PARAMETER CONTROLLED REMOTE PERFORMANCE (PCRP): PLAYING TOGETHER DESPITE HIGH DELAY

Aristotelis Hadjakos *Erwin Aitenbichler* *Max Mühlhäuser*
TU Darmstadt
Department of Computer Science

ABSTRACT

PCRP (Parameter Controlled Remote Performance) is a new method to play music together on the Internet despite high delay. In many musical forms, the score of the piece is known beforehand. PCRP measures the deviations of the players from the given score and describes them as parameters. The parameters, which are transmitted instead of the audio data itself, are used on the opposite side of the network to control the synthesis of the score. We implemented a PCRP system for piano duo. The system analyzes and transmits tempo, dynamics, and articulation parameters to assure synchronization of the players and give the users control over tempo, dynamics, and articulation. An evaluation of the system was performed with pianists. The evaluation showed that musical interaction with PCRP is possible at higher latencies than related approaches.

1. INTRODUCTION

When playing music over the Internet, the musicians' signals are delivered with delays. This is due to properties of the current network technology and physical limitations (speed of light).

Effects of delay on musical interaction have been studied [3, 4, 6]. In an experiment conducted by Chafe et al., two players, who were separated by artificial delays, clapped simple rhythms [3]. The analysis of the recorded rhythms showed that the players tended to slow down at latencies higher than 11.5 ms. This effect increased at higher latencies. A similar setup was used to determine the effects of delay on piano duo performance in an experiment by Chew et al. [4]. The duo played different pieces at different artificially introduced delays. The pianists reported that satisfactory interplay was possible at delays of less than 50 ms. In a following evaluation of the recorded data [6], the tempo variation of the players was determined. At latencies slightly above 50 ms, the players used compensation strategies, which led to an increased variation of tempo. At very high latencies, the players played in a overly steady tempo, because the compensation strategies failed.

In this paper, we present PCRP, a new method that enables a traditional musical interaction in presence of high delays.

The paper is structured as follows. First, related Network Music Performance systems are discussed (Section 2). In Section 3, PCRP is introduced and the analysis-transmission-synthesis process is explained in detail. This provides the ground for the following evaluation in Section 4. In Section 5, the offered possibilities and limitations are discussed. Conclusions are drawn in Section 6, and future work is described in Section 7.

2. RELATED WORK

NMP (Network Music Performance) systems enable geographically separated users to musically interact over a network. Two properties of NMP systems are especially important for the following discussion: (1) the type of the musical interaction and (2) the maximum delay that can be compensated (Figure 1). NMP systems can change the type of musical interaction to allow interaction at higher delays.

Systems without delay compensation transmit audio data over the network and limit delay by minimizing processing time of the application. Early systems were TransMIDI [9] and RMCP [10]. Recent research focuses on low-latency compression techniques [2] and the use of advanced I/O and network hardware to achieve immersive experiences [4, 5, 6, 7]. Systems without delay compensation are only suitable when network delay is low.

Delay compensation systems modify the musical interaction so that the musicians can compensate higher delays. In the reminder of this section different NMP systems with delay compensation will be discussed.

One class of delay compensation systems artificially delay the audio signal of a player to her own headphones. The signal is delayed the same amount as the signal needs to travel to the opposite partner through the network. This method, which will be called Artificial Delay method henceforth, is best suited for instruments that can be muted like the electronic keyboard or the electronic guitar. The Artificial Delay method was developed by Chew et al. and evaluated with a piano duo [5]. The piano duo was able to compensate more delay (ca. 15 ms more) than in a setting without delay compensation. The commercial product eJamming [8] is a commercial implementation of the Artificial Delay

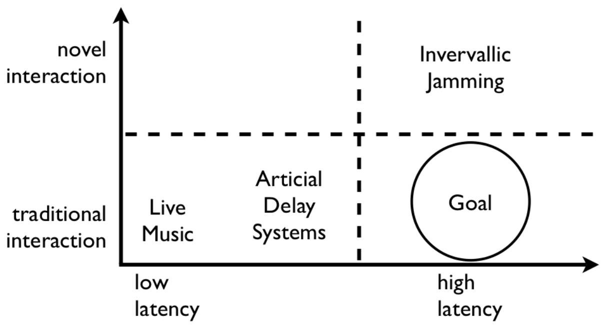

Figure 1. Properties of NMP systems

method. PCRP is a different method for delay compensation that enables the players to musically interact at higher delays than it is possible with Artificial Delay systems.

Intervallic Jamming systems artificially augment the delay between the players to a multiple of a musical time interval, e.g., a beat, a bar, or a 12-bar blues progression. Examples of intervallic systems are the Global Delayed Music system [11], the M. A. S system [13], and the Ninjam system [12]. PCRP is a different method for delay compensation that, in contrast to Intervallic Jamming, keeps the impression of traditional interplay.

The TablaNet system [14] uses pattern recognition and music prediction to enable musicians playing the tabla, an Indian drum, to play together despite high delay. TablaNet identifies rhythmic patterns of the players and maps them on to known musical structures. Symbolic representations of these musical structures are transmitted and are used to create a similar audio output at the receiver. The TablaNet system is specific to Indian tabla music and can not be used by players who want to perform a fixed score. PCRP enables players who want to perform a fixed score to play together on the Internet despite high delay.

3. PCRP

PCRP (Parameter Controlled Remote Performance) is based on an analysis-transmission-synthesis scheme. Deviations of the players from a given score are measured and mapped to parameters. The parameters are transmitted through the network instead of the audio data itself. Based on the transmitted parameters, the music is synthesized on the opposite side. The players are hearing each other mediated by the analysis-transmission-synthesis connection. Therefore, the users do not feel the delay of the network. The delay of the audio signal is exchanged with the delay of the parameter signal. If a player changes her style of play, the opposite player will perceive that change of style after the correspondent pa rameters have been received. This is less critical than a delay on the audio signal.

For the following discussion we assume a scenario where two players, Alice and Bob, want to use PCRP to play together. A recording of Alice's part is also present on Bob's side and vice versa (Figure 2). The computer analyzes Alice's play and extracts parameters that describe her deviation from the score. On Bob's side, the playback of Alice's score is modified according to the transmitted parameters. The same procedure is applied to Bob's play.

The parameters that are extracted by our implementation of the PCRP method are: tempo, dynamics, and articulation.

3.1. Architecture

We implemented a system for piano duo based on the PCRP method. Two Casio CDP-100 keyboards are used by the players and provide MIDI input to the system. The system is composed of two analysis-transmission-synthesis connections. The main components of a connection are a score follower module, a tempo module, a dynamics module, an articulation module, and a synthesis module (Figure 3). The tempo, dynamics, and articulation modules transmit parameters to the synthesis module over the Internet.

The MIDI input of the player is passed on to the score follower, the dynamics module, and the articulation module. The score follower module uses the polyphonic dynamic matching algorithm [1], which allows to use polyphonic scores, to match the MIDI input to the score. The score follower module generates a sequence of timestamped score positions, which are used by the tempo module to generate the tempo parameter.

The dynamics and articulation modules get the current score position from the score follower and calculate the dynamics and articulation parameters. This is done by comparing the dynamics and articulation of the currently played note with the dynamics and articulation of the corresponding score segment. Finally, tempo, dynamics, and articulation parameters are transmitted over the network to the synthesis module.

3.2. Tempo Parameter

The tempo parameter enables Alice and Bob to stay synchronized and influence each other's tempo.

3.2.1. Determination of the Tempo Parameter

The tempo module computes the tempo parameter from the sequence of timestamped score positions, which is generated by the score follower module. The time difference between two successively recognized score positions is compared to the IOI (inter-onset-interval) of

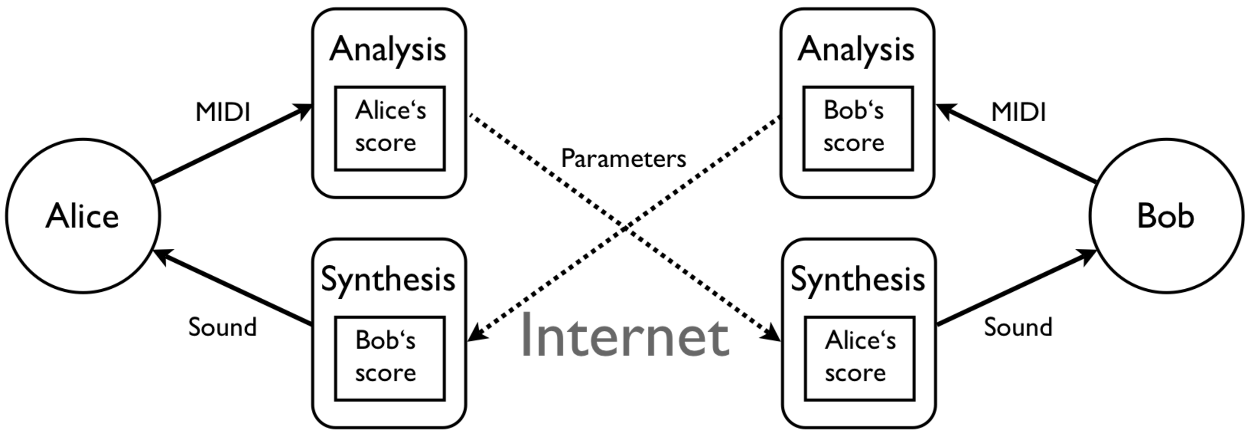

Figure 2. Overview of PCRP

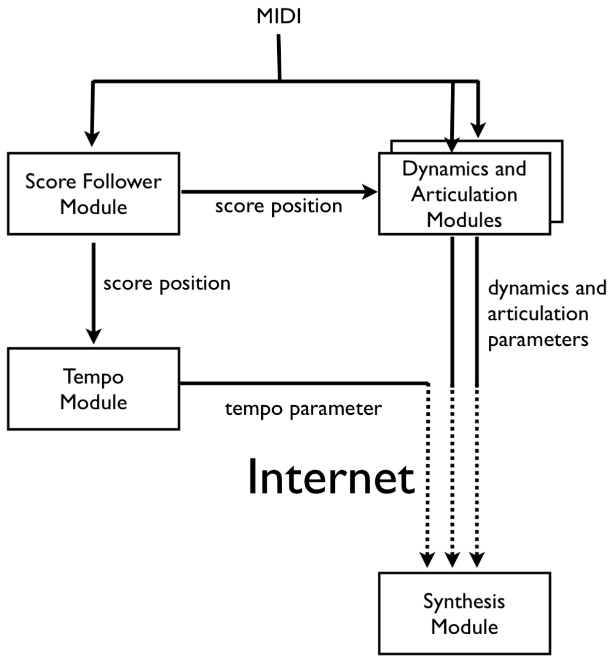

Figure 3. Analysis-transmission-synthesis connection

the corresponding positions in the score. The tempo hypothesis is the fraction of these values:

Tempo hypothesis =
IOI in the score / Timestamp difference

The tempo parameter is the combination of tempo hypothesis, the current score position, and the timestamp. The timestamp is needed so that the synthesis module at the opposite side of the network knows when the tempo of the player was measured. As time will have passed until the parameter is received this information is crucial. The synthesis module uses tempo hypothesis, score position, and timestamp to compute the current position of the distant player, given that she has not changed the tempo since. The tempo parameter is transmitted to the opposite player's computer where it is further processed.

3.2.2. Tempo Parameter Controlled Synthesis

Synthesis is based on rate-controlled playback. Alice's tempo parameter reaches Bob's computer after some delay. The delay is computed by examining the timestamp. Considering the measured delay, the reported position, and the tempo hypothesis, Bob's computer can determine the so-called position hypothesis by linear interpolation. (The computer assumes that Alice has not changed the tempo since.) The current position of the playback is determined and compared to the position hypothesis. The play-rate is adjusted so that Alice's play and the playback will be synchronized at a certain time in the future, given that Alice does not change tempo since the tempo parameter was determined. It was empirically determined that synchronizing to one second in the future obtains good results. If the calculated play-rate is negative, then the playback is stopped until the next tempo parameter is received. (A negative play-rate occurs if the calculated position of the opposite player is more than one second before the current position of the playback.)

The receiver can estimate when the next tempo parameter is due, by taking into account the current tempo, the score, and the measured delay. If no tempo parameter is received one second after due time of the tempo parameter, the playback is paused.

3.3. Dynamics and Articulation Parameters

The dynamics and articulation parameter enables Alice and Bob to control the dynamics and articulation of the remote synthesis.

3.3.1. Determination of the Dynamics and Articulation Parameters

The dynamics and articulation modules calculate the dynamics and articulation parameters. The modules receive MIDI input from the MIDI keyboard and are informed about the score position by the score follower module.

To calculate the dynamics parameter, the system waits for a note-on message and compares the loudness of the note with the loudness of the corresponding note in the score. The dynamics hypothesis is the fraction of played loudness and loudness of the corresponding note in the score.

Dynamics hypothesis =
Played note loudness / Score note loudness

The system keeps track of all sounding notes by listening to note-on and note-off messages. Received note-on messages are timestamped and stored. To calculate the articulation parameter, the system waits for a note-off message and determines the length of the played note. The length of the played note is compared to the length of the corresponding note in the score. The articulation hypothesis is the fraction of played length and length of the corresponding note in the score.

Articulation hypothesis =
Played length / Score note length

In contrast to changes of the tempo parameter, which are often gradual, the determined dynamics and articulation hypotheses often vary abruptly from note to note. To make the system more predictable, the average of the three recent dynamics hypotheses is calculated. This average provides a good compromise between avoiding erratic behavior and providing fast response to deliberate change. The calculated average is called the dynamics parameter. The same averaging procedure is applied to articulation hypotheses and results in the articulation parameter. Dynamics and articulation parameters are transmitted to the opposite player's computer where they control the synthesis of the score.

3.3.2. Dynamics and Articulation Parameter Controlled Synthesis

We developed a software MIDI sequencer that allows to modify play rate, loudness, and articulation of a piece during playback. Play rate, articulation, and dynamics are controlled by specifying floating point modifiers. The modifiers are relative to the original MIDI sequence. A tempo modifier of 0.5, for example, designates the half tempo of the original sequence, a dynamics modifier of 0.5 designates half MIDI velocity, and an articulation modifier of 0.5 designates that the notes are all notes on the piece are half as long as in the original sequence. The sequencer modifies MIDI velocities to change loudness and changes the timing of the note-off messages to change articulation. The MIDI signal of the sequencer is output to the attached MIDI keyboard, which generates the sound.

4. EVALUATION

The PCRP method was implemented for piano duo. To allow local testing, the system allows to introduce artificial delays between the participants.

A first version of the PCRP system, Tempo-PCRP, was evaluated with piano teachers and students of the Academy of Musical Arts Darmstadt. At that time, the system used tempo parameters only. The system has meanwhile been extended and also uses dynamics and articulation parameters. The extended version, Full-PCRP, has been informally evaluated with pianists.

4.1. Selecting the Piece

Because the pianists participating in our user studies could not be expected to prepare a piece in advance, a piece was selected for them. The piece had to fulfill certain requirements:

Difficulty: The users should be able to perform the piece from first sight or after little practice.

Length: The piece should be short so that the experiments could be executed in appropriate time.

Uniformity and Variation: It should be possible to pinpoint variations of tempo, dynamics, and articulation to specific parts of the piece. Other parts should require only little variation of tempo, dynamics and articulation. The different parts of the piece would challenge the PCRP system differently. The ability to pinpoint variations to specific points in the piece also helps in the discussions with the participants.

Diabelli's Scherzo op. 149/6 for four hands fulfills the requirements and was therefore chosen. The piece has uniform parts that are played with steady tempo and little variation of dynamics and articulation. These uniform parts are interrupted by culmination points, where loudness may be increased and tempo may be reduced, caesuras, where the tempo becomes blurred, and contrasting parts, where the player might vary tempo, dynamics and articulation.

4.2. Tempo-PCRP

We evaluated the system with 12 pianists. The pianists were piano teachers and piano students at the Academy of Musical Arts Darmstadt. They formed 6 piano duos. First, each duo practiced the piece (without PCRP) until they played it without perceived errors. After explaining and showing the PCRP system to the users, they would play via PCRP five times with artificially introduced delays of 50, 100 150, 200, and 250 ms. The pianists used headphones so that they could not hear each other directly. Additionally, the pianists played together without PCRP with latencies of 50, 100, and 150 ms. The MIDI data of the performances was recorded.

The pianists used headphones so that they could not hear each other directly. The keyboards faced in different direction with an angle of approximately 90 degrees, so that the pianists could not see each other while playing.

After each run, the two participating pianist were asked to score the ease of playing together with a value from 0 to 5. A rating of 0 means that playing together was impossible. A rating of 1 to 5 means that playing together was difficult to easy. In addition to this subjective measure, the errors made by the players were counted. The determination of errors was done automatically by using the score follower that had been implemented for the PCRP system. Whenever a played note could not been matched, an error was reported (see [1] for details).

PCRP was always rated higher than 0. Playing together via PCRP was possible for all players at all examined delays. The scores of playing together with and without PCRP are presented in Table 1.

Three pianists rated playing together at a delay of 150 ms without PCRP to be impossible. Comparing the score results of the same delay once with PCRP and once without, it is evident that at a latency of 50 ms, the pianists preferred a system without delay compensation to PCRP. However, at latencies of 100 and 150 ms, the pianists preferred to play with Tempo-PCRP. To not unnecessarily frustrate the participants, the system without delay compensation was not used at latencies above 150 ms.

The errors counted when playing via PCRP can be found in Table 2. The error count did not steadily increase when PCRP was used at higher latencies. This is an indication of the resilience of PCRP against delay.

In discussions with the participants, areas for improvement were identified. The participants suggested that the approach should be extended to other musical parameters than tempo. In the meantime, the system has been extended and uses dynamics and articulation parameters to provide richer control. A reoccurring topic in the discussions with the participants was the sensitivity of the system. Some participants felt that the system was exaggerating the tempo changes of their co-player. The new tempo of the score synthesis depends only on the last received

Delay (ms)	50	100	150	200	250
Score with PCRP	3.7	4.5	3.2	4.1	3.5
Score without PCRP	4.1	2.3	1.1	unusable	

Table 1. Average score at different delays with and without Tempo-PCRP

Delay	50ms	100ms	150ms	200ms	250ms
Errors	5.2	2	9.3	7.3	2.9

Table 2. Average errors at different delays using Tempo-PCRP

tempo parameter. Using multiple tempo parameters (averaging) to determine the new tempo would make the tempo changes smoother but would also make the system less responsive to deliberate changes.

4.3. Full-PCRP

To provide a richer experience, the Tempo-PCRP system was extended to support dynamics and articulation parameters. The new system, Full-PCRP, was informally evaluated with pianists. In contrast to the evaluation of the Tempo-PCRP system, where interplay via PCRP was examined, the user study of the Full-PCRP system aimed to evaluate the accuracy of the synthesized performance. Therefore, the participants did not form piano duos but instead performed alone. They heard the performance of the remote synthesis module; their own instrument was muted so that the participants could not hear their own signal. The synthesized performance was discussed with the participants. The participants felt that they had control over tempo, dynamics and articulation. In the participants opinion, the system rendered slow transitions well, e.g., a slow transition from soft to loud or a slow transition from short to long held notes. Abrupt changes of dynamics or articulation between long passages are also rendered well, as the system adapts the new playing style in short time. Abrupt changes in dynamics and articulation between very short segments, e.g., after each note, are not rendered well by the Full-PCRP system. The response time of the system is (principally) too slow to render these changes accurately. This issue has to be addressed in future versions of PCRP. The participants of the user tests felt that the possibility to control dynamics and articulation enhances the system.

5. DISCUSSION

The implemented PCRP system allows control of tempo, dynamics, and articulation of a pre-recorded score. The synthesis of the score is controlled by three parameters that express the relation of the played tempo, dynamics, and articulation to the tempo, dynamics, and articulation of the recorded score. The combination of these parameters can create very distinct renderings of the pre-recorded score. Even more, when considering that the parameters are updated with every played note. In principle the parameters suffice to model delicate variations. For example, the tempo parameter could be used to control note timings. However, the player has to anticipate the note timing ahead in time. The same also applies to the control of dynamics and articulation. The implemented PCRP system therefore offers only limited control over delicate deviations.

Related approaches cannot be used at very high latencies or provide an interaction that is not compatible with traditional ensemble play. PCRP offers a musical interaction that is not as rich as live ensemble play but still provides enough possibilities for meaningful musical interaction between the players.

6. CONCLUSION

PCRP is a new method to play music together under high delay in the Internet. The score of the piece has to be known beforehand. Deviations of the players from the score are analyzed and described as parameters. These parameters are transmitted instead of the audio data itself. On the opposite side of the network, the transmitted parameters are used to control the synthesis of the score. We implemented a PCRP system for a piano duo that supports tempo, dynamics, and articulation parameters. The tempo parameter lets the users play synchronized and influence each other's tempo. The dynamics and articulation parameters allow the users to control dynamics and articulation of the remote synthesis.

Delay can be directly perceived by a user of a system without delay compensation or Artificial Delay systems. A user of PCRP, however, cannot directly perceive the delay. PCRP exchanges the delay of the audio signal with the delay of the parameter signal: Changes in playing style are first perceivable for the other player after the corresponding parameters have been received. Delayed perception of a changed playing style is less critical for interplay at high delay than a lasting high delay on the audio signal. Therefore, PCRP can be used for interplay at higher delays than plain NMP systems without delay compensation strategies or Artificial Delay systems. In an evaluation with pianists, a PCRP system was compared to a system without delay compensation. At latencies of 100 ms and above the users clearly preferred PCRP. As Artificial Delay systems have only a small advantage over systems without delay compensation (only 15 ms in a conducted experiment with a piano duo [5]), PCRP should also be better suited for interplay than Artificial Delay systems at high delay. In contrast to Intervallic Jamming systems, which can be used at very high latencies, PCRP provides an illusion of a traditional musical interaction

7. FUTURE WORK

Abrupt changes in dynamics and articulation from note to note are not rendered well by the system, because the response time is (principally) too slow. As abrupt changes from note to note are not uncommon in music, improvement of their handling is an interesting goal. Fortunately, the abrupt changes often form musical patterns, e.g., a sequence of loud and soft when a grace note is dissolved. Knowledge of musical patterns could be used by a PCRP system to build a model of the user. The system could, e.g., analyze how the user dissolves grace notes and use this information for the synthesis of corresponding parts.

Information about musical processes could also be used to improve the control of the tempo. If a PCRP system had information where a tempo change, e.g., a ritardando (slowing down), could be expected, then the system could anticipate the change and consequently improve the rendering of the change. Furthermore, the system could have a reduced sensitivity where tempo is expected to remain stable. This would make it even more predictable and easier to control.

PCRP could also be applied to other instruments. It would be possible to include instrument specific parameters. For example, for string instruments a vibrato parameter could be used.

8. REFERENCES

[1] Bloch, J. J., and Dannenberg, R. B., Real-Time Computer Accompaniment of Keyboard Performances, Proceedings of the International Computer Music Conference (ICMC 1985), 1985.

[2] Carot, A., Krämer, U., and Schuller, G., Network Music Performance (NMP) in narrow band networks, Proceedings of the Audio Engineering Society 120th Convention, 2006.

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

[4] Chew, E., Zimmermann, R., Sawchuck, A. A., Kyriakakis, C., Papadopoulos, Francois, A. R. J., Kim, G., Rizzo, A.., and Volk, A., Proceedings of the 4th Open Workshop of MUSICNETWORK: Integration of Music in Multimedia Applications, 2004.

[5] Chew, E., Zimmermann, R., Sawchuck, A. A., Papadopouslos, C., Kyriakakis, C., Tanoue, C., Desai, D., Pawar, M., Sinha, R., Meyer, W., A Second Report on the User Experiments in the Distributed Immersive Performance Project, Proceedings of the 5th Open Workshop of MUSICNETWORK: Integration of Music in Multimedia Applications, 2005.

[6] Chew, E., Sawchuck, A., Tanoue, C., and Zimmermann, R., Segmental Tempo Analysis of Performances in User-Centered Experiments in the Distributed Immersive Performance Project, Proceedings of the Sound and Music Computing ’05 International Conference (SMC ’05), 2005.

[7] Cooperstock, J. R., Roston, J., and Woszczyk, W., Broadband Networked Audio: Entering the Era of Multisensory Data Distribution, 18th Internation Congress of Acoustics, 2004.

[8] eJamming Website, http://www.ejamming.com, 2008.

[9] Gang, D., Chockler, G.V., Anker, T., and Kremer, A., TransMIDI: A System for MIDI Sessions Over the Network Using Transis, Proceedings of the International Computer Music Conference (ICMC 1997), 1997.

[10] Goto, M., Neyama, R., and Muraoka, Y., RMCP: Remote Music Control Protocol - Design and Applications -, Proceedings of the International Computer Music Conference (ICMC 1997), 1997.

[11] Nagashima, Y., Hara, T., Kimura, T., Nishibori, Y., GDS (Global Delayed Session) Music – new improvisational music with network latency, Proceedings of the International Computer Music Conference (ICMC 2003), 2003.

[12] Ninjam Website, http://www.ninjam.com, 2008.

[13] Obu, Y., Kato, T., and Yonekura, T., M.A.S.: A Protocol for a Musical Session in a Sound Field where Synchronization between Musical Notes is not guaranteed, Proceedings of the International Computer Music Conference (ICMC 2003), 2003.

[14] Sarkar, M., and Vercoe, B., Recognition and Prediction in a Network Music Performance System for Indian Percussion, Proceedings of the International Conference on New Interfaces for Musical Expression (NIME07), 2007.