Exploring the role of latency and orchestra placement on the networked performance of a distributed opera

Adriana Olmos[1], Mathieu Brulé[1], Nicolas Bouillot[1], Mitchel Benovoy[1], Jeff Blum[1], Haijian Sun[1], Niels Windfeld Lund[2], Jeremy R. Cooperstock[1]

[1]McGill University, Centre for Intelligent Machines
{aolmos|mathieu|nicolas|benovoym|jeffbl|hsun|jer@cim.mcgill.ca}

[2]University of Tromsø
{niels.windfeld.lund@uit.no}

Abstract

Can opera be performed if the singers are standing on different stages in different countries? This basic question motivated the body of work presented here. Our effort was intended to develop and explore the methodology necessary to answer this question in general, and specifically, to examine the effect of delay on tempo variations and the emotional connection between singers that is vital to a successful performance. In carrying out the initial experiment, we also examined the strategies employed to compensate for delay and the effect of orchestra placement on the outcome. This paper presents the results of a preliminary set of experiments, largely from a qualitative perspective, and describes the methodology for quantitatively addressing the questions above.

Keywords--- **network musical performance, audiovisual communication, affective interaction, latency**

1. Introduction

The body of work presented here has been inspired by the World Opera Project[1], a distributed, real-time live opera performance planned to take place simultaneously in several Canadian, U.S. and European cities. Opera is an art form in which singers and musicians perform a dramatic work. It incorporates many of the elements of spoken theatre, such as acting, scenery and costumes and sometimes dance. While singers and musicians strive to coordinate the timing of musical passages, their interaction is affected by both *internal timing variances* and *external latencies* [1]. The former are attributed to performer anticipation and delays due to expressive performance, errors in performance, and random timing variations due to physiological and biological constraints. External latencies are the result of audio propagation through the air, traveling from voice or instrument to the performers' ears. Players in a small chamber ensemble typically experience an external latency of 5-10 ms, whereas a double bass player on the other side of an orchestra could encounter external latency of 80-100 ms. During their training, singers and musicians develop various techniques that allow them to overcome these latencies, following or isolating certain sounds or instruments in order to coordinate the timing of the musical passage that they are performing. From his central position, the conductor has the role of integrating both coordination and expression. He may alter the tempo for artistic reasons, but must also account for delays, taking the attack times and relative amplitudes of the different instruments into account, while adjusting for the desired expression.

Distributed performance poses a range of additional challenges, including the delays in visual cues, changes in the acoustic environment, and the general increase in external latency due to encoding and network delays. Previous work addressing these challenges is discussed in Section 2. As described in Section 3, our study examines various latency conditions applied to both audio and video transmissions, simulating the effect of having the conductor, singers and musicians located in the cities of Montreal, New York, San Francisco and Tromsø. Measurements of tempo, physiological biosignals, gaze patterns and the singer's impressions were used to develop an analysis protocol for assessment of the singer's performance and coping strategies. Our methodology is presented in Section 4, followed in Section 5 by qualitative observations of the emotional connection perceived by the singers, and the role of the conductor in these network-mediated conditions.

2. Related work

Various studies attempting to evaluate the effect of network latency on musical performance can be found in the music and perception literature [1, 2, 3, 4, 5, 6]. Chew et al. [2] proposed the use of tempo difference and tempo ratio

[1] http://www.theworldopera.org

from a baseline performance as objective quantifiers of performance strategy, also taking into account qualitative reports from the musicians. Their findings show an increase in tempo variability when the auditory delay is in the range of 50-100 ms for two slower segments, and 50-75 ms for the most rapid segment. Surprisingly, the tempo difference decreased as auditory delay increased from approximately 50-75 ms to 150 ms. They concluded that the performers normally attempt numerous performance strategies to compensate for delay. However, at latencies exceeding some threshold, such strategies fail and the musicians revert to more stable practice norms, essentially ignoring the remote performer. Similarly, Willey [3] found that larger delays are easier to navigate, but prevent performers from interacting in a meaningful way. Other studies exploring delay thresholds for clapping synchronization [4] established an optimal one-way delay of 11.5 ms, with lower values leading, in general, to tempo acceleration and higher values to tempo deceleration. In an earlier study, Chafe et al. [5] subjected two trumpet players to latencies in the order of 200 ms. The problem of tempo deceleration was avoided only “when one player agreed to play behind the other”. Buillot [6] showed that it is possible to overcome the effect of delay by synchronizing audio streams and adding a fixed amount of audio delay related to the musical piece performed by the musicians. The limitation of this approach is that it only works with pattern based music.

Although there exists a long history of distributed network musical performance trials,[2] with many of these studying the effect of delayed audio signals between musicians, no protocol has been established to investigate the strategies that the singers or the conductor might use to cope with such delays. Of particular interest to us, and thus, the focus of our investigation, are the questions of how delay affects the emotional connection between singers, the choice of strategies singers use to compensate for delay, the effect of orchestra placement on the resulting performance, and finally, how these factors might affect the role of the conductor.

3. Experimental configuration

In order to explore these questions, we worked with six performers from McGill's voice program, one internationally renowned conductor, and a single pianist, representing the orchestral accompaniment. The singers were divided into three groups, each of which performed one of the following pieces: "Il core vi dono...", from Mozart’s Cosi fan tutte (mezzosoprano and baritone voices); "Ah! -- Voi signor" from Verdi’s La Traviata (soprano, tenor and bass-baritone voices); and "Bess you are my woman" from Gershwin’s Porgy and Bess (soprano and bass baritone voices). The selection of these pieces was made in order to experiment with music of varying complexity. Cosi, being more in the style of a *recitativo,* requires the performers and conductor to pay close attention to entrances; the musicians cannot rely on blindly counting the tempo.

Six different combinations of perceived audio and video delays (summarized in Table 1) were selected in order to simulate the various latency conditions between Montreal, New York, San Francisco and Tromsø. Each location was simulated using an isolated room containing two speakers, two cameras and two monitors, with each monitor/camera/speaker set representing the audio and video from a different location, as illustrated in Figure 1. The singers were able to see and hear each other through the video monitors and speakers at all times (Figure 2). To maximize presence, an attempt was made to display the performers at approximately life-size (upper body only).

The perceived delays are a sum of deliberately added delay, simulating measured network conditions, plus inherent delay from the acquisition and display devices (approximately 15 ms for the audio signal path [7], 60 ms for video displayed on CRT monitors, and 80 ms when displayed on LCD monitors). The manually added delay was introduced by a Pure Data patch for audio and by a parameter-controlled client-side buffer for our video transport protocol [8].

Table 1. Video and audio latency combinations used in this study. The synchronized low-latency condition refers to an idealized situation in which video acquisition and display latencies are equivalent to that for audio.

Latency combination	Perceived audio delay	Perceived video delay
Baseline (no added delay)	15 ms	60 ms
L: Low latency (best effort) ***Montreal-New York***	35 ms	80 ms
L_S: Synchronized low latency ***Montreal-San Francisco***	80 ms	80 ms
H: High latency (best effort) ***San Francisco-Tromsø***	135 ms	180 ms
H_V: High latency (video only)	35 ms	180 ms
H_A: High latency (audio only)	135 ms	80 ms

Although our *synchronized low-latency* experimental condition assumed an idealized scenario in which the inherent video delay was on par with that of audio (15 ms); to date, the minimum value we have been able to achieve in this regard is 30 ms for a direct camera connection to an HD-SDI monitor, running at 720p60 (60 frames per second in progressive scan mode). With the video signal passing through a computer interface acting as an intermediary under

[2] See http://www.cim.mcgill.ca/sre/projects/rtnm/history.html

software control, this figure increases to a minimum of 50 ms.

Two experimental sessions were designed to observe the effects of the placement of the orchestra, represented by the pianist. In the first, the pianist was placed in the same room as the conductor, and in the second, which took place on a second day of trials, with the pianist in one of the singers' rooms. For each session, the latency conditions were presented in the order listed in Table 2, thus, with an overall trend of increasing latencies as the experiment advanced. Each condition was run as a single "take" of approximately six minutes, followed by a seven minutes break for completion of questionnaires and rest. In total, each session (of six conditions) took approximately two hours, including post-experiment questionnaires and debriefing.

Individual audio recordings were made of the singers in each room and the piano. Singer A was equipped with a Thought Technology ProComp Infiniti biosignals recording unit, and an additional video camera was used to record the interactions of this performer with the remote singer and the conductor.

Table 2. Conditions tested with each group of singers and musical piece. The voice marked by an asterisk corresponds to the singer in room A who had his or her biosignals recorded.

			LATENCY BETWEEN		
Musical piece and Voice		*Take*	*Singers* **AB**	*Conductor & Singer A* ***CA***	*Conductor & Singer B* ***CB***
Cosi	*Mezzo-Soprano* Baritone*	1	*L*	*L*	*L*
		2	*L*	L_S	*L*
		3	L_S	*L*	*L*
		4	L_S	L_S	*L*
		5	*H*	*L*	*L*
		6	*H*	L_S	*L*
Porgy & Bess	*Soprano Base Baritone**	1	*L*	*L*	*L*
		2	*L*	H_A	H_V
		3	L_S	*L*	*L*
		4	L_S	H_A	H_V
		5	*H*	*L*	*L*
		6	*H*	H_A	H_V
La Traviata	*Soprano Tenor* Base Baritone*	1	*L*	L_S	*L*
		2	*L*	H_A	H_V
		3	L_S	L_S	*L*
		4	L_S	H_A	H_V
		5	*H*	L_S	*L*
		6	*H*	H_A	H_V

3.2. Rehearsals

The day before the actual experiment, the singers were invited to a two-hour preparatory session. This consisted of a brief introduction to the objectives of the World Opera Project, followed by performance and singing exercises led by a stage director, and rehearsals involving the conductor, pianist, and singers all together in the same room. Next, the singers and conductor were separated into the three different rooms and performed through the mediation of our audiovisual technology, under identical baseline conditions across the three rooms (inherent delays of audio = 15ms, video = 60 ms). The purpose of this initial session was to establish a level of rapport between the singers and the conductor, to introduce them to the experience of performing in separate rooms, and to acclimatize them to interacting with each other through a video display. While the singers already knew each other and had interacted through their university program, they had not met the conductor previously. Before the initial rehearsal, a casual conversation was carried out with each of the singers to elicit responses to the following pre-test questions:

- What aspects do you feel are important to establish a close connection with other singers while performing?
- Who do you feel you have to pay most attention to while performing: other singer(s), the conductor or something else?
- What are the factors that you consider important in assessing whether a performance is successful?
- If you feel that your tempo is off (or out of synchrony), who or what do you rely on to correct it?

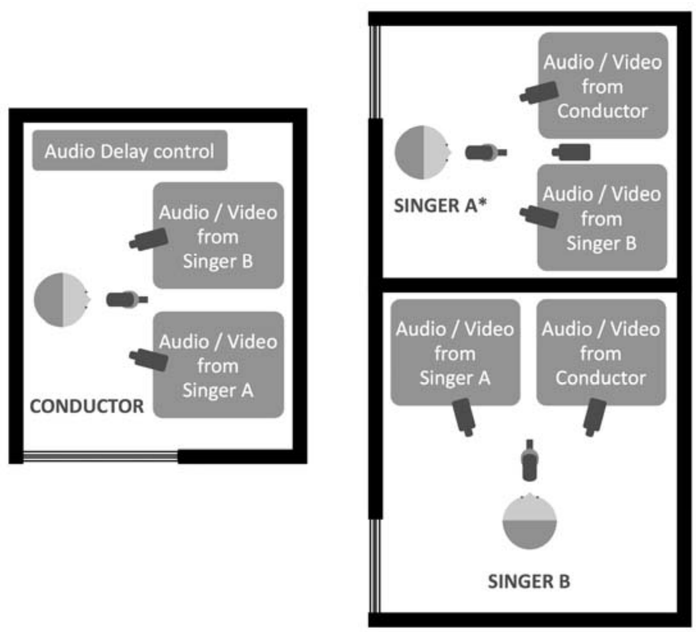

Figure 1. Overview of the hardware topology between the three rooms.

Figure 2. Illustration of the configuration in Room A.

While the answers varied between the singers based on their personal impressions and experience, these hinted at the importance of the rehearsal process as a means to develop a level of connection. The factors mentioned most often by the participants in our experiment were trust, the other singer's breathing, and eye contact. Five out of the six singers replied that they would immediately look at the conductor if something was wrong or they were feeling off tempo. Only one of the singers stated that he would "listen and keep going" until he established a connection, either with the conductor or the other singer, trying to avoid an abrupt change.

The singers were asked to complete a questionnaire, rating their experience during the baseline condition in separate rooms in comparison to the regular situation where they sing together in the same room. Ratings for the following questions were obtained on a Likert scale of 1-7:

1) How satisfied were you with the performance?
2) How would you rate your emotional connection with the remote singer?
3) How would you rate your emotional connection with the conductor?
4) How important was the audio?
5) How important was the video?

Figure 3 presents the average ratings for the singers' responses, with error bars indicating one standard error from the mean. The same questionnaire from the baseline condition was also completed by the singers and the conductor after each of the conditions tested during the study; the insights from this questionnaire are further discussed in Section 5.

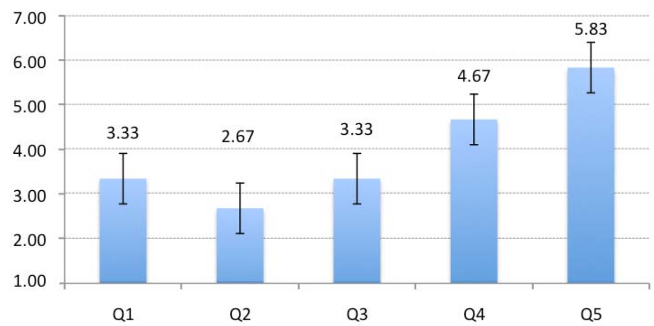

Figure 3. Average rating from the singers, to the baseline condition

4. Analysis Methodology and Initial Results

Audio signals from the singers in each room were recorded on independent tracks and aligned and analyzed as described in the following sections. The audio analysis was performed in parallel with an evaluation of two biometric measures: the average Galvanic Skin Response (GSR), also known as electrodermal activity, and the number of Skin Conductance Responses (SCR). GSR measures the electrical conductance of the skin. Following emotional excitation, the eccrine sweat glands fill sweat channels, which causes a drop in resistance, or conversely, greater conductance. Eccrine sweat glands are concentrated only on the palms of the hands and the soles of the feet of primates, and are innervated by the sympathetic branch of the autonomous nervous system (ANS). As opposed to other sweat glands, these are much less sensitive to variations of temperature. The GSR response is often the primary psychophysiological measure used when assessing emotional arousal as it is highly correlated with sympathetic nervous system excitation. Spikes in the GSR, illustrated below in Figure 4, are termed skin conductance responses (SCR) and are related to phasic emotional arousal, usually triggered by discrete events. The skin conductance level (SCL), shown in red, represents the slow varying tonic component related to the activity of both the perspiratory glands reacting to temperature and the eccrine glands reacting to arousal. The SCL can be interpreted as the generalized excitation level of the performer.

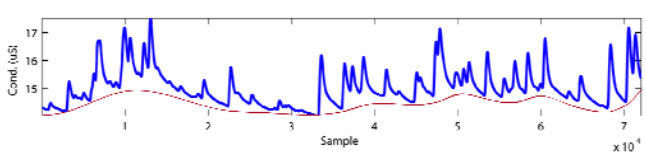

Figure 4. Example of a Galvanic Skin Response (blue) signal with skin conductance level (red).

Separately, we also examined the gaze patterns of the singers. This was accomplished by viewing each video of the performer in room A at half speed and coding the singer's behaviors as follows:

- gaze is to the left of the camera: "looking at the conductor"
- gaze is to the right of the camera: "looking at the remote singer"
- otherwise: "looking elsewhere". This could include looking away to in a corner of the room, closing their eyes for a long period, or peering down. Such behaviors are often observed as a mechanism of theatrical expression while performing.

The Noldus software package *The Observer* was used to record the behaviors. The following section describes the methodology used to study each singer's strategy to cope with the delay conditions, the emotional connection between the singers, and the effect of the orchestra placement in terms of shifts in tempo.

4.1 Latency-coping strategy

Throughout their careers, singers learn to select what to pay attention to and what to ignore while singing. Throughout our experiments, the singers commented, both while on camera or during completion of questionnaires, about the strategies used. They seemed to be able to connect emotionally up to a certain extent and display their emotions visually to the remote singer. However, they could not necessarily sing in synchrony with what they heard or saw. They often mentioned that they ended up following the conductor, the piano, or their inner beat.

The question we first investigate is in which conditions did the singers focus on the audio piano cues, and in which did they focus on the video of the conductor's hands? The answer can be explored through the detection of attacks, defined as the prompt and decisive beginning of a note or passage by the musical performer, in the audio recordings of the singer and piano. If a singer followed the sound of the piano, the singer's attacks would match those of the original piano signal, temporally offset by the perceived delay. The piano recording in the conductor's room was taken as the reference, since this provides the best possible acoustic synchrony with the conductor's hands. This piano waveform was shifted by the amount of perceived audio or video delay experienced by the singer. The associated attack times were then compared between these shifted signals and the singer's recording, as illustrated in Figure 5. A contingency table was used to compare the number of closest matches to the audio and video and further verified through analysis of the gaze patterns of the singer in each condition.

We applied this analysis to the recordings of *Cosi fan tutte* (Cosi) and *Porgy and Bess* (Porgy), which involved two singers per piece. The recording of *La Traviata* was excluded because the piece involved three singers. The analysis was completed for the three conditions shared by both groups of singers (Table 2, takes 1, 3 and 5).

Our analysis found no significant difference in the matches of attack times of the singers with those of the audio signal or the video signal for either Porgy ($t_{(2)}$= 4.6421, p = 0.0982) or Cosi ($t_{(2)}$= 2.4818, p = 0.2891). Similarly, gaze patterns did not differ between conditions for either Cosi ($t_{(2)}$= 1.4687, p = 0.4798) or Porgy ($t_{(2)}$= 5.9858, p = 0.0501). These preliminary findings could be explained by the thresholds reported in the literature, as the perceived latencies in our experiment generally exceed the range at which Chew et al. observed that performers might switch to an "auto-pilot" performance mode [2]. This question requires further investigation with a larger sample size and more finely spaced delay combinations, in order to discover if there is a particular threshold at which the singers change strategies.

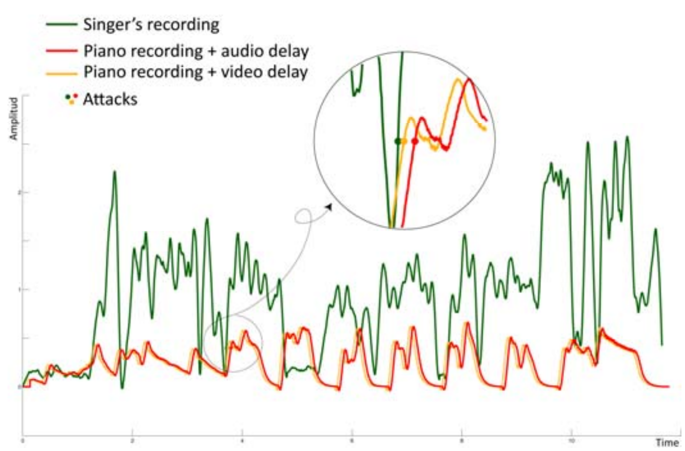

Figure 5. Example of the attacks detected in the singer's recording, matched against the piano signals generated by adding the perceived delay of the audio or video.

4.2 Effect of orchestra placement on singer's tempo

In this analysis, we compared a reference signal with the other conditions using Dynamic time warping techniques. Dynamic time warping (DTW) is a well-known technique for finding an optimal alignment between two time-dependent sequences. For our analysis we used *MATCH* [9], a toolkit that aligns audio recordings of different interpretations of the same piece of music. This gives us a vector with the differences in tempo throughout the song. Figure 6 presents the vectors corresponding to the DTW analysis for *Porgy* and *Cosi*. The first take of each session, as listed in Table 2, was used as a reference signal.

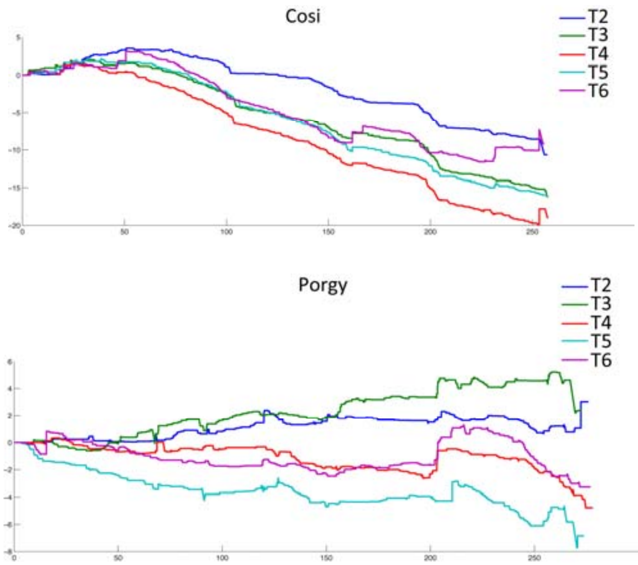

Figure 6. Dynamic time warping analysis from the audio recordings from singer A in each of the music pieces tested (Experiment A). A negative slope indicates increased tempo relative to the first take of each session.

A global measure can be computed that characterizes the overall changes in tempo. We introduce here a measure based on the curvature of an object defined as the derivative of the tangential angle to the curve. This concept [10] is used in image processing at a scale-space level to capture local and global deformations in a shape or curve. Here, we use the sum of curvature zero-crossings (or *inflection points*) at the first scale divided by the length of the analyzed signal as a measure of global change in tempo. An example of the behavior of this measure, C_m, is shown in Figure 7.

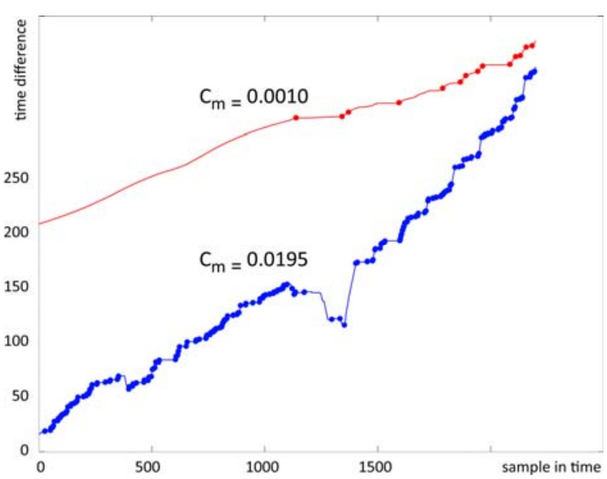

Figure 7. Example of two signals and their curvature measures (C_m) The dots indicate inflection points, of which the blue plot contains a significant number, thus accounting for its higher value of C_m.

We compare the tempo changes resulting from the two placements of the piano by computing the curvature, C_m, from the time-warped audio signals of the singers in room A for *Porgy* and *Cosi* (Figure 8 and 9). A paired t-test was used to compare the curvature of the corresponding latency conditions between the two orchestra placements. For *Cosi*, global curvature was lower when the orchestra (piano) was with the conductor; singer A ($t_{(4)}$= 5.593, p = 0.0016) and singer B ($t_{(4)}$= 6.067, p = 0.0011). However, the opposite effect was observed for *Porgy*; singer A ($t_{(4)}$= 4.1533, p = 0.0058), and singer B ($t_{(4)}$= 3.3660, p = 0.0130). One initial hypothesis was that placing the piano in the same room as singer B would reduce the performance effort of singer A, who would receive both the audio from singer B and the piano delayed by the same amount.

Similarly, we hypothesized that the local piano would provide singer B with temporal support, thereby facilitating the task of tracking the tempo. Although partly supported by the results for *Cosi*, the contradictory results from *Porgy* suggest, instead, that the effect of orchestra location may be highly dependent on the type of music performed. This issue demands further study with musical pieces of varying complexity. Understanding this effect of orchestra placement on the effort required by the singers is of considerable importance as one considers the optimal distribution of musicians to facilitate an effective networked performance.

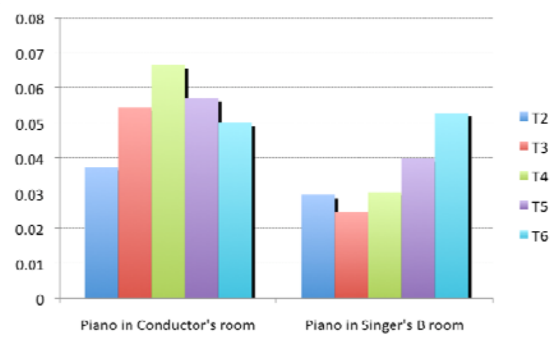

Figure 8. Singer A global tempo measures across orchestra placement conditions in *Cosi*.

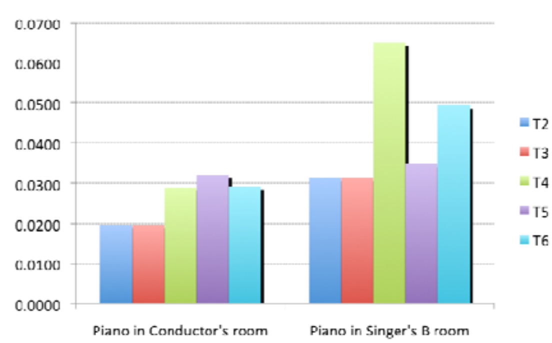

Figure 9. Singer A global tempo measures across orchestra placement conditions in *Porgy*.

Finally, we considered the average Galvanic Skin Responses (GSR) and the number of Skin Conductance Responses (SCR) of singer A in each of the two piano placement configurations. For *Porgy* and *Cosi*, the mean GSR between the two configurations (Figures 10 and 11) was higher across all conditions when the piano was with singer B; *Cosi* ($t_{(4)}$= 4.9965, p = 0.002658) and *Porgy* ($t_{(5)}$= 8.2284, p = 0.000123). For the SCR the difference between the two piano placements was not statistically different; *Cosi* ($t_{(4)}$= 2.6778, p = 0.028773) and *Porgy* ($t_{(5)}$= 3.1001, p = 0.015214). Take 2 from Cosi was excluded from this analysis due to a file error while recording the data. While GSR measurements are associated with emotional arousal, one cannot identify the specific emotion being elicited. Thus, the higher GSR values observed from the placement of the pianist in the same room as the singer, might be due either to greater engagement between singers, or greater stress. An additional confound was that all the trials of the pianist in the same room as Singer B took place on the second day of trials, so increased experience with the technology-mediated experience may have biased the results. Clearly, additional experimentation would be required in order to draw meaningful conclusions in this regard.

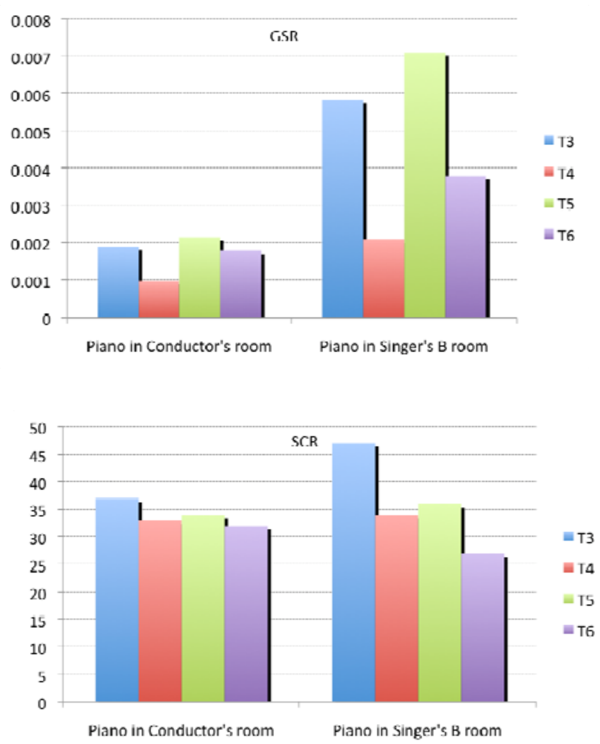

Figure 10. Singer A mean GSR and SCR across orchestra placement conditions in *Cosi*.

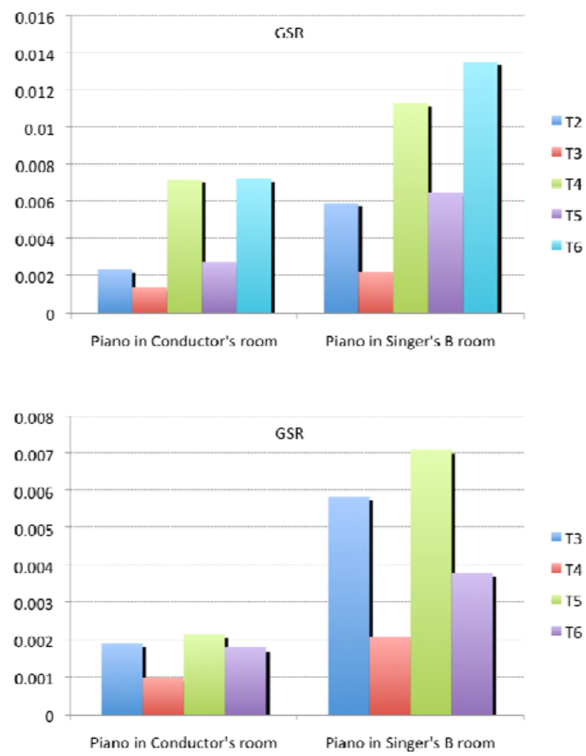

Figure 11. Singer A mean GSR and SCR across orchestra placement conditions in *Porgy*.

5. Observations and discussion

For the most part, the singers that participated in the experiment managed to cope with the various delays and successfully sang their piece under all conditions. The singers mentioned that to a certain extent, they were able to establish emotional connections with each other. Nevertheless, they felt a "disconnect" between what they heard and the events to which they reacted. In this respect, it was remarkable that at the completion of each trial, most of the singers guessed, with considerable accuracy, the different latencies to which they had been exposed.
This was reflected in the comments made by the singers between the conditions:

- *La Traviata*, singer B, experiment A, take 4: "I felt the conductor was behind from the beginning… so I went with the sound of the piano". Delay between the conductor and singer B: audio = 35 ms, video = 180 ms. Piano in the same room as the conductor.
- *Porgy*, singer B, experiment A, take 5: "Followed the conductor. Again, stayed connected with [singer A] but could not sing with him. He was behind the beat this time". Delay between singers: audio= 135 ms, video = 180 ms. Delay between the conductor and each of the singers: audio = 35 ms, video = 80 ms. Piano in the same room as the conductor.

Of possibly greatest interest was the observation that the singers' strategies were highly dynamic, changing between singers and conditions, even while performing a single piece.

All of the singers frequently mentioned that they ended up following the conductor, the piano, or their inner beat, but completely ignored the voice of the remote singer. They also felt that the simulated distributed performance conditions were more demanding than singing together in the same room, as they needed to focus at all times and could not "miss a single millisecond of attention". This is, of course, discouraging from the perspective of presence and interaction.

By the time the experiment reached its conclusion, the singers had largely adapted to the environment and the technology introduced. A typical session would start with comments such as, "This is weird… I relied on my ear more than my visual, especially when you feel disconnected from the other singer physically." As more conditions were tested, comments such as, "This is starting to feel like a normal rehearsal process now", "I felt like I could not connect with [the remote singer] as easily but I felt together with the conductor" became more prevalent, and toward the end of the experiment, we heard comments such as "I looked at the conductor, now that I am getting used to it. It is easier to ignore sounds".

It is important to mention that between trials, the singers felt at ease, joking with each other and conversing naturally through the video and audio connections, illustrating, as expected, that the effects of added latency were challenging only for musical interaction.

Conclusions and Future work

While the delay conditions often led to undesired changes in tempo, this might be overcome, in part, through rehearsal. One can therefore appreciate the importance of the rehearsal process under actual delay conditions when putting together a full operatic production between remote performers. This would not only help to fine tune and balance between the synchrony and asynchrony of the performers, but also would help the singers develop and practice a coping strategy for the very specific delay introduced by their remote location.

As alluded to earlier, the role of the conductor turned out to be important in maintaining the flow of the musical piece and preventing the singers or the pianist from following each other, which would have resulted in the familiar "recursive drag on tempo" [4]. Nevertheless, we observed that the difficulty in judging how to compensate for the audio and video delay precluded any adjustment by the conductor for musical expression. While the initial trial of the entire experiment was difficult for the conductor, after a few more trials, the conductor, along with the pianist, developed a system in which they ignored the audio feed coming from the singers. As the conductor stated, "I do not correct because I can't judge."

Responses from participants to the questions related to their emotional connection with the remote singer indicate that latencies in the range we were testing had little or no impact on perception of presence, even though these may have been significant on the musical result. Rather, we found that the biggest factor influencing presence was familiarity with the technology. The longer the experiments proceeded, the singers indicated less distraction from the camera-monitor mediated experience and a greater sense of emotional connection, regardless of latency. We expect that the effects on presence would be more pronounced as latency values were increased to the range more typical of conventional videoconferencing systems.

Interestingly, and contrary to results from the literature [4], we found that the average tempo actually increased as latency was increased. This is likely due to the mediating role of the conductor, who, in effect, established the tempo, and did so deliberately oblivious to the asynchrony he perceived between his own beat and that of the remote performers. Obviously, the conductor was not a factor in most earlier experiments involving distributed duos or trios, mostly due to the genres of music performed.

The results presented here are, of course, based on a very small sampling of conditions and performers, so cannot be considered as statistically meaningful. Although some of the questions we posed at the outset of this work remain unanswered we believe that this effort has allowed us to develop a useful methodology for analyzing experiments in networked musical performance, assessing the effects of latency and placement of the singers and the orchestra on the resulting performance. It would be particularly valuable to apply this methodology to upcoming tests involving actual distributed rehearsals, spanning a significantly longer duration, and allowing the performers an opportunity to develop specific coping strategies for the longer than normal delay conditions they will encounter. It should also be noted that unlike examples of the traditional repertoire that we included in our experiments, the World Opera will involve new librettos, written specifically for the distributed performance scenario, and which may well place additional demands on the performers or even allow them to play creatively with the artifacts of delay. It is our hope that this work will encourage further investigations of a similar nature and help support the success of a World Opera premiere performance in the near future.

Acknowledgements

This experiment and the comprehensive analysis that ensured would not have been possible without both the generous funding of both the Norwegian Research Council and the Natural Sciences and Engineering Research Council of Canada, who supported the activities leading up to the experiment and the actual execution of the work itself. The effort involved a major investment of time and human resources by the members of the Shared Reality Laboratory, who all contributed many hours to the experiment.

Thanks are also due to Guy Boivine for lending us The Observer, and Mick Wu for the useful discussions on gaze observations and statistical analysis. In addition, we are most grateful to Centre for Interdisciplinary Research in Music, Media and Technology, whose executive members agreed to house our experiments and whose staff invested a tremendous amount of time and provided the physical resources that supported our work.

Special thanks are due to the singers from the McGill Voice Program, pianists Michael McMahon and Jordan de Souza, director Robert Swedberg and of course, to Maestro Niels Muus, who was the consummate professional, always gracious and good-humored when technical problems delayed our start.

References

[1] Bartlette, C., Headlam, D., Bocko, M., and Velickic, G., Effect of network latency on interactive musical performance. Music Perception, 24, 49-59, 2006.

[2] E. Chew, A. Sawchuk, C. Tanoue, and R. Zimmermann. Segmental Tempo Analysis of Performances in User-Centered Experiments in the Distributed Immersive Performance Project. In Proceedings of the Sound and Music Computing Conference, Salerno, Italy, November 24-26 2005.

[3] Willey, R. K. The relationship between tempo and delay and its effect on musical performance. Journal of the Acoustical Society of America, 88 (Suppl. 1), S71, 1990.

[4] Chafe, C., Gurevich, M., Leslie, G., Tyan, S., Effect of the Time Delay on Ensemble Accuracy, Proceedings of the International Symposium on Musical Acoustics, 2004.

[5] Chafe, C., Wilson, S., Leistikow, R., Chisholm, D., & Scavone, G. A simplified approach to high quality music and sound over IP. Paper presented at the COST G-6 Conference on Digital Audio Effects, Verona, Italy, 2000

[6] Bouillot, N. 2007. nJam user experiments: enabling remote musical interaction from milliseconds to seconds. In Proceedings of the 7th international Conference on New interfaces For Musical Expression

[7] Bouillot, N. and Cooperstock J. R., Challenges and Performance of High-Fidelity Audio Streaming for Interactive Performances, In proceedings of the 9th international conference on New interfaces for musical expression Pittsburgh, June 2009

[8] Woszczyk, W., Cooperstock, J. R., Roston, J., Martens, W. Shake, Rattle, and Roll: Getting Immersed in Multisensory, Interactive Music via Broadband Networks, Journal of the Audio Engineering Society, 53(4), April, 336-344, 2005.

[9] Austrian, S. D., and Dixon, S., Match: A Music Alignment Tool Chest, in Proc. ISMIR, 2005

[10] Mokhtarian, A. F., Silhouette-Based Isolated Object Recognition through Curvature Scale Space, IEEE Trans. Pattern Anal. Mach. Intell., 0162-8828, 17, 5, p. 539-544, 1995