Towards the Development of Robot Musical Audition

David K. Grunberg, Alyssa M. Batula, and Youngmoo E. Kim
Music and Entertainment Technology Laboratory (MET-lab)
Electrical and Computer Engineering, Drexel University
{dgrunberg, batulaa, ykim}@drexel.edu

***Abstract*—We are seeking to enable humanoid robots to participate in live human-robot musical ensembles. In ensembles, the musicians must be able to listen to the audio that they and the others are producing in order to identify high-level features of the music (e.g., location in the score, tempo, cues). This information is crucial for the musicians so that they can know what to play next. One requirement for robots capable of participating in musical ensembles is, therefore, the ability to listen to the music and identify some of these features. While other sensory input is also important (vision, for example, could enable a robot to determine what a conductor is indicating), many of these fields already have a significant amount of active research. Robot vision, for example, is a popular topic in the research community. By contrast, a comparatively small amount of research is performed on robot musical audition. We therefore feel it is important for us to pursue research in this field.**

I. Introduction

Our long-term goal is to enable a humanoid to perform alongside humans in a musical ensemble [1]. In order for the robot to be truly useful for this purpose, it would have to be able to extract high-level features from the musical audio and incorporate them into its performance. For example, the robot would have to identify beat locations so that it could synchronize its motions with the audio, and would need to identify pitches so that it could determine if it was playing the correct musical notes [2], [3]. In order to accurately determine these features, the robots require a robust set of music-information retrieval (MIR) algorithms that can function causally and in real-time. These are the algorithms that analyze the audio and identify the features the robot must know.

It is preferable that the musical robots that we use be humanoid in form. Most instruments and dance styles have already been developed for humans, so the robots would be ideally shaped to perform in a variety of musical tasks. While some musical robots have been developed that are specially designed to play one particular instrument, these robots are often only vaguely humanoid, and thus ill-suited to perform any more than one instrument or dance style. In order to keep our system flexible, we thus wish to enable humanoid robots to perform musical tasks.

When determining which humanoid to use, we must balance several considerations. Small humanoids, such as the RoboNova, are rugged and relatively inexpensive [4]. However, these robots are often only marginally humanoid and cannot move with nearly enough control or finesse for our purposes. More advanced robots, such as the Hubo, are generally capable of smoother and more human-like movement [2]. These robots, though, are much more expensive and fragile, so an error which destabilizes (or knocks over) the robot could potentially be very costly. To satisfy these constraints, we are using a multitude of robots for our purposes. We prototype our systems on small robots such as the RoboNova, and once we know our algorithms work, move them to the Hubo.

II. Related Work

Existing MIR algorithms can extract a variety of features reliably from audio. For example, current beat trackers can identify beat locations with a high level of accuracy, particularly on pieces of popular music, which generally have strong beats. One common algorithm is based on the work of Scheirer [5]. In this method, audio is split into multiple subbands, and periodicity and beat locations are estimated by filtering the subband envelopes with a bank of comb filters. This method is quite accurate, particularly for music with heavy drum sections, and is the basis for several modern beat trackers [6]. Another popular beat tracking algorithm is based on the work of Goto [7]. This algorithm also splits audio up into several subbands, but then performs additional processing to determine drum patterns and chord changes.

Numerous musical robots have been developed for various applications. Keepon [8] is a small, cartoon-style robot that bounces its head in response to music. Haile [9] is a drumming robot that can respond to a human drummer by producing a sequence of beats that compliments the human's. However, neither of these robots are humanoid, so both are less capable of playing on a variety of human instruments. The Honda humanoid, Asimo, has been enabled to step, sing, and scat in response to a piece of music [10]. Similarly, the HRP-2 has been enabled to dance in an extremely humanlike manner, though it does not synchronize its motions with the beats [11].

III. Robot platforms

We are utilizing two types of robots in our research. To prototype and test our music and our gesturing algorithms, we use smaller, more rugged robots that are resistant to damage. This allows us to verify that the systems work in a low-risk environment. Once we are confident that our algorithms are robust and will not destabalize the robot, we port everything to Hubo, a larger and more capable humanoid.

Fig. 1. One of the Drexel RoboNovas.

A. RoboNova

The RoboNova is a small humanoid produced by HiTek Robotics (Figure 1). RoboNova is 35 cm tall and has 16 degrees of freedom (DOF): 5 per leg and 3 per arm. It is a rugged robot and is easy to repair if its motors become damaged. Furthermore, the RoboNova is capable of using Bluetooth communication, so some of the computation can be offloaded to an external computer. This helps ensure that the robot will be able to respond to musical audio despite its relatively limited processor. While the robot's body is limited in many ways, such as having only one finger and relatively limited computational ability, it is still capable of serving as a prototyping platform.

B. Hubo KHR-4

The Hubo+ series adult-size (130 cm) humanoid robot is designed and built by the Hubo-Lab at the Korean Advanced Institute of Science and Technology (KAIST) in Daejeon, Korea (Figure 2). Drexel has obtained six Hubo+ robots from KAIST as part of a multinational collaboration between several American and South Korean universities. These robots are also intended for distribution to the other American universities in the collaboration.

The Hubo+ masses 37 kg and is fully actuated. It has 41 DOF and runs on a single 48V Lithium Polymer battery. Each of its legs contains six DOF: three in the hip (roll, pitch, and yaw), one in the knee (pitch), and two in the ankle (roll and pitch). Its arms contain 6 DOF as well: three in the shoulder, one in the elbow, and two in the wrist. Hubo also has independently actuated fingers.

IV. Music Information Retrieval

A. Pitch tracking

For most instruments, there is a direct relationship between a note's pitch and the fundamental frequency in the resulting acoustic signal. Therefore, pitch detection algorithms typically identify notes by finding strong frequencies present in a signal.

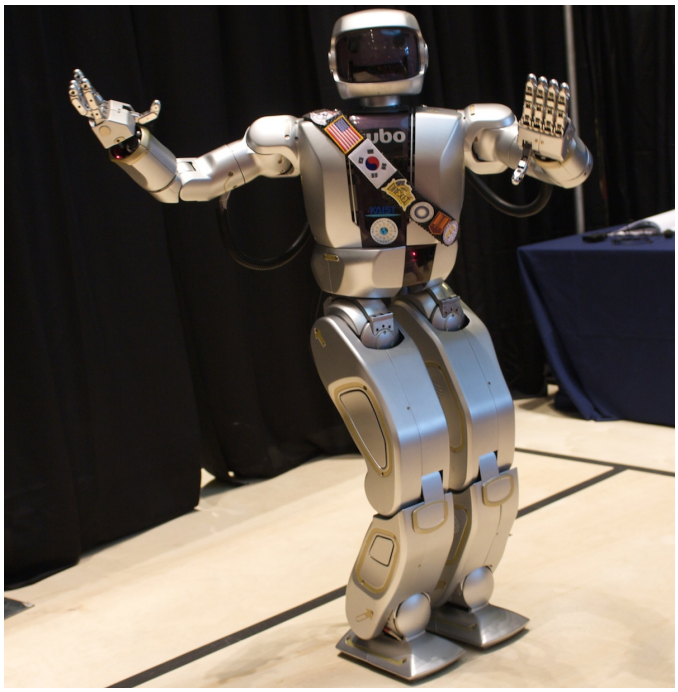

Fig. 2. One of the Drexel Hubos.

Our initial pitch detection method identifies single pitches using the autocorrelation function, as autocorrelation quantifies the similarity of a signal to delayed copies of itself [12]. When only a single note has been played, the autocorrelation of the audio signal can be used to find the signal's period, which corresponds to the note's fundamental frequency. The highest peak (aside from the peak at zero delay) in the autocorrelation will occur at a delay equal to the period of the signal. If there is a single note being played, the period of the audio signal should be the same as the period of the note. However, note detection while playing a song requires the ability to detect multiple notes at once. Because this is difficult to do with autocorrelation, a second algorithm was implemented using the FFT to obtain the frequency spectrum of the signal.

The system looks for a peak in the Fourier spectrum at a specified note's fundamental frequency (with a 4% error tolerance). If it finds such a peak, it assumes the correct note was played. This system requires knowledge of the score, in order to determine which notes should have been played. Figure 3 shows the Fourier magnitude spectrum of two notes played on the keyboard.

B. Beat tracking

We have developed a beat tracker for use on audio taken either from CD or a live acoustic channel [2]. The ability to process live audio is important, because many real-world environments will require the robot to listen with its own microphones over an acoustic channel. Our beat tracker algorithm can function accurately in both environments, and fulfills the

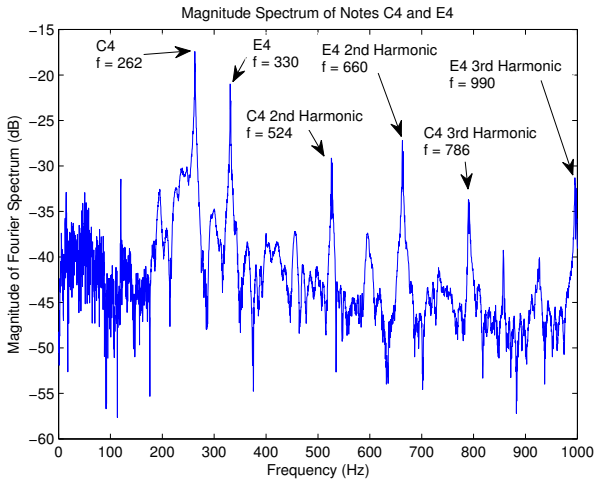

Fig. 3. Spectrum of two keyboard notes.

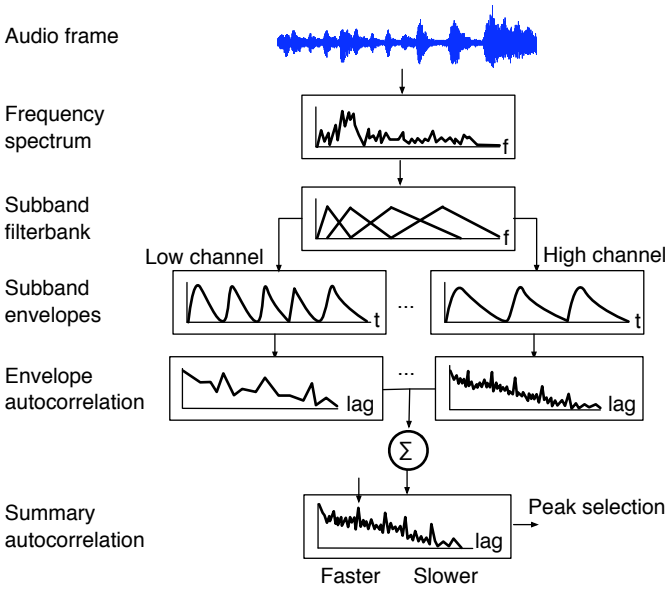

Fig. 4. Flowchart for the beat tracking algorithm.

constraints of real-time performance and causality.

A frame of audio is extracted from the full acoustic signal. (Figure 4). Filters then split the audio into several subbands. The system proceeds to sum the energy in each subband and place them in a buffer that stores values over multiple frames. The buffers are autocorrelated, the autocorrelations are summed, and the peak of the summed autocorrelation is identified. Autocorrelations have large values at lags that are proportional to the periodicity of the original input, and audio can be treated as a somewhat periodic signal, with the period being its tempo. The lag at the location of the maximum value can thus be used to calculate the tempo, or *audio period*. The frames are then analyzed to find sequences of high-energy frames spaced according to the audio period. These frames are likely to contain beats.

Because the audio that the robots listen to may be contaminated by acoustic noise, we have studied some techniques to maintain beat tracking accuracy even in noisy environments [2]. Our beat tracking system achieves an F-Score of .98 with clean audio, and .85 with noisy audio, but by adding noise reduction techniques such as spectral subtraction, accuracy is increased back up to .92.

V. Major Robot Performances

We have used our musical algorithms to enable the robots to participate in robot performances. These demonstrations served both as a large-scale test of our system under unfavorable conditions and as an exhibition to show the public what we have done.

The RoboNova has been enabled to move in response to music by selecting combinations from a gesture corpus of 30 motions and performing them according to the times specified by the beat tracker [13]. It can also play multiple pieces on a small keyboard [3]. These performances are routinely used at demonstrations and open houses to show off the robots' capabilities. The success of these performances also verifies the accuracy and reliability of our MIR algorithms.

In order to demonstrate the ability of our robots to enhance a musical performance, we equipped a Hubo robot with a tambourine and a maraca [1]. The robot then moved its arms according to the dictates of the beat tracker, adding new notes to the piece. It was able to stay in time with the music, so its performance did indeed appear synchronized. We also enabled Hubo to press several piano keys. While this did not directly test the pitch tracking algorithm, this helped us confirm the ability of the Hubo to capably play piano notes.

To test the robot's dancing ability, we then removed the instruments and had Hubo perform a series of arm gestures in response to the audio. This was our first full-fledged Hubo 'dance'. By watching the robot move its arms in synchrony with the music, we confirmed that Hubo could dance to audio.

More recently, Drexel displayed its six Hubo robots during the beginning of its Engineer's Week event on February 20th, 2012 (Figure 5). In addition to the pre-canned demonstrations developed by the KAIST lab, the robot also demonstrated its beat tracking ability. The beat tracking was run on an offboard computer, which sent Universal Datagram Packets (UDP) to each robot to tell it when to move. As the music changed, the robots changed their motions as well to match the new tempos and beats. All six robots were able to successfully respond to the audio in a realistic performance environment.

VI. Conclusion

We have designed several music-information retrieval algorithms to extract useful features from audio to enable a variety of robot platforms to move in response to audio. Our prototyping robots as well as our Hubos are now able to react to music based on high-level features extracted from the acoustic signal. We have thus progressed towards our ultimate goal of enabling humanoids to participate in full-fledged musical ensembles alongside humans.

Fig. 5. Several Hubos at Drexel's Engineering Week ceremony.

We are continuing to study additional algorithms to extract different features from the audio. For example, human dancers and musicians incorporate knowledge of the mood of a piece of music into their performances. The dance motions or musical phrasing performed by humans for a happy and excited piece of audio are generally different from those performed for a morose, depressing piece of music. We are thus examining algorithms that can reliably estimate the mood in audio, so the robots can make use of this information. Similarly, the genre of a piece of music can also influence how a human would react to it, so we are interested in studying genre-detection algorithms for the robots.

We are also interested in using multiple Hubos to make more interesting performances. For instance, we intend to study the effects of having several robots all doing the same thing to see if that has an effect on the perceived performance. There are dance troupes that feature several humans performing the same motions. By enabling robots to move in unison to music as humans do, they could then be used to research the perceptual effects of such performance techniques.

ACKNOWLEDGMENT

This research is supported by NSF Awards DGE-0947936, OISE-0730206, CNS-0960061, and the Graduate Research Fellowship.

REFERENCES

[1] Youngmoo E. Kim, Alyssa M. Batula, David K. Grunberg, Daniel M. Lofaro, JunHo Oh, and Paul Y. Oh, "Developing humanoids for musical interaction," in *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 2010.

[2] David K. Grunberg, Daniel M. Lofaro, Paul Y. Oh, and Youngmoo E. Kim, "Robot audition and beat identification in noisy environments," in *Proceedings of the International Conference on Intelligent Robots and Systems*, 2011.

[3] Alyssa M. Batula and Youngmoo E. Kim, "Development of a mini-humanoid pianist," in *Humanoid Robots, 2010. Humanoids 2010. 10th IEEE-RAS International Conference on*, December 2010.

[4] Robert Ellenberg, David K. Grunberg, Paul Y. Oh, and Youngmoo E. Kim, "Using miniature humanoids as surrogate research platforms," in *Proceedings of the IEEE-RAS Conference on Humanoid Robotics (Humanoids 2009)*, Paris, France, 7-10 December 2009, pp. 175–180.

[5] Eric D. Scheirer, "Tempo and beat analysis of acoustic musical signals," *The Journal of the Acoustical Society of America*, vol. 103, no. 1, pp. 588–601, 1998.

[6] Matthew E. P. Davies and Mark D. Plumbley, "Context-dependent beat tracking of musical audio," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1009 –1020, march 2007.

[7] Masataka Goto, "An audio-based real-time beat tracking system for music with or without drum-sounds," *Journal of New Music Research*, vol. 30, no. 2, pp. 159–171, June 2001.

[8] Marek Michalowski, Selma Sabanovic, and Hideki Kozima, "A dancing robot for rhythmic social interaction," in *Proceedings of the 2nd Annual Conference on Human-Robot Interaction (HRI)*, 2007, pp. 89–96.

[9] Gil Weinberg and Scott Driscoll, "Robot-human interaction with an anthropomorphic percussionist," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, New York, NY, USA, 2006, CHI '06, pp. 1229–1232, ACM.

[10] Kazumasa Murata et al., "A robot uses its own microphone to synchronize its steps to musical beats while scatting and singing," in *Proceedings of the 9th International Conference on Intelligent Robots and Systems*, September 2008, pp. 2459–2464.

[11] Takaaki Shiratori and Katsushi Ikeuchi, "Synthesis of Dance Performance Based on Analyses of Human Motion and Music," *IPSJ Online Transactions*, vol. 1, pp. 80–93, 2008.

[12] Bhagawandas P. Lathi, *Signal Processing & Linear Systems*, Oxford University Press, 2000.

[13] David K. Grunberg, Robert Ellenberg, In Hyeuk Kim, Jun Ho Oh, Paul Y. Oh, and Youngmoo E. Kim, "Development of an autonomous dancing robot," *International Journal of Hybrid Information Technology*, vol. 3, no. 2, pp. 33–44, April 2010.