Creating an Autonomous Dancing Robot

David Grunberg	Robert Ellenberg	Dr. Youngmoo Kim	Dr. Paul Oh
Drexel University	Drexel University	Drexel University	Drexel University
3141 Chestnut Street	3141 Chestnut Street	3141 Chestnut Street	3141 Chestnut Street
Philadelphia, PA, 19143	Philadelphia, PA, 19143	Philadelphia, PA, 19143	Philadelphia, PA, 19143
dkg34@drexel.edu	rwe24g@gmail.com	ykim@drexel.edu	paul@coe.drexel.edu

Abstract

A robot with the ability to dance autonomously has many potential applications, such as serving as a prototype dancer for choreographers or as a participant in stage performances with human dancers. A robot that dances autonomously must be able to extract several features from audio in real time, including tempo, beat, and style. It must also be able to produce a continuous sequence of humanlike gestures. We chose the Hitec RoboNova to use as a robot platform in our work on these problems. We have developed a beat identification algorithm that can extract the beat positions from audio in real time for multiple consecutive songs. Our RoboNova can now produce sequences of smooth gestures that are synchronized with the predicted beats and match the tempo of the audio. Our algorithm can also be easily moved to the HUBO, a large humanoid robot that can move in a very humanlike manner.

1. INTRODUCTION

An increasing number of tasks can be performed by humanoid robots. For instance, the Albert HUBO, from KAIST, could walk, shake hands, grasp objects, and speak with realistic facial expressions [1]. The Asimo robot, from Honda, conducted the Detroit Symphony Orchestra [2]. Wakamaru robots, produced by Mitsubishi, have even performed in an Osaka theater production [3]. However, all of these robots require substantial hard coding of commands. There does not to our knowledge exist a humanoid robot that could, for example, autonomously determine how to conduct a new piece of music, or that could deduce on its own

Figure 1, the RoboNova robot

how to emote in a play.

One particular task that we would like to see robots perform autonomously is dancing. There are myriad uses for a robot that could dance without needing human input. Such robots could perform on stage with bands and could react appropriately to the minor variations that are inherent in live performances. They could aid teachers by demonstrating dance moves to students. These robots could also be very effective in prototyping dance choreographies, as they could quickly perform sequences of moves in specific and unvarying manners. This would make it much easier for choreographers to iterate and improve their dances, since they could rapidly test each iteration on the robot.

An autonomously dancing robot must be able to accomplish several tasks without human help. To line up its gestures with the music, it must be able to identify the music's beats. It must also extract the higher level rhythms and meter of the piece. Simultaneously, it must identify the genre of the music, since dance moves vary between genres. Waltz moves, for example, would look inappropriate when set to rock music, and vice versa. The robot must be able to take this information and use it to produce a string of continuous gestures that are smooth, humanlike, stylistically appropriate, and synchronized with the music's beat. Finally, this all must be done in real time, so that if

the music changes or something else unexpected happens, the robot can continue to dance without requiring human instruction.

We are working on implementing algorithms to solve these problems and enable a Hitec RoboNova (Figure 1) to dance autonomously. The RoboNova is a small humanoid robot that has sufficient degrees of freedom (DoFs) to perform several gestures. The RoboNova is also similar in physical configuration to the HUBO, a larger and more complex humanoid robot designed by KAIST. The similarities between the two robots will make it feasible to move algorithms from the RoboNova to the HUBO, and will thus allow the RoboNova to serve as a prototyping device for the larger robot. Prototyping platforms are essential for larger humanoid robots, as they are very expensive and are often fragile. With the RoboNova, we can test algorithms before implementing them on the HUBO, and can thus ensure that the HUBO will not be commanded to perform any unsafe motions.

2. PRIOR WORK

A previous dancing robot is the Ms DanceR, constructed by Tohoku University, which was designed to perform ballroom dances with human partners [4]. It thus possesses knowledge of ballroom dance styles. However, this robot cannot identify beats in music. It instead requires a human to guide it through dances. This also means that the Ms DanceR is not autonomous.

Two other robots that react to music are the Keepon and the Haile [5] [6]. The Keepon is a small robot consisting of a body and a head, and its head moves in synchrony with the beat of music [5]. This robot was found to elicit a pleasing reaction in child observers, thereby indicating that interactions between music and robots can be constructive and useful. However, this robot can only move its head; it cannot use arms or legs to make humanlike gestures. The Haile is a robot that is programmed to play a drum in a manner that fits music it is hearing [6]. It can operate in real time, and can accompany human players by improvising its own beat. This indicates that it understands rules for generating rhythms. This robot, though, also cannot gesture as humans do. Furthermore, neither Keepon nor Haile are humanoid.

[7] discusses some issues involved in generating a dancing robot. In particular, it mentions the importance of creating humanlike gestures and concatenating them together smoothly. It also mentions that, absent genre information, the robot's gesture database will need to be restricted to one genre.

Music beat identification has been studied extensively by music researchers (for example, [8], [9], and [10]). Beat identification is crucial to an autonomously dancing robot because, if a robot cannot line up its gestures with the beat of audio, its dances will look awkward and unrelated to the music. Modern beat identifiers can successfully locate the pulses that comprise a song's rhythm. Today's best algorithms, however, operate offline, and so are not suitable for a real-time dancer.

More complex beat identifiers, such as those proposed in [11] and [12], can extract the higher level rhythm and meter from audio. This is useful for generating strings of gestures that not only land on beats, but flow logically from one to the next. Human dances make use of metrical concepts like downbeats, measures, and syncopation, and so dancing robots should also understand how to incorporate knowledge of these concepts into their dances. For example, a more expressive motion might occur on the downbeat of a measure, and then a less expressive one on the offbeat. The best performing of these algorithms also operate

offline, and so they too cannot be directly used in an autonomously dancing robot.

Our own prior publication on this project describes some of the algorithms we use for the robot [13]. It discusses our robot platform and our beat identifier. However, this system is less capable than our current one.

3. ALGORITHM COMPONENTS

The dancing robot algorithm can be divided into four parts: a beat identifier, a robot platform, gestures, and a genre identifier.

3.1. Beat Identifier Objectives

At the most basic level, our beat identifier must correctly predict beats in audio. It also must be causal, because an autonomously dancing robot may not know a song in advance, and it needs to run in real time. We also wanted it to identify the tempo of the audio, so that we could scale our gesture speed to match the song's pace. Detection of full rhythms was outside the scope of this paper.

3.2. Robot Platform Objectives

Our robot must have the ability to gesture to music. Therefore, it must have enough DoFs to perform several motions with its arms and legs. It also requires a communications system that can command the robot to gesture so that the robot's motions line up with the beat identifier's beat predictions. Finally, the robot platform should be similar to a more complex robot, the HUBO, so that algorithms and gestures can be tested on the inexpensive prototype before trying them on the more expensive robot. Making the HUBO itself dance was outside the scope of this paper, but making a communications algorithm that can communicate gestures to either robot was within it.

3.3. Gesture Objectives

Gestures for this project were required to be short enough to be performed in one or two beats, and had to be performable with the available DoFs of the robot platforms. Gestures also had to be flexible so that the robots could cross-fade different gestures to make smooth movement sequences. Flexibility also mattered for enabling the robots to perform gestures at varying speeds (such as the tempo of the music) rather than at a fixed speed. Creating more realistic human gestures, such as those extracted from humans with motion capture technology, was outside the scope of this paper.

3.4. Music Genre and Style Objectives

Constructing a style detector was outside the scope of this part of the project. Music with strong beats from a variety of genres was used. Examples of such songs include "That's All," by Genesis, "Achy Breaky Heart," by Billy Ray Cyrus, and "I Turn My Camera On," by Spoon. Style identification is part of our future work.

4. BEAT IDENTIFIER

Our beat tracker is based on the algorithms proposed by [8] and [9], because [14] indicates the strength of their techniques. Our system, however, functions in real time. It is depicted in Figure 2 and functions as follows:

- As the audio is passed into the system, it is split into frames. Each frame is passed through a Cochlear

filterbank. This filterbank splits the audio into frequency subbands that are similar in perception to the human ear. (Figure 3b)

- The subbands are downsampled by a factor of 512, half-waved rectified, and smoothed with a lowpass filter. (Figure 3c)

- The smoothed subbands are passed through a bank of comb filters. When music passes through a comb filter that has a certain delay, resonance will occur if its delay matches the music's tempo. We exploit this by passing the subbands through many comb filters, identifying the one that produces the most resonance across all subbands, and estimating the tempo from that filter's delay. (Figure 3d)

- The phase of the next beat is determined from the delay states of the selected comb filter, and the position of the beat is then determined from the tempo and the changes in the phase.

- To remove artifacts of two beats in rapid succession, estimated beats within .25 seconds of each other are merged into one beat.

5. ROBOT PLATFORMS

The Hitec RoboNova-1 is about 14 inches tall and has 16 DoFs, which are listed in Table 1. Its DoFs enable it to make numerous arm and leg gestures, and therefore allow it to dance. It thus provides a convenient platform to test our autonomous dancing algorithms. In addition, the robot is inexpensive and easy to repair, so if it is damaged while dancing, repairs are cheap and fast. Finally, it has sensors (such as a gyroscope) to enable feedback and thus improve its balance.

In order to dance, the robot must have knowledge of the beats predicted by the beat identifier. The robot's processor does not have enough memory to perform the beat determination onboard, so an offboard laptop is used to communicate with the robot via Bluetooth. The robot does not do any timing calculations itself; rather, it performs each motion immediately upon receiving and processing a gesture command. The computer is responsible for timing the gesture transmission so that the gestures apex on the next predicted beat. Gestures are split into pieces before transmission so that the RoboNova can perform different parts of each gesture at different speeds; for example, it can decelerate gradually in the last part of the gesture. This makes the gestures look more humanlike.

One important goal for the dancing robot is that it must keep up with the music. Initially, the RoboNova's update rate was only about 5 Hz, meaning that it could only check to see if it had received a new gesture every .2 seconds. This caused a variable latency of up to .2 seconds in the gestures, and sometimes delayed

Table 1, DoFs for the RoboNova and HUBO

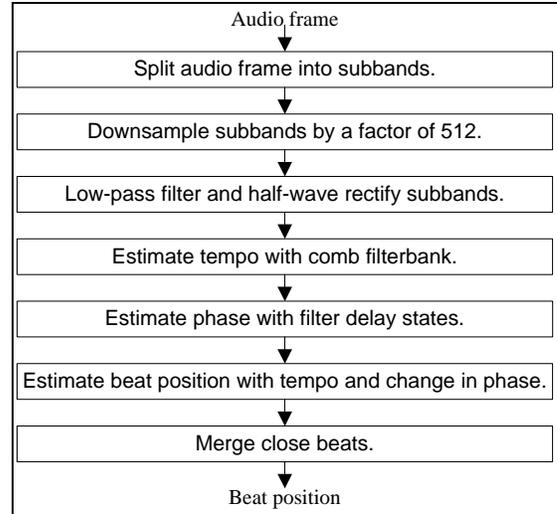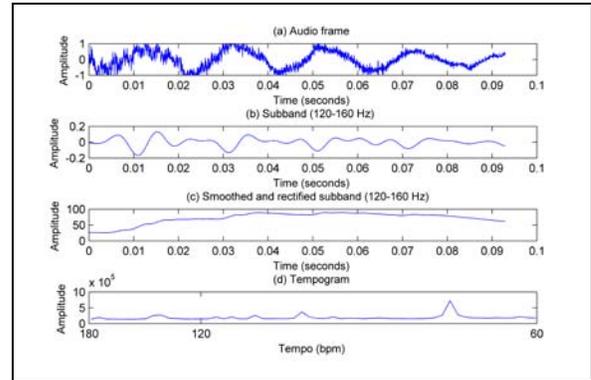| Limb	RoboNova	HUBO
Each arm	3	6
Each leg	5	6
Head and waist	0	2
Total	16	26

Figure 2, beat identifier flowchart

Figure 3, audio frame from "Fire Wire" by Cosmic Gate (a), subband of the audio frame (b), smoothed and rectified subband (c), and tempogram over several audio frames (d)

a gesture enough that it could not be played on the next beat. A new operating system for the RoboNova was provided by the first author of [15]. This increased the update rate to 100 Hz, and allowed the robot to gesture much more closely to the beats than was possible with the old operating system.

The new operating system also improved the accuracy of the dances, as Figure 4 shows. The RoboNova performed a right arm raise several times to synthesized audio with a tempo of 60 beats per minute. We used sensors in the robot to determine the exact position of the joints during this performance, and then compared those positions to the ones that the computer calculated. As seen in Figure 4, the gestures match the predicted models almost exactly. The variable latency is far smaller than the .2 seconds that the old system had.

While the RoboNova is a very useful prototype dancer, it does not have the DoFs or the computational ability to truly mimic human dancing. Another robot is required for this. We selected the HUBO (Figure 5) as our full scale dancer for several reasons. It has more DoFs than the RoboNova, so it can express gestures more realistically (Table 1). It has a variety of sensors that can help it balance and can precisely control the position of the robot, a factor that is crucial in generating high quality dances [16]. It also has an extremely detailed programming environment that
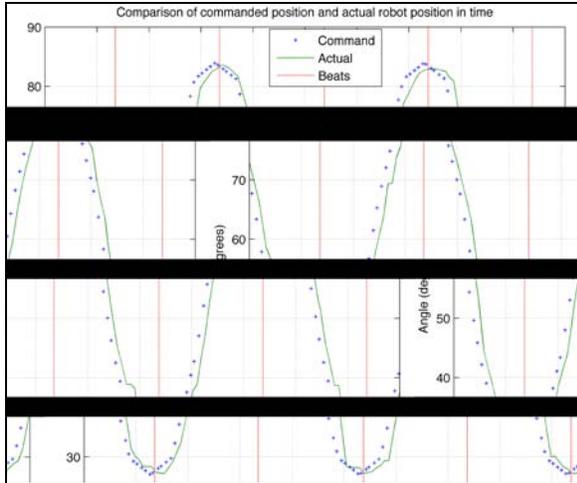
Figure 4, commanded and actual right arm raise

Figure 5, the HUBO robot

allows the programmers to determine the parameters used in the HUBO's motion. This will be useful when making the HUBO dance because the well-described parameters will allow us to define the HUBO's motion much more precisely than is possible with the RoboNova.

The HUBO has already demonstrated significant motion capabilities. It can walk while maintaining stability at a speed of 1.25 km/h and it can shake hands [16]. Its walk in particular has been precisely modeled in a simulation environment to simulate human movement [17]. These existing features will be very helpful when it comes to making the robot able to dance autonomously. Any robot needs to have the ability to remain stable if it will dance without falling over; the HUBO already having such a system will save us time later on. Furthermore, as the HUBO's walking ability is already well explored and simulated, it will be easier for us to determine what sort of dance steps the HUBO is capable of.

The HUBO is similar enough to the RoboNova that the same algorithms used on the smaller robot can be used with almost no modifications on the more expensive, larger one. For example, we took the system used on the RoboNova and transmitted a right

Figure 6, the HUBO GUI

arm raise to the HUBO, then recorded what modifications had to be made to the system before the gesture was successfully performed. Only three changes had to be made before gesture transmission succeeded:

1. The HUBO has more DoFs than the RoboNova, so the gesture had to be augmented with positions for the additional joints. Furthermore, the gesture had to be recoded into the HUBO environment. This was a one-time cost for the gesture. When the robot actually dances, this modification would be performed before the dance, and so it would not influence the real-time control of the robot at all.

2. The offboard computer had to break gestures up into small pieces before sending them to the RoboNova, whereas gestures could be sent to the HUBO in their entirety and then split up onboard the robot. The changes to the gesture transmission algorithm to adjust for this were trivial.

3. The HUBO serial port has a higher Baud rate than the RoboNova. Changing the Baud rate of the offboard laptop was also trivial.

Though work on making the HUBO dance autonomously has not yet begun, some dances have been programmed into the HUBO and show off its ability to make smooth, humanlike gestures. It can already do tai chi, for instance. Figure 6 shows part of the HUBO's Graphical User Interface (GUI), which lists several gestures it can perform. This is more evidence that the HUBO will be a suitable platform for a full-scale dancing robot.

6. GESTURES

30 dance moves were created and placed in a motion database. Each move consists of one limb transiting from one position to another. Arm motions can be combined into continuous sequences when performed by the RoboNova, but leg motions must return to their home position after moving to avoid destabilizing the robot. These moves include simple raises, simple steps, arm swings, and arm and foot taps.

In order to make the gestures look more humanlike, the motion paths were determined with cycloid functions instead of linear interpolation. These functions solves for each joint angle θ_i in terms of time. They have a parameter C that controls how linear

the displacement is, with lower values of C being more linear. One such function is shown below:

$$\theta_i = \frac{2\pi t - C\sin(2\pi t)}{2\pi} \qquad (1)$$

The effects of the cycloid interpolation can also be seen in Figure 5. The gesture position curves, both simulated and actual, are very smooth. They are not the angular curves that one would see if linear interpolation was used to generate the gestures. Furthermore, the curves in Figure 5 show minimal speed at the apexes of the gestures on the beat, and maximal speed during the gestures between the beats, which also matches human dance style.

We also enabled the gestures to cross-fade, or blend, into one another. This was done by interpolating between the ends of gestures and the beginnings of subsequent gestures. As gestures produced by humans also flow into one another, enabling the robot to perform smooth gesture overlaps makes the robot's gestures and dances look even more like human ones. A video of the robot using this ability is available on our website. [1]

Because human dancers adjust the speed of their movements to the speed of the music, we wanted to program the robot to match its gesture speed to the tempos predicted from the beat identifier. This makes the dances look more humanlike.

7. EXPERIMENT OBJECTIVES

The first iteration of the dancing robot, produced last year, can be seen online [13].[2] It can dance to music in a stable manner. However, it presents some opportunities for improvement.

We modified two aspects of the dancing algorithm based on ideas developed in this paper. We then ran experiments with the robot to verify that our improvements were effective towards our goal of moving towards an autonomously dancing robot that can emulate human dances. We tested the following key aspects:

1. In the original design, the beat tracker could run in real time, but only for one song. This is a deficiency, as explained by [10]. We modified the system to handle consecutive songs, and then ran a test to see if the beat tracker could keep up and could adjust to the new music.

2. The robot initially performed all gestures at the same speed. As the video shows, while the gestures land on the beat correctly, they start in seemingly random places and do not fill the beats. We modified the algorithm to have the robot perform gestures at different speeds based on the tempo of the music, so that the gesture length would match the time between beats. We tested the system to make sure these modifications improved the look of the robot's dances and did not destabilize the robot.

Because the objective of this project is to make a robot that dances like a human, we decided that video of the robot demonstrating our upgrades would be an important component of our tests. The web address of these videos is included in the 'Results' section of this paper.

8. EXPERIMENTS AND RESULTS

8.1. Beat Identification of Multiple Songs

We passed multiple songs into our beat identifier. These songs had different tempos and were of different genres. Selected genres included ragtime, soundtrack, and pop, and examples of songs used include "Achy Breaky Heart," by Billy Ray Cyrus, "Combination Rag," by Scott Joplin, and "The Ballad of Czolgosz," by Stephen Sondheim. We tested to make sure that the beat identifier could keep up with the new songs and that the beat tracker could shift to identifying the new source of audio.

An iPod was used to pass multiple songs in succession into the beat tracker. The songs were occasionally fast-forwarded for several seconds just after beginning to see if the robot could handle starting a song from the middle.

This test was a success. We tested ten songs that the beat identifier performed well on individually, and it could still identify their beats when they were played in sequence. An excerpt from our test is available on our website. [3] The beat identifier took a few seconds to begin tracking each new song well, but it did not freeze up or fall behind, and in each case was able to identify the same beats that it could when the songs were played individually within 10 seconds.

8.2. Tempo-Dependent Gestures

We modified our gesture algorithm to take into account the tempo of the audio, then had it run several gestures and watched it to see if the gestures seemed to fit the music better. It was a success; the tempo-dependent gestures did result in a more continuous string of gestures that looked smoother and more humanlike. A video of the robot varying its gesture speed is available on our website. [4]

9. CONCLUSIONS AND FUTURE WORK

We have made substantial improvements in our dancing RoboNova since [13]. The beat identifier can now track many songs in succession, without the need for a human to split them up. While there is a few seconds after a song switch during which the beat identifier is off, this is acceptable because humans also must pause for a few moments to adjust to new music when dancing. The RoboNova can also gesture much more smoothly than before, now that it can adjust its gesture speed to match the tempo of the music. This enables it to produce gestures that are much more connected and fit better to the music than the previous, one-speed gestures.

These results are of great importance to this project. A robot that dances autonomously must be able to dance to multiple songs, and it must continue to dance even if songs that it is not expecting are played. Live performances are inherently unpredictable, and if a band decides to extend one song, or suddenly segue to another, the robot needs to be able to cope. We have demonstrated that the robot can handle this sort of sudden change. Also, the robots'

---

1   http://dasl.mem.drexel.edu/~robEllenberg/Projects/Dance/Media/Gesture%20Control.mov

2   http://schubert.ece.drexel.edu/research/Robotdance

3   http://dasl.mem.drexel.edu/~robEllenberg/Projects/Dance/Media/iPod.mov

4   http://dasl.mem.drexel.edu/~robEllenberg/Projects/Dance/Media/Gesture%20Control.mov

dances now look much smoother and more like human ones because its gesture speed matches that of the music it is dancing too. As we want to eventually enable robots to dance in a manner indistinguishable from humans, this is an important step along that path.

Future work will focus on four different areas. The first will be further improvements in the beat identifier, such as making it able to identify a variety of higher order rhythms. This is essential for an autonomously dancing robot. When humans dance, they perform sequences of moves based not just on the sequence of beats but on beat patterns. These can range from very formal patterns for ballroom dances to more general patterns such as making more emphatic movements on a downbeat. Once the beat identifier is able to identify high level rhythms, it will be possible for the robot to perform sequences of gestures that more closely mimic those a human would perform.

Second, we will continue making our gestures more humanlike. We have gone about as far as we can go with the simple gestures we created. Our next step will be obtaining more realistic gestures, perhaps by using motion capture techniques. We wish to obtain gestures that real humans performing real dances make, and then pass these on to the robot. Once we accomplish this, the robots will be limited in their ability to dance like humans only by their construction. As the HUBO has a body structure with joints similar to those of humans, this different will hopefully be minimal.

Third, we will begin looking into style detection. We will need to identify features in audio that can give a good indication as to what genre of music it is. We will then need to construct an algorithm capable of identifying the genre of music in real time that works in conjunction with the beat identifier. Accomplishing this could allow us to diversify our gesture database and enable our robot to produce gestures that match the music it is using to dance.

Finally, we will do more work on making the HUBO into an autonomous dancing robot. We will focus on getting the HUBO to the point that the RoboNova is in – capable of performing gestures safely and accurately. The HUBO is much more capable of producing humanlike gestures than the RoboNova, so enabling it to dance will result in a robot that more closely approximates human dances.

10. ACKNOWLEDGEMENTS

11. REFERENCES

[1]   L. Zyga. Einstein Returns From OuterSpace. InventorsSpot. http://inventorspot.com/articles/einstein_returns_outer_ space_ 9455.

[2]   Honda 2008. Honda's Asmio Robot to Conduct the Detroit Symphony Orchestra. Press Release. http://asimo.honda.com/ ASIMO_DCTM/News/renditions/pdf/ASIMO_at_DSO_rel ease_rev.pdf.

[3]   British Broadcasting Corporation. Actor Robots Take Japanese Stage. British Broadcasting Corporation U.K. http://news.bbc.co.uk/2/hi/asia-pacific/7749932.stm.

[4]   K. Kazahiro et al, 2003. Dance Partner Robot -Ms DanceR-. In Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems 4 (Las Vegas, October, 2003), 3459–3464.

[5]   Michalowski, M.P., Sabanovic, S., and H. Kozima, A Dancing Robot for Rhythmic Social Interaction. In Proceedings of the 2nd Annual Conference on Human-Robot Interaction (March, 2007), 89-96.

[6]   Weinberg, G. and Driscoll, S. 2007. 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 (New York, March, 2007), 97-104.

[7]   Shinozaki, K., Iwatani, A. and Nakatsu, R. 2007. Concept and Construction of a Robot Dance System. In The International Journal of Virtual Reality 6 (2007), 29-34.

[8]   Scheirer, E.D. 1998. Tempo and Beat Analysis of Acoustic Musical Signals. In Journal of the Acoustical Society of America 103 (Melville, New York, 1998), 588-601.

[9]   Klapuri, A.P., Eronen, A.J., and Astola, J.T. 2006. nalysis of the Meter of Acoustic Musical Signals. In IEEE Transactions On Audio, Speech, and Language Processing 14-1 (2006), 342-355.

[10] Murata, K., et al. 2008. A Beat-Tracking Robot for Human-Robot Interaction and its Evaluation. In Proceedings of the 8th IEEE-RAS International Conference on Humanoid Robots (New York, 2008), 79-84.

[11] Davies, M., and Plumbley, M. 2007. Context-Dependent Beat Tracking of Musical Audio Signals. In IEEE Transactions on Audio, Speech, and Language Processing 15-3 (March, 2007), 1009-1020.

[12] Goto, M. 2001. An Audio-based Real-time Beat Tracking SyStem for Music With or Without Drum-sounds. In Journal of New Music Research 30-2 (Routledge, New York, 2001), 159-171.

[13] Ellenberg, R., Grunberg, D., Oh, P., and Kim, Y. 2008. Exploring Creativity through humanoids and dance. In Proceedings of the 5th International Conference on Ubiquitous Robotics and Ambient Intelligence (November, 2008).

[14] Gouyon, F., et al. 2006. An Experimental Comparison of Audio Tempo Induction Algorithms. In IEEE Transactions On Audio, Speech, and Language Processing 14-5 (2006), 1832-1844.

[15] Kushleyev, A., Garber, B., and Lee, D.D. 2008. Learning humanoid locomotion over rough terrain. In Proceedings of the 5th International Conference on Ubiquitous Robots and Ambient Intelligence (November, 2008).

[16] Park, I.W. et al. 2005. Mechanical design of humanoid robot platform KHR-3 (KAIST HUMANOID ROBOT-3: HUBO). In Proceedings of the 5th IEEE/RAS International

Conference on Humanoid Robots (New York, 2005), 321-326.

[17] Lee, J., and Oh, J. 2008. Walking Pattern Generation for

[18] Planar Biped Using Q-Learning. In Proceedings of the 17th World Congress of the International Federation of Automatic Control (IFAC) (Laxenburg, Austria, July, 2008), 3027-3032.