Temporal Scaling of Upper Body Motion for Sound Feedback System of a Dancing Humanoid Robot

Takaaki Shiratori† Shunsuke Kudoh† Shin'ichiro Nakaoka‡ Katsushi Ikeuchi†

***Abstract*— This paper proposes a method to model the modification of upper body motion of dance performance based on the speed of played music. When we observed structured dance motion performed at a normal music playback speed and motion performed at a faster music playback speed, we found that the detail of each motion is slightly different while the whole of the dance motion is similar in both cases. This phenomenon is derived from the fact that dancers omit the details and perform the essential part of the dance in order to follow the faster speed of the music. To clarify this phenomenon, we analyzed the motion differences in the frequency domain, and obtained two insights on the omission of motion details: (1) High frequency components are gradually attenuated depending on the musical speed, and (2) important stop motions are preserved even when high frequency components are attenuated. Based on these insights, we modeled our motion modification considering musical speed and joint limitations that a humanoid robot has. We show the effectiveness of our method via some applications for humanoid robot motion generation.**

I. INTRODUCTION

Synthesizing and reproducing human-like motion with CG characters and humanoid robots is currently an important topic in CG and robotic research. In particular, recent research has been conducted on applying human's improvisational aspects to motion generation for CG characters [1], [2] and humanoid robots [3], [4]. This research showed that people first receive external signals such as visual or audio information, then recognize essential information or feel some emotions from the obtained information, and finally perform movements.

Toward reproducing this ability, we are developing a *sound feedback system*, in which a humanoid robot mimics a human's *dancing-to-music* ability for entertainment. Synchronizing recorded human motion data with currently played music is an important part of the sound feedback system. In this paper, we propose a novel method to temporally scale upper body motion involved in dance performance for synchronization with music.

Acquiring motion capture data is very time consuming, and many researchers have attempted to efficiently synthesize human motion from a single motion sequence through such procedures as editing motion capture data by signal processing techniques [5], retargeting motion to new characters [6], and modifying human motion to make it funny [7]. However, there are no methods to generate temporally scaled human motion except McCann's method [8] that aimed at temporal scaling of jumping motion by considering physical laws. Their method does not work well for non-jumping motion.

†Takaaki Shiratori, Shunsuke Kudoh, and Katsushi Ikeuchi are with Institute of Industrial Science, The University of Tokyo, Tokyo, 153-8505, Japan {siratori, kudoh, ki}@cvl.iis.u-tokyo.ac.jp

‡Shin'ichiro Nakaoka is with National Institute of Advanced Industrial Science and Technology, Ibaraki, 305-8568, Japan s.nakaoka@aist.go.jp

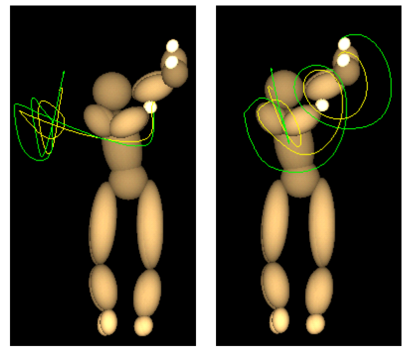


Fig. 1. Comparison of hand trajectory differences depending on music playback speed. The green and yellow curves represent the hand trajectories at a normal musical speed, and 1.3 times faster musical speed, respectively.

To achieve this goal, we first observe how dance motion is modified based on played musical speed, and then model the modification based on the acquired insights. When we observe structured dance motion performed by humans at normal music playback speeds versus motion performed using music that is 1.3 times faster, we find that the details of each motion sequence differ slightly, though the whole of the dance motion sequence is similar in both cases. An example of this type of motion modification, natural in humans, is shown in Fig. 1. This phenomenon is derived from the fact that dancers omit details of a dance, but retain its essential aspects if this is necessary to follow faster music. If we therefore observe motion differences in dances performed at different speeds in the frequency domain, we can obtain useful insights about the omission of motion detail. Based on these insights, we propose a new modeling method and develop some applications useful for the generation of humanoid robot motion.

This paper is organized as follows: Section II describes our observation method using the hierarchical B-spline technique. Section III describes the proposed algorithm for upper body motion generation based on our insights. Section IV shows experimental results, and Section V concludes this paper by mentioning possible future work.

II. OBSERVATION OF HUMAN DANCE MOTION

This section describes how to analyze human dance motion. We first calculate each joint angle using quaternion algebra, and then convert it into a 3D logarithmic space. We denote the 3D angle representation as a 3D vector $\mathbf{v}$ whose unit vector represents a rotation axis and whose norm represents half the joint rotation.

A. Motion Decomposition Using Hierarchical B-spline

Hierarchical B-spline consists of a series of B-spline curves with different knot spacings; higher layers of a hierarchical B-spline are based on finer knot spacing that can preserve the higher frequency components of the original sequence. This technique can interpolate and approximate scattered sparse data sets [9], or its hierarchical structure can easily and effectively retarget motion to new characters [10]. Each subject motion sequence has a different underlying musical rhythm, and a B-spline allows us to control frequency resolution by only setting its control points at desired temporal intervals. In our analysis, we consider musical rhythm for knot spacing, and we normalize temporal frames of motion sequences into the knot space with musical rhythm.

We use cubic B-spline in order to preserve continuity of motion acceleration. A cubic B-spline curve $\mathbf{f}_0(t)$ can be represented as

$$\mathbf{f}_0(t) = \sum_{i=0}^{3} B_i(t - \lfloor t \rfloor)\mathbf{Q}_{\lfloor t \rfloor + i - 1}, \tag{1}$$

where $\mathbf{Q}_t$ represents a control point at knot t, and $B_i(t)$ for $0 \leq t < 1$ represents a basis function of a cubic B-spline defined as

$$B_0(t) = \frac{1}{6}(1-t)^3, \tag{2}$$

$$B_1(t) = \frac{3t^3 - 6t^2 + 4}{6}, \tag{3}$$

$$B_2(t) = \frac{-3t^3 + 3t^2 + 3t + 1}{6}, \tag{4}$$

$$B_3(t) = \frac{1}{6}t^3. \tag{5}$$

A least-squares solution of Eq. (6) using pseudo inverse provides control point sets $\{\hat{\mathbf{Q}}_0, \cdots, \hat{\mathbf{Q}}_n\}$ that can roughly approximate an input joint angle trajectory:

$$\begin{pmatrix} \mathbf{v}(t_1) \\ \mathbf{v}(t_2) \\ \vdots \\ \mathbf{v}(t_m) \end{pmatrix} = N^{pos} \begin{pmatrix} \hat{\mathbf{Q}}_0 \\ \hat{\mathbf{Q}}_1 \\ \vdots \\ \hat{\mathbf{Q}}_n \end{pmatrix}, \tag{6}$$

where $\mathbf{v}(t)$ represents a joint angle calculated from an input motion sequence at knot t and N^{pos} represents a matrix whose elements are the basis function represented as

$$N^{pos}_{ij} = \begin{cases} B_{j+1-\lfloor t_i \rfloor}(t_i - \lfloor t_i \rfloor) & \text{if } j \leq t_i < j+1 \\ 0 & \text{otherwise} \end{cases}. \tag{7}$$

An overview of hierarchical B-spline construction is illustrated in Fig. 2. The coarsest layer of hierarchical B-spline, $\mathbf{f}_0$, cannot contain the high frequency components of an input joint angle trajectory. So the difference between the input

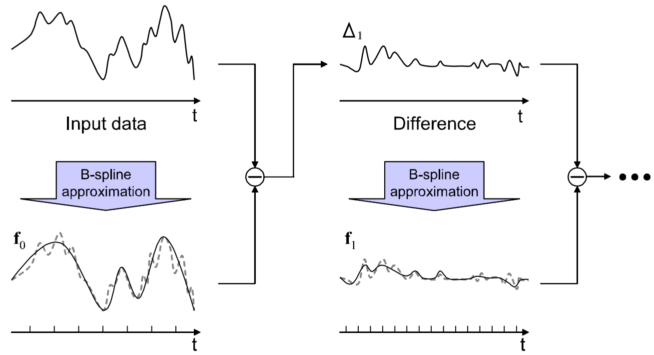


Fig. 2. Illustration of hierarchical B-spline construction.

joint trajectory and the coarsest layer, $\Delta_1(t) = \mathbf{v}(t) - \mathbf{f}_0(t)$, is calculated and the n approximated with a B-spline curve $\mathbf{f}_1(t)$ by solving Eq. (6). The knot spacing of $\mathbf{f}_1$ is half the $\mathbf{f}_0$ knot spacing. A hierarchical B-spline is constructed by doing the same process iteratively until reaching the specified finest layer.

B. Observation Using Hierarchical B-spline

Using an optical motion capture system, we captured the *Aizu-bandaisan* dance, a classical Japanese folk dance, at three varying musical speeds for observation: the original speed, 1.2 times faster speed, and 1.5 times faster speed. Motion sequences at each speed were captured five times in order to investigate motion variance, so a total of 15 datasets were considered in this experiment. We set the knot spacing to the musical rhythm, and then applied a hierarchical B-spline decomposition technique. We used up to five layers in our motion decomposition and observed the mean and variance of each reconstructed motion. Our choice of five layers was arbitrary, but it was empirically found to be enough to reconstruct high-frequency components of human motion.

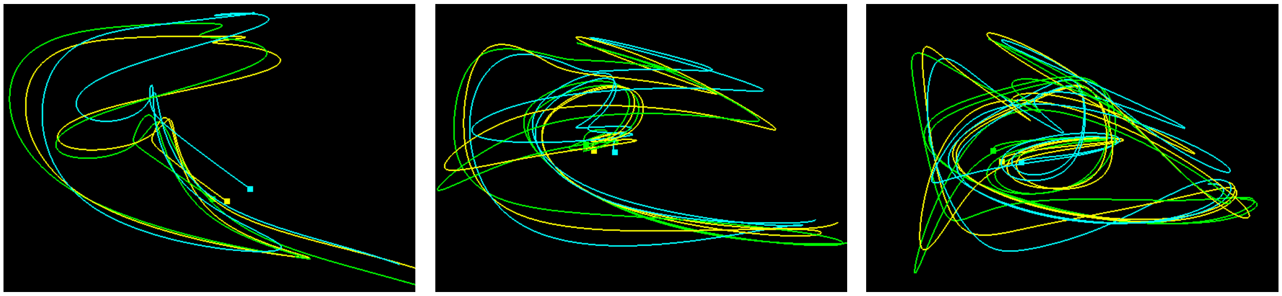
The mean $\bar{\mathbf{v}}$ and the variance d of the j-th joint angle are calculated as

$$\bar{\mathbf{v}}_j(t) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{v}^i_j(t) \tag{8}$$

$$d_j(t) = \frac{1}{N-1}\sum_{i=1}^{N}\left(1 - \mathbf{h}\big(\mathbf{v}^i_j(t)\big) \cdot \mathbf{h}\big(\bar{\mathbf{v}}_j(t)\big)\right), \tag{9}$$

where $\mathbf{v}^i_j$ represents the j-th joint angle of the i-th motion sequence, N represents the number of input motion sequences (in our case, $N = 5$ for a given music playback speed), and $\mathbf{h}$ converts an input 3D vector to 4D vectors in a homogeneous coordinate as $\mathbf{h}(\mathbf{a}) \equiv (\mathbf{a}^T, 1)^T / |\mathbf{a}^T, 1|$. The variance metric accounts for both magnitude and direction differences [11].

Fig. 3 shows mean the joint angle trajectories of the left shoulder; (a) mean motion using a single-layer B-spline, (b) mean motion using a two-layer hierarchical B-spline, and (c) mean motion using a three-layer hierarchical B-spline. The green, yellow, and light blue lines represent the mean joint angle trajectories at the original musical speed, 1.2 times



(a) (b) (c)

Fig. 3. Comparison of mean joint angle trajectories at the original musical speed (green), 1.2 times faster speed (yellow), and 1.5 times faster speed (light blue). (a) mean motion using a single-layer B-spline, (b) mean motion using a two-layer hierarchical B-spline, and (c) mean motion using a three-layer hierarchical B-spline. These trajectories are in the logarithmic space of a quaternion.

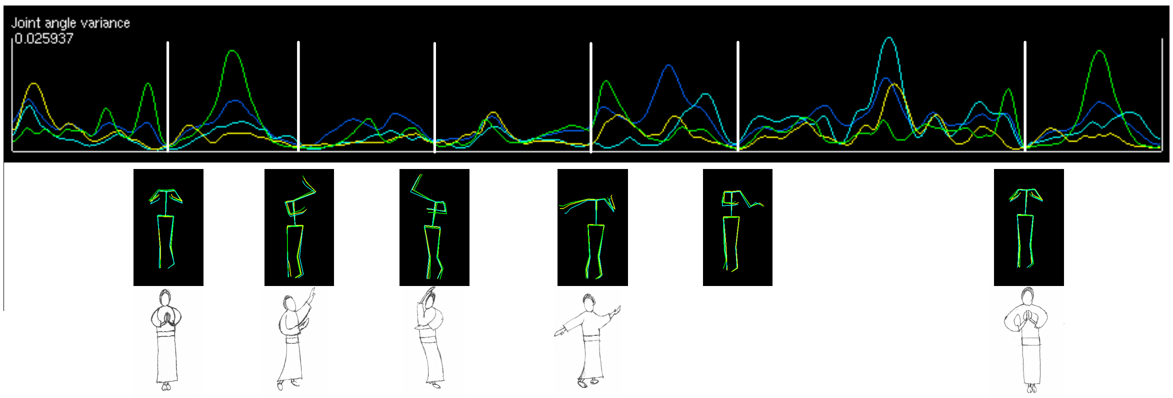


Fig. 4. Comparison of variance sequences. Top row: variance sequences at the original musical speed (green), 1.2 times faster speed (yellow), and 1.5 times faster speed (light blue), and the blue line represents the variance calculated from all the motion sequences. Middle row: postures corresponding to the common local minima. Bottom row: important stop motions detected by dance masters.

faster speed, and 1.5 times faster speed, respectively. With regard to motion reconstructed from a single-layer B-spline (Fig. 3 (a)), the motion at the 1.2 times faster musical speed is quite similar to the motion at the normal musical speed. The motion at the 1.5 times faster musical speed is also similar to the motion at the normal speed, but their details such as curvature differ slightly from each other. With regard to motion reconstructed from a two-layer hierarchical B-spline (Fig. 3 (b)), the shape of the joint angle trajectory at the normal musical speed differs slightly from that of the 1.2 times faster musical speed, especially in the trajectory's sharpest curves. On the other hand, the shape of the joint angle trajectory at the 1.5 times faster musical speed appears to be a smoothed version of the normal music playback speed trajectory. With regard to motion reconstructed from a three-layer hierarchical B-spline (Fig. 3 (c)), the differences among the joint angle trajectories become more noticeable. The shape of the joint angle trajectory at the 1.5 times faster musical speed is a much smoothed version of the trajectory at the normal musical speed, whereas the shape at the 1.2 times faster musical speed is just a slightly smoothed version of the trajectory at the normal musical speed. As for motion reconstructed from a four-layer hierarchical B-spline and a five-layer hierarchical B-spline, these phenomena appear more clearly. Fig. 3 shows only the left shoulder joint angle, but we found the same phenomena in other joint angle trajectories of the upper body, and in other dance masters' joint angle trajectories.

Fig. 4 shows the comparison of variance sequences: the green, yellow, and light blue lines represent the variance sequences of the left shoulder joint angle at the normal musical speed, 1.2 times faster musical speed, and 1.5 times faster musical speed, respectively, and the blue line represents the variance sequence calculated from all the motion sequences. The joint angles for the variance calculation were reconstructed with a five-layer hierarchical B-spline, and normalized by adjusting the knot of the estimated control points. From these variance sequences, it is confirmed that there are some valleys where all variance sequences commonly have local minima. This means that the postures at these valleys (the middle row of Fig. 4) are preserved even if the musical speed gets faster and the high frequency components are attenuated. We found that most valleys represent the important stop motion (called keypose) specified by the dance masters (the bottom row of Fig. 4).

From these observations, we obtained the following two insights:

Insight 1 High-frequency components of human motion will be attenuated when the music playback speed becomes faster.

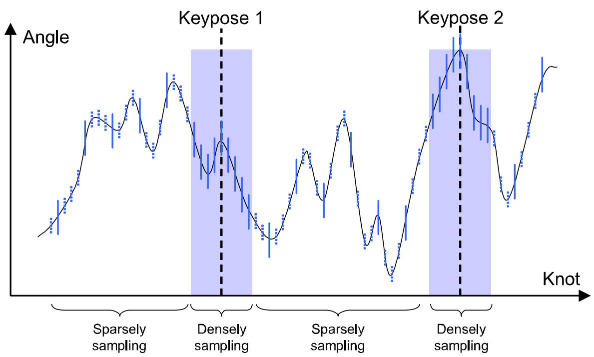


Fig. 5. Illustration of our sampling method to consider keypose information for hierarchical motion decomposition.

Insight 2 Keyposes will be preserved even if high frequency components are attenuated.

Based on these insights, we propose a method to model the temporal scaling of human dance motion.

III. UPPER BODY MOTION GENERATION BY TEMPORAL SCALING

In this section, we propose a method to temporally scale a dance motion based on the acquired insights and joint limitations that a humanoid robot has. In the following, θ represents a 1D joint angle of a humanoid robot, and control points of B-spline can be represented in one dimension. The proposed method consists of two phases:

1) Hierarchical motion decomposition using keypose information
2) Motion generation based on joint limitations

A. Hierarchical Motion Decomposition Using Keypose Information

According to Insight 2, keypose information, including posture and velocity components, is preserved even if the musical speed is fast. Therefore, low frequency components of dance motion sequence must contain the keypose information. Remembering this insight, we can improve the method of motion decomposition described in Eq. (6). To achieve this, our motion decomposition method should consider the posture and velocity information of the keyposes.

To consider posture information, we densely sample input motion sequence around keyposes, we sparsely sample it in other parts, and then we use these samples to form a linear system of equations. Figure 5 provides an illustration of our data-sampling method for motion decomposition. All vertical lines in this illustration represent originally sampled data, and our method uses only the solid lines shown among them.

With regard to velocity information, the movements of a dancer's arms and hands stop around keyposes: the velocity of the hands and arms are approximately zero at keyposes. We exploit this useful property of keyposes as velocity information in our motion decomposition method. From all the keyposes, we form a linear system of equations to satisfy

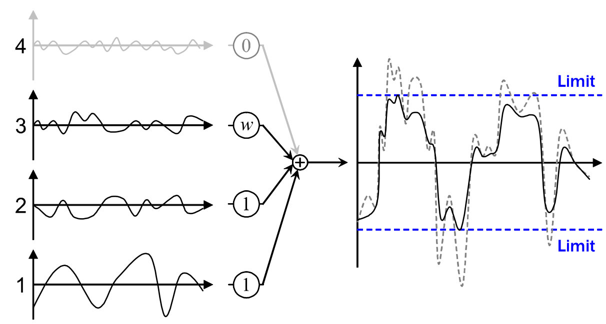


Fig. 6. Motion reconstruction considering joint limitations. Our optimization process gradually attenuates the weighting factors from the finest layer.

the velocity constraints:

$$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = N^{vel} \begin{pmatrix} \hat{Q}_0 \\ \hat{Q}_1 \\ \vdots \\ \hat{Q}_n \end{pmatrix}, \tag{10}$$

where N^{vel} represents a (the number of keyposes) $\times (n+1)$ matrix whose elements are given as

$$N_{ij}^{vel} = \begin{cases} \frac{d}{dt} B_{j+1-\lfloor t_i \rfloor}(t_i - \lfloor t_i \rfloor) & \text{if } j \le t_i < j+1 \\ 0 & \text{otherwise} \end{cases}. \tag{11}$$

Considering both the posture and velocity constraints, the motion decomposition method is modified as

$$\begin{pmatrix} \theta(t_1) \\ \vdots \\ \theta(t_m) \\ \hline \mathbf{0} \end{pmatrix} = \begin{pmatrix} N^{pos} \\ \hline N^{vel} \end{pmatrix} \begin{pmatrix} \hat{Q}_0 \\ \hat{Q}_1 \\ \vdots \\ \hat{Q}_n \end{pmatrix}, \tag{12}$$

where N^{pos} represents a coefficient matrix of B-spline basis functions modified by our densely/sparsely sampling method. For each layer of hierarchical B-spline, we can estimate the control points by solving Equation (12) and decompose the input motion sequence.

B. Motion Generation Based on Joint Limitations

The final step is to generate temporally-scaled motion for a humanoid robot. Simple temporal scaling can be done by adjusting the temporal frame of B-spline control points with the specified scaling ratio. However, the resulting motion may violate angular limitations such as joint angular velocity. To solve this, we consider Insight 1 and joint limitations that a humanoid robot has, and we modify upper body motion.

In this step, we first segment the motion sequence to correspond to music rhythm frames, and then we optimize weighting factors for each hierarchical B-spline layer in each motion segment so that a resulting joint angle θ_{opt} must satisfy certain joint limitations:

$$\theta_{\min} \le \theta_{opt}(t) \le \theta_{\max} \tag{13}$$

$$\dot{\theta}_{\min} \le \dot{\theta}_{opt}(t) \le \dot{\theta}_{\max} \tag{14}$$

where $\theta_{\min}$ and $\theta_{\max}$ represent minimum and maximum joint angles, respectively, and $\dot{\theta}_{\min}$ and $\dot{\theta}_{\max}$ represent minimum and maximum joint angular velocities, respectively. Finally the resulting joint angle is represented as

$$\theta_{opt}(t) = \sum_{i=1}^{N} w_i f_i(2^{i-1} st), \tag{15}$$

where s represents a temporal scaling factor (i.e. the resulting motion is s-times faster than the original motion), N represents the number of hierarchical B-spline layers, and f_i represents the i-th layer of the constructed hierarchical B-spline. $w_i \in [0, 1]$ represents the weighting factor for the i-th layer to be detected via this optimization process.

According to Insight 1, the high frequency component is attenuated when the motion is beyond joint angle limitations. Therefore, this optimization process is done by attenuating the weighting factors from the finest layer. When the weighting factor reaches zero and the resulting motion does not satisfy joint limitations, the weighting factor for next coarser layer is then gradually attenuated. Finally, when the resulting motion consists of n layers, the weighting factors from the 1st to the $(n-1)$-th layers are 1, the factor for the n-th layer is $(0, 1]$, and the factors from the $(n+1)$-th to the N-th layers are 0. This is illustrated in Fig. 6.

In this process, a discontinuity might develop between neighboring motion segments if there ends up being a difference in the weighting factors. So we apply motion blending around the discontinuities. Let $\mathcal{A}$ and $\mathcal{B}$ be neighboring motion segments. The interpolated joint angle θ' can be calculated as

$$\begin{aligned}\theta'(t) =& \alpha\left(\frac{t-(t_b-0.5L)}{L}\right)\theta^{\mathcal{A}}_{opt}(t) \\ &+ \left(1-\alpha\left(\frac{t-(t_b-0.5L)}{L}\right)\right)\theta^{\mathcal{B}}_{opt}(t),\end{aligned} \tag{16}$$

where t_b represents a discontinuity frame existing between $\mathcal{A}$ and $\mathcal{B}$, L represents the duration of interpolation, and $\alpha(t)$ is a quintic polynomial equation given as

$$\alpha(t) = -6t^5 + 15t^4 - 10t^3 + 1. \tag{17}$$

This quintic polynomial equation is a C^2 continuous function such that $\alpha(0) = 1$, $\alpha(1) = 0$, $\frac{d}{dt}\alpha(0) = \frac{d}{dt}\alpha(1) = 0$, and $\frac{d^2}{dt^2}\alpha(0) = \frac{d^2}{dt^2}\alpha(1) = 0$.

Through this interpolation process, there is a possibility of going beyond mechanical limitations. So we iteratively do the optimization and interpolation procedures until the resulting motion does not violate the joint limitations.

IV. EXPERIMENTS

In this section, we show the results of the experiments that evaluated our method. We tested our algorithm by modifying the Aizu-bandaisan dance data through our algorithm. All motion data was captured at 120 fps by an optical motion capture system produced by Vicon. We applied the proposed method to the upper body motion, and applied Nakaoka et al.'s method to generate leg motion [12]. Our experimental platform was HRP-2.

A. Result of Original-Speed Motion Generation

We first tested our algorithm by generating the dance motion for the HRP-2 at the normal speed. In this experiment, we compared our method with Pollard et al.'s method that can adapt motion capture data for a humanoid robot using PD filter [13].

Fig. 7 shows the experimental result with the actual HRP-2. It is confirmed that the robot can stably imitate the human dance motion. Fig. 8 shows the resulting joint angle trajectories of the left shoulder yaw. The red, green, and blue lines represent the trajectories of the original captured motion, the result of Pollard et al.'s method, and the result of our method. As for the joint angular velocity (Fig. 8 (b)), our method has two advantages. One is that our method can preserve more details than the trajectories resulting from Pollard et al.'s method. The trajectories resulting from Pollard et al.'s method often lack high frequency components, due to the PD control. This phenomenon is shown in Fig. 8 (b.1). The other is that the speed around constraint-violating motion frames generated by Pollard et al.'s method is a constant value. This phenomenon is shown in Fig. 8 (b.2). This can create two problems. One is that the humanoid robot cannot clearly reproduce a keypose if the posture and angular speed around the keypose violate kinematic constraints. The other is that the humanoid robot may fall because of the rapid changes in acceleration.

B. Simulation Result of 1.2 Times Faster Motion Generation

Next, we tested our algorithm by generating the dance motion whose speed was 1.2 times faster than the original speed in simulation. The upper body motion was generated by our proposed method, and as for leg motion, we first applied simple temporal scaling to the motion capture data and then applied Nakaoka et al.'s method. Fig. 9 shows the simulation results, and the red sphere represents a Zero Moment Point. Our simulated motion satisfied the criterion for balance maintenance, and the humanoid robot successfully performed the dance.

V. CONCLUSION

In this paper, we proposed a method to model temporal scaling of upper body motion for a sound feedback system of a dancing humanoid robot. We analyzed motion data captured at varying musical speeds by using a hierarchical motion decomposition technique. Through this observation, we obtained the following two insights:

1) High frequency components of human motion will be attenuated when music playback speed becomes faster.
2) Keyposes will be preserved even when high frequency components are attenuated.

We applied these insights to model motion modification that can generate motion satisfying joint limitations of a humanoid robot. Our experimental results show the effectiveness of our method.

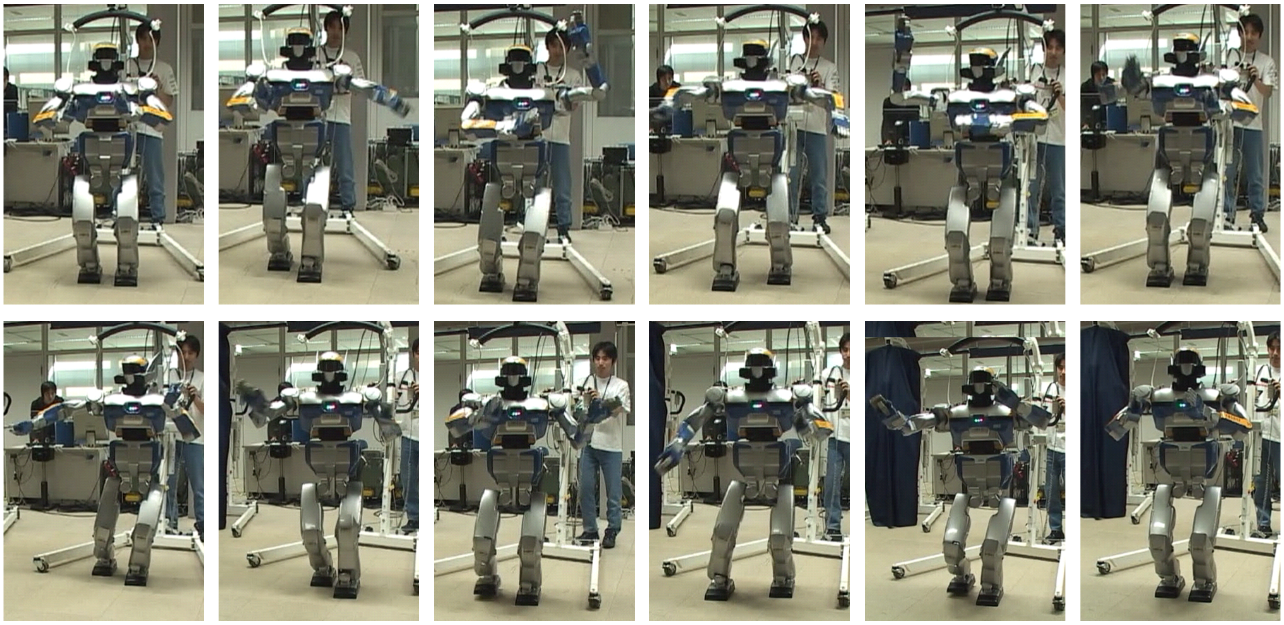
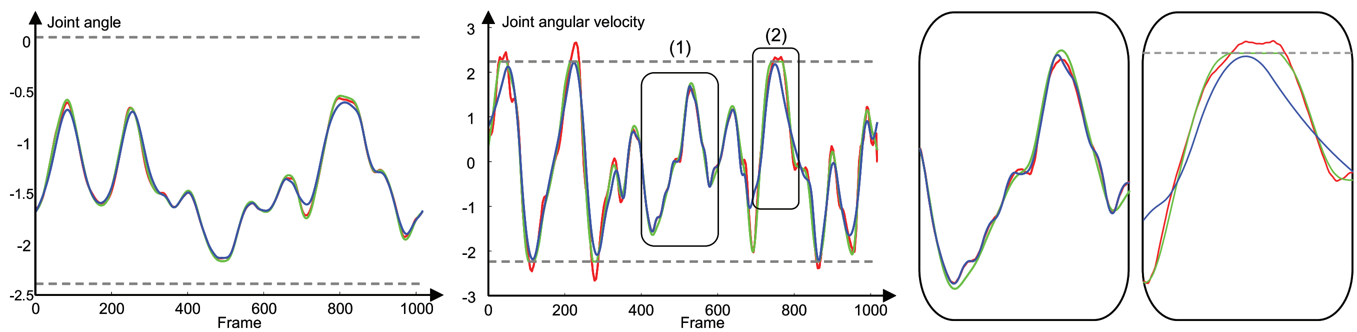


Fig. 7. Result of generating the Aizu-bandaisan dance motion at the musical original speed.



(a) (b) (b.1) (b.2)

Fig. 8. Comparison of left shoulder yaw angle trajectories of the original motion (red), generated by Pollard et al.'s method (green), and generated by our method (blue). (a): joint angle trajectories, and (b): joint angular velocity. (b.1) and (b.2) represent the zoomed-in graph of part (1) and (2) in (b), respectively.

As a future work, we plan to extend this method for temporally-scaled leg motion generation. When feet touch the ground, their high frequency components are suddenly much larger [14]. But these high frequency components are not derived from a performer's style. The violation of high frequency components in synthesized foot motion produces an unnatural and unstable motion called *foot-skating*. Nakaoka et al. [12] proposed a method to recognize the states of leg motion and to extract the style components of lower body motion. We believe that our method is applicable to leg motion if Nakaoka et al.'s method is applied so that possible states of "swing sole" for a leg are properly recognized.

ACKNOWLEDGEMENT

Takaaki Shiratori is supported by the Japan Society for the Promotion of Science. This work is supported in part by the Japan Science and Technology Corporation (JST) under the CREST project.

REFERENCES

[1] M. Stone, D. DeCarlo, I. Oh, C. Rodriguez, A. Stere, A. Lees, and C. Bregler, "Speaking with hands: Creating animated conversational characters from recordings of human performance," *ACM Transactions on Graphics (Proc. ACM SIGGRAPH 2004)*, vol. 23, no. 3, pp. 506–513, 2004.

[2] T. Shiratori, A. Nakazawa, and K. Ikeuchi, "Dancing-to-music character animation," *Computer Graphics Forum (Proc. Eurographics 2006)*, vol. 25, no. 3, pp. 449–458, 2006.

[3] K. Kosuge, T. Hayashi, Y. Hirata, and R. Tobiyama, "Dance partner robot -Ms DancerR-," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2003, pp. 3459–3464.

[4] K. Yokoyama, H. Handa, T. Isozumi, Y. Fukase, K. Kaneko, F. Kanehiro, Y. Kawai, F. Tomita, and H. Hirukawa, "Cooperative works by a human and a humanoid robot," in *Proc. IEEE International Conference on Robotics and Automation*, 2003, pp. 2985–2991.

[5] A. Bruderlin and L. Williams, "Motion signal processing," in *Proc. ACM SIGGRAPH 95*, 1995, pp. 97–104.

[6] M. Gleicher, "Retargetting motion to new characters," in *Proc. ACM SIGGRAPH 98*, 1998, pp. 33–42.

[7] J. Wang, S. M. Drucker, M. Agrawala, and M. F. Cohen, "The cartoon animation filter," *ACM Transactions on Graphics (Proc. ACM SIGGRAPH 2006)*, vol. 25, no. 3, pp. 1169–1173, 2006.

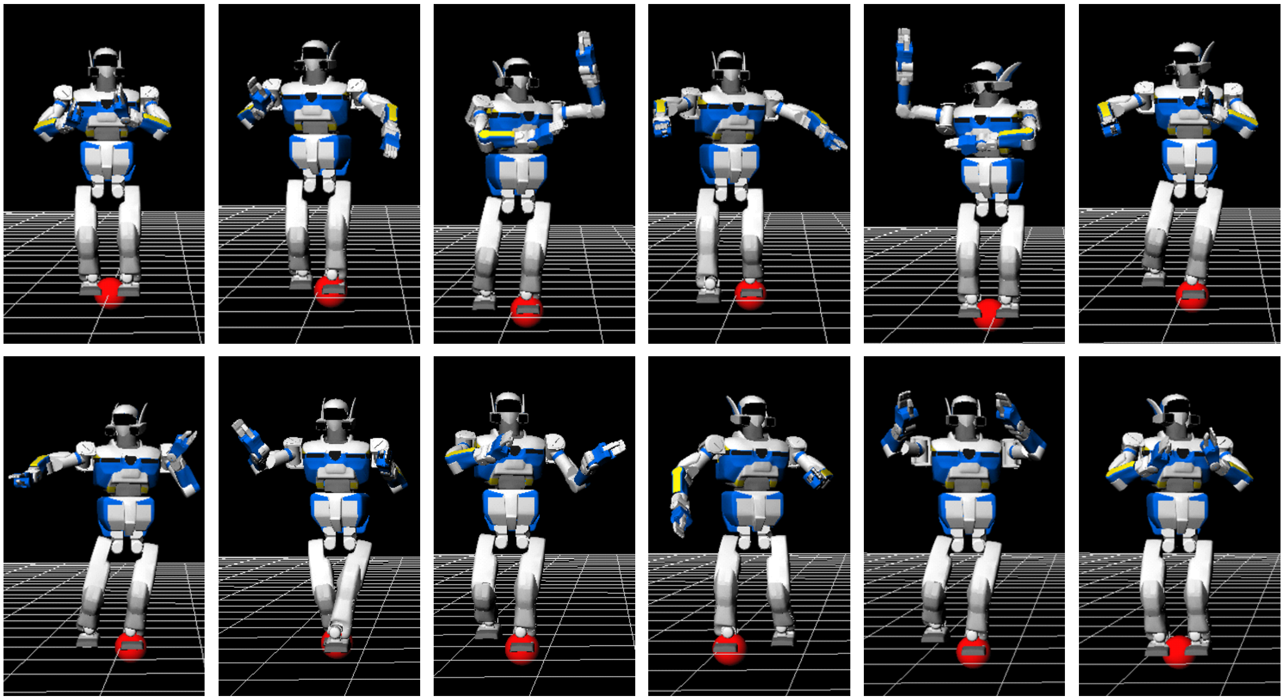


Fig. 9. Simulation result of generating 1.2 times faster dance motion. The red sphere represents the ZMP of the resulting motion.

[8] J. McCann, N. S. Pollard, and S. Srinivasa, "Physics-based motion retiming," in *Proc. ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, 2006, pp. 205–214.

[9] S. Lee, G. Wolberg, and S. Y. Shin, "Scattered data interpolation with multilevel B-splines," *IEEE Transactions on Visualization and Computer Graphics*, vol. 3, no. 3, pp. 228–244, 1997.

[10] J. Lee and S. Y. Shin, "A hierarchical approach to interactive motion editing for human-like figures," in *Proc. ACM SIGGRAPH 99*, 1999, pp. 39–48.

[11] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43–77, 1994.

[12] S. Nakaoka, A. Nakazawa, F. Kanahiro, K. Kaneko, M. Morisawa, and K. Ikeuchi, "Task model of lower body motion for a biped humanoid robot to imitate human dances," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2005, pp. 3157–3162.

[13] N. S. Pollard, J. K. Hodgins, M. J. Riley, and C. G. Atkenson, "Adapting human motion for the control of a humanoid robot," in *Proc. IEEE International Conference on Robotics and Automation*, 2002, pp. 1390–1397.

[14] O. Arikan, "Compression of motion capture databases," *ACM Transactions on Graphics (Proc. ACM SIGGRAPH 2006)*, vol. 24, no. 3, pp. 890–897, 2006.