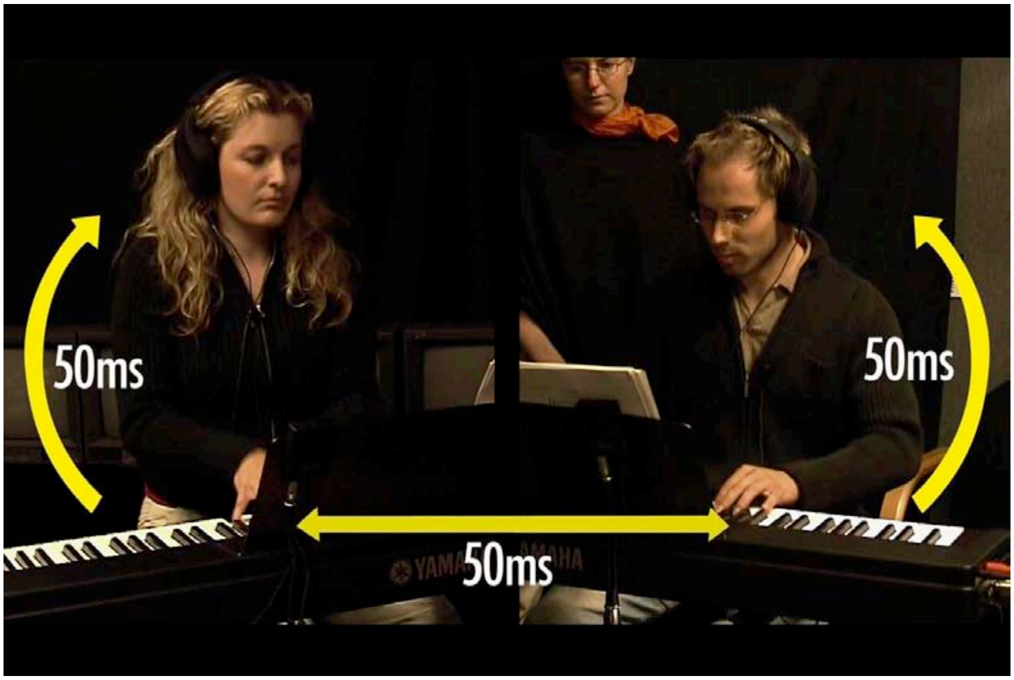


Figure 10. The Tosheff Piano Duo in Experiment D (split screen view) with 50ms audio delay.

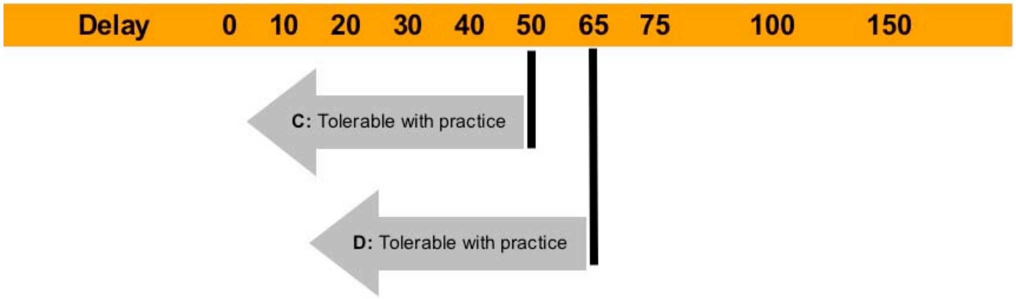


Figure 11. Audio latency tolerance in experiment sets C and D.

The players were noticeably much happier in condition D than in condition C. The overall tolerance threshold, originally at around 50ms for condition C, were shifted to 65ms for condition D (as shown in Figure 11). The explanation for this can be found in Ilia Tosheff's statement that when he is playing, he is not thinking about what his hands are doing. He focuses on what it is the audience hears, creates a mental image of what he wishes to portray and lets his hands do the rest. For a musician, hearing oneself delayed does not appear to be as difficult as hearing an unsynchronized (or unsynchronizable) rendition of one's own performance. In fact, organists in a large cathedral often have to cope with delayed sounds from their keystrokes.

Our preliminary results lead us to conclude that in remote collaborative performance where network delay is unavoidable, players may be willing to tolerate and adjust to delayed feedback of their own actions in order to achieve the experience of a common perspective. Further details of our distributed immersive performance experiments and the analyses of the experiment data are chronicled in Sawchuk et al (2003), Chew et al (2004, 2005), and Zimmermann et al (2005).

What Is the Impact for Musicians?

Ensemble performance over the Internet will promote new modes of musical communication. By systematically studying the effects of network delay, we can better understand collaborative performance. Distributed ensemble playing is already a reality today. The *New York Times*, on October 5, 2004 reported that as the Broadway pit shrinks, some orchestra musicians are sent to a room connected to the conductor only by a video link. By studying musicians' preferences in remote collaboration, we can develop technologies that will alleviate any distress associated with remote ensemble playing.

Acknowledgements

The research has been funded (or funded in part) by the Integrated Media Systems Center, a National Science Foundation (NSF) Engineering Research Center, Cooperative Agreement No. EEC-9529152, and an NSF Major Research Instrumentation Grant No. 0321377 titled "*Equipment Acquisition for Distributed Immersive Performance*." We thank the research assistants who helped develop the technologies that enabled the DIP experiments, including, Dwipal Desai, Frances Kao, Kanika Malhotra, Moses Pawar, Rishi Sinha, Shiva Sundaram and Carley Tanoue. We are grateful to Will Meyer for the video recording (and editing) of the experiments. We greatly appreciate the expert technical help of Seth Scafani and Allan Weber. Last but not least, we thank the musicians who have participated in the previous experiments, including Wilson Hsieh, Ron Leonard, Dennis Thurmond and the students at the New World Symphony, and Christopher Sampson, Associate Dean for New Initiatives at the USC Thornton School of Music, for supporting our efforts and connecting us with the Tosheff Piano Duo.

References

C. Chafe, documentation of recent distributed performance projects at the Center for Computer Research in Music and Acoustics' Soundwire Group at Stanford University: *ccrma.stanford.edu/~cc* .

Center for Computer Research in Music and Acoustics (CCRMA), Stanford University's Soundwire Group, Jam Session, 2002. URL: *www.ccrma.stnaford.edu/groups/soundwire* .

E. Chew, R. Zimmermann, A. A. Sawchuk, C. Kyriakakis, C. Papadopoulos, A. R. J. François, G. Kim, A. Rizzo, and A. Volk, "Musical Interaction at a Distance: Distributed Immersive Performance," in Proceedings of the 4th Open Workshop of MUSICNETWORK: Integration of Music in Multimedia Applications (Barcelona, Spain, September, 2004).

E. Chew, A. Sawchuk, C. Tanoue, R. Zimmermann, "Segmental Tempo Analysis of Performances in User-Centered Experiments in the Distributed Immersive Performance Project," (submitted, 2005).

S. Gresham-Lancaster, "AB_Time," a three-way concert between Vancouver-Marseilles-Troy, opening concert at the International Conference for New Interfaces for Musical Expression, Vancouver, Canada, May 26, 2005. URL: *hct.ece.ubc.ca/nime/2005/concerts.html* .

S. Kanki. Mélange à trios. NOTAM, 1998. URL: *www.notam02.no/warsaw/melange.html* .

C. Kyriakakis, P. Tsakalides, and T. Holman, "Surrounded by Sound: Immersive Audio Acquisition and Rendering Methods", *IEEE Signal Processing Magazine*, **16**(1), 55-66, (1999).

A. Mouchtaris, S. S. Narayanan, and C. Kyriakakis, "Multichannel Audio Synthesis by Subband-Based Spectral Conversion and Parameter Adaptation," *IEEE Trans. Speech and Audio Processing*, **13**(2), 2005.

A. Mouchtaris, S. S. Narayanan, and C. Kyriakakis, "Virtual Microphones for Multichannel Audio Resynthesis", *EURASIP Journal on Applied Signal Processing* (JASP), Special Issue on Digital Audio for Multimedia Communications, vol. 2003:10, pp. 968-979, September 2003.

A. Mouchtaris, P. Reveliotis, and C. Kyriakakis, "Inverse Filter Design for Immersive Audio Rendering over Loudspeakers", *IEEE Transactions on Multimedia*, **2**(2), 77-87, (2000).

A. Sawchuk, E. Chew, R. Zimmermann, C. Papadopoulos, C. and Kyriakakis "From Remote Media Immersion to Distributed Immersive Performance," in Proceedings of the ACM SIGMM Workshop on Experiential Telepresence *(ETP 2003)* (Berkeley, California, November, 2003).

E. Schooler, "Distributed Music: A Foray into Networked Performance, Haydn Piano Trio No.1 in G, the *Finale*," 2001. URL: *www.postel.org/pipermail/end2end-interest/2001-August/001314.html* .

R. Zimmermann, E. Chew, S. Arslan Ay, M. Pawar, "Distributed Musical Performances: Architecture and Stream Management," (submitted, 2005).

Roger Zimmermann, Kun Fu and Wei-Shinn Ku, "Design of a Large Scale Data Stream Recorder," in Proceedings of the 5th International Conference on Enterprise Information Systems (ICEIS 2003), Angers - France, April 23-26, 2003.

Roger Zimmermann, Moses Pawar, Dwipal A. Desai, Min Qin, and Hong Zhu, "High Resolution Live Streaming with the HYDRA Architecture," *ACM Computers in Entertainment*, volume 2, issue 4, October/December 2004.