Shimon: An Interactive Improvisational Robotic Marimba Player

Guy Hoffman

Georgia Institute of Technology

Center for Music Technology

840 McMillan St.

Atlanta, GA 30332 USA

ghoffman@gmail.com

Gil Weinberg

Georgia Institute of Technology

Center for Music Technology

840 McMillan St.

Atlanta, GA 30332 USA

gilw@gatech.edu

Abstract

Shimon is an autonomous marimba-playing robot designed to create interactions with human players that lead to novel musical outcomes. The robot combines music perception, interaction, and improvisation with the capacity to produce melodic and harmonic acoustic responses through choreographic gestures. We developed an anticipatory action framework, and a gesture-based behavior system, allowing the robot to play improvised Jazz with humans in synchrony, fluently, and without delay. In addition, we built an expressive non-humanoid head for musical social communication. This paper describes our system, used in a performance and demonstration at the CHI 2010 Media Showcase.

Keywords

Music, Robotic Musicianship, Gestures, Improvisation, Human-robot Interaction, Anticipation

ACM Classification Keywords

H.5.5 Sound and Music Computing: Systems, I.2.9 Robotics

General Terms

Algorithms, Human Factors

Introduction

Most computer-supported interactive music systems are hampered by not providing players and audiences with physical cues that are essential for creating expressive musical interactions. For example, in human musicianship, motion size often corresponds to loudness, and gesture location to pitch. Other physical gestures are used to communicate turn taking and beat. These cues provide visual feedback and help players anticipate and coordinate their playing. They also create a more engaging experience for the audience by providing a visual connection to the sound. Many computer-based music systems are also limited by the electronic reproduction and amplification of sound through speakers, which cannot fully capture the richness of acoustic sound.

On the other hand, most research in musical robotics focuses mostly on sound production, and rarely addresses perceptual and interactive aspects, such as listening, improvisation, or interaction. Many such devices can be classified as either robotic musical instrument—mechanical constructs that can be played by live musicians or triggered by pre-recorded sequences [2]; anthropomorphic musical robots that attempt to imitate the action of human musicians [9]; or systems that use the human's performance as a user-interface to the robot's performance [8]. Only few attempts have been made to develop perceptual, interactive robots that play autonomously [1].

This work presents Shimon, a new research platform for Robotic Musicianship (RM) [10]. We define RM to extend both the tradition of computer-supported interactive music systems, and that of music-playing robotics, by being simultaneously embodied,

perceptual, and improvisational. Shimon extends our previous work in RM on *Haile*, a robotic drummer [10]. While Haile's instrumental range was percussive and not melodic, and it's motion range was limited to a small space relative to the robot's body, we have addressed these limitations with Shimon, a robot that plays a melodic instrument—a marimba—and does so by covering a larger range of movement [11].

Physical Robot

In designing Shimon, we wanted large movements for visibility, as well as fast movements for virtuosity. In addition our we aimed for wide range of sequential and simultaneous note combinations. The resulting design was a combination of fast, long-range, linear actuators, and two sets of rapid parallel solenoids, split over both registers of the instrument.

The physical robot is thus comprised of four arms, each actuated by a voice-coil linear actuator at its base, and running along a shared rail, in parallel to the marimba's long side. The robot's trajectory covers the marimba's full four octaves (Figure 1).
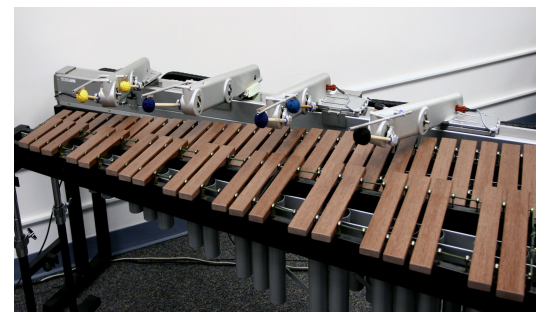

figure 1. Shimon's single-track linear actuators span four octaves. Rotational solenoids control the mallet strikes.

The arms are custom-made aluminum shells housing two rotational solenoids each. The solenoids control mallets, chosen with an appropriate softness to fit the area of the marimba that they are most likely to hit. Each arm contains one mallet for the bottom-row keys, and one for the top-row keys.

The robot also includes a socially expressive head (Figure 2). The head is non-humanoid and was designed to distill the essential movements used in joint musical performance. It is constructed of four harmonic-drive gear motors, two of which are at a 40 degree angle to produce a unique organic movement. In addition, a servomotor-controlled shutter allows the opening and closing of the head to convey emotional state and liveliness. The head also contains a single high-definition digital video camera.

figure 2. The expressive head serves social communication

Shimon was designed in collaboration with Roberto Aimi of Alium Labs.

Performance System

The work described in this paper concerns a human-robot Jazz performance system, allowing the robot to play in conjunction with a human pianist. The performance is made up of a number of different autonomous interactive machine improvisation modules. Some of these modules are calls-and-responses between the human and the robot; some are lead-and-accompaniment; but the bulk of the performance is in the form of joint, fluent, real-time improvisation, where human and robot play together.

The main novel contribution is in the fact that the robot matches the human's playing style, tempo, and harmony in real time, while extending on the human's playing and contributing its own musical phrases and ideas. This results in a back-and-forth inspiration between the human and robot. Also, since our system is using an anticipatory approach, the interaction is concurrent and does not rely on turn-taking. Both players play simultaneously, without noticeable delay.

Our choreographic gesture approach allows an additional visual performance layer unique to robotic music. The robot makes use of its physical presence to not only play music, but also perform as part of its music production, just as one would expect from a human performer. The movement of both the arms and the head help frame the performance both visually and acoustically.

Gestures and Anticipation

In this system we model interactive musical improvisation as *gestures*. Using gestures as the building blocks of musical expression is appropriate for robotic musicianship, as it puts the emphasis on

physical movement instead of on the sequencing of notes. This is in line with an embodied view of human-robot interaction [3].

Moreover, in order to allow for real-time synchronous non-scripted playing with a human counterpart, we also take an anticipatory approach, dividing gestures into *preparation* and *follow-through*. This principle is based on a long tradition of performance, such as ensemble acting [7], and has been explored in our recent work, both in the context of human-robot teamwork [4], and for human-robot joint theater performance [5]. By separating the—potentially lengthy—preparatory movement (in our case: the horizontal movement) from the almost instant follow-through (in our case: the mallet action), we can achieve a high level of synchronization and beat keeping without relying on a complete-musical-bar delay of the system.

Improvisation

As mentioned above, in our system, a performance is made out of *interaction modules*, each of which is an independently controlled phase in the performance. An interaction module runs in a continuous loop until the module's end condition is met.

Each module contains one or more musical *gestures*, which are selected and affected by information coming in from *musical percepts*. Percepts are mid-tier nodes analyzing input from the robot's sensory system. These percepts can include, for example, a certain note density, the appearance of a particular melodic phrase or rhythm, or a sudden beat change.

In the performance described in this paper, we use the following interaction modules:

Module I: Call-and-Response
The first interaction module is the phrase-call and chord-response module. In this module, the system responds to a musical phrase with a chord sequence. The challenge is to be able to respond in time and play on a synchronized beat to that of the human player. This module makes use of the anticipatory structure of gestures. During the sequence detection phase, the robot prepares the chord gesture, mirrored also by a head gesture. When the phrase is detected, the robot strikes the response almost instantly, resulting in a highly meshed musical interaction.

The robot adapts to the call phrase using a simultaneous *sequence spotter* and *beat estimator* percept. We use a Levenshtein distance metric [6] with an allowed distance $d=1$ to consider a phrase detected. Using the beat estimate, the robot responds with the appropriate phrase for the detected sequence. The result is an on-sync, beat-matched call-and-response pattern, a common interaction in a musical ensemble.

Module II: Opportunistic Overlay Improvisation
A second interaction module is called opportunistic overlay improvisation. This interaction is centered around the choreographic aspect of movement with the notes appearing as a side-effect of the performance. The central gesture in this module is a rhythmic movement gesture driven by a beat detection percept tracking the beat of the bass line in the human's performance. In parallel, a chord classification percept is running, classifying the currently played chord by the human player.

Without interrupting the choreographic gesture, this interaction module attempts to opportunistically play notes that belong to the currently detected chord, based on a preset rhythmic pattern. The result is a confluence of two rhythms and one chord structure, resulting in an improvisational gesture which is highly choreographic, can only be conceived by a machine, and is synchronized to the human's playing.

Module III: Rhythmic Phrase-Matching Improvisation
The third interaction module that we implemented is a rhythmic phrase-matching improvisation module. As in the previous section, this module supports play that is beat- and chord-synchronized to the human player. Additionally, it attempts to match the style and density of the human player, and generate improvisational phrases inspired by the human playing.

Beat tracking and chord classification is done in a similar fashion as the in the opportunistic overlay improvisation. In addition, this module uses a decaying-history probability distribution to generate improvisational phrases that are rhythm-similar to phrases played by the human. Using clustering techniques, the module learns rhythmic sequences played by the human, and generates similar sequences in the currently detected chord.

The result is an improvisation system that plays phrases influenced by the human's previous playing rhythm, clustering, and density. However, since the arm positions change according to the *current* harmonic lead of the human, and the robot's exploration of the chord space, the phrases will never be a precise copy of the human improvisation but only rhythmically and

harmonically inspired by it. This leads to a back-and-forth inspiration between human and robot.

Embodied Improvisation
Importantly—in both improvisation modules—the robot never maintains a note-based representation of the keys it is about to play, but instead generates its music solely based on its physical movement, as prescribed by our embodied gesture-based approach.

Musical-Social Communication
As part of the performance we are using the socially expressive robotic head in a number of ways: the head bobs to signal the robot's internal beat, allowing human musicians to cue their playing to the robot's beat and take note of errors in the beat detection percept. The head makes and breaks approximate eye contact—based on fixed band member positions—to signal and assist turn-taking. For example, when the robot takes the lead in an improvisation session, it will turn towards the instrument, and then it will turn back to the human musician to signal that it expects the musician to play next. Also, the head tracks the currently playing arms, by employing a clustering algorithm in conjunction with a temporal decay of active and striking arms. And finally, two animation mechanisms—an occasional blinking of the shutter and a slow breathing-like behavior—convey a continuous "liveliness" of the robot.

In future work, we plan to have the head anticipate playing movements, in order to allow human band members to prepare for upcoming notes and better synchronize their playing with the robot's. In addition, we are working to integrate face detection and tracking on the video image to achieve more accurate eye contact.

Conclusion

Shimon is an interactive improvisational robotic marimba player developed for research in Robotic Musicianship.

In this paper we describe an anticipatory gesture-based musical improvisation system for human-robot joint performances. The design of this system stems from our belief that musical performance is as much about visual choreography and communication, as it is about tonal music generation; and from the belief that anticipation is a key property of any temporally synchronized human-robot interaction. We also describe a socially expressive non-humanoid robotic head for musical communication and coordination

We have implemented our system on a full human-robot Jazz performance, and performed live with a human pianist in front of a public audience.

figure 3. Live human-robot Jazz performance using the system described in this paper (not including the social head).

References

[1] Baginsky, N. The three sirens: a self-learning robotic rock band. `http://www.the-three-sirens.info/` (2004).

[2] Dannenberg, R.B., Brown, B., Zeglin, G., and Lupish, R. Mcblare: a robotic bagpipe player. *Proc. NIME* (2005).

[3] Hoffman, G., and Breazeal, C. Robotic partners' bodies and minds: An embodied approach to fluid human-robot collaboration. *Proc CogRob, AAAI* (2006).

[4] Hoffman, G., and Breazeal, C. Anticipatory perceptual simulation for human-robot joint practice. *Proc AAAI* (2008).

[5] Hoffman, G., Kubat, R., and Breazeal, C. A hybrid control system for puppeteering a live robotic stage actor. *Proc RO-MAN* (2008).

[6] Levenshtein, V. I. Binary codes capable of correcting deletions,insertions and reversals. Soviet Physics Doklady, 10:707 (1966).

[7] Meisner, S. and Longwell, D. Sanford Meisner on Acting. Vintage, 1st edition, (1987).

[8] Petersen, K., Solis, J., and Takanishi, A. Toward enabling a natural interaction between human musicians and musical performance robots: Implementation of a real-time gestural interface. *Proc RO-MAN* (2008).

[9] Solis, J., Taniguchi, K., Ninomiya, T., Yamamoto, T., and Takanishi, A. The Waseda flutist robot no 4 refined IV: enhancing the sound clarity and the articulation between notes by improving the design of the lips and tonguing mechanisms. *Proc* IROS (2007).

[10] Weinberg, G., and Driscoll, S. Toward robotic musicianship. Computer Music Journal, 30:4 (2006).

[11] Weinberg, G., and Driscoll, S. The design of a perceptual and improvisational robotic marimba player. *Proc RO*-MAN (2007).