Robot-Human Interaction with an Anthropomorphic Percussionist

Gil Weinberg  
Georgia Institute of Technology  
840 McMillan St. Atlanta GA 30332  
gilw@gatech.edu

Scott Driscoll  
Georgia Institute of Technology  
840 McMillan St. Atlanta GA 30332  
scott.driscoll@gatech.edu

ABSTRACT
The paper presents our approach for human-machine interaction with an anthropomorphic mechanical percussionist that can listen to live players, analyze perceptual musical aspects in real-time, and use the product of this analysis to play along in a collaborative manner. Our robot, named Haile, is designed to combine the benefits of computational power, perceptual modeling, and algorithmic music with the richness, visual interactivity, and expression of acoustic playing. We believe that when interacting with live players, Haile can facilitate a musical experience that is not possible by any other means, inspiring users to collaborate with it in novel and expressive manners. Haile can, therefore, serve as a test-bed for novel forms of musical human-machine interaction, bringing perceptual aspects of computer music into the physical world both visually and acoustically.

Author Keywords
Human-robot interaction, music, collaborative systems.

ACM Classification Keywords
H.5.2 User Interfaces: Auditory Feedback, H.5.5 Sound and Music Computing: Systems

MOTIVATION
Computer supported interactive music systems are hampered by their inanimate nature, which does not provide players and audiences with physical and visual cues essential for creating expressive and intuitive musical interactions. Such systems are also limited by the electronic reproduction and amplification of sound through speakers, which cannot fully capture the richness of acoustic sound. Our approach for addressing these limitations is to utilize a mechanical apparatus that converts digital musical instructions into acoustic and physical generation of sound. We believe that such a musical robot can bring together the advantages of computation power and algorithmic music with the expression and richness of creating acoustic sound using physical and visual gestures. A musical robot can combine real-time analysis and response algorithms that are not possible by humans with rich sound and visual gestures that cannot be reproduced by speakers. Such novel human-machine interaction can lead to new music and novel musical experiences that cannot be conceived by traditional means.

RELATED WORK
In his book Machine Musicianship [13], Robert Rowe describes interactive systems that demonstrate musicianship as those that analyze, perform, and compose music with computers based on theoretical foundations in fields such as music theory, computer music, music cognition, and artificial intelligence. Several effective approaches for the design of such interactive musical systems have been explored over the years by researchers such as Dannenberg [3], Lewis [7], Cope [1], Pachet [9], and others. Such digital interactive systems, however, are limited by their inanimate and acoustically flat nature, which does not provide players and audiences with physical and visual cues, nor with the rich and varied sound of acoustic instruments. Current research directions in musical robotics, on the other hand, focus mostly on sound production and rarely address social aspects such as listening, analysis, group improvisation, or collaboration. Both “robotic instruments” (mechanically automated devices that can be played by live musicians or triggered by pre recorded sequences as in [6] [12] [15]) and “anthropomorphic robots” (hominoid robots that attempt to imitate the action of human musicians as in [16] [17] and [19]) function mostly as mechanical apparatuses that follow deterministic rules. Our work is also informed by studies in computational modeling of human perception of rhythm. Low-level cognitive rhythmic modeling addresses aspects such as hit onset, tempo, and beat detection, using audio sources [5] [11] [14] or MIDI [23]. Higher-level rhythmic perceptual modeling addresses more subjective percepts such as rhythmic stability, similarity, and tension [4] [10] [18]. We implement some of these models in Haile’s perceptual module as described in detail in [22].

GOALS AND CHALLENGES
Our main goal in this project is to construct a robotic device that listens to and analyzes live musical input in real-time and responds in an expressive acoustic manner which may inspire human players to interact with it in novel manners. This goal addresses challenges in perception, mechanics, and interaction design. In perception, our main challenge is...
challenge was to implement models for low- and high-level musical percepts, allowing Haile to develop a meaningful representation of the music it listens to. In mechanics our challenge was to develop a dexterous robotic apparatus that can translate perceptually based performance algorithms into a rich acoustic and visual performance. In interaction design, we aimed to develop performance algorithms that would enable the robot to collaborate with human players in a meaningful and intuitive manner, using transformative and generative methods both sequentially and synchronously.

THE SYSTEM

Form – In order to support familiar and expressive interactions with human players, Haile’s design is anthropomorphic, allowing it to operate two percussive arms both vertically and horizontally. At the current stage of development only one arm is operational, which is designed to play a Native American Pow-wow drum – a unique multi player instrument that supports the collaborative nature of the project. In order to match the natural aesthetics of the Native American Pow Wow ritual, we chose to construct Haile from wood. A design made by Clint Cope was used as a basis for Haile’s appearance. Wooden parts were made on a CnC wood cutting machine at the Advanced Wood Product Laboratory in the College of Architecture at Georgia Tech. Metal joints were designed to allow arm movement, leg adjustability for different drum heights, and complete disassembly.

Perception – Haile is designed to listen to audio input via a microphone installed on the drum. Its low-level perceptual analysis algorithms address aspects such as note onset detection, pitch detection (which often corresponds to hit location on the drumhead), and amplitude detection (which corresponds to hit strength). For detecting hit onsets we use the Max/MSP [2] object bonk~ while the object pitch~ is used to ascertain pitch and timbre information. In general, a “hit” is characterized as a sharp change in amplitude and spectral composition. Pitch information is obtained by looking at the amplitudes of spectral peaks every time a hit is recorded. Other relatively low-level perceptual modules that we implemented in Haile are beat detection (utilizing the beat~ Max/MSP object based on [14]), density detection, where we look at the number of note onsets per time unit to represent the density of the rhythmic structure, and accuracy, where we compare hits onset times to the detected beat, providing a “rubato coefficient” which represents how metrical the input rhythm is. We also implemented a number of higher-level rhythmic analysis modules for percepts such as rhythmic stability and similarity, based on Desain and Honing’s computational model [4]. This model is based on the relationship between pairs of adjacent note durations that are rated according to their perceptual expectancy (see details of our implementation in [22]).

Mechanics – Haile can adjust the sound of its hits in two manners: pitch and timbre variety are achieved by striking the drumhead in different locations and loudness variety is achieved by hitting harder or softer. The robotic arm’s main hardware components consist of a linear slider driven by a gear motor and a hitting mechanism that utilizes a solenoid. Pulse width modulated (PMW) and digital output signals from the USB controlled Teleo board [8] directly drive the motor and solenoid through H-bridge amplifiers. Slider position is also fed back through the Teleo’s analog inputs using a rotational potentiometer. As opposed to other robotic drumming systems that allow hits at only a few discrete locations, Haile’s arm can be moved to any location within a range of 10 inches. Position control is calculated within Max/MSP using a variant of simple Proportional Derivative (PD) control, with a 100Hz control signal. The hitting mechanism is loosely inspired by a piano action and was designed to enable quick design configuration changes during development. The hitter can strike at about 15 Hz with approximately 5 noticeable volume levels, and can be moved from lowest to highest pitch (strike location) at 3-4Hz.

INTERACTION DESIGN

The main challenge in designing the human interaction with Haile was to combine our perceptual modules with the robot’s mechanical abilities in a manner that would lead to an inspiring human-machine collaboration. Our approach for addressing this challenge is based on our model for interdependent group interaction in interconnected musical networks [21]. At the core of this model is a categorization
of collaborative musical interactions between artificial and human players based on sequential and synchronous operations with centralized and decentralized control schemes. For example, in sequential decentralized interactions players create their musical materials with no outside influence and only then interact with the algorithmic response in a sequential manner (see Figure 3). In a synchronous centralized network topology, on the other hand, players modify and manipulate their peers’ music in real-time, interacting through a computerized hub that performs analysis and generative functions (see Figure 4). More sophisticated schemes of interaction can be designed by combining centralized, decentralized, synchronous, and sequential interactions in different directions, and by embedding weighted gates of influence between participants (see Figure 5).

Figure 3. Sequential decentralized interaction – musical actions are taken in succession without synchronous input from other participants, and with no central coordination.

Figure 4. Synchronous centralized interaction – human and machine players take musical actions simultaneously, interacting through a computerized coordinating hub.

Figure 5. A combination of centralized, decentralized, synchronous and sequential musical actions in an asymmetric topology with weighted gates of influence. Based on [21].

Based on these ideas, we developed six different modes of interaction for Haile: *Imitation; Stochastic Transformation; Perceptual Transformation; Beat Detection; Simple Accompaniment* and *Perceptual Accompaniment*. Each interaction mode utilizes one or more of our perceptual modules. The interaction modes can be embedded in different combinations in interactive compositions and other educational musical activities with Haile. In the first mode, *Imitation*, Haile merely repeats what it hears based on its low-level onset, pitch, and amplitude perception modules. Users can play a rhythm and by pressing a foot pedal, instruct Haile to imitate their input in a sequential call-and-response manner. Haile adjusts pitch and timbre by playing closer to the rim (higher pitch) or closer to the center (lower pitch) as well as hits’ loudness by controlling the striker’s velocity. In the second mode, *Stochastic Transformation*, Haile is designed to improvise based on players’ input. Here, the robot stochastically divides, multiples, or skips certain beats in the input rhythm, creating variations of users’ rhythmic motifs while keeping their original feel. Players can choose between different transformation coefficients, setting the level of similarity between their motifs and Haile’s responses. In the *Perceptual Transformation* mode, Haile analyzes the stability level of users’ rhythms, and responds by choosing and playing other rhythms that have similar levels of stability to the original input. In this mode Haile automatically responds after a specified phrase length, so that users are not required to manage turn-taking procedures manually. *Imitation, Stochastic Transformation, and Perceptual Transformation* are all sequential interaction modes that form decentralized call-and-response routines between human players and the robot. *Beat Detection and Simple Accompaniment* modes, on the other hand, allow synchronous interaction, where humans play simultaneously with Haile. In *Beat Detection* mode, Haile tracks the tempo and beat of the input rhythm. Haile uses the Max/MSP beat~ object to detect the beat for a short period (5-10 seconds) and then locks the tempo before joining in. A much simpler, yet effective, synchronous interaction mode is *Simple Accompaniment*, where Haile plays pre-recorded MIDI files so that players can interact with it by entering their own rhythms or by modifying elements such as drumhead pressure to modulate and transform Haile’s timbres in real-time. This synchronous centralized mode allows composers to feature their structured compositions in a manner that is not susceptible to algorithmic transformation or significant user input. Perhaps the most advanced mode of interaction is the *Perceptual Accompaniment* mode, which combines synchronous, sequential, centralized and decentralized operations. Here, Haile plays simultaneously with human players while listening to and analyzing their input. It then creates local call-and-response interactions with different players, based on its perceptual analysis. In this mode we utilize the amplitude, density, and accuracy perception modules that are described above. While Haile plays short looped sequences (captured during *Imitation* mode) it also listens to and analyzes the amplitude, density, and accuracy of human playing. At periodic intervals it then modifies its looped sequence, using the amplitude, density, and accuracy coefficients analyzed from human playing. This leads to a multi player call-and-response interaction as the robot responds directly to specific players based on their individual perceptual coefficients. As a showcase for these
interaction modes we composed a short musical piece, titled Pow (video clips from Pow can be seen at http://www.cc.gatech.edu/~gilwein/pow.htm).

FUTURE EVALUATION AND DEVELOPMENT
We are currently developing a set of user studies to evaluate the interaction with Haile. As part of our assessment, subjects will be observed and interviewed after playing with Haile in each of the six interaction modes. The observations and interviews will be designed to compare and assess the effectiveness of aspects such as automatic vs. user-controlled turn taking, synchronous accompaniment vs. sequential call-and-response interactions, and centralized vs. decentralized control. We will also question subjects regarding the perceptual modules used in the interaction. In particular we are interested in evaluating the effectiveness of our algorithmic approach for high-level musical percepts such as density and tension, and in comparing the product of these algorithms to the manner in which human subjects perceive such percepts. Based on this evaluation we plan to refine our perceptual modules, and to develop computational models for other higher-level musical percepts such as density and tension. Mechanically, we plan to develop an intricate control mechanism that would emphasize visual cues through larger and more detailed motions. We also intend to further investigate the mechanics of hand drumming. Drums’ timbre is highly dependent on a variety of factors including hand shape, contact area, hit damping duration, pressure on the drum skin, among others. By controlling these parameters human players enrich their timbre richness and intonation. In future work, we plan to develop a second robotic arm that would apply continuous pressure on the drum skin, capturing more drumming nuances, and enriching sonic variety. Through evaluation and improvements of our mechanical, perceptual and interaction modules, we hope to continue to use Haile as a test-bed for novel forms of musical-human-machine interaction, bringing perceptual aspects of computer music into the physical world both visually and acoustically.

REFERENCES