5th Open Workshop of MUSICNETWORK: Integration of Music in Multimedia Applications

A Second Report on the User Experiments in the Distributed Immersive Performance Project

E. Chew, R. Zimmermann, A.A. Sawchuk, C. Papadopoulos, C. Kyriakakis,

C. Tanoue, D. Desai, M. Pawar, R. Sinha, and W. Meyer*

University of Southern California Viterbi School of Engineering,

Integrated Media Systems Center, Los Angeles, CA 90089

{echew, rzimmerm, sawchuk, chrisp, ckyriak, tanoue, dwipalde, mpawar, rishisin, wmeyer}@usc.edu

ABSTRACT

This report is our second one at the MUSICNETWORK open workshop focusing on the user experiments in the Integrated Media Systems Center's Distributed Immersive Performance (DIP) Project. The DIP project explores the creation of a seamless environment for remote and synchronous musical collaboration. We describe here the DIP experiments and findings since the last MUSICNETWORK open workshop. At the last workshop, we introduced the DIP project, our goals of studying the effects of auditory latency on musical ensemble and coordination, and the capture and analysis of user experiment data. We presented the first two sets of user experiments – collaborative performance with musicians facing each other and experiencing controlled auditory delay while performing movements from Poulenc's *Sonata for Piano Four-Hands*. The preliminary results from these first experiments showed that the expert users felt that they could adapt to delays below 50ms.

This year, we describe the next two sets of experiments, in which the first set of experiments required the players to practice performing with levels of delay around the threshold value, ranging from 40ms to 75ms. As before, our expert users are the award-winning Tosheff piano duo, Vely Stoyanova and Ilia Tosheff. Their practicing led to the creating of the next set of experiments, wherein the players requested for additional delay in the feedback of their own playing in order to hear the audience's perspective. This additional delay in the feedback of their own signal, and the resulting common perspective, improved the conditions so much so that the delay threshold was increased from 50ms to 65ms. We describe the sequence of events that led to the defining of these new conditions, and provide some explanations for this counter-intuitive result.

Background to the DIP Project

The goal of the user experiments in the DIP project is to comprehensively and systematically study users' needs and preferences in the DIP environment [1,2,3]. DIP and its predecessor, Remote Media Immersion (RMI) [4], are examples of immersive technology: a complete aural and visual environment that places a participant, or group of participants, in a virtual space where they can communicate naturally even though they are in different physical locations [5]. A goal of the DIP project is to create a framework that will enable comprehensive evaluation and analysis of the technical barriers and psychophysical effects of latency and fidelity on music and

* USC School of Cinema-Television.

other forms of human interaction between two interconnected sites. In this project we plan to explore detailed requirements on presence, latency, synchronization, audio and video resolution, quality and fidelity, and the subsequent effects on the ability of musicians to play together, and generally interact while at distant physical locations.

The DIP project and its goals are designed to further the vision of the Integrated Media Systems Center, which is the creation of seamless immersive distributed environments, transporting people from the present two-dimensional world of computers, television and film, to three-dimensional immersive environments, with visual, aural and haptic capabilities. The DIP system integrates many of the cutting edge research results from IMSC and elsewhere [2], and is built with the high-level principles defined by the Media Immersion Environment (MIE). It supports the creation, usage, distribution, and effective communication of multi-modal information. To enable rigorous user studies, the DIP v.2 framework was created is such a way as to allow comprehensive evaluation and analysis of the technical barriers and psychophysical effects of latency and fidelity on music, and other forms of human interaction between two interconnected sites.

Since many of the system parameters for minimum thresholds for usability are unknown due to the lack of adequate prior studies in distributed live music performances, much of the user-centered development process is formative in nature. As part of the user-centered development process, the DIP v.2 experiments are designed to identify the important system parameters for minimum usability, such as the maximum tolerable auditory delay, the maximum tolerable audio/video delay, minimum display size, appropriate viewing direction, inclusion of multi-channel audio, etc. Results from the experiments not only will shed light on guidelines for designing economically viable RMI and DIP systems, but also are expected to confirm the positive effects of immersipresence.

Other research groups have experimented with distributed collaborative performance environments, for example USC's Information Sciences Institute [6], Stanford's SoundWIRE Group (led by Chris Chafe) [7], McGill University (led by Jeremy Cooperstock) [8], Princeton University (Ajay Kapur et al) [9], and Scott Gresham-Lancaster [10]. Stanford's SoundWIRE group has conducted several experiments to quantify the effects of collaboration over the Internet by analyzing the ensemble accuracy of two persons clapping a short but inter-locking rhythmic pattern [11,12]. These experiments represent a departure from the previous one-time demonstration/performances, and a step towards the rigorous study of performance in the time-delayed environment. To the best of our knowledge, our experiments involving professional musicians performing complex composed pieces is the first such evaluation experiments on a realistic scale.

Overview of User-based Experiments

The objective of our Two-way base-line user experiments is to measure and document qualitatively and quantitatively the effects of delay and other variables on immersion, usability, and quality in the Distributed Immersive Performance scenario. For these experiments we enlisted the help of the Tosheff Piano Duo (*www.tosheffpianoduo.com*), Vely Stoyanova and Ilia Tosheff.

In our two-way baseline user studies, the two pianists were seated facing each other in the same room so as to isolate the effects of auditory delay from that of visual delay. In version 2 of the DIP platform, the audio and MIDI output from each keyboard and video from three high-definition (HD) cameras were streamed to the

HYDRA database developed by Zimmermann et al [2,3]. Christos Papadopoulos and Rishi Sinha created software for low-latency multi-channel audio streaming. Audio delay was controlled from a Protools console.

The Tosheff Duo was asked to play Poulenc's *Sonata for Piano Four-Hands* on two 88-key Yamaha P80 weighted action 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 to rate the following: (1) ease of ensemble playing; (2) ease of creating a musical interpretation; and, (3) the likelihood of their being able to adapt to the condition with practice. Each rating was performed on a scale of 1 to 7, with 1 being the easiest and 7 being the hardest. They are then debriefed and their observations recorded. Elaine Chew is leading the developing of quantitative methods for analyzing the performances captured on MIDI.

The following is the list of experiment sets that were conducted:

(A) first time players perform under delayed conditions – the audio latency was a number randomly chosen from the set {0ms, 10ms, 20ms, 30ms, 40ms, 50ms, 75ms, 100ms, 150ms};

(B) player 1 and player 2 swap parts (symmetry test), delays as in set A;

(C) players practice to compensate for delay (ranging from 40ms to 75ms); and,

(D) players perform with both partner *and* self delayed (ranging from 40ms to 75ms).

We presented videos and descriptions of experiments A and B, and some preliminary results at the last MusicNetwork workshop. 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. The user response results summary shows that delays 50ms and less 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.

We now give an update of the DIP experiments that will include descriptions and videos of experiments C and D, and analysis of users' needs and preferences in the remote collaboration environment.

Research Findings in Experiments C and D

Because the delay tolerance threshold appeared to be around 50ms, our next two sets of experiments focused on the region around 50ms, more specifically, the audio latency was restricted to the range from 40ms to 75ms. In experiment set C, the duo was asked to practice and strategize to compensate for the delay. The experimental setup is shown diagrammatically in Figure 1.

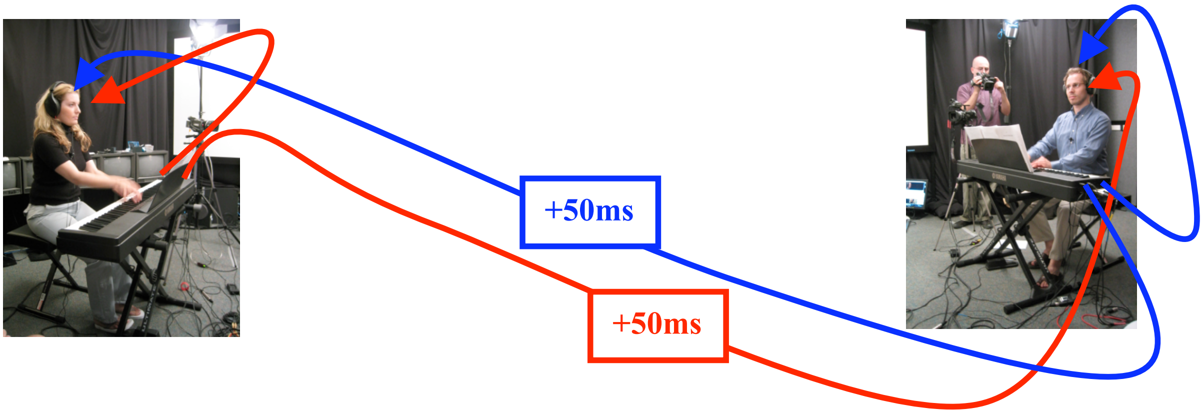

Figure 1. Diagram of setup C: players practice with delay (example shows 50ms delay between players).

In one of the performances of the Poulenc Sonata's third movement, the most challenging of the three to play with delay, with 50ms auditory delays, we have the duo attempting a slower version of the third Poulenc movement to compensate for the 50ms delay. These were their words recorded during the ensuing debriefing session:

Vely: It isn't our normal way of communication.

Ilia: It was too much. It started getting on my nerves. It felt that if we played a little slower, it would be fine. But the slower we played, the slower it gets. And the slower we play, the slower it gets, and at some point, it might stop! But it makes you feel like it's going to be alright if you play slower. But now I have a confirmation and proof that it doesn't work.

Ilia: You're asking if with practicing it would improve. We may just be so fed up with it that we didn't even keep up at some point.

In the practice sessions of experiment set C, the players were generally frustrated with each other's perceived inability to stay together. To relieve some of the tension, the players were given the opportunity to put on the other person's headphones to better understand the different delay situations at both ends, essentially hearing the audio input from a second perspective.

The duo then asked to hear the output from a third perspective, what the audience hears, in which the audio signal from their own keyboard would be delayed in transmission to their own headphones as well.

Vely: In this situation we talked about yesterday, we would never be satisfied at the moment when we play what we hear because it's not what the way the audience hears it, right.

Ilia: Can we both hear what the audience is hearing?

This request resulted in experiment set D where each player's audio input was from the third perspective; both players' audio inputs delayed, as shown in Figure 2.

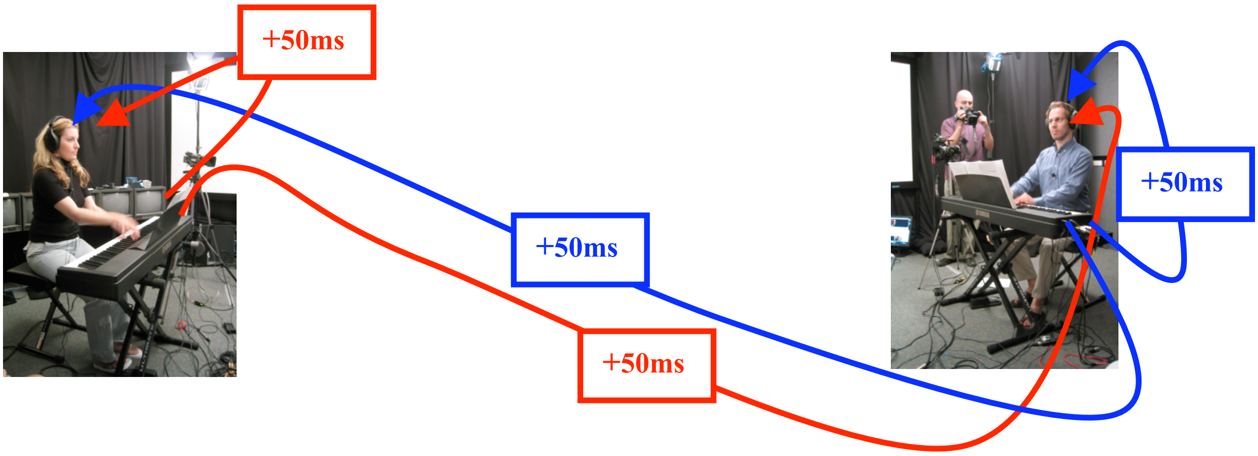

Figure 2. Diagram of setup D, example shows 50ms delay added to all audio streams.

The duo's response during the debriefing session after their first experience in setup D best communicates their preference for setup D:

Ilia: I am so happy. I truly enjoyed it.
Vely: Mm hmm …
Ilia: Now, this we can practice and just …, and even make it perfect
Sandy: Tell me your impression of this setup as compared to the previous one. You said you liked it.
Ilia: With one word, that's what I need, that's what I want. I just want to forget about it (previous takes).

Both players were noticeably much happier in setup D than in setup C. Further experiments in setup D showed that the altered setup (with delays on all streams) increased the overall tolerance threshold. The usability threshold, originally at around 50ms for condition C, was shifted to 65ms for condition D.

What is most important is that the players are able to hear their rendition of the piece as the audience hears it. In Ilia's statement the previous day, he explained 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. The fact that both players are able to hear what the audience hears allows the duo to maintain a common goal of creating a performance for the audience. With this common goal in mind, they can successfully develop strategies for creating an interpretation that is satisfying for both the audience and themselves. Adapting to the delay then fades into the background as a technicality that is easily overcome by their expert training.

Sandy: The delay in your headphones, does it bother you? Are you able to ignore it, mentally?
Ilia: I am able to cope with it.
Vely: Yes, me too.
Ilia: I would 100% prefer this one to the other. The other one is just very confusing … I wouldn't use it.
Vely: It's … there are hundreds of things that confuses you … in the other way.
Ilia: Here, just everything is there, … you just have to adjust your touch a little bit.
Vely: Here, it's like we're on the same track …
Ilia: … we're on the same track, like in a real performance.

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. As the players put it, it is imperative that they can feel that they are "on the same track." This shared reality gives them a common timeframe within which they can operate and create.

Discussion and Conclusions

These case studies supported by the musicians' self-reported experiences indicate that, in remote collaborative performance where network delay is unavoidable, players may be willing to tolerate, and to adjust to, delayed feedback of their own actions in order to achieve common ground. This common reality allows them to agree on the same goals and strategies for achieving them.

It is not surprising that musicians can adapt relatively easily to delayed feedback of their own playing. Organists in large cathedrals often have to cope with delay between the time they depress a key and the time the sound reaches their ear. As Ilia put it, "everything is there, … you just have to adjust your touch a little bit." There are other examples of musicians needing to hear what the audience hears, rather than the sounds in their immediate vicinity. Rock musicians in large concert settings almost always wear in-ear monitors to hear the sounds in the audience in order to react to, and interact with, the audience. Hearing what it is that the other hears is a critical part of communicating, of reaching out to others. It is a vital component of collaborative playing, and of musical performance, that cannot be ignored.

We are in the process of analyzing the performances quantitatively to validate the musicians' self-reported preferences. By systematically studying the effects of network delay, we can better understand collaborative performance, and develop technologies that will alleviate distress and enable better remote musical collaboration over the Internet.

Acknowledgements

We thank Chris Sampson, Associate Dean for New Initiatives at the USC Thornton School of Music, for introducing us to the Tosheff Piano Duo, who have proved to be such wonderful collaborators on this project.

This research was made possible by National Science Foundation grants MRI-0321377 and Cooperative Agreement No. EEC-9529152. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors, and do not necessarily reflect those of the National Science Foundation.

References

[1] Chew, E., Sawchuk, A.A., Zimmerman, R., the Tosheff Piano Duo (Stoyanova, V. and Tosheff, I.), Kyriakakis, C., Papadopoulos, C., François, A.R.J., Volk, A. (2004). “Distributed Immersive Performance”, *Proceedings of the 2004 Annual National Association of the Schools of Music (NASM) Meeting*, San Diego, CA, November 22, 2004.

[2] E. Chew, R. Zimmermann, A.A. Sawchuk, C. Kyriakakis, C. Papadopoulos, A.R.J. François, G. Kim, A. Rizzo, A. Volk, “Musical Interaction at a Distance: Distributed Immersive Performance.” Proceedings of the *MusicNetwork Fourth Open Workshop on Integration of Music in Multimedia Applications*, September 15-16, 2004, Universitat Pompeu Fabra, Barcelona, Spain. Co-located with WEDELMUSIC 2004 Conference.

[3] A. A. Sawchuk, E. Chew, R. Zimmermann, C. Papadopoulos, C. Kyriakakis, “From Remote Media Immersion to Distributed Immersive Performance.” Proceedings of the *ACM SIGMM 2003 Workshop on Experiential Telepresence* (ETP 2003), November 7, 2003, Berkeley, California, USA. In conjunction with ACM Multimedia 2003.

[4] R. Zimmermann, C. Kyriakakis, C. Shahabi, C. Papadopolous, A. A. Sawchuk, and U. Neumann, “The Remote Media Immersion System, *IEEE MultiMedia Magazine*, special issue on “Digial Multimedia on Demand,” April 2004.

[5] D. McLeod, U. Neumann, C.L. Nikias and A.A. Sawchuk. Integrated Media Systems. *IEEE Signal Processing Magazine*, vol. 16, no. 1, pp. 33-76, January 1999.

[6] Eve Schooler’s Musical Distractions – *http://www.async.caltech.edu/~schooler/music.html.*

[7] Stanford’s SoundWIRE Group – *http://ccrma.stanford.edu/groups/soundwire*.

[8] McGill University’s Center for Interdisciplinary Research in Music, Media and Technology – *http://www.music.mcgill.ca/cirmmt.*

[9] Kapur, A., Wang, G., Davidson, P., Cook, P.R., Trueman, D., Park, T.H., and Bhargava, M. "The Gigapop Ritual: A Live Networked Performance Piece for Two Electronic Dholaks, Digital Spoon, DigitalDoo, 6 String Electric Violin, Rbow, Sitar, Tabla, and Bass Guitar," *New Interface for Musical Expression (NIME)*, Montreal, Canada, May 23, 2003.

[10] Gresham-Lancaster, S. Vancouver-Marseilles-Troy concert at NIME 2005 – *http://hct.ece.ubc.ca/nime/2005/concerts.html*

[11] N. Schuett, “The Effects of Latency on Ensemble Performance.” Undergraduate Honors Thesis, Stanford CCRMA, May 2002.

[12] C. Chafe, M. Gurevich, G. Leslie and S. Tyan, “Effect of Time Delay on Ensemble Accuracy.” Proceedings of the International Symposium on Musical Acoustics, 2004, Nara, Japan.