[16] K. Seo, S.-J. Chung, and J.-J. Slotine, “CPG-based control of a turtle-like underwater vehicle,” *Auton. Robots*, vol. 28, pp. 247–269, 2010.
[17] M. Saito, M. Fukaya, and T. Iwasaki, “Serpentine locomotion with robotic snake,” *IEEE Control Syst. Mag.*, vol. 22, no. 1, pp. 64–81, Feb. 2002.
[18] S. Ma, Y. Ohmameuda, and K. Inoue, “Dynamic analysis of 3-dimensional snake robots,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2004, pp. 767–772.
[19] M. Benosman, F. Boyer, G. Le Vey, and D. Primault, “Flexible links manipulators: From modelling to control,” *J. Intell. Robot. Syst.*, vol. 34, pp. 381–414, 2002.
[20] J. Blair and T. Iwasaki, “Optimal gaits for mechanical rectifier systems,” *IEEE Trans. Autom. Control*, vol. 56, no. 1, pp. 59–71, Jan. 2011.
[21] Z. Chen and T. Iwasaki, “Robust entrainment to natural oscillations of asymmetric systems arising animal locomotion,” in *Proc. 48th IEEE Conf. Decision Control*, 2009, pp. 2954–2959.
[22] L. Zhu, Z. Chen, and T. Iwasaki, “Oscillation, orientation, and locomotion of underactuated multi-link mechanical systems,”*IEEE Trans. Control Syst. Technol.*, vol. 21, no. 5, pp. 1537–1548, Sep. 2013
[23] M. N. Islam and Z. Chen, “Natural oscillation control of prototype mechanical rectifiers,” *IEEE Trans. Control Syst. Technol.*, vol. 20, no. 6, pp. 1559–1566, Nov. 2012.
[24] M. Ohsawa and S. Namiki, “Anisotropy of the static friction of plain-woven filament fabrics,” *J. Textile Mach. Soc. Jpn.*, vol. 12, pp. 197–203, 1966.

Toward a Dancing Robot With Listening Capability: Keypose-Based Integration of Lower-, Middle-, and Upper-Body Motions for Varying Music Tempos

Takahiro Okamoto, *Member, IEEE*, Takaaki Shiratori, *Member, IEEE*, Shunsuke Kudoh, *Member, IEEE*, Shin’ichiro Nakaoka, and Katsushi Ikeuchi, *Fellow, IEEE*

***Abstract*—This paper presents the development toward a dancing robot that can listen to and dance along with musical performances. One of the key components of this robot is the ability to modify its dance motions with varying tempos, without exceeding motor limitations, in the same way that human dancers modify their motions. In this paper, we first observe human performances with varying musical tempos of the same musical piece, and then analyze human modification strategies. The analysis is conducted in terms of three body components: lower, middle, and upper bodies. We assume that these body components have different purposes and different modification strategies, respectively, for the performance of a dance. For all of the motions of these three components, we have found that certain fixed postures, which we call *keyposes*, tend to be preserved. Thus, this paper presents a method to create motions for robots at a certain music tempo, from human motion at an original music tempo, by using these keyposes. We have implemented these algorithms as an automatic process and validated their effectiveness by using a physical humanoid robot HRP-2. This robot succeeded in performing the Aizu-bandaisan dance, one of the Japanese traditional folk dances, 1.2 and 1.5 times faster than the tempo originally learned, while maintaining its physical constraints. Although we are not achieving a dancing robot which autonomously interacts with varying music tempos, we think that our method has a vital role in the dancing-to-music capability.**

***Index Terms*—Dancing robot, entertainment robot, temporal scaling.**

Manuscript received September 27, 2013; accepted January 10, 2014. Date of publication January 31, 2014; date of current version June 3, 2014. This paper was recommended for publication by Associate Editor N. Mansard and Editor A. Kheddar upon evaluation of the reviewers’ comments. This work was supported by JSPS KAKENHI under Grant 23240026.

T. Okamoto and K. Ikeuchi are with the University of Tokyo, Tokyo 1530041, Japan (e-mail: tokamoto@cvl.iis.u-tokyo.ac.jp; ki@cvl.iis.u-tokyo.ac.jp).
T. Shiratori is with Microsoft Research Asia, Microsoft Corporation, Beijing 100080, China (e-mail: takaakis@microsoft.com).
S. Kudoh is with the University of Electro-Communications, Tokyo 182-0021, Japan (e-mail: kudoh@is.uec.ac.jp).
S. Nakaoka is with the National Institute of Advanced Industrial Science and Technology, Tokyo 100-8921, Japan (e-mail: s.nakaoka@aist.go.jp).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.
Digital Object Identifier 10.1109/TRO.2014.2300212

I. Introduction

Entertainment is one of the promising application areas of humanoid robots. Entertainment applications such as dance or music, for which even a human performer needs special practice to perform, require special skills for robots, and as a result, push the horizon of robotics technologies. When successful, the performance of entertainment robots fascinates an audience.

Recently, development of such entertainment robots has accelerated as a showcase of robotics technologies [1]–[4]. Toyota introduced musician robots at a recent AICHI expo [3]. Kosuge *et al.* developed a dance partner robot for dance practice [5]. We have been developing a dancing robot based on the learning-from-observation (LFO) paradigm [6], [7].

Our dancing humanoid robot is based on *task models*, defined in the paradigm of LFO [8], [9]. The robot has the capability of observing human dance motion, analyzing such human dance motions using task models, and finally, generating imitation motions with balance maintenance.

This paper focuses on the human capability of dancing to music performances of varying tempos, and proposes an algorithm to realize this capability in a humanoid robot. The previous algorithm, proposed by Nakaoka *et al.* [6], realized a predefined static interaction with the environment; the resulting robot can only perform a dance at a prefixed tempo. This paper considers dynamic interactions for the robot to be able to modify its motion according to tempos of a given music piece. Especially, in faster tempos, physical robots will be required to omit insignificant details from the motion according to their motor limitations.

Interaction via musical expression has been investigated in the robotics field. Mizumoto *et al.* [10] developed a robot musician that mutually interacts with a human musician using musical instruments, such as a flute. Yoshii *et al.* and Murata *et al.* [11], [12] have described a humanoid robot that can sing and step to musical tempos using robot audition. Their work focuses on catching the target sound in a noisy environment including ego-noise, and differs from our goal on that point. Tanaka *et al.* [13], [14], Kozima *et al.* [15], and Kosuge *et al.* [5], [16] have developed robots that react to human motions via simple dance. Oliveria *et al.* [17]–[19], Grunberg *et al.* [20], Sun and Cheng [21], Gao *et al.* [22], and Xia *et al.* [23] have developed robot audition techniques and miniature robots that dance autonomously synchronizing to musical stimuli such as tempo, beat, and musical mood. However, most of their dance is designed for the robots by simple interpolation of physically realizable poses. They could skip the target beats coinciding with the motions according to the limitations. On the other hand, our target dance is designed for human dancers and contains more stylistic movements in trajectories. Additionally, a specific pose needs to be coincided with a specific beat like lyrics for a song. Although our life-size humanoid robots are designed in the image of humans, they are difficult to move at the same level as human joints. Therefore, to achieve our goal, an appropriate modification strategy to adapt to dance in arbitrary tempos is necessary; this differentiates our work from theirs.

Adjusting timings or time warping of motions is popular in the field of computer animation. Many researchers have developed motion editing tools for designers [24]–[30]. Although some studies take into account physical constraints, most of them are for creating natural motions of CG characters; the application to biped robots based on zero moment point (ZMP) will require additional cares even if it is

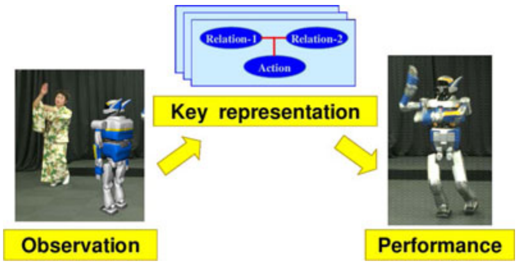

Fig. 1. Learning from observation.

possible. In our temporal scaling, the simultaneous pursuit of keeping the artistic/meaningful expression of the dance motions and satisfying physical limitations is the challenge. Our approach differs from theirs in terms of being based on the observation of actual human dancers.

This paper first analyzes human dancing, and extracts modification strategies to adjust motions to music tempos. Then, the paper presents motion modification strategies for the humanoid robot to dance to arbitrary musical tempos. Here, before modification, the algorithm assumes that dance motion at a certain music tempo is learned by using the task model representation in the Nakaoka system [6]. The validation of our proposed algorithm is conducted using a physical humanoid robot HRP-2.

This paper is organized as follows. First, Section II briefly reviews our dance generation paradigm based on LFO. Section III analyzes motion variations of human dance along music tempos and extracts motion variation strategies. Based on this analysis, Section IV proposes motion modification strategies for a humanoid robot. Section V validates our proposed system via experiments using a physical humanoid robot HRP-2, and Section VI concludes this paper.

II. Dance Motion Generation Based on the Learning-From-Observation Paradigm

We developed a paradigm referred to as LFO, which enables robots to learn how to perform various tasks from observing human performance [8], [9]. As shown in Fig. 1, the LFO generates robot actions through the following three steps.

1) A human dancer performs actions in front of the robot [see Fig. 1(left)].
2) The robot recognizes those demonstrated actions based on predefined abstract task models and constructs a series of task models [see Fig. 1(middle)].
3) The robot converts those recognized task models into robot physical actions [see Fig. 1(right)].

In general, performing the same action does not require mimicking the entire original action performed. It is difficult, if not impossible, to repeat the same trajectories to be mimicked, because the humanoid robot has different dimensions, proportion, and mass distribution from those of the human dancer. Instead, for this purpose, characteristics or important features of the actions are extracted and performed.

Essential and nonessential parts in each action are defined based on the knowledge of task domains. This top-down approach of designing domain-specific task models distinguishes our approach from other bottom-up learning approaches such as those developed by the Nakamura group [31]–[33] or the Kawato group [34]–[36]. Our top-down approach first defines task domains such as polyhedral-world operations [37], flexible rope handling [38], grasping motions [39], and whole-body motions [40]. Then, we define task models to represent all necessary essential actions based on the domain knowledge.

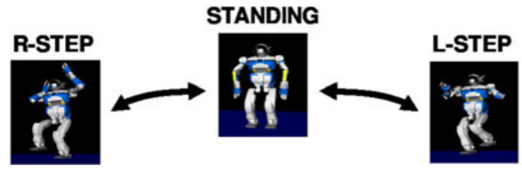

Fig. 2. Lower-body task models.

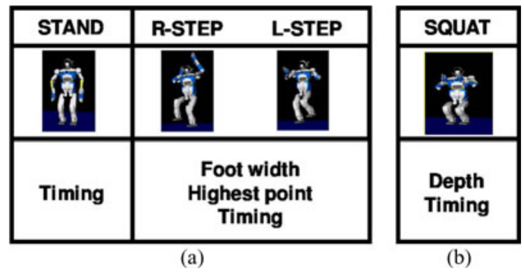

Fig. 3. Lower-body and middle-body skill parameters.

The LFO introduces abstract task models to represent essential parts in a sequence of actions. Each abstract task model describes the task, i.e., what to do. Each task model also contains skill parameters that explain how to do the specific task. Usually, recognizing tasks and extracting skill parameters are done automatically from input data.

Our dancing humanoid robot is based on this LFO paradigm. In particular, our method handles upper-body, middle-body, and lower-body motions separately by defining different types of task models. This separation is natural because although the whole body dance motion is conducted simultaneously, the lower, middle, and upper bodies have different roles or constraints to play in the performance of the dance. Considering how human dancers are taught to dance in lessons, those body motions are often taught separately. Thus, this separation is natural also for humans, and does not destroy the basic structure of the dance.

The constraint of lower-body motion is to stably support the whole body while performing a dance. The lower-body task model [6] is defined based on two foot–floor contacting conditions, STEP and STAND tasks (see Fig. 2). For the lower-body motion, a continuous foot motion is segmented and recognized using these defined task models. The skill parameters defined for each task model characterize the trajectory of the foot, such as the highest positions and length of a stride [see Fig. 3(a)]. The obtained skill parameters modify the default trajectory of STEP tasks, while stably supporting the whole body. Inverse-kinematics provides the joint angles of the robot's foot.

The aim of middle-body motion is twofold: expression of the dance and balance maintenance. For dance expression, the SQUAT task [6] is defined to lower the waist position. The skill parameters attributed to this SQUAT task include how deep the squat is, and the duration of each squat [see Fig. 3(b)]. The horizontal trajectory of the middle body is generated by computing the balance of the whole body [41]. Although some dance categories may include artistic expressions with horizontal movements of the middle body, those are out of our current scope to maintain dynamic stability.

The purpose of upper-body motion is to express the dance. We introduce keyposes [42] to represent such dance characteristics. A keypose is defined as a fixed posture of a dancer for the purpose of providing the viewers with expressions and meanings of a dance. Fig. 4 shows some of the keyposes in the Aizu-bandaisan dance, a Japanese traditional folk dance, depicted by a dance teacher. Some expert dancers indicate that these keyposes are the main points during the dance, and to mimic these keyposes is one of the important tasks in showing the beauty of the

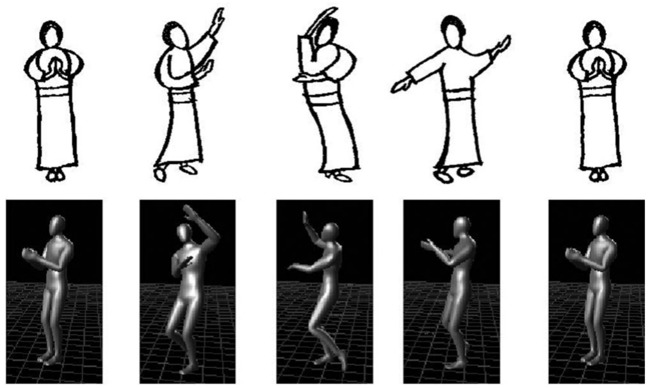

Fig. 4. Keyposes in Aizu-bandaisan dance. Upper row: keyposes depicted by a dance teacher. Bottom row: brief stop motions of dancers corresponding to music beats extracted from [42].

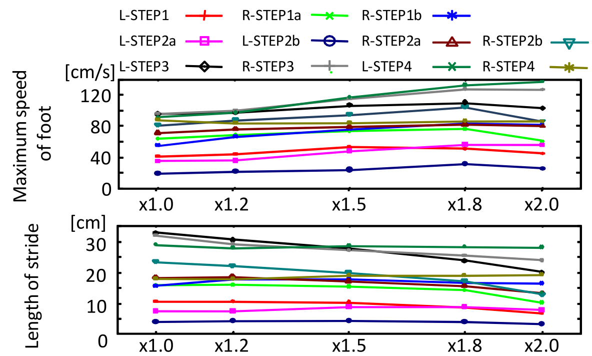

Fig. 5. Maximum speed of foot tip and length of a stride: Each marker represents the average maximum speed of foot tip and length of a stride in STEP tasks at each musical tempo.

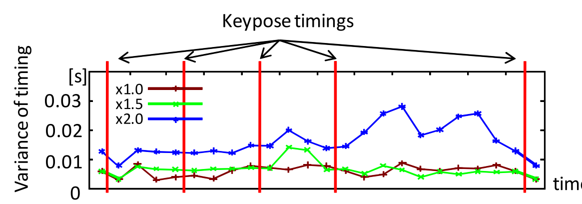

Fig. 6. Variance of start/end timings of each STEP task: Red, green, and blue markers provide variances at the original tempo, 1.5 times faster, and 2.0 times faster.

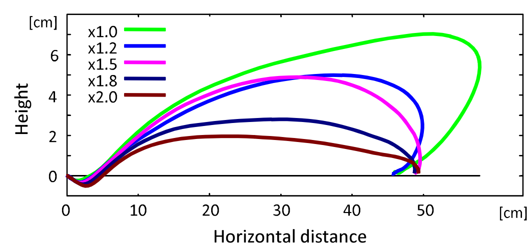

Fig. 7. Trajectories of a foot tip in a STEP task labeled as R-STEP4 at each tempo. The STEP task has a special trajectory like kicking up at the last part of each cycle of the dance.

dance. Thus, we define the performance of keyposes as the upper-body task models. We have developed a method of extracting keyposes from continuous dance motions by detecting brief stop motions of dancers corresponding to music beats [42]. Upper-body motions of a robot follow exactly its configurations at the keypose timing. The trajectory between keyposes, regarded as a skill parameter, is represented with a hierarchical B-spline.

By concatenating lower-body, middle-body, and upper-body motions using the Nakaoka system [6], we can obtain the entire robot motion. In our current implementation, human motions are captured using a motion capture system instead of an on-board camera system.

III. Analyzing Human Dance Performance With Varying Music Tempos

The LFO method provides a new way for a robot to learn how to dance at a fixed tempo. When the music tempo increases/decreases, the robot should dance more quickly/slowly. The research described in this paper aims to build such a capability.

The variation strategy needs to be consistent with that of humans. This is because we aim to design a robot that gives an impression similar to that of a human dancer. For this purpose, we first observed and analyzed how a human dancer modifies his or her motions along with the music tempos [43]–[45].

Dance performances by human dancers at several different music tempos were captured through an optical motion capture system, VICON. We sampled them at the original tempo, 1.2 times faster, and 1.5 times faster. We used the Aizu-bandaisan dance as a dance example. This dance consists of cyclic patterns, each of which takes about 10 s. Three dancers performed the 10–15 cycles of the dance to each music tempo.

As was done in the previous task model design, we assumed that the lower body, middle body, and upper body would have different modification strategies.

A. *Lower-Body Motions*

We observed start and end timing, length of a stride, and the maximum speed of foot tips as well as the trajectories for STEP tasks, as was done in our preliminary experiments reported in [43].

The Aizu-bandaisan dance consists of cycles of a sequence of tasks. In Fig. 5(top), we extracted 11 STEP tasks for one cycle. Here, R-STEPn and L-STEPn denote the nth right and left foot steps, respectively. From observation, we learned that the maximum speed of the foot tip does not vary as much as that of the musical tempo, probably because of the physical limitation of the dancer. Thus, the following discussion focuses on timing, length, and trajectories.

1) Timing: Fig. 6 shows the variance of start and end timings of step motions depending on music tempos. Lines of different colors in the graph represent different tempos. Here, we have normalized the horizontal axis so that one cycle of dance is always depicted from 0 to 1, independent of the music tempos. As can be seen in the graph, when the tempo of music increases, the variance of nonkeypose steps becomes larger, while keypose steps have lower variance. We take it that dancers' physical capabilities required them to change the durations of STEP tasks heuristic in faster tempos while preserving important timings. Thus, we can conclude the following.

L-1 Timing of a STEP near a keypose will be maintained.

L-2 Timing of a STEP far from any keypose will be adjusted when necessary to accommodate music tempos.

2) Length of a Stride: Fig. 5(bottom) shows how the stride varies with the music tempo. The length of a stride is maintained even though the tempo increases. We have also observed a couple of examples of breakdown at a higher tempo. Thus, we can summarize the following.

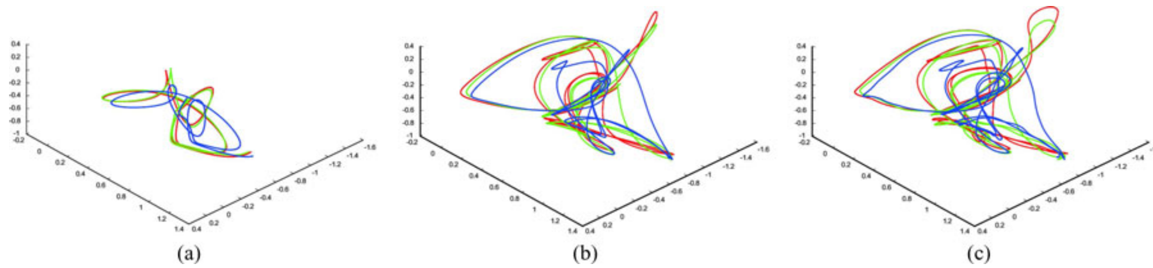

Fig. 8. Comparison of mean joint angle trajectories of the left shoulder in the logarithmic space of a quaternion. Lines of different colors represent different tempos; the original musical tempo (red), 1.2 times faster tempo (green), and 1.5 times faster tempo (blue). (a) Mean motion using a single-layer B-spline, (b) mean motion using a three-layer hierarchical B-spline, and (c) mean motion using a five-layer hierarchical B-spline. Variation of trajectory according to tempos in (c) is greater than that in (a); higher order motions are omitted preferentially with increased music tempo.

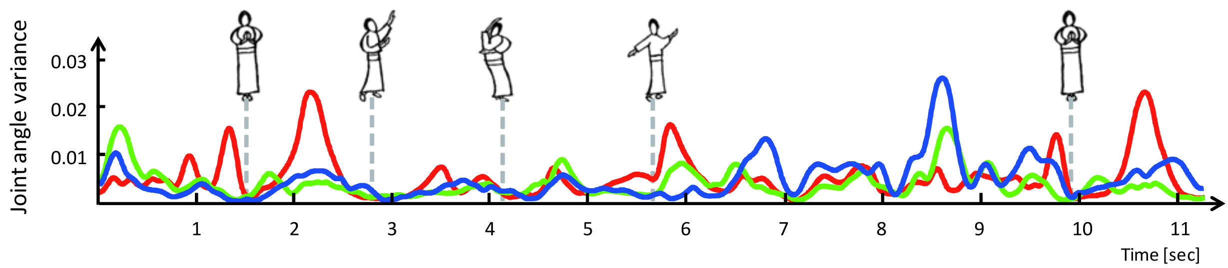

Fig. 9. Comparison of variance sequences of joint angle trajectories of the left shoulder. Lines of different colors represent different tempos; variance sequences at the original musical tempo (red), 1.2 times faster tempo (green), and 1.5 times faster tempo (blue). Those sequences are temporally normalized for comparison. Postures corresponding to the common local minimum of variances are depicted in the top row. Variance sequences for each speed tend to become a local minimum at keyposes.

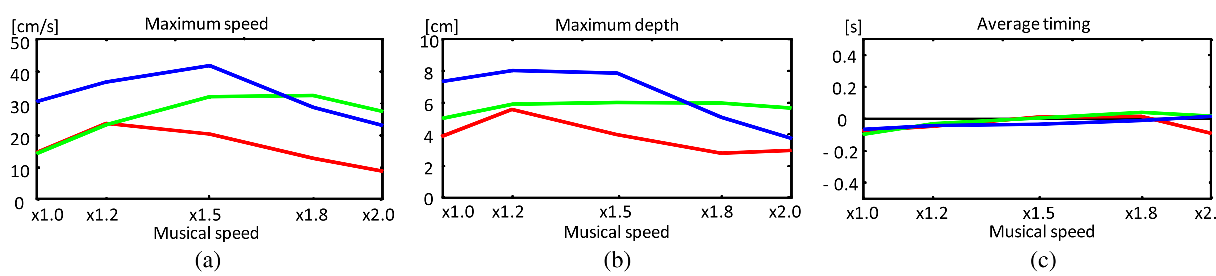

Fig. 10. Maximum speed, maximum depth, and average timing of SQUAT task with varying music tempos: Red, green, and blue lines represent three dancers, respectively.

L-3 The length of a stride will be maintained as much as possible up to a certain threshold. Over this threshold, it will be reduced accordingly.

3) Trajectories of a Foot Tip: Fig. 7 shows the trajectories of a foot tip in STEP tasks at each tempo. The STEP task has a special trajectory like kicking up at the last part of each cycle of the dance. The trajectories become smaller when the music tempo increases.

L-4 Trajectories of the foot tip become compact with increased music tempo.

B. Upper-Body Motion

We decomposed a dancer's motion using the hierarchical B-spline technique [44]–[46]. At the first layer, by using a certain number of knot point sets based on the original music tempo, the dancer's motion was represented by using the B-spline as shown in Fig. 8(a). Then, the difference between the original motion and the resulting B-spline is further represented by using a B-spline with a finer interval of knots as shown in Fig. 8(b). This process is repeated iteratively as shown in Fig. 8(c). As we expected, when the music tempo becomes faster, the higher order motion is omitted.

For a different aspect, we plotted how the variance of motion coincides with the music tempo, as shown in Fig. 9. The bottom row shows the variance sequences of a joint angle at various musical tempos, and the top row shows a sequence of corresponding postures to the common local minimum. This figure shows that the local minimum of variance occurs at certain musical points, and in fact, those postures at those timings correspond to the keyposes defined by a human dancer.

We can summarize our findings as follows.

U-1 Keypose timings and postures will be preserved, even if the musical tempo becomes faster.

U-2 High-frequency components of motion will decrease when the musical tempo becomes faster.

C. Middle-Body Motion

For SQUAT tasks, we observe timing to reach the maximum depth and the maximum speed of the waist. For the timing, as shown in

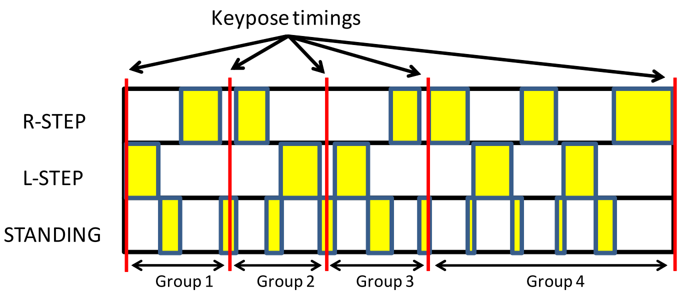

Fig. 11. STEP and STAND tasks grouped by keypose timings.

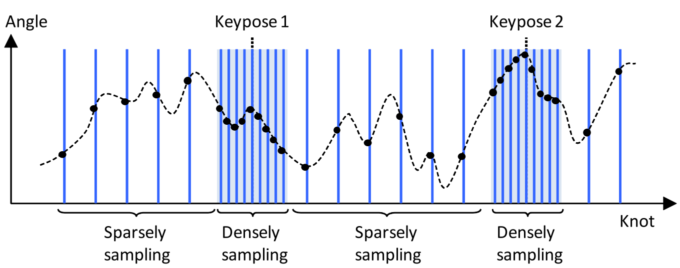

Fig. 12. Sampling method to consider keypose information for hierarchical motion decomposition. Vertical lines represent sampled time instants, and a dashed curve represents ground truth of a continuous joint angle trajectory. Data sampled by our method (black dots) are used.

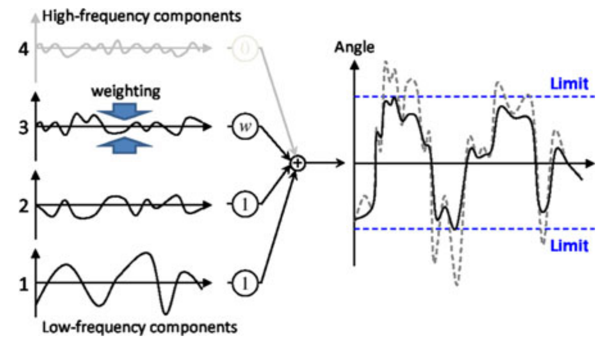

Fig. 13. Skill parameter adjustment for temporal scaling of upper-body motion. This adjustment process gradually decreases the weighting factors from the finest layer of the hierarchical B-spline. Both speed limitations and angular limitations are considered simultaneously in this process.

Fig. 10(c), we find that even though the music tempo increases as depicted in the horizontal axis, the average timing does not change. The dancers try to maintain the SQUAT timing as much as possible. In the Aizu-bandaisan dance, this SQUAT corresponds to the keypose. The speed of the SQUAT and the depth of the SQUAT are depicted in Fig. 10(a) and (b), respectively. As shown in Fig. 10(a) as the music tempo increases, the speed of the SQUAT increases and the depth of the SQUAT is maintained. However, beyond a certain point, when it becomes difficult to increase the speed, the depth of the SQUAT gradually decreases.

From observation, we found the following characteristics.

M-1 The timing of the SQUAT, which usually occurs at the keypose, is maintained independent of the music tempo. There is no difference in timing, even when the music speed increases to twice the original.

M-2 Dancers try to maintain the depth of the SQUAT by increasing the speed of the waist up to a certain music tempo. However, beyond this threshold tempo, the dancers accommodate the faster music tempos by reducing the depth.

IV. Keypose-Based Integration

This section presents modification strategies of lower-, middle-, and upper-body motions of a humanoid robot to accommodate various music speeds. The assumption before we begin this motion adjustment is that the human dance motions at a standard music tempo have been learned as a task sequence based on the LFO method [6]. We present a modification of trajectories of these motions as music tempos are changed.

When a music tempo becomes slower than the standard tempo, modifying the trajectories is relatively easy; we simply make each joint rotate more slowly by adjusting skill parameters of start/end timings and reconstructing whole-body motions based on the skill parameters. When the music tempo becomes faster than the standard tempo, a robot needs to make joints rotate faster in the same way. The payload of motors increases and sometimes may exceed the limit of the motor. In order to avoid this situation, we derive modification strategies based on the observations in the previous section.

From the observation results in the previous section, it would appear that the keypose is an essential factor for the dance performance as was found in L-1, M-1, and U-1. As stated earlier, a keypose is defined as a fixed posture of a dancer for the purpose of providing the viewers with expression and meanings of the dance. Other Japanese dances, such as Nou and Kabuki, also have keyposes that are often referred to as *Kime*, *Tome*, or *Mie*. In these traditional dances, dance masters regard it as very important to represent these keyposes in appropriate timings. A dance performance with sophisticated keyposes is considered a skillful performance. The dancers tend to keep keypose postures in the appropriate relative timings in the cycle as much as possible, even if they have to attenuate the motions to follow a faster tempo of music. Thus, we use these keyposes as anchor points when synchronizing the lower-, middle-, and upper-body motions to generate whole-body motions on a humanoid robot.

A. Lower-Body Motions

Proportional temporal shrinkage of performance durations along with the music tempo is applied to the task sequence of lower-body motions. This operation is based on the observation of L-1. As a result, all the lower-body motions corresponding to keyposes occurred at the appropriate music timings. However, such shrinkage causes an overload of joint motors. Avoidance of such overload is the issue in this section.

Tasks between two adjacent keypose times are considered as a group. Such a group consists of several STEP and STAND tasks as shown in Fig. 11; the start of the first STAND task and end timing of the last STAND task within the group are fixed so as to maintain the keypose timings. For all the STEP tasks in a group, the system increases the speeds of the joint motors to achieve shortened execution periods. First, the speeds of the joint motors are computed using the inverse kinematics method at each newly created start and ending time. Then, those speeds are examined to determine whether they exceed the motor capability limit or not.

From among those STEP tasks in the group, the duration of the task, with the maximally exceeding speed limitation, is extended so as to satisfy the motor limitation. This is achieved by first reducing the period of the following STAND task. If this is not enough, the following STEP and STAND tasks with capacity allowance in the group are considered as candidates for duration reduction. This operation is conducted iteratively along the descending order of the exceeding tasks within the group. This process is based on the observation of L-2.

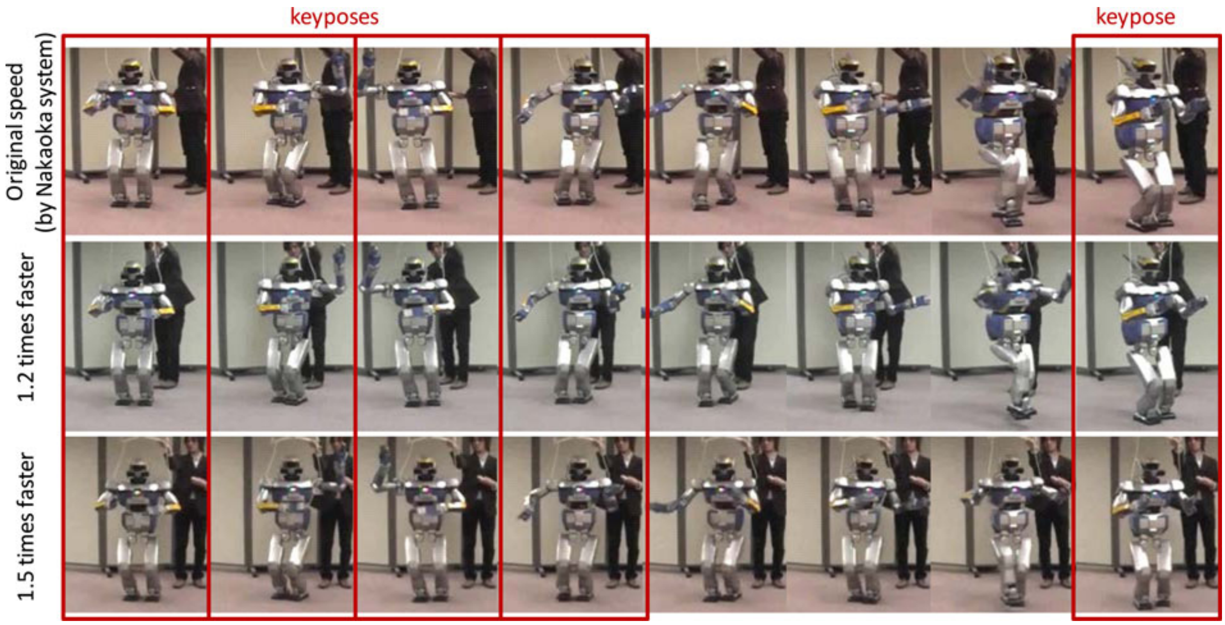

Fig. 14. Experiment of whole-body dance motion for the Aizu-bandaisan dance with a physical humanoid robot HRP-2. Top row: reference of the sequence of postures for the original tempo of music generated using the Nakaoka system [6]. Middle and bottom rows: the sequence of postures for a tempo 1.2 and 1.5 times faster than the original.

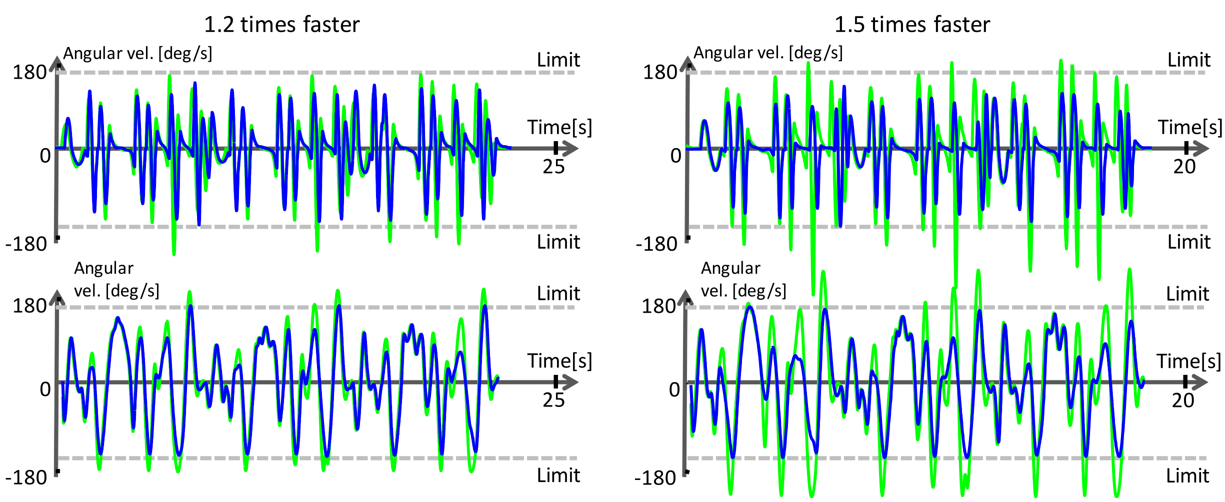

Fig. 15. Velocity sequences of the right knee angle (upper row) and left shoulder pitch angle (lower row) generated for 1.2 (left row) and 1.5 (right row) times faster tempos than the original tempo. The green lines represent the corresponding joint angular velocities of the motion generated by simple temporal scaling; proportional temporal shrinkages are applied to task sequences and motion generation is done using the Nakaoka system [6]. The blue lines represent the joint angular velocities of the motion generated using our proposed system. The gray lines represent the upper/lower limits of the velocity. Motions generated using our method satisfy the limitations and are feasible for the physical humanoid robot.

If this period adjustment is not enough, the strides and trajectories of all the exceeding tasks are reduced iteratively so as to satisfy the motor limitation. This stride reduction satisfies the limitation because eventually the strides of the robot will become zero, and the robot is simply standing with upper-body motions only. This process is based on the observations of L-3 and L-4.

B. Upper-Body Motions

Joint trajectories are represented by hierarchical B-splines. Here, in order to preserve posture information of keyposes in the following operations, we sample joint trajectories around keyposes more densely than those in other parts. Our sampling method is illustrated in Fig. 12. Then, proportional temporal scaling, fitted to the new music tempo, is applied to each layer of B-spline representation so that the resulting motion is consistent with the new music tempo. This operation satisfies the observation U-1; whatever the music tempo is, keyposes occur at particular music timings. However, the resulting motion may exceed motor capability. Thus, this excess joint speed is amended in the hierarchical manner.

We examine the motor capacity by examining the layers of the hierarchical B-spline iteratively from higher to lower layers. First, we recalculate the motor load by reducing the amplitude of the highest level of the hierarchical B-spline. If we cannot achieve the motor load within the capacity by setting the amplitude of this highest level to zero, then we iteratively repeat the same operation in the next layer of the hierarchical B-spline. This adjustment is illustrated in Fig. 13.

C. Middle-Body Motions

For vertical movements, abstracted as SQUAT tasks, we found that the most important factor is to maintain the starting and ending periods (M-1) while adjusting the depth (M-2). Thus, the depth of SQUAT is gradually reduced iteratively up to a certain value to satisfy the motor limitation. If the reduction of the depth to the limit does not provide the allowance of motor capability, then we expand the period of the SQUAT task. If this is still not sufficient, we eliminate the SQUAT task from the task sequence. These actions are not based on the observation, but they are inevitable in order to avoid exceeding the payload limit of the motors.

For horizontal movements, the desired horizontal trajectories of the waist are calculated using a ZMP compensation filter [41]. This calculation is done considering upper- and lower-body motions.

V. Experiments

In our experiments, whole-body motions of the Aizu-bandaisan dance were generated to increase tempos of music, and HRP-2 [47] was selected as a physical robot platform. Target tempos of music were 1.2 times faster and 1.5 times faster than the original tempo. To achieve our goal, temporal scaling up to 1.2 times faster will be enough because musical tempos normally do not fluctuate so much. In this section, we tried out 1.5 times faster tempos as a cushion.

Given a dance motion and a music tempo as inputs, our learning system automatically detects each task in the motion and extracts a task sequence that is based on the LFO method [6]. The modification algorithm changes the speed of the dance motion by adjusting skill parameters of each task iteratively, according to the excess of limitations. Self-collision avoidance [6] and a ZMP compensation filter [41] automatically correct the generated motions in the iterative process. The process is continued as long as a motion checker detects the exceeding of physical limitations.

Each result is shown in the middle and bottom of Fig. 14. The top figures are the sequences of Aizu-bandaisan dance at the original tempo generated by the Nakaoka system [6] as the reference. Poses that are surrounded in red boxes are keyposes in each sequence. The robot expressed the keyposes using the whole body in appropriate timings in the sequence, and provided viewers with an artistic dance pattern in which upper body and leg motions were fully harmonious. Although the dance motions are modified separately according to musical tempos using different strategies, differences between keyposes at each music tempo are difficult to find. In both music tempos, our proposed system generated feasible motions within the joint limitations, and the HRP-2 could perform without falling down.

The joint angular velocities in the aforementioned two experiments are shown in Fig. 15. Velocity sequences of the right knee angle (upper row) and left shoulder pitch angle (lower row) that are generated for 1.2 (left row) and 1.5 (right row) times faster tempos than the original one are shown. The blue lines represent the joint angular velocities of the motions generated using the proposed system. The green lines represent the motions generated from a simply scaled task sequence. The gray lines represent the upper/lower limits of the velocity. Motions generated using our method satisfy the limitations and are feasible for the physical humanoid robot.

To demonstrate stability, the robot performed each dance motion ten times in a row. Because physical humanoid robots have some noise in motor control, we equipped a security crane in case of failure. Our dancing robot successfully performed without a failure and the security crane was not used.

VI. Conclusion

This paper presented an algorithm to generate whole-body motions feasible for physical humanoid robots by the integration of individual temporal scaling algorithms for lower-body, middle-body, and upper-body motions. These algorithms were obtained from modeling the human capability of dancing to various musical tempos. We found that keyposes are essential in the dance and can be employed as anchor points for integration of whole-body dance motions. Our integrated system modifies skill parameters for lower-, middle-, and upper-body motions so that they can maintain keyposes and provide whole-body motions. We validated our algorithm via experiments using the physical humanoid robot HRP-2.

For future work, we aim to develop a dancing robot which autonomously interacts with varying music tempos. To achieve this, integration with the state-of-art technology of music analysis, and connection of dance motions continuously generated for different music tempos, will be our next focus.

Additionally, analysis of other dances is vital to validate the generality of our method. Considering the fact that there are previous studies which analyze other dances and suggest the importance of keyposes, we assume that our method can be applied to those other dances and thus be further validated.

Acknowledgment

The authors would like to thank KAWADA INDUSTRIES, INC. The experiments using a physical humanoid robot in this paper were supported by the company.

References

[1] Y. Kuroki, M. Fujita, T. Ishida, K. Nagasaka, and J. Yamaguchi, "A small biped entertainment robot exploring attractive applications," in *Proc. IEEE Int. Conf. Robotics Autom.*, Sep. 2003, vol. 1, pp. 471–476.

[2] Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki, and K. Fujimura, "The intelligent ASIMO: System overview and integration," in *Proc. Int. Conf. Intell. Robots Syst.*, 2002, vol. 3, pp. 2478–2483.

[3] A. Goto, "Musical instrument playing robot: Toyota partner robots for the 2005 world exposition, Aichi, Japan," in *Proc. IEEE Int. Conf. Robotics Autom. Workshop Art Robots*, Nov. 2007, pp. 41–45.

[4] K. Kaneko, F. Kanehiro, M. Morisawa, K. Miura, S. Nakaoka, and S. Kajita, "Cybernetic human hrp-4c," in *Proc. Int. Conf. Humanoid Robots*, Dec. 2009, pp. 7–14.

[5] K. Kosuge, T. Hayashi, Y. Hirata, and R. Tobiyama, "Dance partner robot—Ms DancerR," in *Proc. Int. Conf. Intell. Robots Syst*, Oct. 2003, vol. 3, pp. 3459–3464.

[6] S. Nakaoka, A. Nakazawa, F. Kanehiro, K. Kaneko, M. Morisawa, H. Hirukawa, and K. Ikeuchi, "Learning from observation paradigm: Leg task models for enabling a biped humanoid robot to imitate human dances," *Int. J. Robot. Res.*, vol. 26, no. 8, pp. 829–844, Aug. 2007.

[7] K. Ikeuchi, T. Shiratori, S. Kudoh, H. Hirukawa, S. Nakaoka, and F. Kanehiro, "Robots that learn to dance from observation," *Intell. Syst.*, vol. 23, no. 2, pp. 74–76, Mar./Apr. 2008.

[8] K. Ikeuchi and T. Suehiro, "Toward an assembly plan from observation—Part I: Task recognition with polyhedral objects," *IEEE Trans. Robot. Autom.*, vol. 10, no. 3, pp. 368–385, Jun. 1994.

[9] T. Suehiro and K. Ikeuchi, "Towards an assembly plan from observation—Part II: Correction of motion parameters based on fact contact constraints," in *Proc. Int. Conf. Intell. Robots Syst.*, Jul. 1992, vol. 3, pp. 2095–2102.

[10] T. Mizumoto, A. Lim, and T. Otsuka, "Integration of flutist gesture recognition and beat tracking for human–robot ensemble," in *Proc. Int. Conf. Intell. Robots Syst. Workshop Robots Music Exp.*, Oct. 2010, pp. 159–171.

[11] K. Yoshii, K. Nakadai, T. Torii, Y. Hasegawa, H. Tsujino, K. Komatani, T. Ogata, and H. G. Okuno, "A biped robot that keeps steps in time with

musical beats while listening to music with its own ears," in *Proc. Int. Conf. Intell. Robots Syst.*, Oct. 2007, pp. 1743–1750.

[12] K. Murata, K. Nakadai, K. Yoshii, R. Takeda, T. Torii, H. G. Okuno, Y. Hasegawa, and H. Tsujino, "A robot uses its own microphone to synchronize its steps to musical beats while scatting and singing," in *Proc. Int. Conf. Intell. Robots Syst.*, Sep. 2008, pp. 2459–2464.

[13] F. Tanaka and H. Suzuki, "Dance interaction with QRIO: A case study for non-boring interaction by using an entertainment ensemble model," in *Proc. Int. Workshop Robot Human Interactive Commun.*, Sep. 2004, pp. 419–424.

[14] F. Tanaka, B. Fortenberry, K. Aisaka, and J. Movellan, "Plans for developing real-time dance interaction between QRIO and toddlers in a classroom environment," in *Proc. Int. Conf. Dev. Learn.*, Jul. 2005, pp. 142–147.

[15] H. Kozima, M. P. Michalowski, and C. Nakagawa, "Keepon: A playful robot for research, therapy, and entertainment," *Int J. Soc. Robot.*, vol. 1, no. 1, pp. 3–18, Jan. 2009.

[16] K. Kosuge, T. Takeda, Y. Hirata, M. Endo, M. Nomura, K. Sakai, M. Koizumi, and T. Oconogi, "Partner ballroom dance robot—PBDR," *SICE J. Control, Meas., Syst. Integr.*, vol. 1, no. 1, pp. 74–80, 2011.

[17] J. Oliveira, F. Gouyon, and L. P. Reis, "Towards an interactive framework for robot dancing applications," in *Proc. Int. Conf. Digital Arts*, 2008, pp. 52–59.

[18] C. B. Santiago, J. L. Oliveira, L. P. Reis, and A. Sousa, "Autonomous robot dancing synchronized to musical rhythmic stimuli," in *Proc. Inf. Syst. Technol.*, Jun. 2011, pp. 1–6.

[19] C. B. antiago, J. L. Oliveira, L. P. Reis, A. Sousa, and F. Gouyon, "Overcoming motor-rate limitations in online synchronized robot dancing," *Int. J. Comput. Intell. Syst.*, vol. 5, no. 4, pp. 700–713, Aug. 2012.

[20] D. Grunberg, R. Ellenberg, Y. Kim, and P. Oh, "Creating an autonomous dancing robot," in *Proc. Int. Conf. Hybrid Inform. Technol.*, 2009, pp. 221–227.

[21] J. Sun and J. Cheng, "A robot dance system based on real-time beat prediction," in *Proc. Autom. Control Artif. Intell.*, Mar. 2012, pp. 287–291.

[22] Q. Gao, B. Zhang, X. Wu, Z. Cheng, Y. Ou, and Y. Xu, "A music dancing robot based on beat tracking of musical signal," in *Proc. Int. Conf. Robot. Biomimetics*, Dec. 2010, pp. 1536–1541.

[23] G. Xia, J. tay, R. Dannenberg, and M. Veloso, "Autonomous robot dancing driven by beats and emotions of music," in *Proc. Auton. Agents Multiagent Syst.*, 2012, vol. 1, pp. 205–212.

[24] A. Bruderlin and L. Williams, "Motion signal processing," in *Proc. ACM SIGGRAPH*, 1995, pp. 97–104.

[25] G. Alankus, A. A. Bayazit, and O. B. Bayazit, "Automated motion synthesis for dancing characters," *Comput. Animat. Virtual Worlds*, vol. 16, no. 3–4, pp. 259–271, Sep. 2005.

[26] H. C. Lee and I. K. Lee, "Automatic synchronization of background music and motion in computer animation," *Comput. Graph. Forum*, vol. 24, no. 3, pp. 353–361, Sep. 2005.

[27] E. Hsu, K. Pulli, and J. Popovic, "Style translation for human motion," in *Proc. ACM SIGGRAPH*, 2005, vol. 24, pp. 1082–1089.

[28] R. Heck, L. Kovar, and M. Gleicher, "Splicing upper-body actions with locomotion," *Comput. Graph. Forum*, vol. 25, no. 3, pp. 459–466, Sep. 2006.

[29] J. McCann, N. S. Pollard, and S. Srinivasa, "Physics-based motion retiming," in *Proc. ACM SIGGRAPH*, 2006, pp. 205–214.

[30] T. Shiratori, A. Nakazawa, and K. Ikeuchi, "Dancing-to-music character animation," *Comput. Graph. Forum*, vol. 25, no. 3, pp. 449–458, Sep. 2006.

[31] T. Inamura, H. Tanie, and Y. Nakamura, "Keyframe compression and decompression for time series data based on the continuous hidden Markov model," in *Proc. Int. Conf. Intell. Robots Syst.*, Oct. 2003, vol. 2, pp. 1487–1492.

[32] T. Inamura, H. Tanie, I. Toshima, and Y. Nakamura, "An approach from motion generation/recognition to intelligence based on Mimesis principle," in *Proc. Adapt. Motion Animals Mach.*, Mar. 2003, pp. SaA-II-1.

[33] T. Inamura, H. Tanie, and Y. Nakamura, "From stochastic motion generation and recognition to geometric symbol development and manipulation," in *Proc. Int. Conf. Humanoid Robots*, 2003, pp. 1b–02.

[34] H. Miyamoto and M. Kawato, "A tennis serve and upswing learning robot based on bi-directional theory," *Neural Netw.*, vol. 11, no. 7–8, pp. 1331–1344, Oct. 1998.

[35] S. Schaal, "Is imitation learning the route to humanoid robots?" *Trends Cognit. Sci.*, vol. 3, no. 6, pp. 233–242, Jun. 1999.

[36] A. Ude, "Filtering in a unit quaternion space for model-based object tracking," *Robot. Auton. Syst.*, vol. 28, no. 2–3, pp. 163–172, Aug. 1999.

[37] M. Kawade, T. Suehiro, and K. Ikeuchi, "Assembly task recognition with planar, curved and mechanical contacts," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 1993, vol. 2, pp. 688–694.

[38] J. Takamatsu, T. Morita, K. Ogawara, H. Kimura, and K. Ikeuchi, "Representation for knot-tying tasks," *IEEE Trans. Robot.*, vol. 22, no. 1, pp. 65–78, Feb. 2006.

[39] S. B. Kang and K. Ikeuchi, "Toward automatic robot instruction from perception-mapping human grasps to manipulator grasps," *IEEE Trans. Robot. Autom.*, vol. 13, no. 1, pp. 81–95, Feb. 1997.

[40] K. Ogawara, J. Takamatsu, H. Kimura, and K. Ikeuchi, "Extraction of essential interactions through multiple observations of human demonstrations," *IEEE Trans. Ind. Electron.*, vol. 50, no. 4, pp. 667–675, Aug. 2003.

[41] K. Nishiwaki, S. kagami, Y. Kuniyoshi, M. Inaba, and H. Inoue, "Online generation of humanoid walking motion based on a fast generation method of motion pattern that follows desired ZMP," in *Proc. Int. Conf. Intell. Robots Syst.*, Oct. 2002, vol. 3, pp. 2684–2689.

[42] T. Shiratori, A. Nakazawa, and K. Ikeuchi, "Detecting dance motion structure through music analysis," in *Proc. Int. Conf. Autom. Face Gesture Recog.*, May 2004, pp. 857–862.

[43] T. Okamoto, T. Shiratori, S. Kudoh, and K. Ikeuchi, "Temporal scaling of leg motion for sound feedback system of a dancing humanoid robot," in *Proc. Int. Conf. Intell. Robots Syst.*, Oct. 2010, pp. 2256–2263.

[44] T. Shiratori, S. Kudoh, S. Nakaoka, and K. Ikeuchi, "Temporal scaling of upper body motion for sound feedback system of a dancing humanoid robot," in *Proc. Int. Conf. Intell. Robots Syst.*, Oct. 2007, pp. 3251–3257.

[45] T. Shiratori and K. Ikeuchi, "Synthesis of dance performance based on analyses of human motion and music," *IEEE Trans. Comput. Vis. Image Media*, vol. 1, no. 1, pp. 34–47, Jun. 2008.

[46] J. Lee and S. Y. Shin, "A hierarchical approach to interactive motion editing for human-like figures," in *Proc. ACM SIGGRAPH*, 1999, pp. 39–48.

[47] K. Kaneko, F. Kanehiro, S. Kajita, H. Hirukawa, T. Kawasaki, M. Hirata, K. Akachi, and T. Isozumi, "Humanoid robot HRP-2," in *Proc. Int. Conf. Robot. Autom.*, 2004, vol. 2, pp. 1083–1090.