The Creation of a Multi-Human, Multi-Robot Interactive Jam Session

Gil Weinberg
Center for Music Technology
Georgia Tech
Atlanta, GA 30332
gilw@gatech.edu

Brian Blosser
Center for Music Technology
Georgia Tech
Atlanta, GA 30332
bblosser3@gatech.edu

Trishul Mallikarjuna
Center for Music Technology
Georgia Tech
Atlanta, GA 30332
tmallikarjuna3@gatech.edu

Aparna Raman
College of Computing
Georgia Tech
Atlanta, GA 30332
araman9@gatech.edu

Abstract

This paper presents an interactive and improvisational jam session, including human players and two robotic musicians. The project was developed in an effort to create novel and inspiring music through human-robot collaboration. The jam session incorporates Shimon, a newly-developed socially-interactive robotic marimba player, and Haile, a perceptual robotic percussionist developed in previous work. The paper gives an overview of the musical perception modules, adaptive improvisation modes and human-robot musical interaction models that were developed for the session. The paper also addresses the musical output that can be created from increased interconnections in an expanded multiple-robot multiple-human ensemble, and suggests directions for future work.

Keywords: Robotic musicianship, Shimon, Haile.

1. Introduction

A common goal in human-computer interactive music systems is to take advantage of the memory and real-time processing power of computers in conjunction with human expression and creativity to create novel and inspiring music. In such settings, a computer exhibits *machine musicianship* [1] in that it takes in musical input from a human, analyzes the input using human perceptual framework (such as rhythmic stability [2]) and generates improvisational responses informed by the human input. The algorithmic computer-generated response can then inspire the human to create music in novel ways. However, musical interaction with a computer is often limited by the unanimated nature of the computer, and by the fact that the musical outcome can only come out of loudspeakers.

Therefore, we've attempted to explore how musical robots can overcome these limitations by having a physical, anthropomorphic form, which humans can see and synchronize with, and by actualizing the musical sounds acoustically. Similar works with robotic musicians and human-robot interactive music include [3] and [4].

In our earlier work, Haile (/ˈheɪli/), a robotic percussionist, was conceived to interact rhythmically with two human percussionists on a Native American pow-wow drum [5]. In order to explore the melodic dimension of robot-human musical interaction, Haile was modified to play a toy xylophone, generating melodic responses based on a genetic algorithm [6]. While it served as an initial platform for melodic experiments, the system was limited to one octave and was restricted to score-based interaction.

Shimon (/ˌʃɪˈmoʊn/), a newly-developed robotic marimba player, extends our previous work by expanding the range of melodic actualization, introducing new modes of melodic perception and improvisation, and incorporating social interaction schemes using a social anthropomorphic head. Shimon has four striking arms installed on a 7' horizontal slider, which can move across a range of two octaves in under half a second, covering altogether the full four-octave range of the marimba. Each arm is fitted with two striking mallets driven by rotational solenoids, capable of hitting in a continuum of strike velocities. Shimon can play up to four simultaneous notes with a frequency of over 10 Hz per mallet. Unlike Haile, Shimon has an anthropomorphic face, designed by Andrea Thomaz's group at the Georgia Tech RIM Center. The head is currently rendered in animation while the hardware is under development. It can nod, turn, and control facial expressions, which are used to provide musical and social cues to the human collaborators. We believe that the incorporation of these cues, not yet extensively addressed in the space of human-robot musical interaction, would greatly improve the fluidity of interaction.

In our current research, we incorporate both Haile and Shimon in a human-robot improvisatory jam session. In the session, Haile plays a pow-wow drum, Shimon plays a marimba, and two humans play a keyboard and darbuka drum (see example video clips online at [7] and [8]). The jam session configuration offers musical flexibility for robot-human musical collaborations, as the interaction in the session is not driven by a pre-determined score, and the robots and musicians dynamically adapt to each other's playing. To facilitate this interaction we developed software modules for perception of dynamic aspects of musical performance, robotic improvisation that adapts to any musical input, and fluid models of human-robot musical interaction. The inclusion of multiple robots in the

NIME09, June 3-6, 2009, Pittsburgh, PA

jam session opens the door for new kinds of interactions, including mutual robot-human inspiration and influence across instrumental modalities (rhythmic and melodic) as well as robots influencing the improvisations of other robots, thereby facilitating our end goal of creating novel music through mutual inspiration among humans & robots.

2. Perception

2.1 Interest Perception for Social Head

For facilitating interaction with all participants in the musical jam session, Shimon responds to the musical input by registering interest in the changing musical flow of the session. It does this by turning and looking at the player that it thinks is of the most interest at any instant. We calculate 'interestingness' measures of each sound source, including Shimon itself, on the basis of user-set weights on factors such as note density, volume, beat frequency and instrument proximity and preference. These values are calculated and compared to get the 'interest leader' – the player of most interest at any instant. Shimon turns to look at the leader for a period of time determined by the relative interest value of the leader. If Shimon finds itself to be the leader, it looks down at its playing arms and tracks their progress over the keys to give an impression of concentrating on the music it is creating.

2.2 Beat Detection

A key component in adapting human-robot musical interaction to an improvisatory jam session is the application of a real-time beat tracking algorithm. This algorithm allows Haile to continuously detect the beat of the human drummer based on the acoustic input it receives from a microphone attached to the human's drum. The human drummer, therefore, can play freely as Haile detects the beat and joins in. Unlike discrete and static tempo detection approach in previous work, a dynamic beat detection approach allows Haile to adjust to the tempo of the human drummer at any time during the performance. This allows the human drummer to play in a freer and more expressive manner. The beat detection algorithm used for the project was originally implemented by Scott Driscoll, but had never been used in a free jam session. The approach (based on [9]) converts audio input to an onset-detection function, then uses autocorrelation and comb filterbanks to extract the tempo and beat alignment. The beat information from Haile is also used by Shimon for synchronized head-nodes and for aligning the melodic improvisation to the beat.

3. Improvisation

In order to create a melodic improvisation that adapts to any musical context, Shimon performs Markov Chain statistical analysis on the note input from the human keyboard performer. This creates a stylistic model of the melodic input, and allows for the robot to respond with note sequences that are different from the input, while conforming to the style of the song. As an initial implementation, we had a fixed order Markov chain tracking the succession of notes. The system provides for continuous 'training' of the Markov chain in the background while the robot plays, allowing the keyboard player to join in with the robot and change or add to the melodic organization of the song. The system also provides for a parallel line of melody/harmony on a different octave using the extra arms of the robot, based on simple operations such as direct transposition, inversion, metrically synchronized delay and others, leading to a richer musical outcome.

Haile's rhythmic improvisation results from stochastic transformation of the previously recorded human rhythms. In addition, Haile switches probabilistically between playing dense and sparse rhythms in order to complement the density of its earlier playing.

4. Interaction

4.1 Shimon: Social Cues

To make the robot seem more engaging and responsive, Shimon's social head provides visual cues to the human musicians in the improvisational setting. It also provides a way for the humans to connect and synchronize rhythmically through modeling a common human response to beats – head nodding. Shimon nods its head in time with the beat that is detected from the human drummer, or the keyboardist if no drum input is present.

In addition, Shimon's head direction and gazing provide a means for communication with the human musicians by signaling Shimon's interest in listening to new material. This emulates the kind of visual cues that real musicians exchange for collaboration when playing and improvising without any pre-determined musical score.

4.2 Shimon: Turn-Taking Model

Our earlier work on melodic improvisation with Haile [10] involved transitions between interaction modes that were essentially score-driven, with different modes pre-allocated for different sections of the composition. With Shimon, we have experimented with making the transitions seamless and automatic, driven by real-time input. Figure 1 represents the various perceptual and improvisational states of the musical interaction module. The transitions between these states are governed by timers that are triggered by low-level real-time performance information as well as high-level preset stylistic information. At the low-level, the transitions are aligned in real time to multiples of the discrete beat durations derived out of human input. At the high-level, ranges of optimum durations are preset individually for each of the timers based on the style of music and skills of the co-performers. The robot selects values from within these ranges during the session.

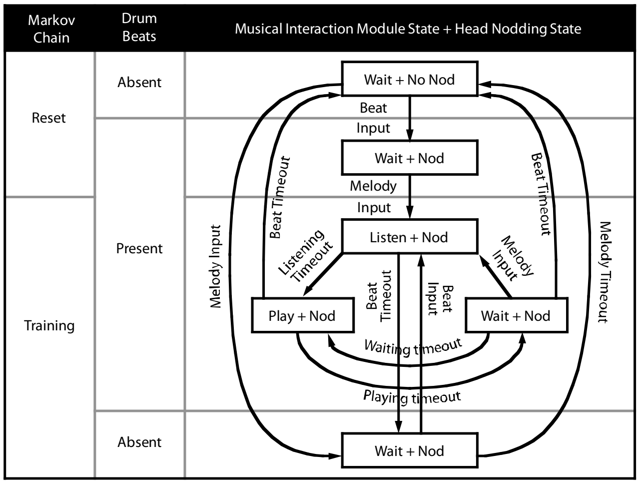

Figure 1: State diagram for the interaction modules

The states of the modules and the various transitions are designed to provide for a number of elements of interaction in a jam session. Songs are demarcated by a gap in the reception of melodic input or the rhythmic beats or both, and the Markov model is reset at this transition to allow for a new style for the new song. The human keyboardist may just play a few notes or a line or two of a song and expect the robot to follow up in a 'call-and-response' manner (typical in the case of an amateur performer), or he may continue playing for a long period of time expecting the robot to 'accompany' him (typical in the case of a seasoned performer and in styles like jazz). In addition, the system provides for distinct periods where the robot stops playing and 'listens' to the keyboard player, and at these times it gives the keyboard player a chance to change the metric structure of the notes to be played later in the session, allowing for 'listen-and-follow' interaction. Furthermore, the system allows for a 'solo' mode, wherein the robot starts playing melodic segments out of its current Markov 'experience' if it detects a prolonged gap in the keyboard input. In this case the silence of the human keyboard player serves as implicit agreement that Shimon should start playing, similar to the turn-taking model in [11].

4.3 Shimon: Dynamic Adjustment to Beat Changes

While Shimon derives the metric structure for its improvisation from that of the keyboard input, it uses the beat data from Haile for synchronization of its improvisation, as the beat of the music can change dynamically. When Shimon is listening to the keyboard input, the relative metric structure of the input, consisting of the note durations, is placed in a timing grid demarcated by the current beat pulses. During improvisation, if the tempo has changed in comparison to the value at the time of listening, a corresponding scaling is applied to the note durations between two beats, or a beat block. The system further ensures that the intended start locations of the beat blocks fall on the beat pulses received in real time during improvisation, thus providing synchronization at the beat level.

4.4 Haile: Leader-Follower Turn-Taking Model

To increase the intuitiveness of interaction between the human drummer and Haile, a leader-follower musical interaction model was developed that governs turn-taking between Haile and the human drummer. The leader-follower model is based on the idea that when musicians interact in an improvisatory session, they switch leadership roles in a continuous loop, taking the lead using techniques such as playing loudly, densely, or in a different tempo or beat. To apply this model to human-robot musical interaction, we had to determine what the leader's role is musically and how the leader communicates his or her leadership to the other participants.

For the jam session interaction, it was decided to use the cues of volume and density to determine leadership. When the human plays loudly and/or densely, the robot concludes that the human is leading. As a follower, Haile continuously detects the beat, listens to and records the human's drumming, and provides accompaniment in synchrony with the beat. When the human plays softly and/or sparsely, the robot infers that it can lead and takes the lead. As a leader, the robot locks the accompaniment tempo in one of its arms and plays an improvised rhythm with the other arm, as shown in Figure 2.

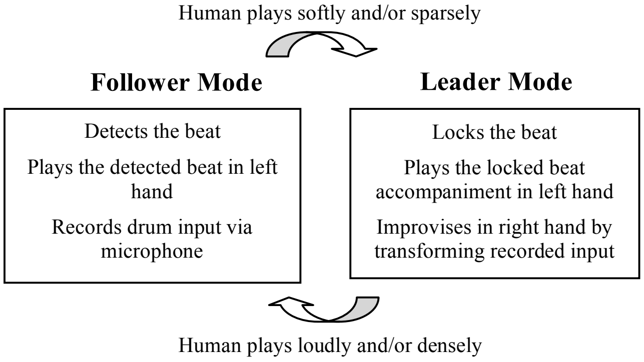

Figure 2: Leader-follower interaction scheme for Haile with volume/density leadership cue

5. Preliminary Observations and Discussion

In our previous work, the human-robot musical sessions included one robot, which played in one modality (rhythmic or melodic) and only took in human input in that modality. In our current multiple-robot ensemble, a robot in one musical modality can influence a human player in another modality. Moreover, each robot can make use of analyzed data from human input in multiple modalities, expanding the musical interconnections within the ensemble. The robots can also send their output data to each other, adding a dimension of robot-robot collaboration to the musical mix.

In a preliminary observation and discussion with participants, a human keyboardist testified to being significantly influenced by the improvisatory rhythms of Haile. This shows that the presence of a rhythmic robot in the ensemble can effect the playing of not only the human percussionist, but also the human melodic player. An example of a robot making use of input data from humans on multiple modalities was the use of the detected human beat by Shimon, the marimba robot. In addition to the MIDI data from the human keyboardist, Shimon also received the beat period and alignment from the human drummer, enabling both robots to stay synchronized to the beat of the human drummer. One can imagine an extension of the system where human rhythmic motifs are passed to Shimon and influence the rhythms of its improvisation. In terms of robot-robot musical collaboration, a module was developed that transferred pitch data from Shimon's current melodic improvisation to Haile, who then mapped it to the relative pitch positions on the drumhead during its rhythmic improvisation. Robot-robot collaboration can also take advantage of uniquely robotic qualities – like, Haile's ability to remember exactly the notes played by Shimon and use these in its own improvisation – to further expand the possibilities of robots to create unique, inspiring musical collaborations with humans.

6. Future Work

In an effort to improve the social cues between robots and humans we plan to add robotic facial expressions to convey the reaction to consonance and dissonance in the musical input. We also plan to improve the turn taking interaction, allowing the robot to show a range of attention levels (as described in [12]) within saliency-based gazing. Currently, Shimon can provide visual cues to other musicians in the ensemble but cannot take any visual input from them. We therefore intend to make the visual interaction two-sided using cameras positioned inside the eyes of the robotic head.

Future research will also address improvements in our beat-detection algorithm. We plan to extend the algorithm to make the detected beat remain stable even in the presence of expressive variation in the human drum input. In an effort to eliminate abrupt beat transitions, we will attempt to use learning-based anticipatory cues, which can be helpful, for example, in scenarios where the human gradually increases or decreases the tempo, allowing the robot to pre-adjust its tempo accordingly rather than analyzing the change of tempo in retrospect. In addition, we plan to build upon the leader-follower interaction model so that it is not strictly determined by the human player but allows for the robot to take the lead even when the human is determined as a current leader.

Lastly, we plan to develop new perceptual models for both melodic and rhythmic improvisation by Shimon to create more engaging musical collaboration. In particular, the current sub-system for melodic improvisation only models the statistics of input note succession. We plan to extend this to incorporate higher-level musical features such as phrasal structure, consonance/dissonance, stability and density, which have been separately treated in earlier work [5][6][10]. Currently, the metric structure of melodic improvisation is linearly related to that of melodic input. We plan to improve this by using a statistical perceptual model for the same. Furthermore, we plan to use these models of metric structure and density to dynamically adjust the ranges for various timers in Shimon's turn-taking model, so that the speed of interaction would match the musical 'pace' of the session more effectively.

7. Acknowledgments

We thank Andrea Thomaz and Maya Cakmak from the Georgia Tech RIM Center and Roberto Aimi from Alium Labs for their continued support with the project.

References

[1] Rowe, R. *Machine Musicianship.* Cambridge, Massachusetts: 2004.

[2] Desain, P., and H. J. Honing. 2002. "Rhythmic Stability as Explanation of Category Size." *International Conference on Music Perception and Cognition*, 2002, University of New South Wales, Sydney, July 17–21.

[3] Singer E., Feddersen J., Redmon C., and Bowen B. "LEMUR's Musical Robots", *Proceedings of the 2004 Conference on New Interfaces for Musical Expression (NIME)*, Hamamatsu, Japan, pp. 181-184

[4] Kapur A., Singer E., Benning M. S., Tzanetakis G., Trimpin "Integrating hyperinstruments, musical robots & machine musicianship for North Indian classical music" *Proceedings of the 2007 Conference on New Interfaces for Musical Expression (NIME),* New York, USA, pp. 238-241

[5] Weinberg, Gil and Driscoll, Scott. "Towards Robotic Musicianship." *Computer Music Journal.* 30:4 (Winter 2006), 28-45.

[6] Weinberg, G. and M. Godfrey, A. Rae, J. Rhoads. "A Real-Time Genetic Algorithm in Human-Robot Musical Improvisation." In *Proceedings of the International Computer Music Conference (ICMC),* 2007, pp. 192-195.

[7] Robotic Musicianship jam session – Jazz. Youtube. http://www.youtube.com/watch?v=MSxeN00C20g

[8] Robotic Musicianship jam session – Eastern. Youtube. http://www.youtube.com/watch?v=5DYOqSTmGDA

[9] Davies, Matthew and Plumbley, Mark. "Context-Dependent Beat Tracking of Music and Audio." *IEEE Transactions on Audio, Speech, and Language Processing*, 15:3 (March 2007), 1009-1020.

[10] Weinberg G., Driscoll S. "Iltur – Connecting Novices and Experts Through Collaborative Improvisation" In *Proceedings of Conference on New Interfaces for Musical Expression (NIME)*, 2005, pp. 17-22.

[11] Hoffman, G. and Breazeal, C. "Collaboration in Human-Robot Teams." *1st AIAA Intelligent Systems Conference*, Chicago, IL, September 2004.

[12] Breazeal, C. "Emotion and sociable humanoid robots". *International Journal of Human-Computer Studies,* Volume 59, Issues 1-2, July 2003, Pages 119-155.