Interactive Jamming with Shimon: A Social Robotic Musician

Gil Weinberg
Center for Music Technology
Georgia Institute of Technology
Atlanta, GA 30322 USA
gilw@gatech.edu

Aparna Raman
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332 USA
aparna.raman@gatech.edu

Trishul Mallikarjuna
Center for Music Technology
Georgia Institute of Technology
Atlanta, GA 30322 USA
trishul.mallikarjuna@gatech.edu

ABSTRACT
The paper introduces Shimon: a socially interactive and improvisational robotic marimba player. It presents the interaction schemes used by Shimon in the realization of an interactive musical jam session among human and robotic musicians.

Categories and Subject Descriptors
J.5 [**Arts and Humanities**]: Performing arts

General Terms
Performance, Design, Algorithms, Human Factors

Keywords
Robot, music, interaction, improvisation, jam, social, rhythm, beat, melody, Markov, Shimon, Haile, marimba

1. INTRODUCTION
We investigate the possibility of providing rich musical collaboration between human and robotic musicians. Our work builds on previous work on robotic musicianship using Haile (/ˈheɪli/) – the robotic percussionist [1]. Shimon (/ˌʃɪˈmoʊn/), the new robotic marimba player, extends our previous work by introducing melodic perception and improvisation modules as well as social interaction schemes using a social head (currently rendered in animation). The paper examines the various interaction paradigms designed for Shimon in a dynamic jam setting with humans and other robots (see Figure 1). Shimon analyzes melodic information from a MIDI keyboard, processes it musically and responds by playing the marimba (an orchestral mallet instrument). It also analyzes beat information based on audio input from a human drum player and integrates it into its musical and social responses. In addition, Shimon can turn and look at different human players, showing its attention based on the level of 'interest' it has in their playing.

2. HARDWARE
Shimon has four striking arms constructed on a 7' horizontal slider. The arms, designed by Roberto Aimi, can move across a range of two octaves in under half a second, covering altogether the full 4 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. In general, Shimon can play up to four simultaneous notes with a frequency of up to 10 Hz per striking 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. The head's animation can currently nod, turn, and control facial expressions, which are used to provide musical and social cues to the human collaborators.

Figure 1: A musical jam session with Shimon, Haile and two human players

3. MODES OF INTERACTION
Shimon has two main modules for its perception and improvisation – the social module that provides visual cues through the expressions and movements of Shimon's robotic head, and the musical module that processes the musical input and generates the musical improvisation.

3.1 Social interaction module
Shimon's social module aims to provide visual cues for human players, modeling the robot's behavior based on how human musicians interact when playing and listening to music in an improvisational setting. The social module has two primary actions – nodding and saliency-based gazing. The first action is based on humans' response to beats while listening to music – the robotic head nods in time with the beat using data sourced from the drum, or if it is not present, the keyboard. By watching the nods, human musicians can stay synchronized with the robot, and connect to the improvising robot in a more expressive manner. The second action – saliency-based gazing – models how musicians and listeners move their focus among sound sources based on the quality of the music played. Improvising musicians tend to look at the musical source that interests them. We, therefore, calculate the level of 'interest' by listening to all instruments and applying user-set weights for factors such as beat frequency, note density, volume and instrument preference. This 'interest' measure is determined constantly for each instrument, the most interesting instrument at any given moment is deemed

the ‘leader’, and Shimon is programmed to turn its head to look at this source accordingly.

The head's gaze is fixed on the ‘leader’ for only a small duration, determined by its relative level of interest. Shimon turns its gaze away to another source that may be the next most interesting. When Shimon determines itself to be the leader, the head looks down at the arms and the eyes move to track the progress of the arms over the keys, giving an impression of concentrating on what it is playing.

3.2 Musical interaction module

This module addresses perception, analysis and improvisation based on musical input from human and robotic co-performers. Implemented with the Max/MSP software platform, it includes a number of perceptual and improvisational states (see Figure 2). Transitions between these states are governed by timers, which are based on low-level real-time performance information as well as high-level stylistic decisions. The real-time based transitions are aligned and quantized to the beat detection module, so that the robot can seamlessly accommodate tempo changes during the session. The stylistic decisions control turn taking based on the style of music and skills of the co-performers.

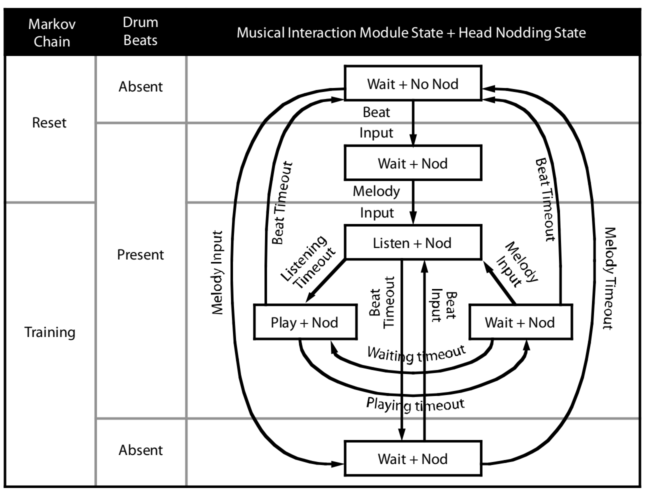

Figure 2: State diagram for the interaction modules

The melodic improvisation is based on Markov Chain statistical analysis performed on the note input from the keyboard performer. The analysis is then used to create musical responses that are stylistically similar to the input, yet introduce new musical ideas. The module also generates a parallel line of melody/harmony based on simple operations such as musical transposition, inversion, metrically synchronized delay and other transformations, which lead to a richer musical outcome. The musical input is adjusted and quantized based on the beat detected from another human drum player.

Turn taking schemes among participants are based on demarcated gaps in the reception of melodic and rhythmic input. The Markov model is reset at certain transitions to allow for a new style for the new input. The system supports three different turn taking modes: In Call-and-Response Mode the human plays a few notes or phrases before the robot provides its response. In Accompaniment Mode, the human can continue playing for longer periods of time while the robot provides accompaniment. In the Solo Mode, the robot introduces melodic segments from its current experience if it detects a prolonged gap in the keyboard input. In addition, the system provides for continuous background training of the Markov chain, which allows the keyboard player to join in and change characteristics of the melodic response in real-time.

4. AN EXAMPLE INTERACTION SESSION

In one demonstration jam session developed for the system, Shimon begins the session in a non-learning state, receiving data from the drum player through the external beat detection module. It turns its head to look at the drum player and starts to nod its head based on the beat information received. When the keyboardist starts playing, Shimon switches to its listening and learning state and adjusts its gaze to look at the instrument it determines to be musically most interesting. It then switches to the playing state for a stochastically determined duration, where it improvises based on the Markov chain model. The state transitions are marked by visual cues from the robotic head, which constantly changes its gaze based on the saliency of the inputs it receives. Shifting between states continues as shown in Figure 2 until both the drum and the keyboard inputs stop and the system moves to a non-listening state with the head shifting its gaze away from the musical instruments. The setting has been tried in a number of interactive sessions involving both musically trained and untrained performers, in styles including western jazz [3] and eastern Indian raga [4] systems.

5. FUTURE WORK

We plan to expand the range of visual cues provided by the robotic head by building more expressions such as eye and eyebrow movements into the system, for responding to consonance/dissonance levels in the music and for portraying interest and attention levels [2] within the current expression sets for saliency-based gazing. We also plan to use more advanced perceptual models for both melodic and rhythmic improvisation to create more engaging musical collaboration.

6. CONCLUSION

A number of social and musical interaction schemes have been developed for the Shimon system in an effort to create rich and inspiring human-robot musical jam sessions. These schemes allow humans and robots to take turns in listening, learning, and improvising with each other, based on the manner in which human players interact in free-flowing interactive music sessions.

7. ACKNOWLEDGEMENTS

We would like to thank Andrea L Thomaz and Maya Cakmak from the Georgia Tech RIM Center, Brian P Blosser from the Georgia Tech Center for Music Technology and Roberto Aimi from Alium Labs for their continued support in the project.

8. REFERENCES

[1] Weinberg G., Driscoll S. “Towards Robotic Musicianship”, Computer Music Journal 30:4, MIT Press, pp. 28-45.

[2] Cynthia Breazeal. “Emotion and sociable humanoid robots”, International Journal of Human-Computer Studies, Volume 59, Issues 1-2, July 2003, Pages 119-155.

[3] YouTube video: Robotic Musicianship jam session - Jazz http://www.youtube.com/watch?v=MSxeN00C20g

[4] YouTube video: Robotic Musicianship jam session - Eastern http://www.youtube.com/watch?v=5DYOqSTmGDA