Development of an Autonomous Dancing Robot

David Grunberg\textsuperscript{1} Robert Ellenberg\textsuperscript{1,2} In Hyuek Kim\textsuperscript{3}
Jun Ho Oh\textsuperscript{3} Paul Y. Oh\textsuperscript{2} Youngmoo E. Kim\textsuperscript{1}
\textsuperscript{1}Department of Electrical and Computer Engineering
\textsuperscript{2}Department of Mechanical Engineering
Drexel University
Philadelphia, PA
\textsuperscript{3}Department of Mechanical Engineering
Korea Advanced Institute of Science and Technology (KAIST)
Daejeon, South Korea

dgrunberg@drexel.edu rwe24@drexel.edu inhyeok@kaist.ac.kr
jho@kaist.ac.kr paul@coe.drexel.edu ykim@drexel.edu

Abstract

We have developed an autonomously dancing humanoid, based on the Hitec Robonova platform, that is able to perform sequences of smooth movements in synchrony with music. Our system automatically determines the locations of beats from the acoustic signal of music in real-time, and this information is used to coordinate the motions of the robot, enabling it to perform rudimentary dances. The small (14 inch-tall) RoboNova serves as a prototype development platform for a much larger and more sophisticated humanoid, the KAIST Hubo, which is capable of more fluid and human-like movements. But by first developing and testing systems and algorithms on the smaller robot, we reduce the risk of damaging the larger and more costly Hubo. Regardless of the humanoid platform, however, it is difficult to synthesize motions that appear expressive and artistic. We developed our prototype dancing humanoid in order to further investigate and, hopefully, quantify some parameters of creative movement.

Keywords: robotics, humanoids, dance, beat tracking, beat extraction, dancing robots

1 Introduction

Humanoids, robots engineered to mimic human form and function, have become increasingly capable in recent years. For instance, the Albert Hubo, from the Korean Advanced Institute of Science and Technology (KAIST), is able to walk, shake hands, grasp objects, and speak with realistic facial expressions [1]. Honda’s ASIMO has walked onto a concert hall stage and conducted the Detroit Symphony Orchestra [2], and Wakamaru robots, produced by Mitsubishi, have performed in a theater production in Osaka,
Japan [3]. The Kawada HRP2 humanoid can perform elaborate traditional Japanese dances [4]. All of these robot performances, however, have employed pre-planned choreographed movements that have been meticulously programmed, essentially by hand.

In order to dance autonomously, a humanoid must be able to accomplish several tasks without assistance. In order to synchronize movements with the music, it must be able to identify beats, preferably from the acoustic waveform in order to allow the broadest possible range of music. The robot must then use this timing information to produce a sequence of continuous gestures, which must also be carefully selected to minimize the risk of destabilization. In order for the humanoid to respond appropriately when the music changes, these tasks should be performed in real time. Ultimately, we would like the robot to select movements similar to those a human would choose for a particular piece of music.

We are working on implementing algorithms to solve these problems and enable humanoids to dance autonomously. Our initial prototype uses a Hitec RoboNova (Figure 1) as our humanoid platform. The RoboNova is a small humanoid robot that has sufficient degrees of freedom (DoFs) to perform a variety of arm and leg movements. Our ultimate goal, however, is to deploy the system on Hubo, a larger and more complex humanoid developed by KAIST. We intend to use the RoboNova to prototype and test methods before deployment on the larger robot, a necessity for research with life-sized humanoids as they are complex and costly. General physical similarities between the two platforms ease the transfer of system implementation from RoboNova to the Hubo.

An autonomously dancing humanoid could be a valuable tool for furthering our understanding of creative movement. We know from experience that certain dance movements seem appropriate for certain types of music, but these same moves appear out of place when performed with different music. The quantification of dance styles remains an open area of research. We envision that our dancing humanoid could be used in perceptual experiments to quantitatively test and adjust gesture variations, to gain further insight into the relationship between creative movement and music.
2 Prior Work

In a recent paper, Shinozaki reviews some of the challenges involved in developing a robot dancer [5]. The synthesis of realistic, human-like gestures is required, but an even greater challenge lies in concatenating these movements smoothly. The paper also emphasizes the relationship between specific gestures and dance styles and the need for a robot dancer to be aware of differences in genre, perhaps through music analysis.

An existing dancing robot is Ms DanceR, constructed by Tohoku University [6]. This robot was designed to perform ballroom-style dances with human partners, and it possesses knowledge of ballroom dance styles and the ability to produce dance movements. Ms DanceR, however, does not identify beats to determine movement timing, but instead requires a human to meticulously guide it through dances.

Haile is a robot drummer programmed to accompany a piece of music [7]. Using rules for generating rhythms, Haile can synthesize and perform drum sequences in real-time and can accompany human players by improvising its own beat. Although vaguely human in form, Haile is not humanoid and lacks the ability to generate human-like gestures.

Keepon is a small cartoon-like robot consisting of a soft body and head, which can perform movements in a rhythmic manner [8]. But lacking arms or legs, it cannot perform human-like gestures. This robot is particularly appealing to children and was designed specifically to explore the use of music and interactive movement therapeutically for children with autism.

In order for a robot to dance autonomously, it must be aware of musical rhythm and beat. Automatic beat detection and identification is a subject that has been studied extensively by audio researchers (for example, [9] and [10]). Modern beat detection algorithms can successfully locate the time-specific pulses comprising a song’s rhythm. The best-performing methods, however, operate offline and are not suitable for the real-time analysis required for a robot dancer.

Other research aims to extract higher level rhythmic features and meter from audio [11][12]. Such detailed rhythmic data could potentially be useful for generating sequences of motions appropriate for the style of the music. For example, a more intense movement might occur on the downbeat of a measure, followed by a more subtle gesture on the offbeat. Once again, many of these algorithms are designed to operate offline, making them difficult to incorporate into a dancing robot.

3 Algorithm Components

Our dancing robot system is comprised of three primary components: the beat identifier, the robot platform, and the gesture generation and control system. Currently, we constrain our music selections to modern popular music (the detection of different musical genres remains part of our future work). Popular music usually employs strong beats, which enables robust beat identification from the acoustic signal.

3.1 Beat identifier

Our autonomously dancing robot will not have advance knowledge of music selections, so our beat identifier must be efficient enough to operate in real time and must operate in a causal manner. In addition to beat identification, estimation of the overall tempo of the music is needed so that the robot’s gestures can be scaled to match the pace of a song. Our beat identifier is an optimized and scaled version of a frequency-based algorithm, first proposed in [9].
3.2 Robot platform
The robot dancer must be able to perform motions similar to those performed by humans. This requires a humanoid platform with sufficient DoFs to perform a range of human-like motions. Dancing is a highly dynamic exercise, and the robot will be vulnerable to destabilization and loss of balance. Generally, inexpensive and durable humanoids lack the DoFs to mimic human movements accurately, while humanoids with higher DoFs are expensive and fragile. To address this problem, we employ two robot platforms: an inexpensive humanoid, the Hitec RoboNova, for prototyping algorithms and a highly advanced (and costly) humanoid, Hubo, for final implementation of algorithms that have been developed and tested on the prototyping platform.

3.3 Gesture generation and control
The robot dancer employs movements from a database containing many human-like gestures, specifically designed to maintain smoothness and balance and to minimize the energy needed to overcome gravity while moving [13]. The gestures are also short enough to be performed within one or two beats of music so that they can be easily integrated into dance sequences. The RoboNova does not have sufficient DoFs or fine enough motor control to perform extremely accurate gestures, such as those obtained from motion-capturing, so we have not incorporated such methods into our movement database.

4 Beat identifier
A general overview of our beat tracking system is provided in Figure 2, and a more detailed description follows.

- Audio is split into short-time analysis frames upon entering the system. Each frame is passed through a Cochlear filterbank, a set of filters that divides the audio into frequency sub-bands similar to those perceived by the human ear.

- Each sub-band is downsampled by a large factor (512 for 11kHz audio), half-wave rectified them, and smoothed with a lowpass filter. This provides an estimate of the time-domain energy envelope for each sub-band at a much lower data rate (∼20 Hz) more suitable for detecting musical beats (∼1-2 Hz).

- The energy envelopes are then analyzed by a bank of comb filters. These filters compare a signal with a delayed copy of itself, which produces resonance if the delay of the filter is equal to the period of the input signal. When the processed music sub-bands pass through such a filter, resonance occurs if the music’s tempo matches the filter’s delay. We exploit this effect by passing the sub-band envelopes through a range of comb filters and identifying the filter producing the greatest resonance across all sub-bands. We estimate the tempo from the corresponding delay of that filter.

- We determine the phase (exact time location) of the most recent beat by examining the delay states of the selected comb filter, indicated by a peak in the delay buffer. The position of the next beat is predicted from this location and prior tempo estimates.
Figure 2: The processing flow of the beat tracking system. Input audio is divided into sub-bands, downsampled and smoothed, and analyzed by a comb filterbank, the outputs of which are summarized in a “tempogram”.

5 Robot Platforms

The Hitec RoboNova-1, used for initial development and testing of the dancing system, has 16 DoFs. These are described in Table 1.

The RoboNova provides sufficient DoFs to perform numerous arm and leg gestures, and it is able to dance in a rudimentary fashion, making it a convenient platform to test our dancing algorithms. In addition, the robot is relatively inexpensive, durable, and easy to repair, so there is minimal risk in permanently damaging the robot. The RoboNova also is configured with sensors (such as a gyroscope) to enable feedback and thus improve its balance.

This robot’s microcontroller is not capable of performing beat detection on-board, so we employ an off-board computer that communicates with the RoboNova wirelessly.
Table 1: RoboNova and Hubo degrees of freedom

<table>
<thead>
<tr>
<th>Limb</th>
<th>RoboNova</th>
<th>Hubo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Each arm</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Each leg</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Head and waist</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>26</td>
</tr>
</tbody>
</table>

via Bluetooth. The off-board computer is also responsible for transmitting the gesture commands, which are executed immediately by the robot upon receipt of a command over the Bluetooth connection. Furthermore, the movement command transmissions are timed so that the motions reach their apex on the next predicted beat. Gestures are divided into sub-commands prior to transmission so that the RoboNova is able to perform different parts of each movement at different speeds. For example, the robot can decelerate gradually as a motion concludes, making the gestures appear more human-like.

The RoboNova’s maximum command update rate was originally \( \sim 5 \) Hz, meaning that it was able to receive a new movement command only 5 times per second. In this configuration, the robot occasionally received gestures too late to perform them in songs with fast tempos, so it occasionally fell behind the audio. Furthermore, motions performed within 1/5 second of each other had to be sent and performed simultaneously, or the robot would sometimes drop one of them. This problem, due to the limitations of the RoboBasic programming language, resulted in movements that were noticeably desynchronized with the audio beats. We have implemented a novel solution to this problem, which is described below in Section 6.

5.1 Intended platform: Hubo

Our eventual target platform for the robot dancer is the KAIST Hubo, a highly advanced life-sized humanoid. Hubo has more DoFs than the RoboNova, so it is able to express gestures in a more realistic manner. It has a variety of sensors for maintaining balance and precisely controlling the position of the robot, which are crucial for generating dance movements [15]. Hubo has already demonstrated impressive movement capabilities: it can shake hands and walk at a speed of 1.25 km/h [15], and its walk has been precisely modeled in a simulation environment to simulate human movement [16]. Furthermore, as Hubo’s walking ability is already well explored and simulated, it will be easier for us to determine the types of dance steps Hubo is able to perform.

Drexel University is the only institution in the United States possessing Hubo, and our experience with both platforms enables the implementation of the same beat identification and movement algorithms on this robot. Developing and testing these algorithms on RoboNova, of course, makes it much less risky to use Hubo. Complex movements, such as a tai chi sequence (from human motion-captured data), have been implemented on Hubo as pre-programmed choreography, demonstrating its gesturing capabilities. We have made preliminary progress towards enabling Hubo to dance autonomously, described in the experiments below.
6 GESTURE GENERATION AND CONTROL

We initially developed a library of 30 dance motions, forming a movement database. Each move consists of the motor control sequences enabling one limb to transit from one position to another. Arm motions may be combined into continuous sequences when performed by the RoboNova, but the legs are constrained to return to their home position after moving to avoid destabilizing the robot.

Due to the limitations imposed by the slow control update rates of the initial system, one of our research collaborators developed an alternative implementation, reducing the load on the RoboNova’s microcontroller by generating direct motor control commands on an external computer [14]. Reprogramming the microcontroller to simply distribute these motor control sequences increases the update rate to 100 Hz, allowing us to send individual gestures to the robot independently at whatever time the beat identifier found. In our preliminary results using this system, the robot’s movements appear much more synchronized with the music.

This significant improvement also enabled the use of cycloid functions, as opposed to linear interpolation, in determining motion paths (Equation 1).

\[ \theta_i = \frac{2t\pi - C \times \sin(2t\pi)}{2\pi} \]  

(1)

This function can be used to specify each joint angle with respect to time. The parameter C controls the degree of linearity of the displacement (lower values of C being more linear). The use of cycloid interpolation results in gestures that appear smooth, fluid, and human-like, much more so than their linearly interpolated counterparts. The improved control structure also enables the gestures to “cross-fade” into one another (interpolating between the ends of gestures and the beginning of subsequent gestures). As human movements similarly flow into one another, this makes the robots gestures appear more realistic. Human dancers move proportionally to the speed of the music they are listening to, and our system is programmed to match gesture speeds with the tempo predicted from the beat identifier.

7 EXPERIMENTS AND RESULTS

A demonstration of the first iteration of robot dancer can be viewed online [17]. This system, based on the RoboNova, is able to dance along with music in a stable manner. This preliminary work revealed several opportunities for improvement, and we modified several aspects of the system using the techniques described above. We performed experiments with the system to test the efficacy of our improvements, targeting five key areas to test the performance of our algorithms. Video examples of several of our tests are available online as described below.

7.1 Beat identification of multiple songs

In the original system design, the beat tracker could run in real time, but only for one song. We modified our system to handle consecutive songs and then conducted tests to determine if the beat tracker could keep pace and adjust to new music. If so, the system could be used in a simulated concert or dance environment in which consecutive pieces of music would be presented to the robot, more closely modeling real-world conditions.

1. http://music.ece.drexel.edu/research/Robotdance
We tested the beat identifier using multiple consecutive songs played from an iPod to ensure that the system could keep up with the new songs and that the beat tracker could shift to identifying the new source of audio. Using ten songs that the beat identifier performed well on individually, the system could still accurately identify beats when the songs were played in sequence. The beat tracker initially took a few seconds to synchronize with each new song, but it did not stall or fall behind. An excerpt from this test is available on our website.2

7.2 Gesture cross-fading and cycloid interpolation

The gestures produced by the RoboNova originally did not “cross-fade” (interpolate) between movements. These gestures also employed linear interpolation as opposed to cycloid interpolation. We wanted to verify that the new cycloid interpolated gestures worked appropriately and made the robot’s gestures look more human-like.

We ran an experiment to confirm that the cycloid-interpolated and cross-faded gestures appeared smoother than the previous system. Our qualitative judgements were that the gestures appeared significantly smoother and more continuous than those of the original system, and a video demonstrating the improved system is available on our website.3

7.3 Improved control environment

The upgraded RoboNova control environment is theoretically much faster than the old RoboBasic environment. The previous system was slow enough that a random 0.2 second variable latency was added to all gestures. This gave the robot a corresponding delay, and when the random delay was near the 0.2 second point, the delay was quite noticeable. We tested the new system to determine if latency was significantly reduced from the prior system.

We passed a synthetically-generated periodic audio signal with a tempo of 60 beats per minute to the system with both the original RoboBasic environment and the improved control environment. We directed the RoboNova to perform a right arm raise repeatedly, and we used sensors in the servos to communicate the arm’s true position over time back to the computer. We compared these numbers to the calculated joint positions produced by the linear and cycloid interpolation functions.

Our results demonstrated significant improvement over the original system, with the gestures using the new system matching the predicted positions almost exactly. The movement latency is far smaller than the 0.2 seconds characteristic of the original system. Figure 3 displays our data for the original and improved versions of the robot’s control environment; the original version clearly displays a much higher latency, verifying the speed and latency improvements offered by the new environment.

7.4 Tempo-dependent gestures

In the original system, the RoboNova performed all gestures at the same speed. As shown in the demonstration video, while the gestures land on the beat correctly, they start in random locations and do not fill the time between the beats. We modified the algorithm to have the robot perform gestures at different speeds based on the tempo

2. http://dasl.mem.drexel.edu/~robEllenberg/Projects/Dance/Media/iPod.mov
3. http://dasl.mem.drexel.edu/~robEllenberg/Projects/Dance/Media/Gesture%20Control.mov
of the music, so that the gesture length would correspond more accurately to the pulse of the music. We tested the system to make sure these modifications were successful and did not destabilize the robot.

We modified our gesture calculation algorithm to incorporate the tempo of the audio, and the tempo-dependent gestures appeared qualitatively more continuous, smoother, and more human-like.

7.5 Hubo gesture development and transmission

We wished to assess the utility of the RoboNova as a prototyping robot for Hubo. In particular, we wanted to know how easy it would be to enable Hubo to dance using the same system, both for a simple single gesture and for a dance sequence. We also wanted to determine the difficulty of developing gestures for Hubo, given familiarity with the RoboNova gestures.

Implementing a gesture, one originally designed for the RoboNova, on Hubo required three significant efforts:

- Hubo has significantly more DoFs than the RoboNova, so the gesture has to be augmented with positions for the additional joints. Furthermore, the gesture parameters had to be transcoded for the Hubo development environment, although this is a one-time cost for each gesture.

- It was necessary for the off-board computer to break gestures up into smaller components (for control efficiency) before sending them to the RoboNova, whereas gestures can be sent to the Hubo in their entirety and then split on-board the robot.
The changes to the gesture transmission algorithm to adjust for this were trivial.

- Hubo’s serial port is capable of a higher Baud rate than the RoboNova, allowing for much faster communication. Setting the Baud rate of the serial port on the offboard laptop to match was also trivial.

From the original library of 30 RoboNova gestures, we chose 16 to develop and implement for Hubo. These gestures were recalculated and transcoded for the Hubo development environment, and the system was implemented on a set of Hubo arms for testing. The same beat identification and Bluetooth communication system as used with the RoboNova was employed for this experiment.

The most time-consuming aspect of this effort was simply re-coding the gestures. This required empirical determination of some hardware constraints of the arms (such as the maximum motion speed and the minimum gesture time). Once these gestures were developed, there was no additional difficulty in producing gestures in exactly the same manner as with the RoboNova system. A video of the dancing arms experiment is available online.  

8 Conclusions and Future Work

We have made substantial improvements since our original implementation of a robot dancer [17]. The beat identifier is now able to track music continuously and does not require human intervention between songs. Although algorithm takes a few seconds to synchronize with a new piece of audio, we believe this to be acceptable because humans also pause for a few moments to adjust to new music when dancing. The RoboNova can gesture much more smoothly and accurately than it could previously, as it can cross-fade and employ cycloid interpolation to more closely mimic human movements. Finally, we have verified that the system architecture can be translated to Hubo with minimal changes.

Future work will focus on four areas. The first is further improvements to the beat tracker, such as enabling the identification of higher-order rhythmic information (strong vs. weak beats, time signatures, and other patterns). When humans dance, they choose sequences of moves based not only upon the sequence of beats, but on beat patterns. These can range from formal patterns for ballroom dances to more general rules, such as making stronger movements on a downbeat. If the beat identifier is able to identify these types of rhythmic features, it may be possible for the robot to autonomously create sequences of gestures that more closely match those of a human.

Second, we will continue to improve our gestures so they appear more realistic. Creating gestures synthetically (“from scratch”) is tedious and inaccurate. Our next step will be to obtain more realistic gesture data, perhaps by using motion capture techniques with a human dancer. Since Hubo has a body structure with joints similar to those of humans, the translation of motion-captured data to the robot is more straightforward. Hubo is, of course, has an upper limit of dance ability due to its construction (weight, speed of motors, etc.), but we are far from approaching these limits in practice.

Third, we will begin pursuing automated musical style detection. Past research on genre identification has investigated acoustic features that provide a strong indication of musical genre. Our focus will be to develop a system capable of real-time identification.

of musical genre, working in conjunction with the beat identifier. Accomplishing this could enable our system to produce gestures appropriate to the style of the music playing.

Finally, we plan to continue efforts towards making Hubo a truly autonomous dancing robot. We will focus on bringing the dancing-Hubo system to the current level of the dancing-RoboNova system, that is, capable of performing gestures safely and accurately. Already the Hubo arms can produce specific motions in time with music, but we will continue to implement motions to provide a full-range of human gestures. This effort will require the incorporation of additional constraints in gesture synthesis, such as balance and energy minimization.

9 ACKNOWLEDGEMENTS

This research is supported by the National Science Foundation (NSF) under grants Nos. 0730206 and IIS-0644151. We gratefully thank them. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF. We also thank Alex Kushleyev and Dr. Daniel Lee (University of Pennsylvania) for sharing their improvements to the RoboNova development environment.

REFERENCES


