Developing Humanoids For Musical Interaction

Youngmoo E. Kim, Alyssa M. Batula, David Grunberg, Daniel M. Lofaro, JunHo Oh[†], and Paul Y. Oh*

Electrical & Computer Engineering
Drexel University
Philadelphia, PA USA
{ykim,batulaa,dgrunberg,dlofaro,pyo22}@drexel.edu

[†]Mechanical Engineering
Korea Advanced Institute of Sci. & Tech.
Daejeon, Korea
jhoh@kaist.ac.kr

Abstract— **For many people, playing and enjoying music are integral activities in their daily lives, and the development of musically-aware robot systems provides a unique opportunity for richer forms of human-robot interaction. A robot participating in an ensemble musical performance requires a wide variety of skills in order to perform alongside humans. For instance, human musicians make use of substantial auditory and visual information throughout a performance; they watch the conductor and other musicians for rhythm and other musical cues, and they listen to their own instrument as well as the overall ensemble in order to adjust their own performance. Our current work focuses on providing such capabilities (e.g., audio and visual beat detection, note onset and pitch detection, and basic control for music keyboard performance), with the long-term goal of enabling a large humanoid to be an interactive participant in a live music ensemble. We use miniature humanoids to first prototype and refine many of these systems before deploying them on the life-sized KAIST Hubo humanoid.**

I. INTRODUCTION

Playing and enjoying music are integral activities in the daily lives of many people, and robots have been incorporated into several high-profile musical performances. Honda's ASIMO, for example, conducted the Detroit Symphony Orchestra in 2008, and while reviewers described the event as a "technological marvel"[1] and "more realistic than they expected"[2], the same writers noted that ASIMO "can't respond to the musicians"[2], and that "it was conducting in only the most limited definition"[1]. In this particular performance, ASIMO served as "a very expensive metronome"[3], following a predetermined set of choreographed motions and gestures. This article goes further to state that "the orchestra's musicians would have had a similar experience if they had followed a pre-recorded human conductor. There was no opportunity for improvisation"[3].

But developing systems with a true *understanding* of music and performing arts is becoming increasingly important as humans and robots interact more frequently. The more robots are able to replicate uniquely human activities and behavior, such as music performance, the easier it becomes for people to interact with them [1]. Simply rendering notes exactly as they are printed on a sheet of music, however, is not enough to make a performance *musical.* Without subtle expressive variations in timing, dynamics, and articulation, a performance is perceived as flat, boring, and emotionless. Similarly, dancing in response to music involves much more than simply gesturing synchronously with the beats.

Furthermore, cooperation, interaction, and improvisation are important and necessary qualities for performing ensembles, and human musicians make use of substantial auditory and visual information throughout a performance. Instrumentalists watch the conductor and listen to each other while playing, and a conductor is able to tell whether or not an ensemble is on pitch and in tempo. Dancers adjust their gestures, speed, and style based on the tempo and genre of music. Additionally, human performers generally have some prior knowledge of a piece's musical score or a dance's choreography and have practiced it in rehearsals.

Robots capable of playing musical instruments can be designed for the requirements of a specific instrument (several well-known robots have been designed in this manner), but our work focuses on humanoid robots. In general, musical instruments have been designed to be played by humans, making humanoids a natural choice for developing robot performers. By definition, humanoids have limbs and joints similar to human anatomy and could potentially play an instrument in a manner similar to that of a human performer. Furthermore, these similarities make humanoids ideal for investigating and quantifying the importance of human gestural control in musical expression. A human performer's precise motions provide important insight into musical expression, affecting not only the visual performance, but also impacting the sound of an instrument by altering the manner in which individual notes are articulated. Because a robot's gestures can be controlled precisely and repeated accurately many times, a humanoid instrumentalist could serve as a platform to investigate how specific parameter changes (large or small) in the execution of gestures affect the perceived musicality and expressiveness of the resulting performance.

This paper highlights our recent work towards developing

---

*P. Y. Oh is with the Department of Mechanical Engineering & Mechanics

This work is supported by NSF Award 0730206

[1] M. Stryker, "DSO led by robot maestro," *Detroit Free Press*, 14 May 2008.

[2] D. Durbin, "Honda robot conducts Detroit Symphony," *Associated Press*, 14 May 2008.

[3] J. Strickland, "How does a robot conduct a symphony?" *HowStuffWorks.com*, 28 July 2008.

a humanoid robot capable of performing as a member of a musical ensemble alongside human performers. Such a robot must be able to understand and respond to changes throughout the performance (tempo, dynamics, and other variations). Like a human performer, a robot should be able to determine whether or not the correct (intended) notes and articulations have been executed and incorporate this feedback to make adjustments to its performance. In pursuing these goals, our work has focused on following tempo and beat from musical audio as well as visual tracking of a human conductor. We have also implemented detection of other auditory cues, such as whether or not the robot (or the ensemble) is playing in tune. These sources of information are incorporated into the control of the humanoid's movements and performance.

The target platform for these efforts is Hubo, an adult-sized humanoid developed by the Korea Advanced Institute for Science and Technology (KAIST), and our eventual goal is for Hubo to play an instrument or dance alongside human musicians. Many of our systems are first developed for a miniature humanoid, the Hitec Robonova-1. This robot provides a relatively robust and inexpensive prototyping platform before we apply our algorithms to the more capable, but more costly, Hubo. Once methods have been thoroughly tested and demonstrated to run safely without damaging the robot, our system allows for straightforward adaptation and implementation on the Hubo platform. Demonstrations with life-sized humanoids have generated significant prior interest, and Hubo has been used effectively in outreach events with the public. Eventually, collaboration between robot and human performers may help make robotics and engineering more compelling and accessible to the public.

II. RELATED WORK

The development of robotic systems for musical interaction and performance has been an active topic of research for many years, beginning with the first player pianos [2]. Here, we classify these robots into several groups, depending on their capabilities and applications.

A. Rhythmic interactivity

One group of musical robots focuses on sensitivity to musical beats. Some early robots were built to study motion kinematics and played drumbeats. Two such robots are Harvard's drumming arm [3] and MIT's robot arm [4]. These robots could play drum beats but they did not interact with music, since all of their motions were preprogrammed.

Later beat-sensitive robots were designed for human interaction. One such robot is the University of Hertfordshire's Kaspar, a small, doll-like humanoid [5]. Kaspar can either listen to and play back the drumbeats that it hears or can 'lead' a human by playing its own beats [6]. It has been used to study emergent features in turn-based drumming by playing with adult drummers. Another robot, Keepon, is small cartoon-like robot designed for studying interaction with autistic children. Keepon can follow and move in time with musical beats [7]. The Ugobe Pleo[4] is a small toy robot in the shape of a dinosaur that can detect and react to beats in music. One of the more interactive robots, Georgia Tech's Haile, can listen to human drum sequences and synthesize its own accompaniments. Haile has performed live onstage in an ensemble with human performers [8].

Musical ensembles often use visual cues, such as a conductor, to follow the tempo. The robot Nico has demonstrated the ability to follow a conductor visually and determine beat positions based on the conductor's tempo [9].

B. Instrument performance

A variety of custom robotic instruments have been developed in which the sound producing mechanism is part of the robot, such as the League of Electronic Musical Urban Robots (LEMUR) [10]. LEMUR robots include GuitarBot, comprised of electrified slide guitars, and ForestBot, which consists of shakers attached to thin rods. Other robots of this type include the robot rock band called the Three Sirens[5] and the Cybraphon. Cybraphon tracks how often people search for it on the Internet and produces music based on a popularity metric[6]. Another similar group is the Absolut Quartet, a collection of large, immobile robots that play separate pitched instruments. One robot group member fires ping-pong balls onto a marimba to produce music based on a motif [11]. The Quartet is able to play instruments with very high accuracy, but it is too large to be easily set up for use in live performances alongside human performers.

Another branch of research has developed robots to play standard (human-playable) musical instruments, such as the Toyota Musical Robots [12]. These humanoids include a flutist, a violinist, and a trumpeter. The Waseda musical robots are also capable of using human instruments, such as flutes [13] and saxophones [14], but are only vaguely humanoid in form. Shimon, a non-humanoid robot developed by Georgia Tech, can play the marimba [15]. All of these robots, able to use the same instruments as humans, can interact with and even perform with human musicians without needing specialized instruments.

Of particular interest to us are piano playing robots, which range from a single "hand" mounted on a rail [16] to full humanoids [17]. One example of a "humanoid" pianist is Waseda's Wabot-2, which can use both its hands and feet to play a variety of classical and modern pieces on pianos and organs. The Wabot-2 can also sight read, enabling it to play novel music without having its performance choreographed in advance. However, it cannot adjust its technique or position in the middle of a piece if it makes a mistake.

C. Dance performance

Other researchers have developed robots with the ability to dance. Some of these robots were designed specifically to replicate human dances or motions. Ms DanceR, for example, can participate in ballroom dances when it is led by a human [18]. Sony's QRIO can follow and then reproduce certain human dance motions [19], and the HRP-2 has been enabled

---

[4] http://pleohq.com/pleo-101/

[5] http://www.the-three-sirens.info

[6] http://www.cybraphon.com

to reproduce traditional folk dances [20]. These prior systems do not detect audio beat information, but other dancing robots can listen to music and then formulate their gestures based on the audio. Honda's ASIMO, for instance, can step in place with music based upon the beats that it "hears" [21]. Our prior research with the Hitec Robonova mini-humanoid is another example; we have enabled the robot to perform motions in synchrony to beats in music [22].

III. SCALABLE HUMANOID RESEARCH PLATFORM

Our efforts thus far have used two different robot platforms, the Robonova mini-humanoid and the Hubo KHR-4 adult-sized humanoid. Previous efforts have demonstrated that rapid prototyping with miniature humanoids has substantial benefits [23]. Development on large robots is more time-consuming and poses much higher risks in terms of cost and safety. The relatively low cost of mini-humanoids also broadens the potential audience for musically-aware robots.

A. Hitec Robotics Robonova-1

Much of our development was prototyped on a modified version of the Robonova-1, a small (36 cm tall) humanoid by Hitec Robotics with 16 degrees of freedom (DOF). The robot's 16 motors are HSR-8498HB servo motors controlled by a MR-C3000 controller with an ATmega 128 MPU.

Due to the relatively low processing capability of the robot, we reprogrammed the onboard microcontroller to allow individual servo motor positions to be specified by an external computer. Offloading the generation of movement commands to a much faster CPU enables more complex and fluid movements (commands are sent via serial connection every 20 milliseconds to set the position of the robot's 16 joints). This configuration provides much more direct and refined control over the robot's motions [23].

We have also modified the robot's hands to improve its ability to play the piano. The new hands, formed using a 3D rapid prototyping printer, are 1.27 cm longer than the original hands, extending the robot's reach. They also curve to a small (0.64 cm) square tip that allows it to hit keys more accurately without also striking neighboring keys.

B. Hubo KHR-4

The Hubo KHR-4 series adult-size (130 cm) humanoid robot is designed and built by the Hubo-Lab at KAIST in Daejeon, Korea. Currently three KHR-4 model Hubos have been created, and Drexel's Hubo ("Jaemi") is the only one permanently housed outside of Hubo-Lab, placed at the Drexel Autonomous Systems Lab (DASL) in Philadelphia, PA, USA. The name "Jaemi Hubo" comes from the Korean phrase "Jaemi Kyopo", meaning Korean-American, and "Jaemi" also means "fun." Literally translated, the name means "Korean-American Hubo", illustrating the international nature of the project.

Jaemi Hubo is the primary focus of the Drexel-KAIST collaboration, which is supported through the US National Science Foundation's Partnership for International Research and Education program. This grant includes participation and cooperation between five universities in the United States and three in Korea. The other participating U.S. universities are University of Pennsylvania, Virginia Tech, Colby College, and Bryn Mawr Women's College, and the additional Korean institutions are Seoul National University and Korea University. The ultimate goal of the collaboration is to create a system for studying humanoid robots safely and robustly. This system includes a 3D virtual version of Hubo, a miniature version of Hubo, and an on-line system that gives other researchers full access to Jaemi Hubo to enable simple and useful remote research on an adult-size humanoid robot.

C. Hubo Technical Specifications

The Hubo KHR-4 is a 37 kg fully-actuated humanoid robot. The majority of the robot's frame is aluminum, cut on a three axis CNC milling machine. It has 41 DOF and runs on a single 48V 5Ah Lithium Polymer (LiPo) 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). Table I shows a comparison of Robonova and Hubo's degrees of freedom currently used for musical interaction.

TABLE I

ROBONOVA AND HUBO DEGREES OF FREEDOM

Limb	RoboNova	Hubo
Each arm	3	6
Each leg	5	6
Head and waist	0	2
Total	16	26

Each of the motors, as well as the joints in the shoulders, elbows and waist are actuated by one or two brushless DC servo motors using quadrature optical encoders with indexing. These motors are connected to Harmonic-Drive gear boxes with gear ratios ranging from 120:1 to 160:1. Each of the servo motors are actuated by custom 400W motor drivers that support up to two motors per controller. The controllers are used as position servos but are able to control velocity and feedback current/torque. It is important to note that the torque cannot be accurately calculated due to the Harmonic-Drive gear's non-linear nature.

Each of the motor drivers are controlled over a 1Mbps Closed Area Network (CAN). CAN was chosen because it is a two line half-duplex system which allows multiple devices to communicate on the same communications line, reducing the number of wires needed to run through the robot's frame. The motor drivers are controlled via two CAN busses, where all of the top joints are controlled by one and all of the bottom joints are controlled by the other. This allows the control loop frequency to be increased to the target of 100Hz.

Hubo balances using the popular zero moment point balancing method. A 1.0Ghz x86 Pentium computer running Windows XP is used to control and balance the robot. The Real-Time Extensions for Windows are used to ensure accuracy of Hubo's control loop frequencies. There are two control loops: one running at 100Hz which sends out the motor commands, and one running at 500Hz which takes readings from all of the sensors. It balances using a 6-axis inertial measurement unit located at the center of mass and two 3-axis force torque (FT) sensors located in each ankle. Jaemi Hubo is also equipped with two 2-axis accelerometers located in each foot, force torque sensors in each wrist, and a vision system consisting of a single 60Hz black and white camera and binaural stereo microphones. Along with balancing, the latter sensors are used for human-humanoid interaction and humanoid-world interaction.

D. *Outreach*

One important aspect of the project is to use humanoids to study social interaction and facilitate education. Unlike most adult-size humanoids, such as the proprietary ASIMO, KAIST developed Jaemi Hubo as a research platform and is more willing to share the inner workings of the robot. The mechanical and electrical schematics, as well as the source code, are available for research purposes.

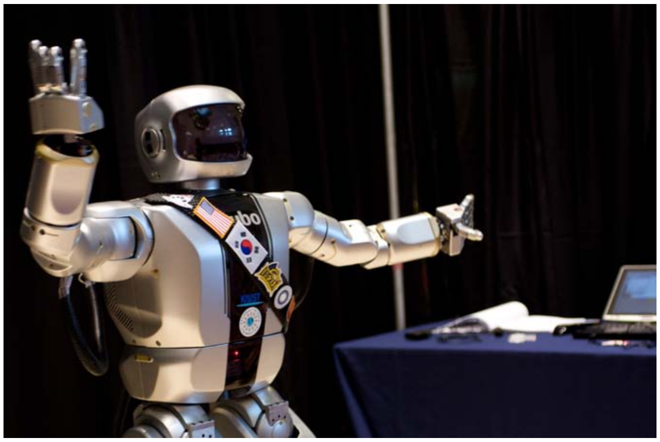

Fig. 1. Jaemi Hubo, displayed at the Please Touch Museum in Philadelphia.

Presenting Hubo to the public is an important part of the project. In the past year Jaemi Hubo performed at the Philadelphia Please Touch Museum for children between the ages of 4 and 7. At any of Jaemi's events access is not limited to watching; human-robot interaction is also encouraged. Fig. 1 is a photograph taken at the event at the Please Touch Museum on May 28th, 2009. This photo depicts how exploring science, technology, and engineering through physical interaction with Jaemi is encouraged.

IV. AUDIO AND VISUAL SENSING

In order to perform interactively with humans, robots need to utilize visual and auditory feedback. This feedback provides the robot with important information about the current tempo as well as any mistakes it may have made.

A. *Audio*

Audio feedback is useful for maintaining correct pitch and timing. In order to synchronize motions to the music's tempo, such as when the robot is dancing, the robot needs to be able to detect beats in the music. Analyzing the frequency spectrum of audio signals also allows the robot to detect misplayed notes during a performance.

1) Auditory Beat Tracking: Detecting and tracking the beat (or tempo) of music is an important aspect of musical performances. One way people follows music is by listening and then identifying the beats and tempo of the audio. Our system uses auditory beat tracking to mimic this process.

Musical beats generally occur at a regular period throughout sections of a song. Our system uses autocorrelation to look for periodicity in the audio signal. The strongest periodicity usually corresponds to the tempo of the music [24]. The tempo can then be used to identify beat locations.

First, the signal frequencies are divided into sub-bands using triangular bandpass filters, with higher weighting on the low frequencies due to the increased likelihood that they contain beat information. The energy of each sub-band is calculated, and then an autocorrelation of the energies is taken. The autocorrelation values are highest when the lag matches the period of the signal or its multiples. The most recent autocorrelations are then summed and the delay corresponding to the highest value (called the *audio period*) is found. The tempo is calculated based on that delay. Fig. 2 shows the flow of the process.

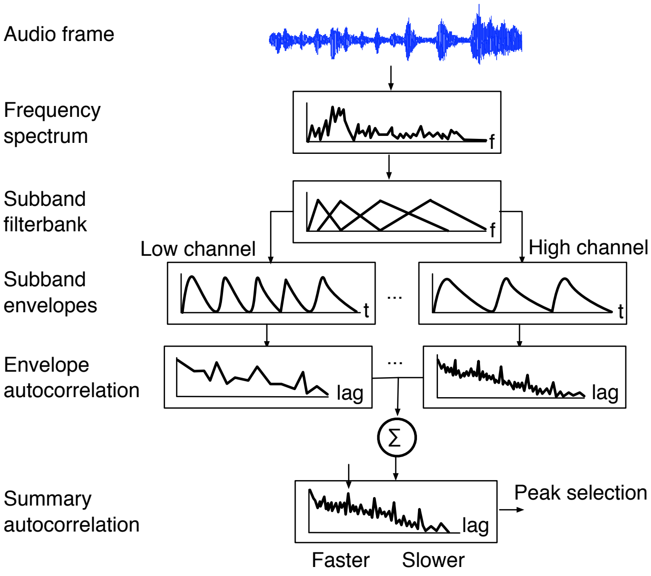

Fig. 2. Flowchart of autocorrelation-based beat-identifying algorithm. The arrow in the Summary Autocorrelation box indicates the point that will be used for estimating the tempo.

The audio period and subband energies are used to determine which frames contain a beat. As the system processes a frame, it sums its subband energies to obtain the frame's total energy. It then sums this total energy with the energies from frames spaced one, two, and three audio periods previous to determine a multi-frame energy value. If a given frame contains a beat, not only will it likely have a high energy value, but the frames that are periods behind it are also likely to contain a beat and have a high energy value. The multi-frame energy for the current frame is compared with that of all the previous frames in the audio period. If the current frame's multi-frame energy is at least 80% of the maximum value, the system determines that a beat is in that frame.

2) Pitch Detection: There is a direct relationship between the fundamental frequency of a note's acoustic signal and the pitch of the note for most instruments and sounds. Pitch detection algorithms use this to determine notes by finding the dominant frequencies present in a signal. Because instrumental notes are not single frequency tones there are also overtones present, which occur at integer multiples of the fundamental frequency [25].

This system assumes that the desired notes are known, and looks for their fundamental frequencies in the frequency spectrum of the audio signal. If a peak is found at the note's fundamental frequency, it determines that the note is present. Fig. 3 shows the notes E4 and C4 on a musical staff along with the Fourier magnitude spectrum of a 2-second clip of the notes played together.

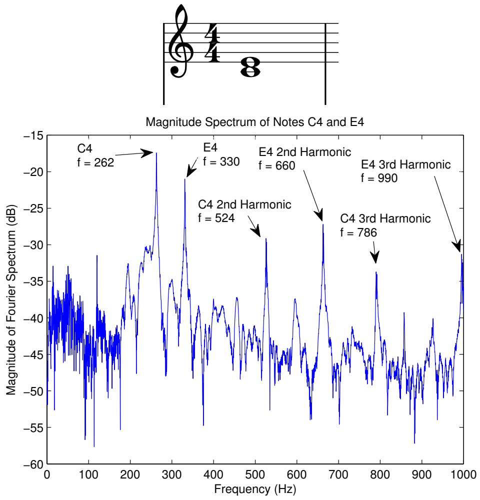

Fig. 3. Notes C4 and E4 displayed on a musical staff (TOP) and magnitude spectrum of the notes played together (BOTTOM)

3) Note Onset Detection: When the robot performs, it occasionally hits a note twice because the force of the hit causes the robot to move backwards, release the key, and then activate it again as it settles. This problem has been observed on electronic keyboards that have a played/not played threshold for the keys. Throughout this paper, this is referred to as a "double-hit" note.

In order to correct this error, our system uses note onset detection. Onset detection finds the beginning, or onset, of notes. If two or more onsets are detected during a single note, the robot can adjust its motion to play the keys less forcefully and hopefully eliminate the problem in the future.

The energy of the audio signal for the duration of the note is calculated in sections of approximately 23 ms. The derivative of the energy is then calculated and the peaks of the derivative are found. The beginning of a note usually corresponds with a sudden increase in signal energy, so if the derivative is above a threshold it is considered an onset. Fig. 4 shows the note C4 during a double-hit along with the signal energy, the derivative of the signal energy, and the threshold value for the derivative.

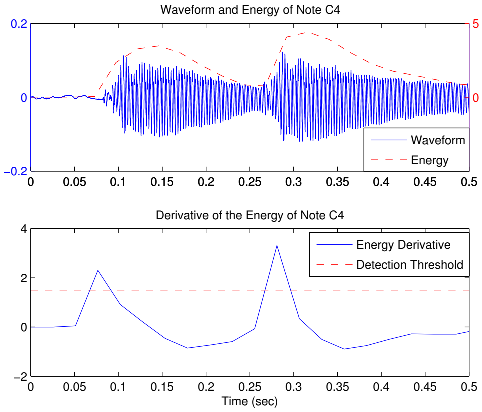

Fig. 4. Signals used in double-hit detection. The double-hit note is shown along with its energy (TOP), and the derivative of the signal energy is shown with the detection threshold (BOTTOM).

The robot is capable of playing notes with both hands, and occasionally two simultaneous notes will be detected as slightly offset. Therefore, the peaks must be at least 0.1 seconds apart in order to be considered a double-hit.

B. Vision

Musicians generally use visual cues in order to find and follow a tempo. In large ensembles, such as an orchestra, there is usually a conductor to visually provide a tempo for all the musicians to follow. Examples of conducting patterns for different meters are shown in Fig. 5.

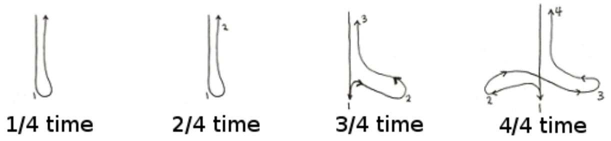

Fig. 5. Diagram of conducting timings/meters. For each different timing the conductors hand follows the arrow. The conductors hand will be where the number is located at the beginning of each beat. The conductors hand starts at the beginning of the arrow at the beginning of each measure.

In order to implement visual beat tracking we used cameras mounted on Jaemi Hubo's head, as shown in Fig. 6. The live video input is equalized to compensate for different lighting conditions and the image size is reduced in order to improve computation time. The relative motion between consecutive frames is then found using the computer vision Horn-Schunck Optical Flow method [26]. The spectrum of this motion's mean magnitude and angle is analyzed to determine the tempo. The possible tempos are weighted using a triangular filter with a peak at 128 beats per minute (BPM), the average tempo of popular music [27]. This reduces the effects of harmonics and helps ensure the correct tempo is chosen. Using this method, Hubo is able to accurately track the beat when watching a trained musician conduct a steady tempo.

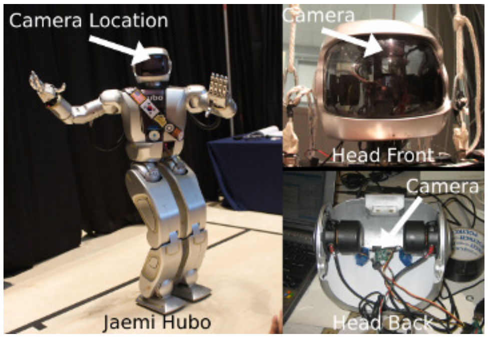

Fig. 6. (LEFT) Jaemi Hubo full body view and camera location. (TOP RIGHT) Jaemi Hubo front head view and camera location. (BOTTOM RIGHT) InsideJaemi Hubo's head and camera location.

C. Combining Audio and Visual Processing

Currently, we have functional methods for musical beat tracking using audio processing techniques and musical beat tracking using purely visual techniques.Combining these systems is key to taking the next step in musical awareness and becoming an interactive performer. The current sensing systems operate independently, but network communications can be used to combine the estimates in order to obtain, for example, more accurate and robust beat detection. On Hubo, all of this can be implemented within the robot's outer shell.

The preliminary system used to combine the visual and auditory tempo data is a confidence-based sliding window averaging system. The system has multiple inputs, including: calculated tempo, tempo confidence, start beat location, and a system time stamp. These inputs are for both the visual and the auditory tracking systems. The time stamp of each tracker is sent in the packet to scynchronize the systems which output at different rates. Before the system starts, the integrating computer logs the current times of the systems and uses those values to synchronize received data.

Once the vision and auditory systems are properly synchronized we average the calculated tempo from both systems using a confidence-based algorithm which acts on the past few seconds. The beat locations determined by the different systems are weighted based on the confidence level of each system. These weighted beat locations are used to calculate the overall predicted beat location. We line up the optimal beat locations based on the calculated tempo with the past N calculated beats, allowing the system to predict the next beat. The start beat location is the time step corresponding to the start of the beat as calculated by the combined vision and auditory beat tracking systems.

V. SYSTEM CAPABILITIES

This section describes the current abilities of our system. The robot is able to play two-finger melodies on a music keyboard. The system can detect missed notes and double-hit notes using audio input. The system is also able to dance synchronously with popular music in response to the audio signal. Tempo can be determined either by listening to the audio input or by watching a human conductor.

A. Piano Performance

Due to the small size of the mini-humanoid, it has a limited range of playable notes. Although it can play a real piano, the demonstration system uses a CASIO SA-75 electronic keyboard with 37 "mini-size" keys. These smaller keys are well-suited to a small humanoid with a limited playing range. The hands were replaced with longer, hooked "fingers" that extend the robot's reach and make it easier for the robot to hit the intended key. The new hands can be seen in Fig. 7.

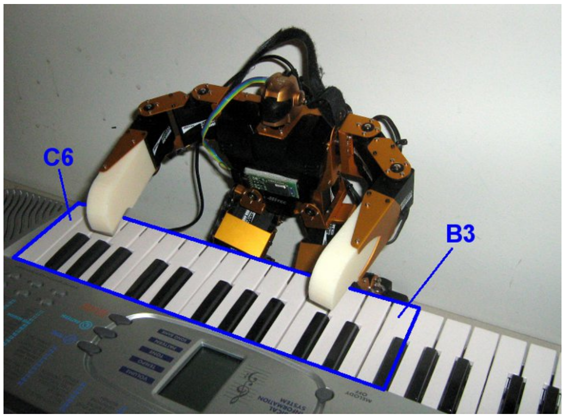

Fig. 7. Setup for the piano-playing robot with the playable range outlined.

The robot can currently play a one octave range with each hand. The arms are too short to allow them to cross over each other, so each octave is playable by only one hand. The left arm plays keys from the B below middle C (B3) up to B4 and the right hand plays notes from C5 up to C6. The range spans about 30 cm. This setup is shown in Fig. 7.

The robot can play a scale or a song using audio feedback. The notes can be specified in several ways, including reading from a MIDI file. As the robot plays, it uses the pitch detection algorithm to check that each note was played correctly. If any notes are not detected, the system checks to see if either neighboring note was played. If one of these notes are detected, the robot can adjust its playing positions to shift the arm away from the incorrect note the next time. Pitch detection also allows the robot to play in harmony with a human. The robot can detect played notes and either repeat the note or play in harmony at a fixed interval from the detected note. It can also use onset detection to compensate for double hit notes by adjusting how far it presses the key.

One of the challenges with this project is determining how far to lower the robot's hand to press a key. In order to make the system more robust and eliminate the need to adjust these parameters manually every time it is set up, a self-calibration program is used. This algorithm goes through every key and directs the robot to lower its hand slowly until the correct note is heard. It then stores the current arm position for future use and moves on to the next note.

B. Dance

The dancing system allows a robot to dance in time to music (Fig. 8). Like the piano-playing robot, the dancing robot

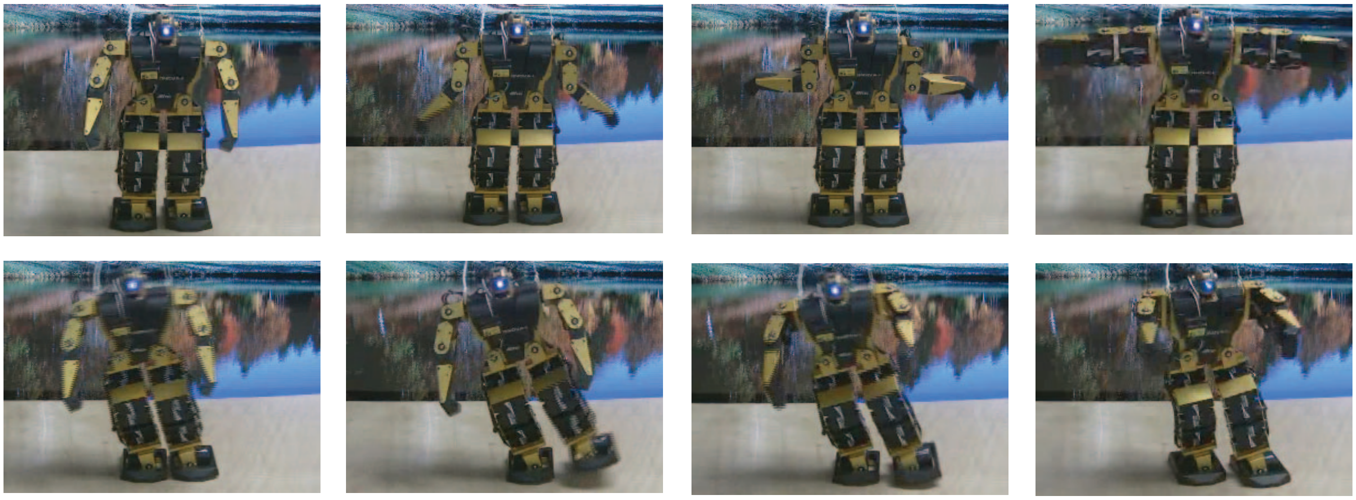

Fig. 8. TOP: The Robonova producing arm motions. BOTTOM: The Robonova producing arm motions and a leg motion.

receives new position commands every 20 ms. Currently, the Robonova can perform smooth, accurate dance gestures with its arms to the beat of popular music.

The dance system uses audio beat detection to find the beat in live music. The beat detection algorithm functions as described in Section IV-A.1. As the music plays, the most recent short time section is analyzed to determine the beat locations and predict the next few beat locations in the music. Analyzing short time segments allows the system to adapt quickly to tempo changes in the song.

Once the system has a set of predicted beats, it synchronizes the dance gestures so they reach their apex on the beat. It determines the start and end time for each gesture based on the beat locations and the end time of the last gesture. It then scales the timing of the gesture in order to change its length to fit into the allotted time period.

The 100Hz update rate allows precise control over the robot's motion, resulting in smoother gestures. In order to test the gestures, we had the robot swing its arm up and down in time to the beat as we recorded its position. The result was a relatively smooth and consistent sinusoidal motion.

VI. RESULTS

A. *Gesture smoothness*

We verified that the Robonova was capable of producing gestures smoothly enough to serve as a useful prototyping platform for Hubo. The robot was programmed to perform ten arm-swing motions in which the two arms rose and then lowered in opposite directions. We recorded these motions at a frame rate of 30 frames per second, determined the positions of the arms in all frames, and then took the derivatives of these positions. The derivatives of abrupt motion sequences contain long strings of zeros (when the robot is not moving) interspersed with some large magnitude values, while smooth motion derivatives contain mostly small, but non-zero, values. We first calculated the average speed the robot took to complete its swings, and then defined a smoothness threshold as being between one-half and three times that value. We calculated the percent of each sequence that was within the smoothness threshold. As a baseline, we performed the same experiment on a Robonova with the original, slower operating system. We found that the motions produced by the Robonova with the new environment were within the smoothness threshold 84% of the time, while only 17.5% of the motion sequences from the RoboNova with the original environment were within the threshold.

B. *Incorrect Note Detection*

To verify that the system can correctly detect missed notes, we ran two experiments. The first tested the system's ability to identify correctly played notes by having the robot play the first 23 notes of "Twinkle Twinkle Little Star" using both hands. This was performed 10 times for a total of 230 correctly played notes, of which only one was misidentified as incorrect. A high accuracy in detecting correct notes is important because believing a played note was missed is more detrimental to system performance than missing a wrong note when using the feedback to adjust playing.

The second experiment tested whether or not the system correctly identified missed notes. The same procedure was repeated, except that six notes were intentionally played incorrectly. Of the 60 total incorrect notes, the system detected 54 of them, for an accuracy of 90%.

VII. CONCLUSIONS AND FUTURE WORK

In its current state the overall system is somewhat limited, but we believe it represents several steps towards developing a fully musically-aware humanoid. With the benefit of external computing power, our humanoids have the ability to interpret sensory information from the environment, including visual and auditory beat tracking as well as pitch and note onset detection. The overall system is also able to incorporate these sources of information to calibrate performance control parameters and adjust accordingly in response to incorrect or missed notes. In the near future, all of the computing resources for these tasks will be integrated within the Hubo body, making the system fully self-contained.

We plan to continue improving the basic audio feature extraction (beat, onset, and pitch detection) to assess trade-offs between accuracy and computational complexity, which may facilitate the implementation of all computation on the robot's hardware. Improving the robustness of these features would enhance the system's ability to determine correctly and incorrectly performed notes, further advancing the overall system. Additionally, we plan to work towards the detection and understanding of higher-level musical features, such as meter, rhythmic styles, and musical genre, which further inform the expressive playing of human performers.

Currently, the adjustment and error correction in response to feedback is quite primitive and could easily be more sophisticated. Rather than always adjusting its movements by a pre-defined constant and relying mostly on dead reckoning, the piano performing robot could be programmed to make more intelligent decisions for correcting its performance based upon, for example, the frequency of occurrence of incorrect notes. Video processing could be expanded to enable tasks in addition to visual beat tracking of a conductor. The piano performing system could be enhanced by determining the location of the keys and the robot's hands visually, obviating the need for an operator to precisely position the robot prior to playing. Continually incorporating visual feedback to assess robot and keyboard position could improve the robustness of the system.

The ultimate target for full system implementation is the adult-sized Hubo humanoid, but we believe rapid algorithm prototyping with miniature humanoids to have substantial benefits. Although Hubo's significantly greater DOF enable a wider range of human-like motions, development on the larger platform is time consuming and poses much higher risks in terms of cost and safety. Developing robust systems for mini-humanoids also broadens the potential audience for this work, making musically interactive robots much more accessible to students and other researchers.

VIII. ACKNOWLEDGMENTS

We would like to thank Robert Ellenberg for his assistance with the Robonova, and Alex Kushleyev and Dr. Daniel Lee (University of Pennsylvania) for sharing their microcontroller code improvements for the Robonova.

REFERENCES

[1] S. Sugano and I. Kato, "Wabot-2: Autonomous robot with dexterous finger-arm coordination control in keyboard performance," in *Proceedings of the International Conference on Robotics and Automation*, 1987, pp. 90–97.

[2] A. Kapur, "A history of robotic musical instruments," in *Proceedings of the International Computer Music Conference*, 2005.

[3] A. Hajian, D. Sanchez, and R. Howe, "Drum roll: Increasing bandwidth through passive impedance modulation," in *Proceedings of the 1997 IEEE International Conference on Robotics and Automation*, Albuquerque, April 1997.

[4] M. Williamson, "Robot arm control exploiting natural dynamics," Ph.D. dissertation, Massachusetts Institute of Technology, 1999.

[5] H. Kose-Bagci *et al.*, "Drum-mate: A human-humanoid drumming experience," in *Proceedings of the IEEE-RAS International Conference on Humanoid Robots (Humanoids2007)*, Pittsburgh, November 2007.

[6] H. Kose-Bagci, K. Dautenhahn, and C. Nehaniv, "Emergent dynamics of turn-taking interaction in drumming games with a humanoid robot," in *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive communication*, Munich, 2008.

[7] M. Michalowski, R. Simmons, and H. Kozima, "Rhythmic attention in child-robot dance play," in *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication*, Toyama, Japan, September-October 2009.

[8] G. Weinberg and S. Driscoll, "The interactive robotic percussionist: New developments in form, mechanics, perception, and interaction design," in *Proceedings of the ACM/IEEE International Conference on Human Robot Interaction (HRI)*. New York: ACM, March 2007, pp. 97–104.

[9] C. Crick, M. Munz, T. Nad, and B. Scassellati, "Robotic drumming: synchronization in social tasks," in *Proceedings of the 15th International Symposium on Robot and Human Interactive Communication*, 2006.

[10] E. Singer *et al.*, "Lemur's musical robots," in *Proceedings of the 2004 Conference on New Interfaces for Musical Expression*, Hamamatsu, Japan, 2004.

[11] J. Bohannon, "Play it again, robot," *Science*, vol. 319, no. 5870, p. 1613, 2008.

[12] A. Kapur *et al.*, "Collaborative composition for musical robots," *Journal of Science and Technology of the Arts*, no. 1, pp. 48–52, 2009.

[13] K. Petersen, J. Solis, and A. Takanishi, "Musical-based interaction system for the waseda flutist robot," *Autonomous Robots*, vol. 28, no. 4, pp. 471–488, May 2010.

[14] J. Solis *et al.*, "Development of the waseda saxophonist robot and implementation of an auditory feedback control," in *Proceedings of the 10th Conference on New Interfaces for Musical Expression*, Sydney, Austraila, June 2010, pp. 156–161.

[15] G. Weinberg, A. Raman, and T. Malikarjuna, "Interactive jamming with shimon: A social robotic musician," in *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*, 2009, pp. 233–234.

[16] D. Zhang *et al.*, "Design and analysis of a piano playing robot," in *IEEE International Conference on Information and Automation*, June 2009.

[17] C. Roads, "The tsukuba musical robot," *Computer Music Journal*, vol. 10, no. 2, pp. 39–43, 1986.

[18] K. Kazuhiro *et al.*, "Dance partner robot - ms dancer," in *Proceedings of the 2003 IEEE/RSJ Conference on Intelligent Robots and Systems*, Las Vegas, 2003, pp. 3459–3464.

[19] F. Tanaka and H. Suzuki, "Dance interaction with qrio: A case study for non-boring interaction by using an entrainment ensemble model," in *Proceedings of the 2004 IEEE International Workshop on Robot and Human Interactive Communication*, Kurashiki, Okayama, Japan, September 2004.

[20] K. Shinozaki, A. Iwatani, and R. Nakatsu, "Concept and construction of a robot dance system," *The International Journal of Virtual Reality*, vol. 6, no. 3, pp. 29–34, 2007.

[21] K. Yoshii *et al.*, "A biped robot that keeps steps in time with musical beats while listening to music with its own ears," in *Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, San Diego, CA, October-November 2007.

[22] R. Ellenberg *et al.*, "Using miniature humanoids as surrogate research platforms," in *Proceedings of the 9th International IEEE/RAS Conference on Humanoid Robots*, 2009, pp. 175–180.

[23] D. Grunberg, R. Ellenberg, Y. Kim, and P. Oh, "Creating an autonomous dancing robot," in *Proceedings of the International Conference on Hybrid Information Technology*, 2009.

[24] A. J. Eronen and A. P. Klapuri, "Music tempo estimation with k-nn regression," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 1, pp. 50–57, January 2010.

[25] P. R. Cook *et al.*, *Music, Cognition, And Computerized Sound: An Introduction to Psychoacoustics*, P. R. Cook, Ed. The MIT Press, 2001.

[26] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–204, 1981.

[27] D. Moelants, "Dance music, movement and tempo preferences," in *Proc. 5th Triennal ESCOM Conference*, Ghent University, Belgium, 2003.