A Jazz Session System for Interplay among All Players

— VirJa Session (Virtual Jazz Session System) —

Masataka Goto Isao Hidaka Hideaki Matsumoto
Yosuke Kuroda Yoichi Muraoka

School of Science and Engineering, Waseda University
3-4-1 Ohkubo Shinjuku-ku, Tokyo 169, JAPAN.
{goto, hidaka, matumoto, ykuroda, muraoka}@muraoka.info.waseda.ac.jp

Abstract This paper presents a jazz session system in which each player is independent and can interplay with all other players. In most previous systems, computer players only reacted to human player's performance with a fixed leader-follower relationship. Our system enables computer players to listen to other computer players' performances as well as human's performance and to interact with each other. Moreover, all players can communicate not only by listening to other players' performances but also by seeing each others' bodies and gestures. In our current implementation, the system deals with a jazz piano trio consisting of a human pianist, a computer bassist, and a computer drummer. These computer players have been implemented as separate processes on several distributed workstations.

1 Introduction

It is important in jazz that all players improvise together while reacting to other players' performances with no leader-follower relationship. This is called *interplay*, and all players communicate by musical sounds and additional visual information such as gestures. The purpose of this research is to simulate the actual interaction that occurs among human players on a computer system that enables interplay among humans and computers.

Most previous jazz session systems [Nishijima and Watanabe, 1992; Kondo *et al.*, 1993; Wake *et al.*, 1994] and automatic accompaniment systems [Dannenberg, 1984; Vercoe, 1984; Baird *et al.*, 1989; Rowe, 1993; Horiuchi and Tanaka, 1993; Hidaka *et al.*, 1995] generated several accompaniment parts together, which only reacted to the soloist's (the human player's) performance. There was a fixed relationship where the human player was the leader and the other computer players were the followers. Since computer players only listened to the soloist's performance, they were not able to interact with each other. Although one previous paper [Kanamori *et al.*, 1993] mentioned interaction among computer players, it did not discuss the generation of musical performance and only focused on the music-listening process.

Our jazz session system enables a human player and computer players to interact with each other without a fixed leader-follower relationship. In our system, a computer player not only reacts to the human player's performance but also plays even a kind of solo. Computer players can react to each other by listening to other computer players' performances just as a human player listens to all other players in an actual jazz session.

This paper moreover presents an advanced jazz session system called VirJa Session (Virtual Jazz Session System). VirJa Session enables all players to communicate not only by listening to other players' performances but also by seeing each others' bodies and gestures that indicate repetition or the end of a song. The human player sees the bodies and gestures of other computer players shown on computer graphics (CG), and thus can feel their presence as if they were actually playing together. In addition, each computer player sees the gestures of the human player through a video camera, and recognizes the virtual gestures of the other computer player through a computer network. Thus, all players can interact with each other using both auditory and visual information.

In our current implementation, our system deals with a four-beat jazz standard with a constant tempo performed by a piano trio consisting of a human pianist, a computer bassist, and a computer drummer. To perform this computationally intensive task in real time, the system has been implemented on several distributed workstations connected to a computer network. The computer bassist and drummer are implemented as separate processes to facilitate system implementation and expansion. Each of them generates and sends MIDI data of his performance reacting to MIDI data received from both the human pianist and the other computer player. In our experiment, we achieved a jazz session in which all players improvised and interacted using both musical sounds and gestures.

2 Jazz Session Model for Interplay among All Players

Figure 1 shows a jazz session model in which all players listen and react to all other players. Although the figure only depicts a piano trio configuration consisting of a human pianist, a computer bassist, and a computer drummer, this model is applicable to various sessions consisting of several players.

In the model, either human or computer can be selected for each player; only system implementation restricts the configuration. We can imagine sessions in which all players are human players and we can imagine sessions in which all players are computer players. The former are useful for remote sessions via computer networks. The latter are effective when various design-

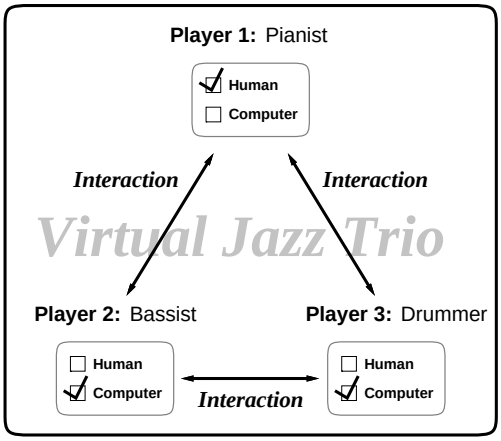

Figure 1: Jazz session model for piano trio.

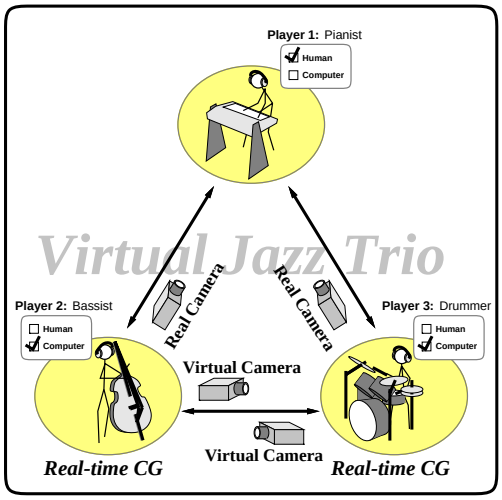

Figure 2: VirJa Session.

ers implement different computer players with various characteristics. These computer players may be thought of as the designers' substitutes, and these substitutes can interact with each other.

We built a jazz session system for the piano trio as illustrated in Figure 1. Both player 2 and 3 are computer players, and it is important that these computer players are executed as separate processes; the system should not determine their performances together according only to the performance of player 1, the human player. Computer player 2, for example, should listen and react the performance of player 3 as well as the performance of player 1. Player 2 (bassist) and player 3 (drummer) thus interact with each other through their musical performances.

3 Virtual Jazz Session System with CG and Camera – VirJa Session

Figure 2 shows the advanced jazz session system, VirJa Session, in which all players communicate both by listening to other players' performances and by seeing each others' bodies and gestures. Each player presented in the last section has only his ears to listen to the performances and only his hands and feet to play the musical instruments. Each player in the VirJa Session is additionally able to use his eyes to see other players and his body to show his gestures. Since each player thus exchanges gestures in addition to musical sounds, we can achieve multimodal interaction among all players using both auditory and visual information. This enables the players to feel the presence of other players as if they were actually playing together, and enables audiences to feel the presence of all players as if they were attending a live concert, and not listening to a compact disc.

The human player sees the bodies and gestures of the computer bassist and drummer shown through three-dimensional CG animation in real time. Each computer player recognizes gestures of the human player through a real video camera, and receives gestures of the other computer player through a virtual camera that means passing messages corresponding to the gestures via computer network. This exchange of gestures enables the players to cooperate well through information that is difficult to exchange only using musical sounds.

3.1 Visualization of Computer Players on CG animation

The entire body of each computer player and his musical instruments (bass or drums) are shown on a CG screen. The CG player makes the following motions: (1) playing musical instruments according to his musical sounds, (2) keeping time to musical beats by foot-tapping or rocking his body, (3) making two kinds of gestures described in Section **3.2**, (4) nodding to show that he understood a gesture, and (5) turning his eyes to another player.

Visualization of a musician on CG was reported in [Katayose *et al.*, 1993; Kamei *et al.*, 1995], but these papers only focused on hand movements of a virtual guitarist. Although an earlier paper [Goto and Hashimoto, 1993] discussed visualization of musicians, the system only generated a CG dancer, not CG musicians.

3.2 Multimodal Interaction

In VirJa Session, we introduce a song form called *scenario*, which allows a player to change his way of playing according to where he plays in it. A scenario is given in advance of a performance as a combination of song parts such as *theme*, *player solos*, and *four verse*[1]. The number of repetitions of each song part in the scenario is not pre-arranged, and is determined dynamically by interaction using gestures among players, as in actual sessions among human players. Our current implementation only limits the maximum number of repetitions.

We introduce two gestures, one indicating that the next player should begin his solo and one indicating that the players should return to the theme (reprise). The first gesture is done by a player leaning to the left or

[1] Players, for example, may perform according to the following scenario: theme ⇒ piano solo ⇒ bass solo ⇒ four verse (piano solo and drums solo alternate every four measures) ⇒ theme (reprise).

right, and directs the player on the selected side[2] to take up the solo. The second gesture is done by a player pointing to his head.

Toward the end of each song part in a scenario, each computer player frequently turns his eyes (which show where his attention is directed) to another player to watch for gestures indicating whether the current song part should be repeated or the next part should be taken up. The human player should make a gesture while computer players are looking at him. When a computer player understands a gesture, he nods to the player who made the gesture.

4 Computer Bassist and Drummer

Each computer player performs in three stages: *Music Listening*, *Session Understanding*, and *Performance Improvising*. In the *Music Listening* stage, he analyzes the musical performance of each of the other players by using the intention parameters described in [Hidaka *et al.*, 1995] according to the character of musical instruments, and estimates how much each player leads the session in the whole ensemble (*leadership percentage*). In the *Session Understanding* stage, he considers the whole musical relationships among players by predicting the next leadership percentage for each player based on recent transitions of the leadership percentage. He then determines how much he tries to lead the session based on where he plays in a scenario; the bassist, for example, tries to take a stronger leadership role during the bass solo. In the *Performance Improvising* stage, he chooses a pitch pattern, a loudness pattern, and a rhythm pattern from pre-registered pattern databases. These patterns are chosen in such a way that he can take the leadership percentage determined in the *Session Understanding* stage. He then combines these patterns on every beat to generate his musical performance.

5 Implementation in Distributed Computing Environment

The following six tasks are necessary in VirJa Session:

1. understand and improvise musical performance for each computer player (*computer player's brain*)
2. keep the tempo of the whole performance
3. output sounds of the computer players' performances for the human player
4. take the human player's performance as input for the computer players
5. display each computer player through CG animation (*computer player's body*)
6. recognize the human player's gestures through a video camera (*computer player's eyes*)

To perform these computationally intensive tasks in real time, each of these tasks is implemented as a separate software process, and these processes communicate on the basis of the server-client model. This implementation makes it possible to allocate these tasks on several distributed computers on the network, and facilitates system implementation and expansion.

[2] The triangle arrangement of the three players facing each other is given in advance.

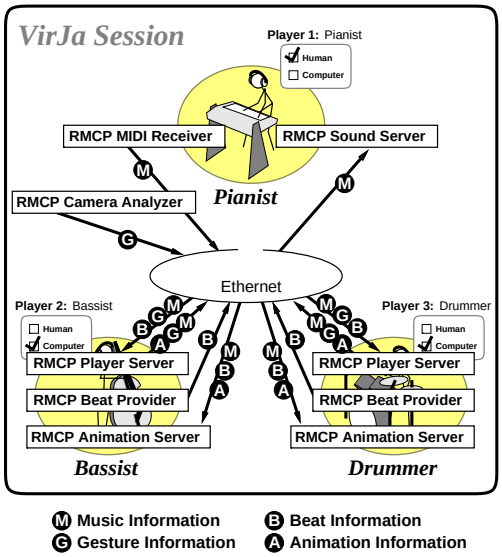

Figure 3: System configuration of VirJa Session *on RMCP.*

In our current implementation the system works on three workstations (SGI Indigo2 Impact, Extreme) connected to the Ethernet. We use MIDI (Musical Instrument Digital Interface) to handle the players' performances and RMCP (Remote Music Control Protocol) [Goto and Hashimoto, 1993] for inter-process communication.

5.1 RMCP

RMCP is a communication protocol on the UDP/IP between servers and clients in a distributed cooperative system that integrates MIDI and LAN. RMCP was designed to transmit symbolized musical information through networks. In VirJa Session, the following four kinds of information are broadcast as RMCP packets: (1) music information (MIDI messages), (2) beat information (temporal positions of quarter notes, tempo), (3) gesture information (leaning to the left or right, pointing to his head), and (4) animation information (direct a computer player on CG to make a gesture, to nod, and to turn his eyes). This broadcast enables several computers on the Ethernet to utilize and share the information in various ways at the same time. Since each RMCP packet has a time stamp (msec-resolution) in its header, packets that are received by RMCP servers before their time stamps can be processed on time and in order.

5.2 Implementation of VirJa Session

Figure 3 shows the system configuration of VirJa Session on RMCP. The above-mentioned six tasks are executed as follows: task 1 (music understanding and improvising) is executed by two *RMCP Player Servers* for both computer players, task 2 (tempo keep) is executed by the *RMCP Beat Providers*, task 3 (sound output) is

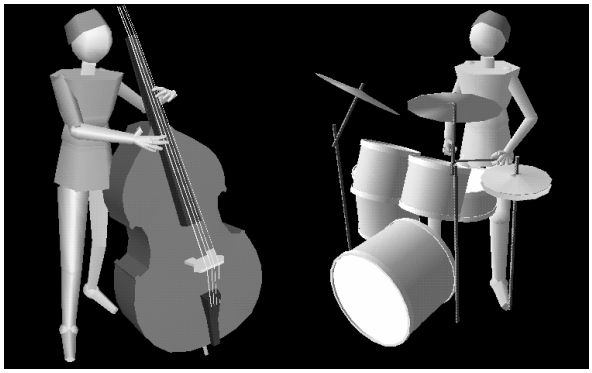

Figure 4: An example of CG output (bassist and drummer).

executed by the *RMCP Sound Server*, task 4 (performance input) is executed by the *RMCP MIDI Receiver*, task 5 (CG output) is executed by the *RMCP Animation Servers*, and task 6 (gesture recognition) is executed by the *RMCP Camera Analyzer*.

The kinds of RMCP packets that these processes broadcast or receive are also shown in Figure 3. Each *RMCP Player Server*, for example, receives music information broadcast by the *RMCP MIDI Receiver* and the other *RMCP Player Server*, and broadcasts music information of his improvised performance, which is used by the corresponding *RMCP Animation Server* to generate CG motions.

6 Experiments and Results

The VirJa Session was tested on a four-beat jazz standard named '*Take the "A" Train*' by a pianist who has been playing jazz piano for 5 years (and classical piano for 21 years). We provided the theme, chord progression, key signature, and tempo (187-230 M.M.) for the system in advance, and tried several scenarios that included theme, bass solo, piano solo, and four verse (piano and drums solo). A MIDI synthesizer (Korg 01/W FD) was used for the sound output and for the keyboard on which the pianist played.

In our experiment, we achieved a jazz session in which all players interacted with each other without the fixed leader-follower relationship. The computer players as well as the pianist led the session and played solos by improvising. The interaction among all players made the entire performance expressive and interesting. As compared with sessions using only musical sounds, the CG animation of the computer players (Figure 4 shows an example of the CG output) gave the pianist a greater feeling of being at a live performance. We also achieved multimodal interaction among players using the gestures in addition to musical sounds.

7 Conclusion

We have described a jazz session system called VirJa Session in which all players listen to the other players' performances and see each others' bodies and gestures. In this system, the players were able to improvise without a fixed leader-follower relationship and to interact using both musical sounds and gestures. We have also presented our implementation on distributed workstations using the RMCP, which facilitated system implementation and expansion.

We plan to upgrade the system so that it can follow tempo changes and to support other configurations, such as a piano trio in which all players are computer players, and sessions with different numbers of players. Future work will include the study of a remote jazz session, since our implementation facilitates use of the system in settings where the players are not in the same physical location.

Acknowledgments

We thank Mineko Ogata, Kaoru Asatani, Tetsuya Ohmori, and Shin'ichiro Mayuzumi for their cooperation in our experiments and their helpful comments.

References

[Baird *et al.*, 1989] Bridget Baird, Donald Blevins, and Noel Zahler. The artificially intelligent computer performer on the Macintosh II and a pattern matching algorithm for real-time interactive performance. In *Proc. of ICMC 1989*, pages 13–16, 1989.

[Dannenberg, 1984] Roger B. Dannenberg. An on-line algorithm for real-time accompaniment. In *Proc. of ICMC 1984*, pages 193–198, 1984.

[Goto and Hashimoto, 1993] Masataka Goto and Yuji Hashimoto. A distributed cooperative system to play MIDI instruments – toward a remote session – (*in Japanese*). *IPSJ SIG Notes*, 93(109):1–8, 1993.

[Hidaka *et al.*, 1995] Isao Hidaka, Masataka Goto, and Yoichi Muraoka. An automatic jazz accompaniment system reacting to solo. In *Proc. of ICMC 1995*, pages 167–170, 1995.

[Horiuchi and Tanaka, 1993] Yasuo Horiuchi and Hozumi Tanaka. A computer accompaniment system with independence. In *Proc. of ICMC 1993*, pages 418–420, 1993.

[Kamei *et al.*, 1995] K. Kamei, K. Sato, H. Katayose, and S. Inokuchi. Animation of human motion by editing real images and employing 3-D hand model (*in Japanese*). *Trans. of IPS Japan*, 36(2):374–382, 1995.

[Kanamori *et al.*, 1993] T. Kanamori, H. Katayose, S. Inokuchi, H. Hirai, and Y. Niimi. Interpretation of musicality in jazz improvisation using multi-agent model. In *Proc. of IAKTA/LIST Intl. Workshop on Knowledge Technology in the Arts*, pages 107–114, 1993.

[Katayose *et al.*, 1993] H. Katayose, T. Kanamori, K. Kamei, Y. Nagashima, K. Sato, S. Inokuchi, and S. Simura. Virtual performer. In *Proc. of ICMC 1993*, pages 138–145, 1993.

[Kondo *et al.*, 1993] Kinya Kondo, Haruhiro Katayose, and Seiji Inokuchi. The session system reacting to the player's intention (*in Japanese*). In *Proc. of the 46th Annual Convention IPS Japan*, pages 1:373–1:374, 1993.

[Nishijima and Watanabe, 1992] Masako Nishijima and Kazuyuki Watanabe. Interactive music composer based on neural networks. In *Proc. of ICMC 1992*, pages 53–56, 1992.

[Rowe, 1993] Robert Rowe. *Interactive Music Systems*. The MIT Press, 1993.

[Vercoe, 1984] Barry Vercoe. The synthetic performer in the context of live performance. In *Proc. of ICMC 1984*, pages 199–200, 1984.

[Wake *et al.*, 1994] S. Wake, H. Kato, N. Saiwaki, and S. Inokuchi. Cooperative musical partner system using tension parameter: JASPER (jam session partner) (*in Japanese*). *Trans. of IPS Japan*, 35(7):1469–1481, 1994.