Motion Analysis using Hilbert-Huang Transform and Its Applications to Robot Motion Synthesis

March ２０２０

Dong Ran

Motion Analysis using Hilbert-Huang Transform and Its Applications to Robot Motion Synthesis

Graduate School of Systems and Information Engineering
University of Tsukuba

March ２０２０

Dong Ran

Abstract

Motion data (captured motions, animated motions, and simulated motions) are widely used in different research fields in medical research, entertainment, and industry. However, most researches about motion analysis and synthesis using motion data have been carried out only in the time-domain. To deal with these complicated motion data (especially motion data collected from the real-world), there is a significant need to analyze and edit them in the instantaneous frequency domain. In order to overcome this problem. In the present dissertation, we present novel insights to motion analysis and synthesis in the instantaneous frequency domain using Hilbert-Huang transform (HHT). HHT can decompose a real-world signal into several pseudo monochromatic signals called intrinsic mode function (IMF) using empirical mode decomposition (EMD). HHT is used for various applications and researches such as blood pressure change measurement, arrhythmia detection, submarine detection, and earthquake countermeasure construction inspection.

In this dissertation, we propose a framework in the instantaneous frequency domain of motion analysis and synthesis using HHT. We inputted motion capture data as multiple variable signals into EMD to get decomposed IMFs. After applying the Hilbert transform (HT) to each IMF, the instantaneous frequencies of motions were obtained. Our proposed research framework can be used in motion analysis for real-world motions that are nonlinear and non-stationary signals obtained from a motion capture system. Also, our framework can apply to the motion synthesis in computer graphics and to robot motion design in robotics. To show the applications of our proposed framework using HHT, we applied our framework to three different applications: (i) Dance motion analysis and editing; (ii) Bunraku ("文楽" in Japanese) puppet motion analysis; and (iii) Robot motion synthesis using Bunraku puppets.

First, we applied our proposed framework to (i) dance motion analysis and editing. Dance motions consist of several choreographies. Using our framework, these choreographies can be analyzed in the instantaneous frequency domain to show the features of different dance styles. As an example, we discuss the feature differences between the dance styles of Japanese techno-pop group Perfume, salsa, waltz, and hip-hop using the spectrum analysis results. For dance editing, our framework decomposes dance motions into choreographies, which can be edited, extracted, blended, exaggerated and exchanged with other dance styles. In this dissertation, we extracted two different choreographies from Perfume dance and blended them separately into a salsa dance.

Second, we used our proposed framework for (ii) Bunraku puppet motion analysis. “Ningyo Joruri Bunraku” (“人形浄瑠璃文楽” in Japanese) is one of intangible world cultural heritages selected by UNESCO (United Nations Educational, Scientific and Cultural Organization). The affective motions of Bunraku puppets attract audiences so deeply that they contribute to interactions between the audiences and puppets. We used three processes in this application: (1) We collected the affective motions of Bunraku using both optical and magnetic motion capture systems; (2)We used our framework using HHT to analyze the interactive mechanism of Bunraku motion called “Jo-Ha-Kyū” (“序破急” in Japanese), which is a Japanese traditional performing art principle; and (3) We verified the difference in mechanism of the interaction techniques between Bunraku and Western dance (Perfume dance) based on the previous Jo-Ha-Kyū researches of Noh and Bunraku using HHT. Thus, the mechanism extracted from Bunraku puppets can be applied to a Human-Robot interaction.

Finally, we adapted our proposed framework to (iii) robot motion synthesis using Bunraku puppets. We present a robot motion design framework using HHT for sparsing motion that can fit the limitation of motor speed. As an example, we use Bunraku affective motions that are based on Jo-Ha-Kyū to create affective motions for a robot. We converted a few simple Bunraku motions into robot motions using deep learning methods. Our primitive experiments showed that Jo-Ha-Kyū can be incorporated smoothly into robot motion design, and some simple affective robot motions can be designed automatically by our proposed framework. Then, a robot or vocaloid (CG) can express affective motion synchronized with narration (story) based on Jo-Ha-Kyū.

Using our proposed framework, complicated motion data (e.g., dance motions, Bunraku puppet motions) can be analyzed and edited in the instantaneous frequency domain. Our research framework can also be applied to Human-Robot interaction.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter is an introduction to the background and motivation of this research. We discuss related researches of motion analysis and synthesis based on motion capture data. After reviewing related researches in motion capture systems, we suggest that Hilbert-Huang transform (HHT) can provide another view of motion processing in the instantaneous frequency domain. We carefully explain the difference between HHT and a short-time Fourier transform (STFT) by comparing their mathematical definitions in section 1.1. At the end of this chapter, we summarize the purpose of our research in section 1.2.

1.1 Research background and motivation

In recent years, human motions have been captured as digital data using cameras and motion capture systems. Particularly, for a motion capture system, it is possible to record joints position including the rotation angle by placing a mark at the joint position of a person. Therefore, when a person is moving, joint angle data can be obtained. There have been many researches on various motions, such as character animation, human and robot motion. Holden et al. [1] proposed a method to reconstruct and stylize human motion using deep learning methods. Choi et al. [2] presented an approach for the exaggerated editing of character motions. Aberman et al. [3] conducted research on motion retargeting with the learning of character-agnostic motions. Kim et al. [4] conducted research about a motion editing method for interactive editing of users using constraints. By using motion beats, Kim et al. [5] presented a scheme for blending motions (e.g. dance motions) rhythmic patterns into other motions. PonsMoll et al. [6] used a 4D capture system and 3D scan to propose a method for modifying the deformations of a character's motions. Kwon et al. [7] presented a motion-labeling method using a learning-based classification for fighting interactive actions. However, most researches have been carried out only in

the space-time domain, because noisy and complicated motions like human kinematic motions are very difficult to analyze and edit in the frequency domain.

Hilbert transform (HT) is used to transform a nonlinear real monochromatic signal into its imaginary part to obtain the instantaneous frequency and amplitude for analyzing signals in the instantaneous frequency domain [8]. The advantage of HT is that an instantaneous frequency and amplitude from a nonlinear chromatic signal can be obtained, and they can be visualized more clearly than by the STFT [9]. However, because most real-world data are not monochromatic signals, there are few data suitable for the HT. Huang et.al [10] proposed an empirical mode decomposition (EMD) first in 1998, which can decompose an original signal into several pseudo monochromatic waves so-called intrinsic mode function (IMF), and a residual so-called trend. Decomposed IMFs are approximately multivariable monochromatic waves that can be applied to the HT. However, please note that no mathematical proof exists that chromatic signals can be always decomposed into chromatic signals using the EMD. The whole process is called HHT. As a result, we can get the instantaneous frequency and amplitude from a nonstationary and nonlinear real-world chromatic signal. Compared with Fourier transform (FT), which decomposes the signal into a series of sine or cosine waves in a linear way, HHT decomposes the signal into finite IMFs and a Trend (residual), empirically and nonlinearly. There are also other decomposition methods such as Generalized Harmonic Analysis (GHA). GHA is minimizing residual energy from the original waveform within the observation interval to extract the monochromatic wave [11, 12]. The advantage of GHA is that this method can deal with frequency fluctuations in a short time interval than FT. However, GHA also decomposes a signal in a linear way as same as FT, different from EMD. Furthermore, many researches have been conducted to expand the EMD from one variable to multiple variables, i.e. bivariate EMD (BEMD), trivariate EMD (TEMD), multivariate EMD(MEMD) and noise-assisted Multivariate EMD(NA-MEMD) [13-16]. Besides, recent studies have reported that HHT is a useful tool for sparse coding in deep learning [17-22]. As a result, HHT is applied widely to various applications now, and its Hilbert spectral analysis (HSA) is getting more and more attention among researchers due to its advantages in the instantaneous frequency domain.

Thus, for nonlinear, nonstationary, and multi-channel signals like motion capture data, HHT is more convenient than FT and can be considered as a powerful tool to analyze and synthesize motion capture data.

Meanwhile, the technology for voice-recognition AI (artificial intelligence) assistants has been explosively spreading. However, for most people, AI assistants are used only for a music player instead of a usual AI [23]. For example, the AI assistants such as Apple Siri, have been used for a voice recognition interface in practical, but few people use Siri regularly except for a music player. The reason for this limited use is not clearly understood, but one possible reason is a lack of the feelings associated with interpersonal conversations that can lead to a human discomfort.

The UNESCO Intangible Cultural Heritage "Ningyo Joruri Bunraku (人形浄瑠璃文楽" in Japanese)", whose affective motion is regarded as "the most beautiful motion in the world" by UNESCO, and Bunraku play can create empathy among the audiences [24]. Thus, it may play an important role in resolving the AI assistant's lack of relatability in interpersonal conversations with humans. There are still many unexplained aspects of how emotions are formed through human-machine interaction and communication. In the case of communications between humans and robots, the more robot resembles humans, the more uncomfortable interaction occurs between humans and vocaloids (CG) / robots, which is so-called "uncanny valley" [25]. Thus, analyzing the mechanism of emotional interaction techniques of the Bunraku play in the instantaneous frequency domain using HHT can be a novel solution to resolve the above problem.

1.2 Research purpose

In this dissertation, we use HHT to decompose these motions into several pseudo monochromatic waves, i. e., IMFs that are nonlinear monochromatic signals and can be analyzed in the instantaneous frequency domain [8]. First, we applied our proposed framework to dance motions. Next, we focused on traditional performing art techniques of Bunraku ("文楽" in Japanese) using our proposed framework with HHT. Following these, we took the motion data and extracted the Jo-Ha-Kyū ("序破急" in Japanese) principle, the mechanism of Bunraku puppet interactive technique, that is usually the change of speeds in motions and music. Finally, we developed a framework using the Jo-Ha-Kyū principle extracted from Bunraku play music and puppet motions to overcome the communication discomfort uncanny valley. We use this framework for robot motion design to express emotions during human-robot interactions as showing in Figure 1.1.

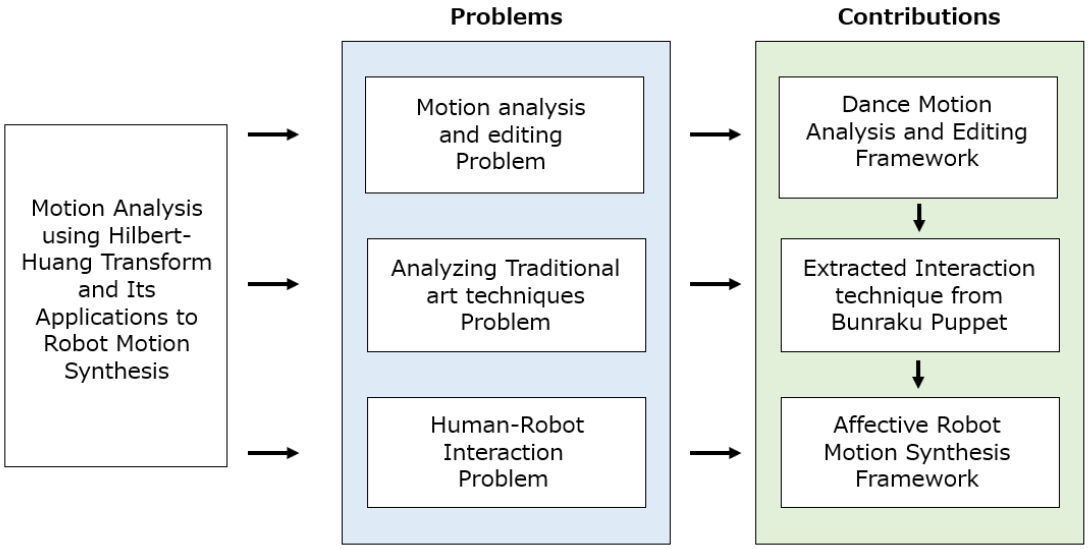

Figure 1.1 A framework for the research work performed in this dissertation

Thus, our researches include the design and development of a motion editing framework using HHT, and the applications of the framework to the Bunraku puppet analysis and robot motion synthesis.

1.2.1 Dance motion analysis and editing using HHT

There have been very few researches conducted on motion analysis and editing in the instantaneous frequency domain. We propose a motion analysis and editing framework using HHT for complicated dance motions. Using our framework, dance motions can be decomposed into each choreography in the instantaneous frequency domain. This allows dancers and animators to analyze dance motion clearly. For example, harmonic motions, and a Fibonacci sequences can be detected by our proposed framework in the instantaneous frequency domain. Furthermore, these decomposed choreographies can be edited, exchanged and blended into other dances to create new dance styles. This is useful for animators to analyze and design dance motions.

1.2.2 Bunraku puppets motion analysis using HHT

It is difficult to analyze the mechanism of Japanese traditional arts such as Bunraku puppets. Our proposed framework can verify the Jo-Ha-Kyū and "Ma" ("間" in Japanese) mechanism of Bunraku plays. Here, Ma is the intervals among Jo, Ha,

and Kyu, which is crucial in the Jo-Ha-Kyū principle. We decomposed Bunraku motions using our framework and showed that the motions of Bunraku puppets are divided into Jo-Ha-Kyū and Ma. These extracted mechanisms and unique techniques can be used for design performing art like actors and robots and CG characters.

1.2.3 Applying HHT to robot motion synthesis using bunraku puppets

"Uncanny Valley" is used to describe the phenomenon in which a human-robot interaction is becoming similar to that with a real human [25]. This makes humans uncomfortable and, thus, prohibits the prevalence of humanoids with AI assistants. We adapted the framework using deep learning and Jo-Ha-Kyū principle to create affective motions of a robot. Using our proposed framework, we can create affective motions to interact with humans. This can be considered as a tool to improve the interaction between humans and humanoids with AI assistants.

In Chapter 2, we review the HHT. In Chapter 3, we introduce our proposed framework of dance motion analysis and synthesis using the HHT. In Chapter 4, we show the results of the Bunraku puppet motion analysis using the HHT. In Chapter 5, we demonstrate the results of robot motion synthesis using HHT and Bunraku puppets. In Chapter 6, we summarize our contributions and discuss future research.

Chapter 2

Hilbert-Huang transform

In this chapter, we introduce the Hilbert-Huang transform (HHT). The key idea of HHT is using empirical mode decomposition (EMD) that decomposes a real-world signal into several intrinsic mode functions (IMFs) that are pseudo monochromatic waves. After decomposed signal into each IMF, their instantaneous frequency and amplitude were obtained with an analytical signal using Hilbert transform (HT). Thus, we first introduce the analytical signal in section 2.1 and HT in section 2.2. Next, we explain the definition and algorithm of EMD in section 2.3. There are different types of EMD, then we introduce one variable EMD and its algorithm to show the mechanism of decomposing nonlinear signals into pseudo monochromatic waves. Besides, we discuss multivariate EMD (MEMD) and noise-assisted multivariate EMD (NA-MEMD) for decomposing multivariate data such as motion capturing data. Because the EMD algorithm uses spline interpolation to extract envelop from the original signal, we discuss the spline implementation for EMD in section 2.4. In section 2.5, we show an example of Hilbert spectral analysis (HSA) for analyzing decomposed signals by EMD. Also, we introduce the weighted average frequency algorithm (WAFA), which is a method for HHT to get average frequencies of each IMF in section 2.6.

2.1 Analytical signal

The analytical signal is a signal analysis theory widely used in the signal processing field. An analytical signal (Figure 2.1) is defined as $z(t) = z_r(t) + iz_i(t)$, where $z_r(t)$ denotes the real part, $z_i(t)$ denotes the imaginary part.

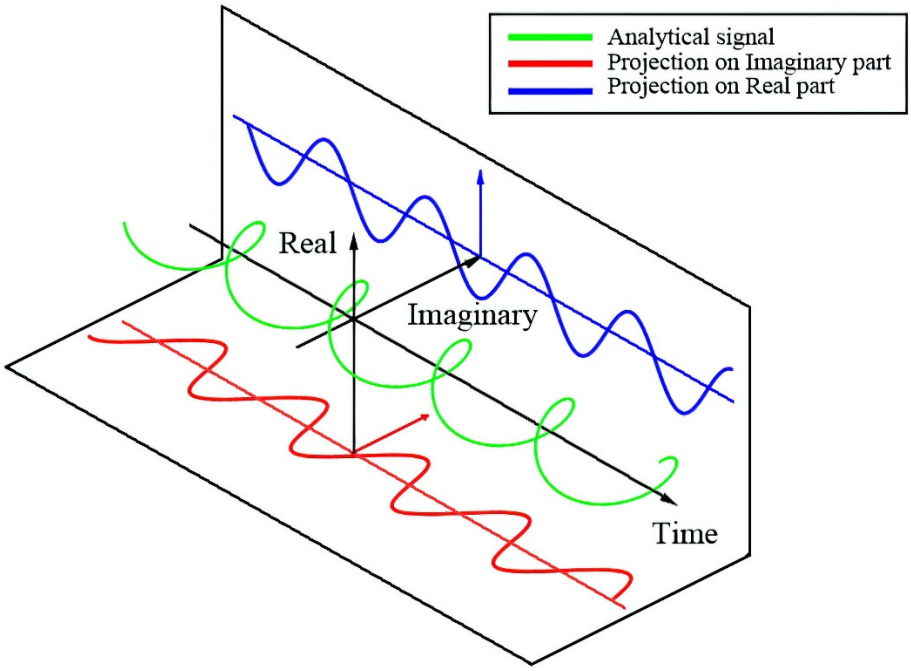

Figure 2.1 Analytical Signal

As shown in Figure 2.1 [9], the signal observed in the real world is a projection from the complex plane onto the real axis as time passed. Therefore, the instantaneous frequency A and the instantaneous amplitude ω_0 can be obtained from the analysis signal by the following formula [9].

$$\mathrm{A(t)} = \sqrt{{z_r}^2(t) + {z_i}^2(t)} \qquad (1)$$

$$\omega_0(\mathrm{t}) = \frac{d}{dt}\tan^{-1}\frac{z_i(t)}{z_r(t)} \qquad (2)$$

2.2 Hilbert transform

Because signals observed in the real world only have the real part, it cannot be determined whether the changing of the signal is caused by amplitude or phase. To obtain the imaginary part of an observed signal, Hilbert transform (HT) converts the real part of the observed signal into its imaginary part. Thus, the real part $z_r(t)$ of the analytical signal can be observed in the real-world. The imaginary part $z_i(t)$ can

be obtained using HT.

Hilbert transform assumes that observed signals are a monochromatic signal [9]. Thus, the real part of the analytical signal can be expressed as $z_r(t) = x(t) = A(t)cos(\omega(t)t)$. Its Fourier transform is:

$$F[x(t)] = F[A(t)cos(\omega(t)t)] = \mathrm{F}\left[A\frac{e^{i\omega_0 t}+e^{-i\omega_0 t}}{2}\right] = \mathrm{A}\frac{\delta(\omega-\omega_0)+\delta(\omega+\omega_0)}{2} \qquad (3)$$

On the other hand, the Fourier transform of the analytic signal $z(t) = z_r(t) + iz_i(t) = Ae^{i\omega_0 t}$ is $F[z(t)] = A\delta(\omega - \omega_0)$. In order to obtain the analytic signal, it is necessary to remove the negative frequency component of the original signal, double the remaining signal, and perform an inverse FT. As a result, the analytical signal is as follows [9]:.

$$z(t) = x(\mathrm{t}) + i(\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(\tau)}{t-\tau}d\tau) \qquad (4)$$

Here, x(t) is the real part $z_r(t)$ observed in the real world. Then, HT can be defined to obtain the imaginary part from the real part is as follows [9]:

$$z_i(t) = y(\mathrm{t}) = \frac{1}{\pi}PV\int_{-\infty}^{\infty}\frac{x(\tau)}{t-\tau}d\tau = \frac{1}{\pi t} * x(t) \qquad (5)$$

Here, PV denotes the Cauchy principal value [9].

Since an analytical signal can be obtained by transforming the observed signal in the real world to the imaginary part, the instantaneous frequency, and instantaneous amplitude can be obtained by equation (1,2).

2.3 Empirical mode decomposition

As we introduced in section 2.2, HT assumes signals are monochromatic wave $A(t)cos(\omega(t)t)$ to obtain their instantaneous frequency and instantaneous amplitude. However, the real-world signals are almost chromatic waves that HT cannot be applied to.

As shown in Table 2.1, when assuming a signal as $s(t) = \beta + \cos(t)$, there are three cases depending on the different values of β.

Table 2.1 Three different cases of Beta

Signal	β
$s(t) = \cos(t)$	$\beta = 0$
$s(t) = 0.5 + \cos(t)$	$1 > \beta > 0$
$s(t) = 5 + \cos(t)$	$\beta > 1$

Figure 2.2 shows the three different signals in the complex plane. The real axis signal is $\cos(\theta)$, and the imaginary axis signal is $\sin(\theta)$. Here, the phase θ corresponds to t. As shown in Figure 2.2, when β is nonzero, the signal rotates away from the origin as time passes.

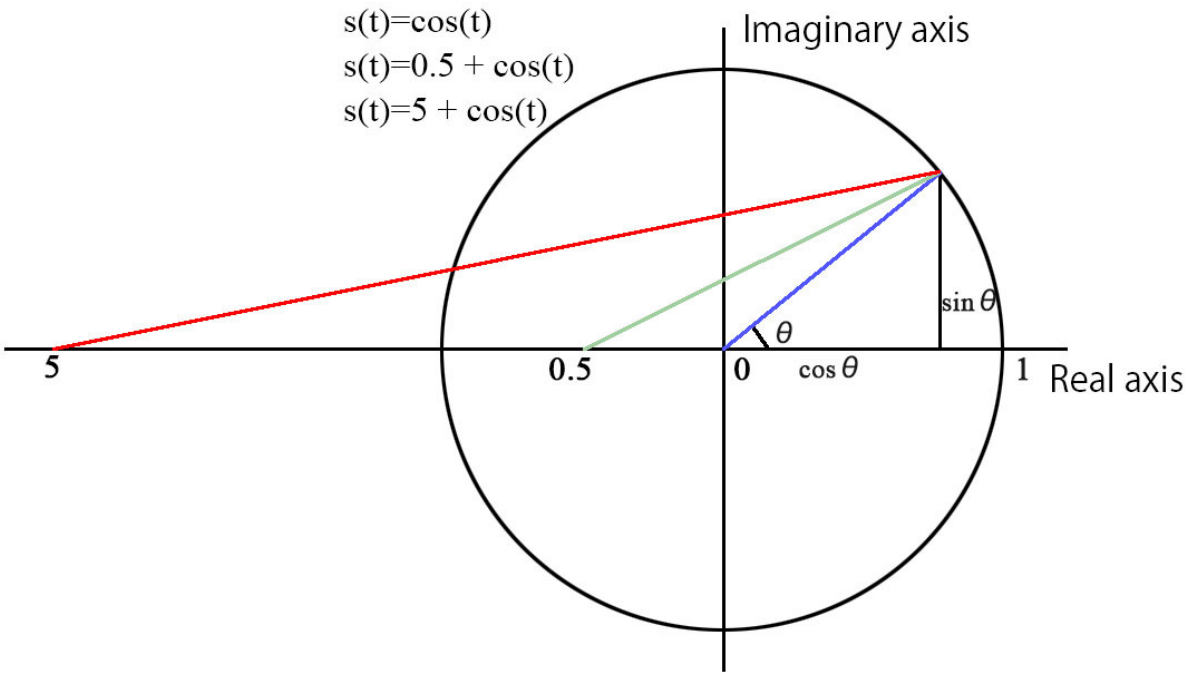

Figure 2.2 Signals in the complex plane

Then, the phase of each signal is obtained as shown in Table 2.2.

Table 2.2 Signal phase

Signal	Phase
$s(t) = \cos(t)$	$\theta = arctan\left(\frac{sin\theta}{cos\theta}\right) + n\pi(n = 0,1,2 \dots)$
$s(t) = 0.5 + \cos(t)$	$\theta = arctan\left(\frac{sin\theta}{0.5 + cos\theta}\right) + n\pi(n = 0,1,2 \dots)$
$s(t) = 5 + \cos(t)$	$\theta = arctan\left(\frac{sin\theta}{5 + cos\theta}\right) + n\pi(n = 0,1,2 \dots)$

When β is zero, the phase θ increases linearly. However, when β is nonzero, the phase θ oscillates as shown in Figure 2.3.

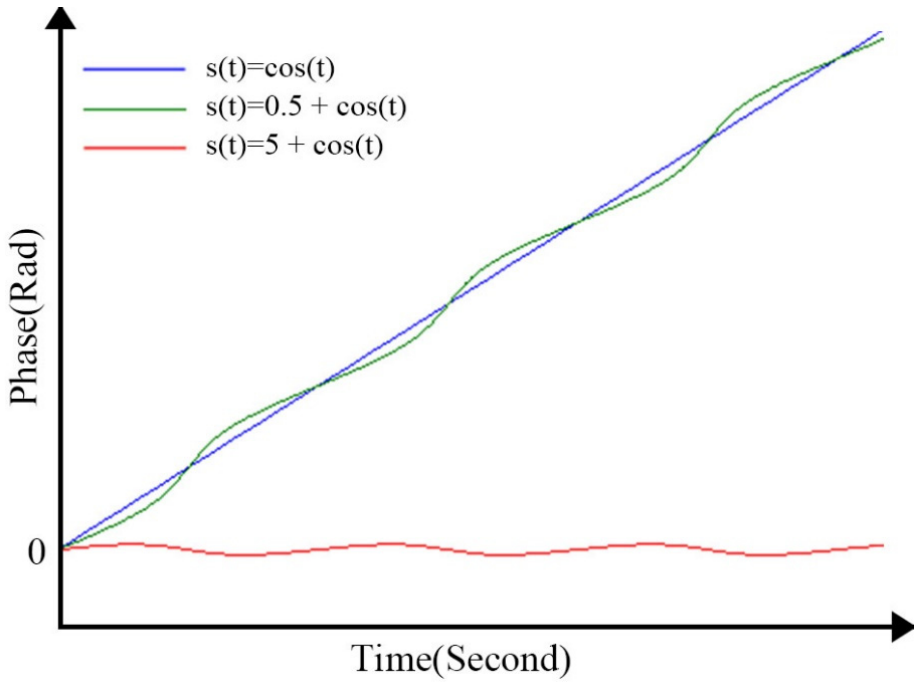

Figure 2.3 Instantaneous phase

Figure 2.4 shows the instantaneous frequency calculated from the instantaneous phase shown in Figure 2.3. Since the phase has not been calculated correctly, the instantaneous frequency is also cannot be obtained correctly. As shown in Figure 2.4, even a negative frequency observed when $\beta = 5$. Because the negative frequency does not have any physical meaning, the instantaneous frequency cannot be calculated directly when β is nonzero.

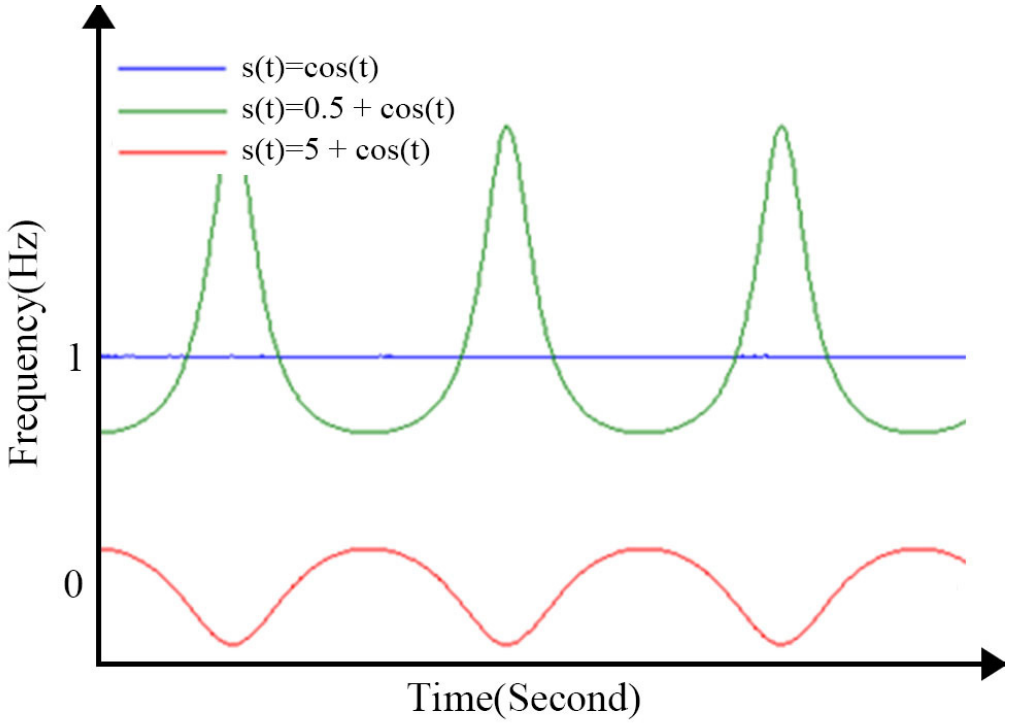

Figure 2.4 Instantaneous frequency

Therefore, we cannot apply signals observed from the real world by HT because HT assumes the signal as $A(t)cos(\omega(t)t)$ to calculate the instantaneous frequency. To solve this problem, empirical mode decomposition (EMD) was proposed by Huang [8, 10]. First, EMD decomposes the signal in an empirical way, which is removing β and return the signal back to the origin. Then, the decomposed signal can be considered as a pseudo monochromatic wave. Thus, we can apply HT to the decomposed signals to obtain instantaneous frequency correctly. However, since there is no mathematical proof to guarantee the chromatic signal can be always decomposed into monochromatic signals, the procedure for cutting β is called EMD and the method is only "empirical."

2.3.1 Intrinsic mode functions and trend

Here, EMD decomposes a chromatic signal into a set of pseudo monochromatic waves so-called the Intrinsic Mode Functions (IMFs) and a residual so-called "trend" [8, 10]. Thus, the real-world observed signal $x(t)$ can be defined as follows:

$$x(t) = \sum_{n} c_n(t) + r(t) \qquad (6)$$

Here, $c_i(t) | \, i = 1, \ldots, n\}$ is the set of IMFs, $r(t)$ is the trend [8, 10]. The definition of Intrinsic mode function is as follows [8, 10]:

- The number of signal extrema and the number of zero crossings points are equal, or their difference is 1; and
- At any time, the average value of the envelope made by the maximum, and the envelope made by the minimum is zero.

After a real-world data decomposed into several IMFs, their instantaneous frequency and instantaneous amplitude can be respectively obtained by HT.

2.3.2 One-variable EMD

Here, we introduce EMD that is the key process of decomposing real-world signals to nonlinear functions entitled IMFs [8, 10]. The algorithm of EMD for a one-variable signal is as follows [8, 10]:

1. Calculate residual $r(t)$ (Let $r(t) = x(t)$ in the first time);

$$r(t) = x(t) - \sum_{n} c_n(t) \qquad (7)$$

2. Initialize $c(t) = r(t)$and extract the IMF
   a) Find maximum envelope $u(t)$ and minimum envelope $l(t)$ of $c(t)$ using cubic spline functions
   b) Subtract the average envelope from $c(t)$

$$c_{new}(t) = c_{old}(t) - \frac{u(t) + l(t)}{2} \qquad (8)$$

   c) If the convergence condition $(0.3 \geq SD \geq 0.2)$ is satisfied, add $c(t)$ into the IMF set; and

$$SD = \sum_{n} \frac{(c_{old}(t) - c_{new}(t))^2}{{c_{old}}^2(t)} \qquad (9)$$

3. Repeat step 1 and 2 to extract all IMFs

2.3.3 Multivariate EMD

The multivariate EMD (MEMD) is proposed for multi-channel or multivariate signals [13]. MEMD uses Quaternion to create the n dimensional sphere and project multivariate signals onto the sphere, to decomposes the signals in MEMD. By finding maximum and minimum envelope covering the n dimensional sphere, we obtain the average envelope and subtract from the original multivariate signals. Using this algorithm, we can decompose multivariate signals like human motions. Here, MEMD Algorithm is as follows [13]:

1. Perform a n dimensional sphere projection properly;

2. Make the projection $\{p^{\theta_k}(t)\}_{k=1}^{K}$ of the input signal $\{v(t)\}_{t=1}^{T}$ for all of k based on the directional vector X^{θ_k};

3. Determine the maximum positions from the projection $\{p^{\theta_k}(t)\}_{k=1}^{K}$;

4. Create multi-dimensional envelopes $\{e^{\theta_k}(t)\}_{k=1}^{K}$ from $[(t_i^{\theta_k}), v(t_i^{\theta_k})]$;

5. Calculate the mean $m(t)$ from the directional vector of all k;

$$m(t) = \frac{1}{K}\sum_{K=1}^{K} e^{\theta_k}(t) \qquad (10)$$

6. If the convergence condition of $d(t) = x(t) - m(t)$ is satisfied, add it into the IMF set; and

7. Repeat steps 1-6 until extracts all IMFs.

2.3.4 Noise-assisted multivariate EMD

Due to the Filter bank property of MEMD, we can improve the accuracy of decomposition using white noises. noise-assisted multivariate EMD uses this property by adding a gaussian white noise (GWN) into MEMD [16]. It is known that NA-MEMD is the most accurate decomposition method among various EMDs [16]. The algorithm can be expressed as follows:

1. Add m-channel GWN to n-channel multivariate signals. According to the principle of thumb, the best amplitude of GWN is about 8% to 10% [16];

2. Decompose $n + m$ multiple signals using MEMD; and

3. Remove m-channel GWN.

Motion capturing data are multivariate signals. For example, the human skeleton has multiple joints with a root position. For multi-channel or multivariate analytical signals, it is impossible to use the univariate EMD to decompose multivariate "chromatic" signals. Then, we take all joint angles into NA-MEMD to decompose motion capturing data.

2.4 Spline implementation for EMD

All decomposed IMFs are determined by spline functions because local means are calculated by spline envelopes [8]. According to empirical statistics, it is found that higher-order splines need more subjective parameters, which violates the adaptive knowledge of the method. In addition, higher-order spline functions are also more time-consuming in the calculation. Therefore, the cubic spline is chosen to make the envelope. Furthermore, the convergence of EMD is also a key problem [8]. All intuitionistic reasoning and experience show that the process of IMF decomposition by EMD is convergent. But at present, there is no strict proof whether the envelope made by cubic spline can converge [8].

Although the research of Chen et al. [26] shows that the B-spline curve can be used to shift and reduce the extremum, which can solve some problems such as flat peaks, this method still needs to be proved clearly. Therefore, for the method of making an

envelope, at this stage, the cubic spline function is the most appropriate only based on experience [8]. Whether the higher-order spline function can play a more effective effect remains to be further studied.

2.5 Hilbert spectral analysis

After original signals are decomposed by EMD, we can apply HT to each IMF to obtain their instantaneous amplitudes and frequencies. Figure 2.5 shows an example of a Hilbert spectral analysis (HSA) using the instantaneous amplitudes and frequencies of all IMFs [8]. As shown in Figure 2.5, the results are displayed in time in the horizontal axis, frequency in the vertical axis, and the amplitude are presented by color.

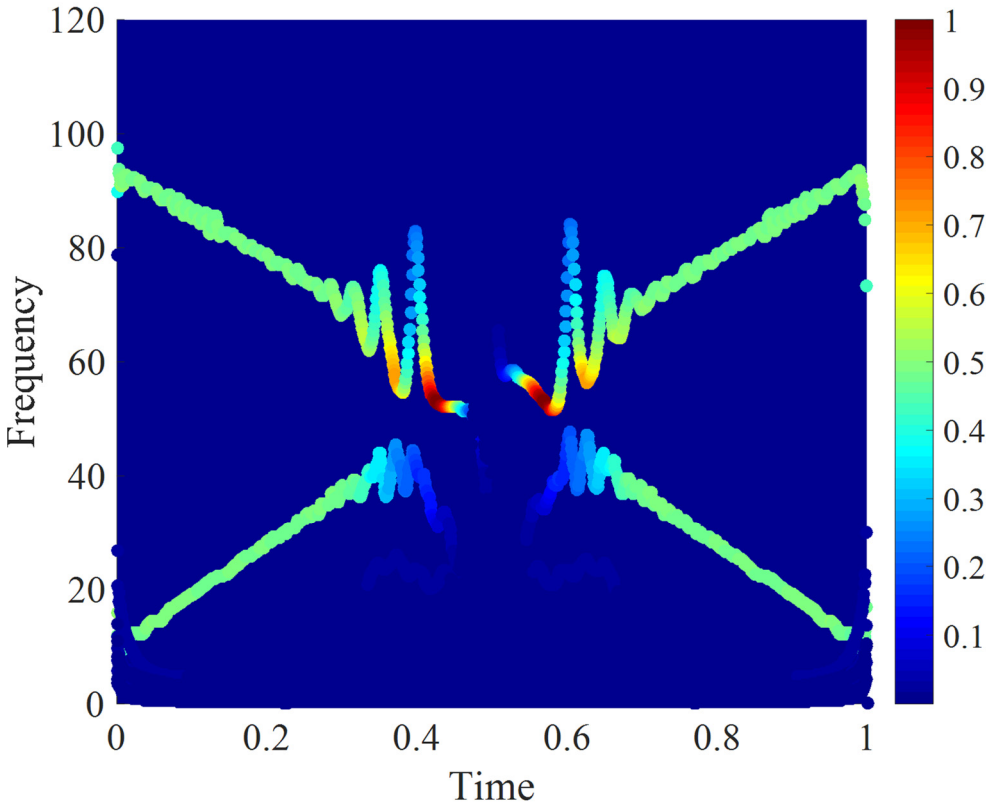

Figure 2.5 An example of HSA

Figure 2.5 shows a signal $x(t) = \cos(90\pi^2 t + 20\pi t) + \cos(-90\pi^2 t + 200\pi t)$ composed of two monochromatic waves. As shown in the figure, the instantaneous frequency is obtained clearly in HSA.

Considering the spectral symmetry of frequency conversion, twice the maximum frequency is sufficient for the sampling frequency. However, since the instantaneous frequency of HHT includes calculation of derivative, the differential approximation method becomes a bottleneck in analyzing accuracy when the sampling frequency is too low. For example, when the fourth-order differential method is used in the HHT of the signal $x(t)$ in Figure 2.5, it is necessary for convergences to use a sampling frequency about four times higher than the maximum frequency of this signal.

2.6 Weighted average frequency algorithm

Boashsah shows that smoothing is possible if the instantaneous frequency is changing slowly enough [27]. Thus, the instantaneous frequency of each IMF obtained by HT can be smoothed. Niu et al. [28] developed a weighted average frequency algorithm (WAFA) that performs a smoothing operation for each IMF. The algorithm uses the instantaneous amplitude and instantaneous frequency obtained for each IMF decomposed by EMD and smoothes the instantaneous frequency in three parts by the specified window length as shown in Table 2.3.

Table 2.3 The equations corresponding to each window

k	Average frequency equations
$1 \le k \le \frac{(m-1)}{2}$	$\bar{\omega}_j(k) = \frac{\sum_{n=1}^{k+(m-1)/2} \omega_j(n) A_j(n)}{\sum_{n=1}^{k+(m-1)/2} A_j(n)}$
$\frac{(m+1)}{2} \le k \le \frac{(2N-1-m)}{2}$	$\bar{\omega}_j(k) = \frac{\sum_{n=k-(m-1)/2}^{k+(m-1)/2} \omega_j(n) A_j(n)}{\sum_{n=k-(m-1)/2}^{k+(m-1)/2} A_j(n)}$
$\frac{(2N-1+m)}{2} \le k \le N$	$\bar{\omega}_j(k) = \frac{\sum_{n=k-(m-1)/2}^{N} \omega_j(n) A_j(n)}{\sum_{n=k-(m-1)/2}^{N} A_j(n)}$

Here, $k = 1,2,\dots,N$ (N is the number of data), n indicates sampling data, ω indicates the instantaneous frequency of the decomposed IMF, A indicates the instantaneous amplitude of the decomposed IMF, and m is the length of the window.

Because human motion is also treated as a signal, previous research has stated

that WAFA can also be applied to human motion data [28]. Thus, WAFA can be used for smoothing IMFs of motion capturing data decomposed by NA-MEMD. Therefore, in the analysis of dance motion, it is necessary to take an average frequency for a series of motions in beat units or keyframe units.

Chapter 3

Dance motion analysis and editing using Hilbert-Huang transform

This chapter is organized as follows. Section 3.1 introduces the research background of motion analysis and editing regarding the problems faced in the instantaneous frequency domain. Section 3.2 describes the proposed framework of dance motion analysis using HHT. Section 3.3 shows the results of the proposed framework from section 3.2. Section 3.4 discusses the results shown in section 3.3. Section 3.5 describes the proposed framework of dance motion editing using HHT. Section 3.6 shows the results of the framework from section 3.5. Section 3.7 discusses the results shown in section 3.6. Section 3.8 explains the limitations of our proposed framework. Section 3.9 summarizes and concludes this chapter.

3.1 Research background

Recently, many motion analyses and syntheses researches have been performed because of developments in motion capture technology. Analyzing and editing dance motions is especially difficult, because its motions are complicated, and the dance data are very noisy due to the motion capture system. Meanwhile, the use of Hilbert-Huang transform (HHT) is becoming more and more popular among researchers because HHT is suitable for multi-channel data. Therefore, HHT can be a very powerful tool to analyze and edit multi-channel data such as motion data in the instantaneous frequency domain.

In this dissertation, we propose a framework using HHT with beat tracking in the instantaneous frequency domain. We propose two frameworks for analyzing and editing dance motions separately in the instantaneous frequency domain. As an example, first, we analyzed the dance motions of Japanese techno-pop group Perfume, salsa, waltz, and hip-hop using our proposed framework. Next, by our

proposed framework, we extracted two different Perfume choreographies and blended them separately into the salsa dance. Thus, our proposed framework can help dancers and animators to analyze and edit dance motion in the instantaneous frequency domain.

3.2 Proposed dance motion analysis framework

There have been many researches about analysis and synthesis of dance motions. Li et al. [29] suggested a motion texture method for synthesizing complicated dance motion techniques using a statistical model. Shiratori et al. [30] proposed a dance motion analysis framework based on a musical rhythm by splitting motion into each motion primitive. Shiratori et al. [31] also used the proposed analysis framework [30] to program the dance motions of humans into a humanoid robot. Chan et al. [32] proposed a dance training system with motion capture and virtual reality (VR) technologies. However, most dance motion researches have only been discussed in the time domain. To analyze features of different dance styles more clearly, we propose a dance motion analysis framework using HHT.

3.2.1 Dance motion analysis framework using HHT

In this section, we propose a method to calculate the instantaneous frequency from capture data of dance motion using HHT and beat tracking. Figure 3.1 shows the proposed framework of dance motion analysis.

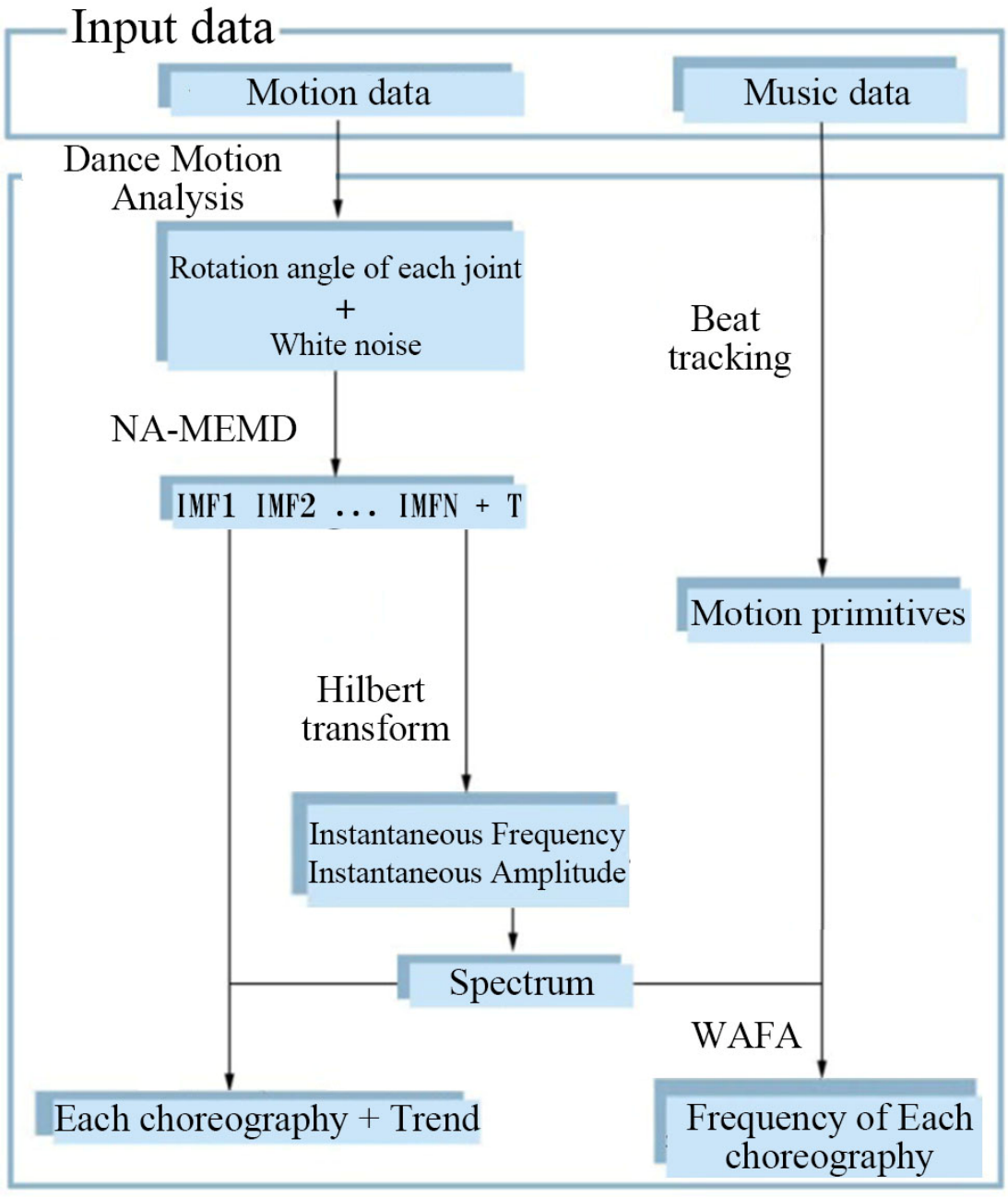

Figure 3.1 The framework of dance motion analysis using HHT

We briefly explain the proposed framework as follows:

(1) Prepare 3 Euler angles θ_x, θ_y, θ_z of each joint obtained from the dancer's body. Here, the number of channels can be obtained by (number of dancers) × (degrees of freedom) × (number of joints) + (Gaussian noises). For example, in the case of the Japanese dance group Perfume, there were 3 members, 3

degrees of freedom of each joint, 23 joints per person, and 1 Gaussian noise, giving 208 channels;

(2) Apply NA-MEMD to all prepared data to obtain the N-mode IMFs covering the N-dimensional (N=208 for Perfume) sphere, and a trend;

(3) Output these IMFs and the trend back as motion data, and confirm if the vibration components (choreographies) are completely separated;

(4) Check whether the N-mode has been sequentially decomposed from high frequency to low frequency and whether there is any singularity. If the IMFs are not in a frequency order of decompositions, and a higher frequency than the previous one appears (this is the so-called "singular IMF" [33]), fix the singular IMF in sequence;

(5) Apply HT to each IMF to obtain the instantaneous frequencies and amplitudes.

(6) Apply WAFA to smooth the instantaneous frequencies of each IMF using the beat window calculated by beat tracking; and

(7) Analyze dance motion using Hilbert spectrum plotted with smoothed instantaneous frequencies and instantaneous amplitudes, and perform HSA in both long team (whole dance) and short term (motion primitive) as necessary.

3.2.2 Motion primitive extraction with beat tracking

The motion primitive is the most basic motion unit of a dance. There have been many studies on the extraction and segmentation of motion primitives from dance motion capture data [30, 31, 34]. Dance motion is often synchronized with the rhythms of the music [35]. Therefore, it is possible to segment motion primitives by detecting the beats. Dan Ellis [36] proposed a beat tracking method by first assuming a certain fixed tempo, beats per minute (BPM), exists in the music. Second, using the present beat detected from music with the BPM, the next beat position is assumed. Third, the assumed beat position is verified and determined by calculating the relationship between the detected beat position and the assumed beat position [36]. This beat tracking method is suitable for extracting beats of western music with tempo

intervals [36]. In our research, we extracted motion primitives using this beat tracking algorithm. Then, the average frequencies of a set of motion primitives can be obtained by WAFA using beat positions and their time intervals.

3.2.3 Dance motion weighted average frequency

In this research, we used the beat interval obtain by beat tracking as the window size for WAFA to calculate the average frequencies of a dance motion. In dance motion analysis of HHT, all choreographies are periodically decomposed into nonlinear monochromatic waves $A(t)cos(\omega_0(t)t)$ according to the beat. Thus, for some monochromatic waves which have no choreography performed, their amplitude $A(t)$ of IMF is almost zero.

3.2.4 Input data and joint angles θ_x, θ_y, θ_z

As shown by the example elbow joint in Figure 3.2, each joint of the body moves with 3 degrees of freedom θ_x, θ_y, θ_z.

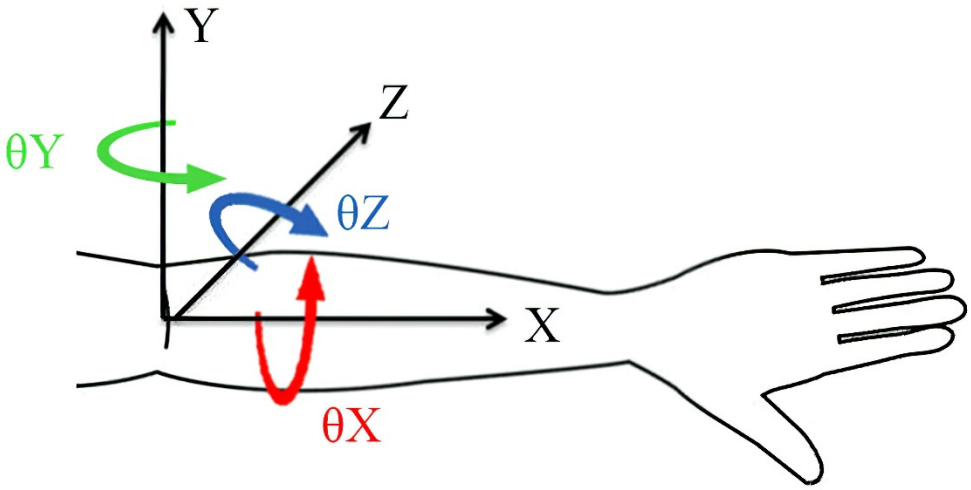

Figure 3.2 Example of Euler angles of an elbow joint.

In dance performing, each motion primitive is considered to move at an Euler angle θ_x, θ_y, θ_z. Therefore, if only one motion primitive is measured, it is enough to analyze instantaneous frequencies only at one of degrees. However, a series of dance performances consist of multiple motion primitives. As a result, different degrees of each joint angle are used for different choreographies. Therefore, when decomposing the whole dance motion, even when a joint angle has a small amplitude,

it is necessary to decompose the joint angle with 3 degrees of freedom θ_x, θ_y, θ_z and analyze all angles. The results are shown in section 3.3.

3.2.5 Intrinsic mode function and trend analysis

A dance motion generally has a basic posture that maintains its style. For example, there is a comparative study of basic posture between oriental dance and western dance [37]. Therefore, by the definition of EMD, each IMF corresponds to dance choreographies as vibration components. Then, the trend (residual) without any vibration components corresponds to the basic posture of the dance.

After decomposed all joints 3 Euler angles by NA-MEMD, we can write out each IMF and trend (residual: the basic posture of dance) according to skeleton structure (e.g. BVH) for analysis. Please notes that, in IMF motion analysis, it is necessary to add the trend to all IMFs to reconstruct each choreography.

3.3 Dance motion analysis using the proposed framework

In this section, we report the results of the dance motion analysis using the proposed framework in section 3.2. In subsections 3.3.1 and 3.3.2, Perfume dance motion was analyzed as an example in a short-term (about 3 seconds). In subsection 3.3.3, based on the results in subsections 3.3.1 and 3.3.2, we analyzed four types of dances, Perfume, waltz, hip-hop, and salsa dance in a long term (about 30-60 seconds). In subsection 3.3.4, we used Perfume dance to show the trend analysis by comparing the difference between the three members.

3.3.1 The skeleton structure of Perfume motion data

As a first example, we analyze BVH data for about 1 minute of the three female techno pop units named Perfume, from the internet [38] in Table 3.1. The reason for using this data is that there is motion capture data for three people in one dance, which is suitable for decomposed by NA-MEMD as a multi-channel data. Besides, music beats are relatively clear, and segmentation is simple.

Table 3.1 shows the details of the motion and audio data. As a result of beat tracking, the music “Enter the Sphere” (composition: Yasutaka Nakata) has 64 measures, and its BPM is 130 [39]. One measure is about 1 second, and the time interval between strong and weak beat is about 0.5 seconds.

Table 3.1 Perfume dance motion capture data

	Motion (BVH)	Audio (WAV)
Time(s)	70.5	64.2
Sampling rate (Hz)	40	44100
Data number	2820*3	2833310*2
Number of joints	23	—
Beat number	—	128

Figure 3.3 shows the skeleton model of these BVH data.

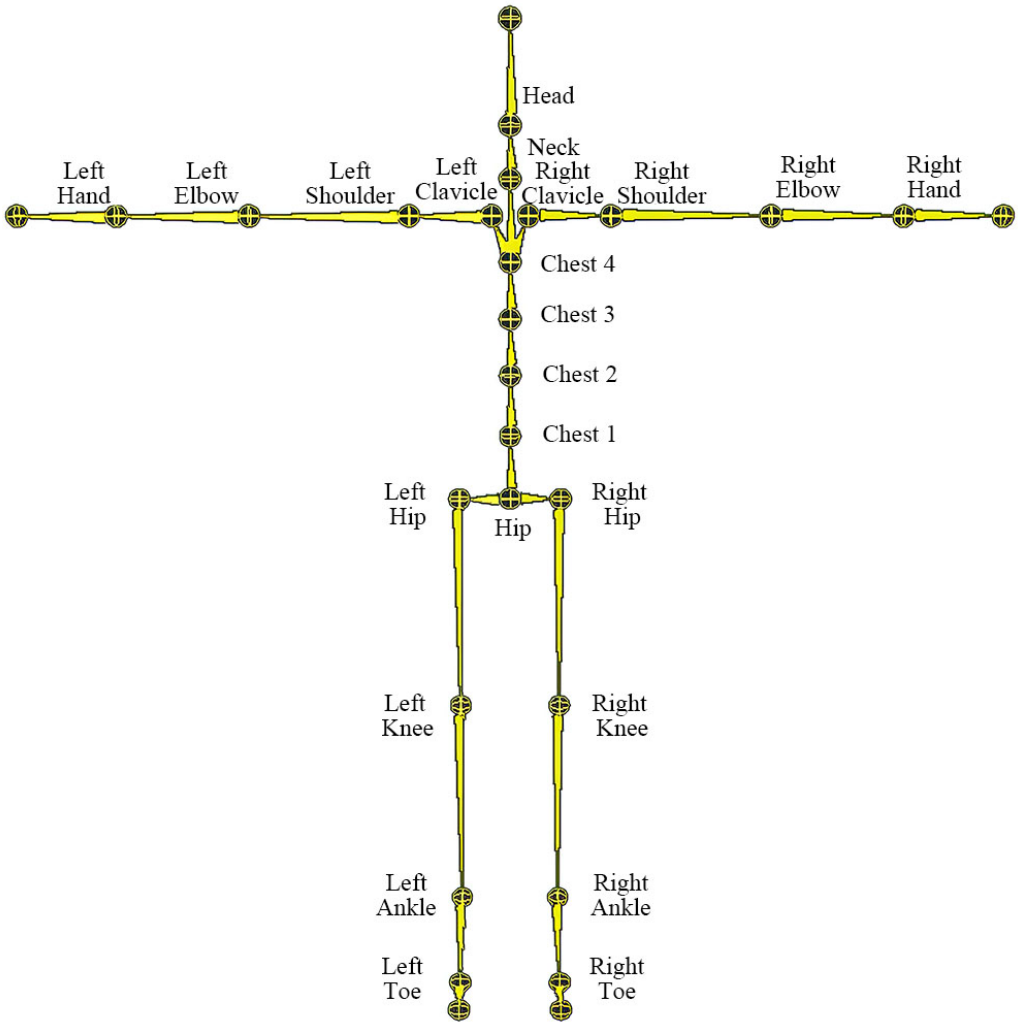

Figure 3.3 T-pose joint structure of Perfume BVH

The dance motion was recorded at 23 joints throughout the body. The figure shows the origin pose (T-pose) for which all joint angles θ_x, θ_y, θ_z are zero. Here, the skeleton model used a hierarchical recording method in which the root origin is placed on the hip.

3.3.2 A Hilbert spectral analysis example of Perfume turning motion

A dance turning motion is a large rotation/movement occurring within a short time, and it is difficult to detect the start and end positions of the movement. Therefore, we apply the proposed framework to the turning of Perfume dance for 3 seconds shown in 3.4. This analysis focused on the motion primitives by Ayaka Nishiwaki, the center dancer.

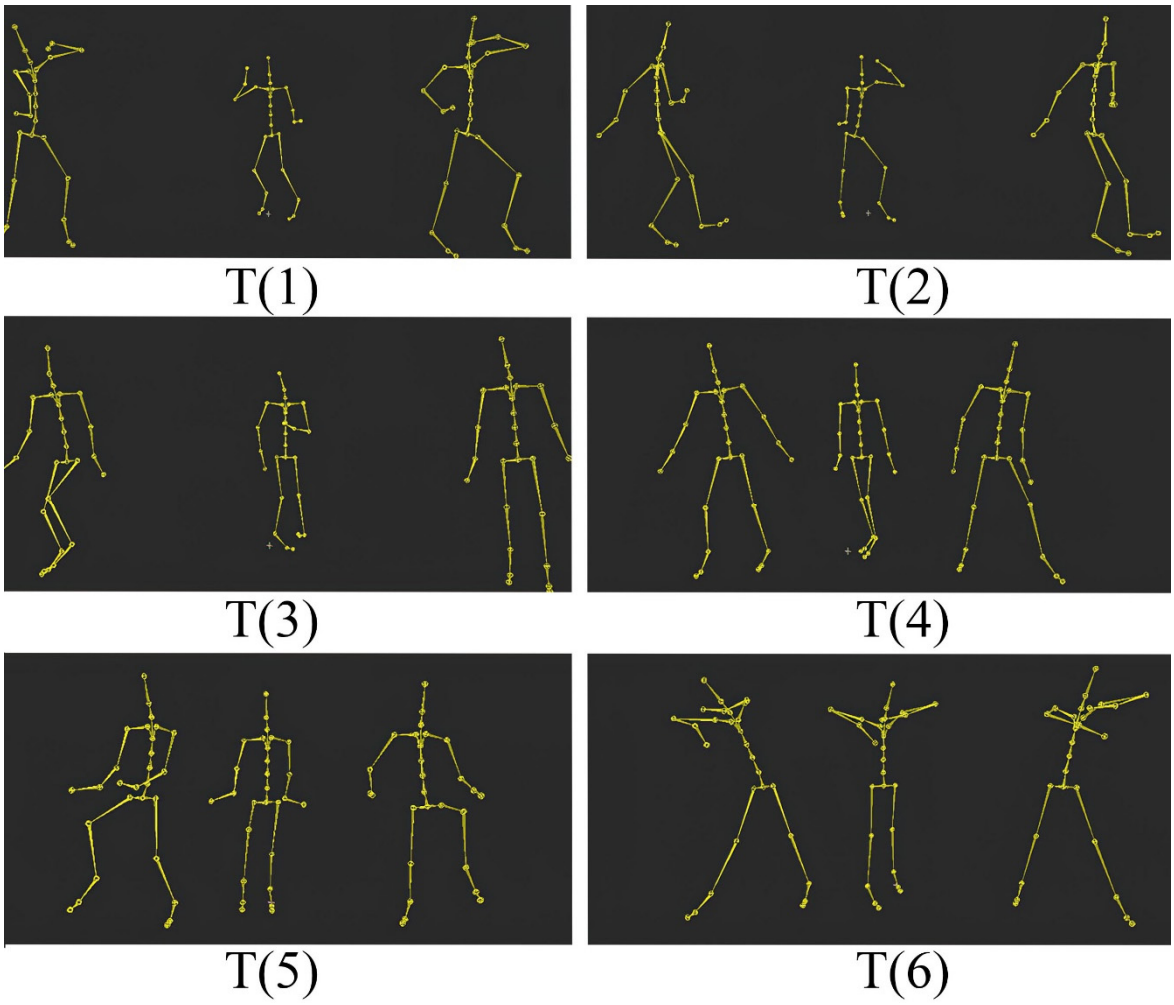

Figure 3.4 The turning dance motion by Perfume

Figure 3.4 shows a dance turning motion from about 56 seconds to 59 seconds in this dance. All three dancers rotate for the same 1 second between T (2) and T (4). Here, NA-MEMD and HSA were performed on these motion data.

In order to analyze the motion of Figure 3.4 [38] in units of motion primitive, we used beat tracking described in Section 3.2.2, to detect all beats from the music. Table 3.2 shows the time positions of strong and weak beats. Six beats were detected in about 3 seconds.

Table 3.2 Positions of beats (55s～59s).

Weak beat	Strong beat
56.28s（START）	56.74s
57.21s	57.67s
58.13s	58.59s（END）

Figure 3.5 shows the corresponding x, y, and z-axis definitions for the hip joint angle θ_x, θ_y, θ_z, which is the root of the hierarchical skeleton model in Figure 3.3.

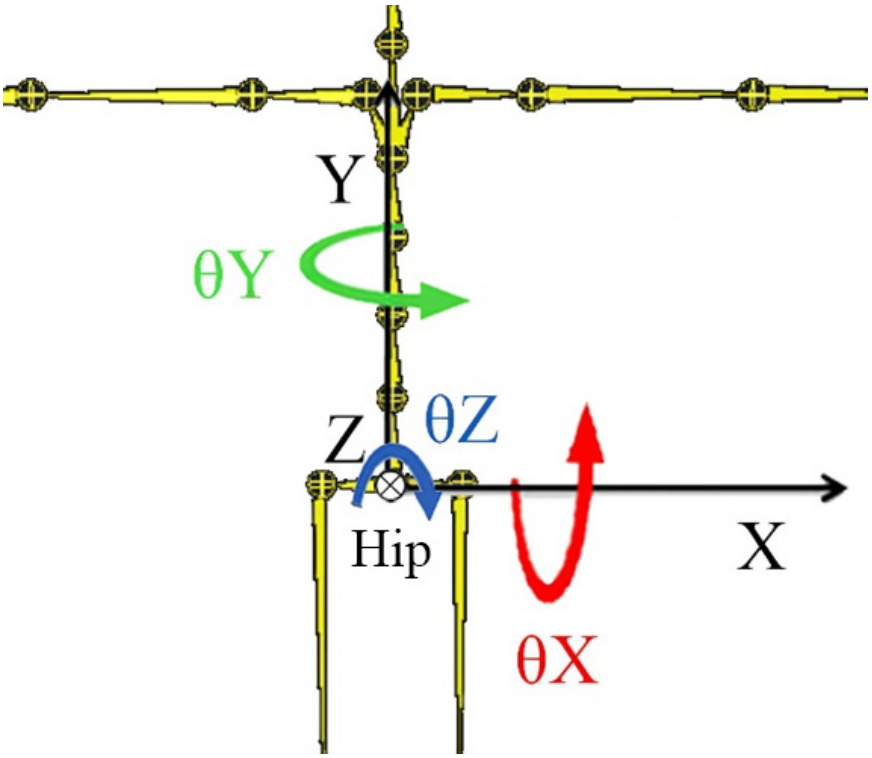

Figure 3.5 Euler angles of the Perfume hip joint

Here, θ_x indicates the rotation of the choreographies the body front and back, θ_y indicates the rotation that arounds the center axis of the body, and θ_z indicates the rotation that choreographies the body left and right.

Figure 3.6 is a three-dimensional plot of the hip joint θ_x, θ_y, θ_z in the analysis interval. The green dots indicate the position of the strong beat, and the red dots indicate the position of the weak beat. The black dots indicate the start and end positions of the motions. Beats are located near the rotation point, which means that dance motions are switched in the beat positions.

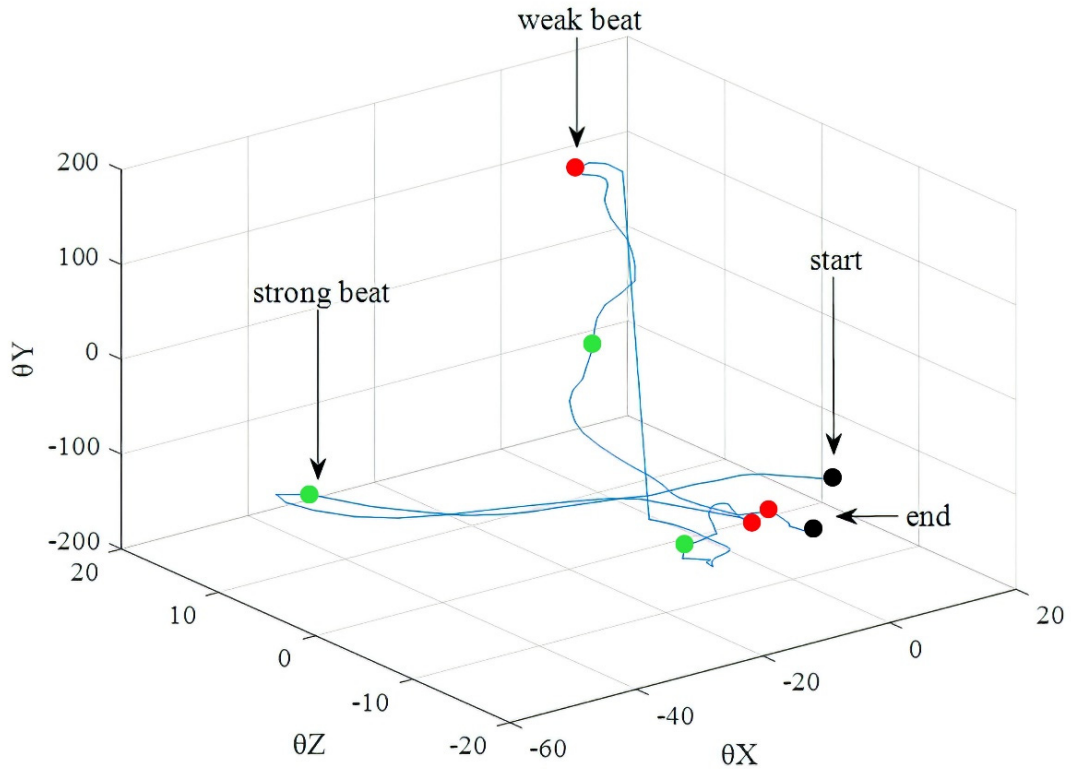

Figure 3.6 The 3-dimensional dance motion trajectories $\theta_x, \theta_y, \theta_z$ by Ayaka Nishiwaki (hip)

Figure 3.7 shows the results of NA-MEMD decomposition of the Euler angles θ_x, θ_y, θ_z from Figure 3.6. The three-dimensional signal of the hip was decomposed into IMF1-6 and one trend. The original signal was decomposed from high to low frequency. The main turning movements IMF1-3 were mainly composed of θ_y, θ_z directions. The major steps in the second half of the turning motion were IMF5 and 6, which were composed of θ_x, θ_z directions. And IMF4 was the middle of both

movements.

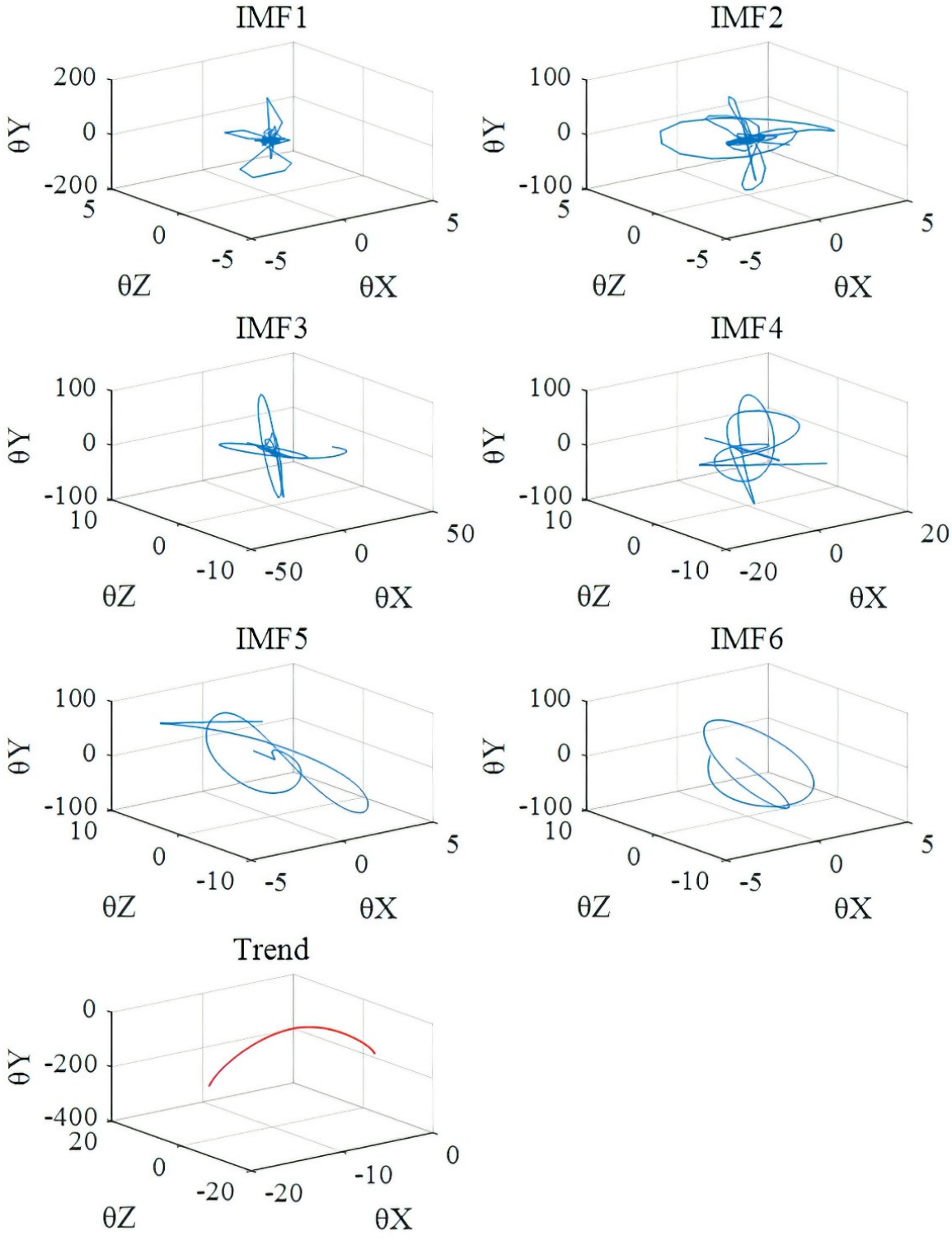

Figure 3.7 NA-MEMD decomposition of the dance turning motion.

Figures 3.6 and 3.7 show that the Perfume turning dance motion was decomposed into several different frequency IMFs. In order to understand this complicated motion, it is necessary to perform HSA. Because the movements are concentrated in the hips, we focused on the hip joint to analyze decomposed motion with Hilbert spectrum.

Figure 3.8 shows the Hilbert spectrum of the dance motion from Figure 3.4. The white line represents strong beats, and the red line represents weak beats. From the results shown in Figure 3.8 (c), NA-MEMD decomposes a high-speed turning dance motions with IMF 1-3. For example, the high frequency turning IMF 2 and 3 were detected around 4.5 Hz for 57-57.5 seconds. The IMF of other, related motions were also decomposed clearly.

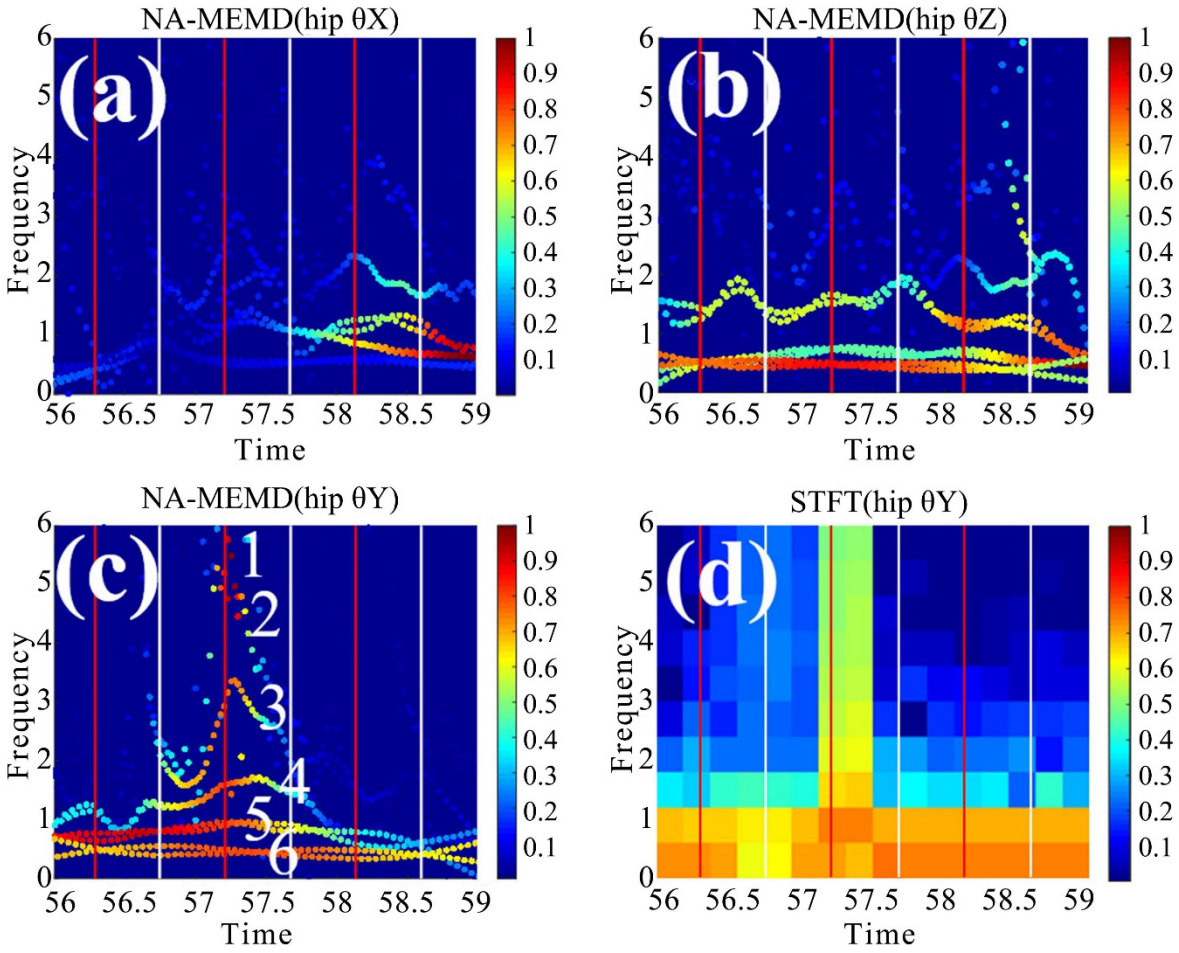

Figure 3.8 The turning dance motion of Perfume. The hip joint spectrogram of (a) θ_x, (b) θ_z, (c) θ_y by HHT, and (d) θ_y by STFT

Figure 3.8 (d) shows the STFT spectrum corresponding to (c), for comparison. In

Figure 3.8 (d), the STFT could not increase the time resolution and frequency resolution simultaneously due to the uncertainty principle. The Hilbert spectrum of Figure 3.8 (c) showed the decomposed signals of each motion are more clearly than the STFT spectrum. Thus, the HHT is more suitable than the STFT in motion analysis field.

3.3.3 Comparison with other dance motions using Hilbert spectral analysis

In this subsection, we use HSA to make a simple comparison of four dances of available data: Perfume, waltz, hip-hop, and salsa. We discuss how the characteristics of each dance can be visualized by our proposed framework using HHT. We do not discuss the four types of dance as a general remark but discuss the characteristics of each dance that can be clarified in the analysis of HHT and beat tracking proposed in this study.

Table 3.3 shows the duration (excluding unrelated actions, such as bows), minimum frequency (1 / time length), and BPM of each dance.

Table 3.3 Four sets of dance motion data for comparisons

Dance motion	Perfume	Waltz	Hip-hop	Salsa
Time(s)	57	63.3	75	27.3
Minimum frequency (Hz)	0.017	0.016	0.013	0.037
BPM	128	80	90	130

The waltz dance motion data was a C3D data converted to BVH from the motion capture database [40]. For hip-hop, we used data published on a website from [41]. The type of dance is not mentioned on the website, but the data was determined to be hip-hop from the opinions of a dance expert (private communication). The salsa BVH data was obtained from the Internet [42]. The BPM of each dance was calculated by beat tracking using the corresponding music. The BPM of the dance without music data was estimated from the average tempo of their music [43-45]. In this comparison, we use the whole dance motion shown in Table 3.3 to HSA. In order to compare results under the same conditions, we unified the frame rate to 30Hz, and reconstruct the hierarchical skeleton model to the same number of joints.

Figure 3.9 shows the HSA results of the four dance motions listed in Table 3.3.

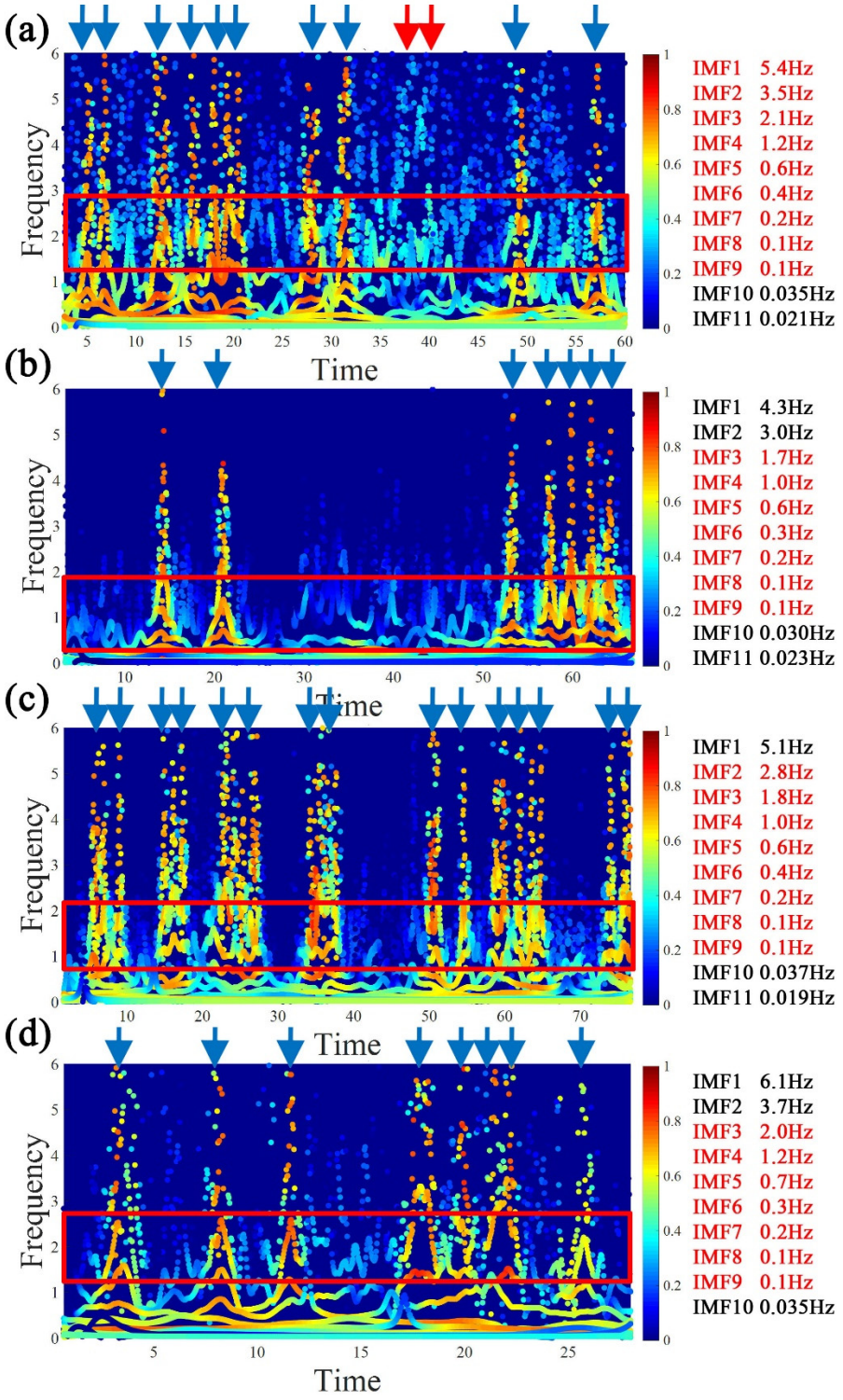

Figure 3.9 NA-MEMD Hilbert spectrum of each dance style (hip θ_y). (a) Perfume (b) waltz (c) hip-hop (d) salsa

In this comparison, we analyzed the Hilbert spectrum of the rotation angle θ_y of the hip joint (subsection 3.3.2), which contains the turning motion, for example. Because the turning motion spectrum intensity was relatively large, the spectrum amplitude was normalized to a range of 0-1 by taking the natural logarithm. This allowed other motions to be visualized. Also, the beats were too many to be plotted, so they were not displayed in the Hilbert spectrums. The average frequency of each IMF is displayed on the right side of the spectrums. In IMF1-9, the frequencies are rounded to the first decimal place. For IMF 10 and larger, the frequencies are rounded to the third decimal place. The blue arrows at the top of Figures 3.9 (a)-(d) indicate the turning motion described in Section 3.3.2. The red box in the figure shows the main motion (BPM / 60 ± 0.5Hz) synchronizes to the beats of each dance. The spectrum (light blue to 0.4) indicated by the two red arrows in Figure 3.9 (a) shows a stepping motion in which the knees and toes of Perfume are alternately turned in and out.

Figure 3.9 (a) is a Hilbert spectrum of Ayaka Nishiwaki, the center dancer. There are 10 high-frequency peaks associated with intense turning motion of the body. Compared with other dances, the Perfume dance clearly shows a spectrum intensity of 0.4-0.6 in a frequency range of 0.1-6 Hz. Even in the main motion part (2.1 ± 0.5Hz, red box in the figure), the strong spectral intensity of 0.4-0.6 appears to be significantly different compared to the other dances. Figure 3.9 (b) is the waltz Hilbert spectrum. Seven spike-like turning motions had peak values of only about 4 Hz compared to the other dance styles. The overall frequency was concentrated in the range of 0.1-1.0 Hz. This because the BPM (80) of waltz was the lowest of the four dances. The hip-hop in Figure 3.9 (c) had 14 spike-like turning motions, which shows that hip-hop choreographer incorporates more turning motions than other dances. In the salsa Hilbert spectrum in Figure 3.9 (d), there were 8 spike-like high-frequency peaks. The top frequency has reached 6 Hz indicating that its turning motion is in higher speed than other dances. In addition, in the main motion part (2.1 ± 0.5Hz, red square frame in the figure), which corresponds to the beats, the spectrum intensity of 0.4-0.6 Hz was remarkable. As with Perfume, the overall frequency was higher than waltz and hip-hop. In addition, the IMF spectrum intensity of 0.5-1.5 Hz, which seems to be a low-frequency step, was as strong as 0.7-0.8 Hz.

What should be noted here is the relationship between the average frequency of each IMF on the right-hand side of Figure 3.9. The IMF frequency f_i in red observed the relation $f_{i+1} = f_i + f_{i-1}$, which is the Fibonacci sequence (the error was about ±

5%). This is not a coincidence because all four dances were observed at the average frequency of the major IMFs. In the analysis of the four dances, the Perfume dance had the most numerous Fibonacci sequence relation between IMF 1-9 with more IMFs than the other dances.

The number of IMFs that made up the Fibonacci sequence (9), is same as the maximum number of the hierarchical levels (9) shown in Figure 3.3. Then, it can be inferred that the Fibonacci relation of the IMF average frequency was related to the skeleton hierarchy of the human body model. The hip is the joint in which human movements are most concentrated. Thus, we can observe the Fibonacci relation from the hip joint, because the motions of other joints are easily added to the hip joint as noise. In subsection 3.5.2, we discuss in-depth the relationship between the Fibonacci sequence and the human body and use this relationship for dance editing.

3.3.4 Trend analysis example of the Perfume dance

Figure 3.10 shows the result of a trend analysis for three Perfume dancers of whole dance motions (about 60s). By outputting the trend back to BVH data according to skeleton structure. The trends of three dancers are moving very slowly. This means that all dance motion choreographies were decomposed as IMFs, which are nonlinear vibrations, and the basic posture of dance (Figure 3.10) was decomposed as the trend.

By comparing the three trends, Ayaka Nishiwaki, the center dancer, rotates in the right-hand direction, while the other two members rotate in the left-hand direction. This because the choreographer Mikiko designed different dance basic postures between the center dancer and others.

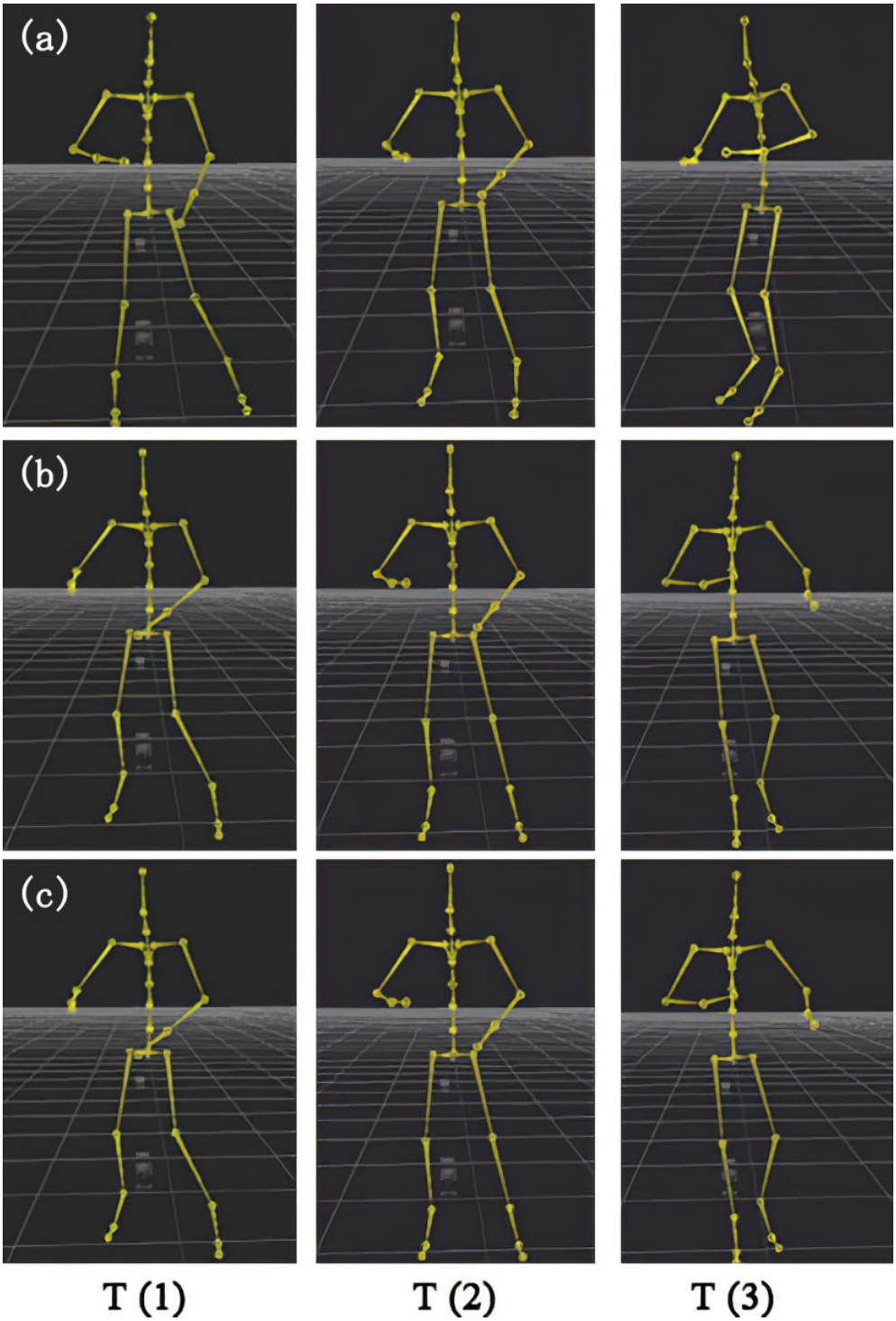

Figure 3.10 Perfume dance trend analysis (a) Nishiwaki (b) Omoto (c) Kashino

3.4 Discussion of dance motion analysis

By our proposed framework, dance motion can be decomposed into IMFs with a trend (dance posture). In so doing, the dance choreographies are segmented into several IMFs corresponding to the joints in the instantaneous frequency domain. Our framework can be a useful tool to analyze dance motions.

In section 3.3, we propose a dance motion analysis framework. First, we apply the HHT to dance motion data, propose a framework shown in Figure 3.1. Second, the results show that using beat tracking, motion primitives can be captured and segmented with Hilbert spectrum. Third, segmented motion primitives are smoothed by WAFA and the Hilbert spectrum was denoised. Fourth, as shown in Figure 3.7 (c), the proposed framework shows that complicated human dance motions like turning motion can be visualized clearly as several IMFs by NA-MEMD in spectrums than STFT. In addition, the Hilbert spectrum shows that the turning motion can be detected and segmented into about three rapidly changing peak IMFs synchronized with the beat as shown in Figure 3.7 (c). Fifth, in Perfume, waltz, hip-hop, and salsa dance, we found that the IMF average frequencies of 70-80% form a Fibonacci sequence. Sixth, the Hilbert spectrum of the four dances was compared, and the characteristics of each dance and their differences were shown in angle θ_y of the hip joint. Perfume's dance has Fibonacci relations in almost all IMFs. The waltz is composed of low-speed steps and high-speed rotation. The hip-hop has high-speed turning motion. The salsa was found that the high-speed motions and the low-speed steps were mixed. In the next section, we show the proposed framework and results of dance motion editing using HHT.

3.5 Proposed dance motion editing method

Numerous researches have been performed on motion synthesis and editing. Yamane et al. [46] presented a computational technique for generating whole-body human character motions using a constraining and de-constraining method. Fragkiadaki et al. [47] proposed an encoder-recurrent-decoder model for recognizing and predicting the posture of the human body for video and motion capture data. Holden et al. [48] proposed the use of an auto-encoder to extract motion manifold that is the correct motion posture of the human body structure. After extracting the motion manifold, the model can be trained to generate a target motion by training a neural network [1].

However, these previous researches depended on massive training data whose

features were previously known. For example, deep learning [1] has been used for simple motion editings such as walking and jumping. However, for some applications, it is difficult to analyze motions such as dance motion choreographies, because the training data of dance motion cannot be classified easily.

Dong et al. [49, 50] proposed a framework for dance motions analysis and editing using HHT. Dance motion can be decomposed into several choreographies that can be used as the training data for deep learning. However, the decomposed dance motions need to be reconstructed and combined to choreographies manually. To resolve the problem, we propose a framework for dance motion choreographic editing using HHT and the Fibonacci sequence to extract complicated dance choreographies automatically. After this extraction, these extracted choreographies can be exchanged, blended and edited into another dance. Also, these choreographies can be considered as sparsified training data for deep learning.

3.5.1 Dance motion choreographic editing framework

Figure 3.11 shows our proposed dance motion editing framework based on previous research using HHT [49, 50]. In the following subsection, first, we introduce a flow chart of each part of the framework. Then, we discuss in detail the relationship between IMFs and choreographies.

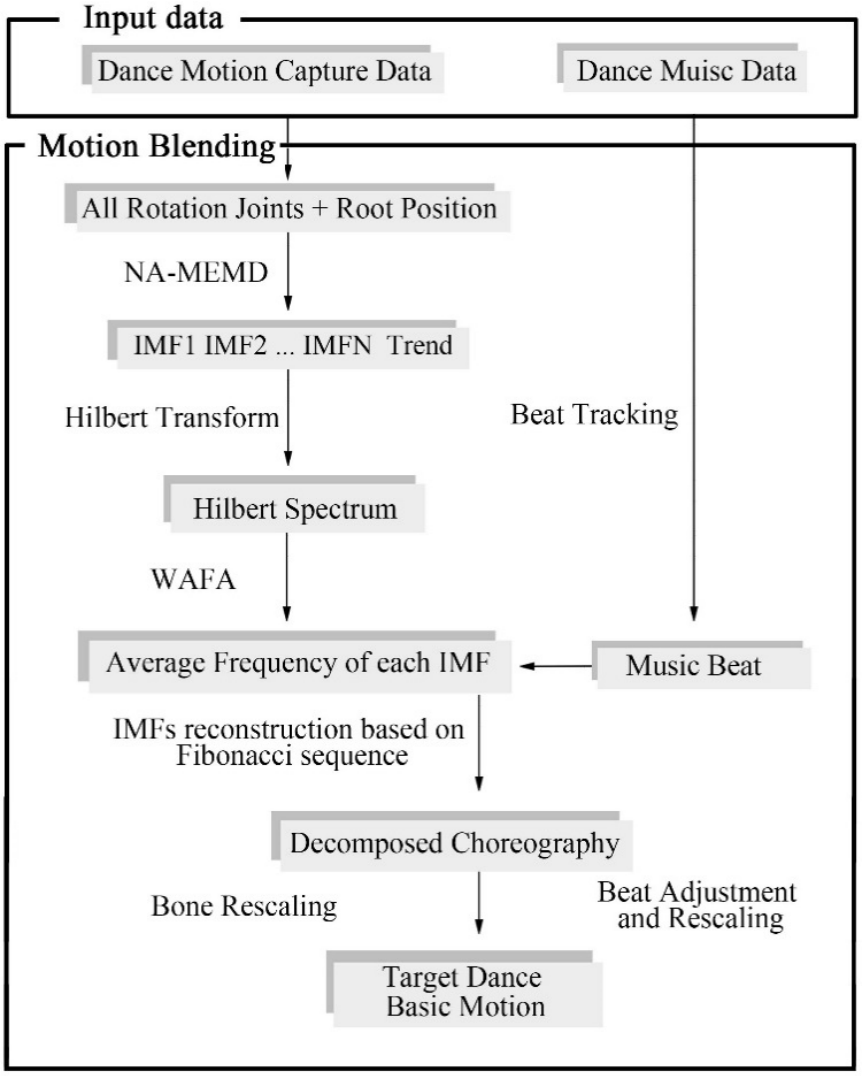

Figure 3.11 Proposed dance motion choreographic editing framework

Our dance motion choreographic editing framework can be summarized as follows:

1. Put dance motion capture data, all of joints Euler angles θ_x, θ_y, θ_z and root position into NA-MEMD as introduced in section 3.2;

2. Apply Hilbert transform to each IMF to obtain the instantaneous frequency and amplitude;

3. Smooth the instantaneous frequency by WAFA using the beat interval as the window length, and obtain the average frequency of each IMF.

4. Use the Fibonacci sequence to reconstruct and combine the high-frequency IMFs to the several correct choreographies, which can be edited; and

5. Extract target dance basic motion from the low-frequency IMF and the trend. Thus, after skeleton and beat adjustment, choreographs extracted in step 4 can be deleted, added, replaced, or blended into the target basic dance motion automatically.

3.5.2 Fibonacci sequence in human joint systems

As we discussed in subsection 3.3.3, the Fibonacci sequence can be observed from the average frequencies of each joint as shown in Figure 3.9. Let the Fibonacci sequence is represented by $i-1$ as the joint to be observed, i as another joint connected to $i-1$, and $i+1$ as the joint connected to it. Let the IMFs vector angular velocities of the joints be constant ω_i, ω_{i+1} as shown in Figure 3.12.

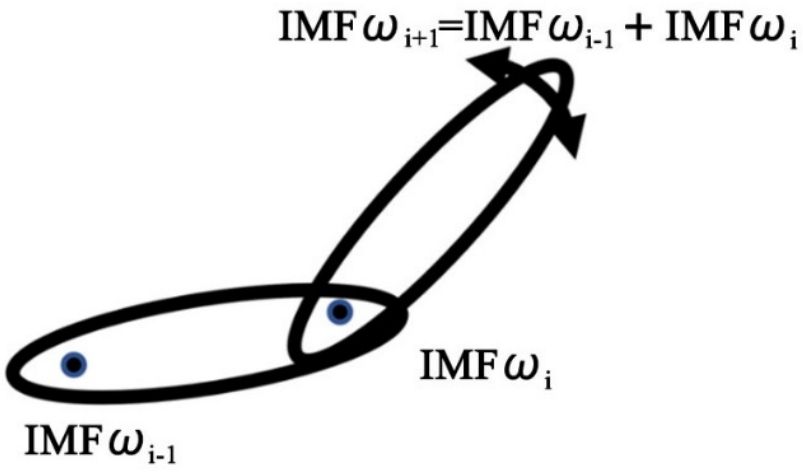

Figure 3.12 Link structure that composes the Fibonacci sequence relation

Thus, the average vector angular velocity of the skeletal element connected to joint $i+1$ viewed from joint $i-1$ is $\omega_{i+1} = \omega_i + \omega_{i-1}$ from the link structure. The angular frequency vector ω_{i+1} is the sum of the previous two angular frequency vectors as equation (11).

$$\omega_{i+1} = \omega_i + \omega_{i-1} \qquad (11)$$

The NA-MEMD decomposes the dance motion into several IMFs corresponding to joints and a posture of the dance as a trend. Thus, the IMF relationship shown in Figure 3.12 can also be assumed from toe to finger in the whole body due to its link-dynamical structure in Figure 3.13.

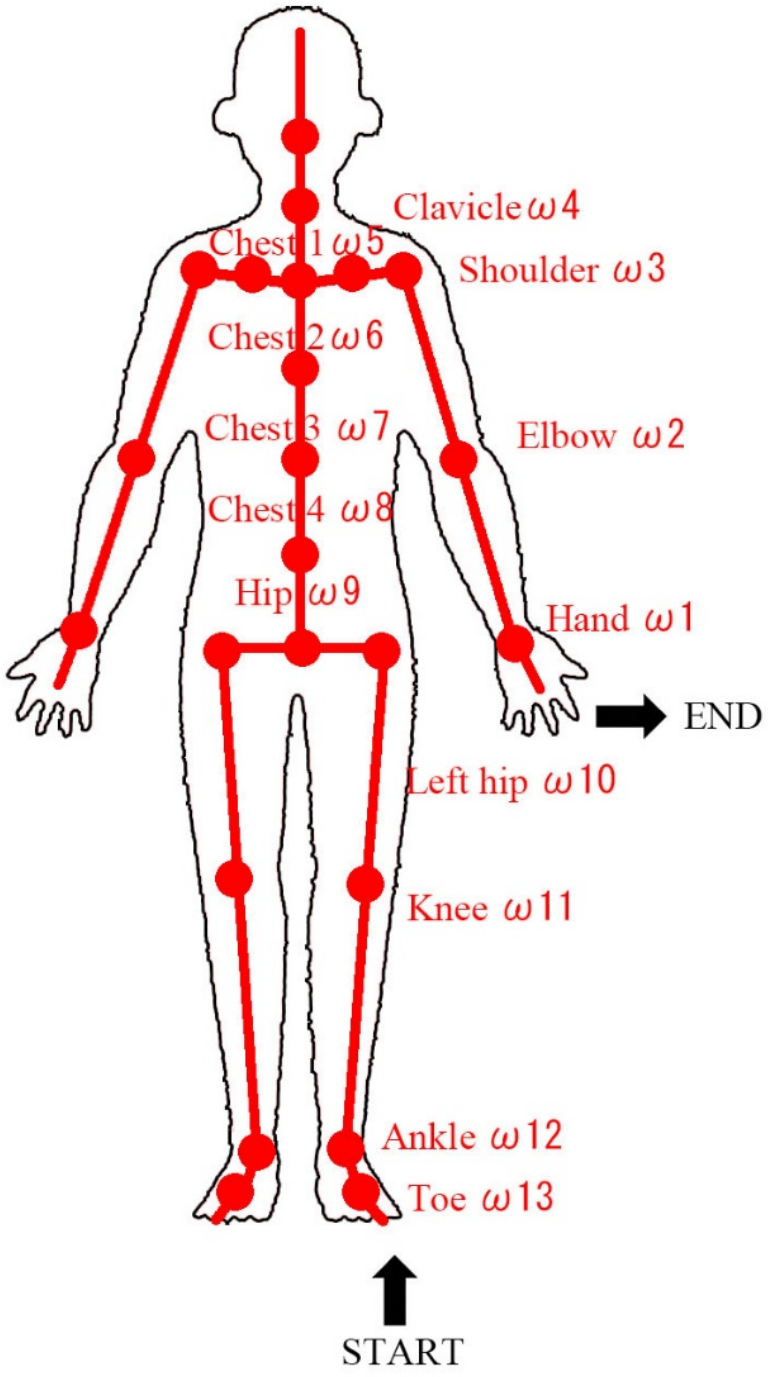

Figure 3.13 Human body link-dynamic system of Perfume

The reason for this Fibonacci relation is simple and is due to the hierarchical human body joint or link dynamical system as shown in Figure 3.13. The human body has a hierarchical joint system from the toes and passes through the hip (root) to the fingers. Figure 3.13 shows the skeletal structure connecting the joints. This body skeletal joint structure can be thought of as a complex network existing in the natural world [51], with each joint as a node and each skeleton element as an edge. In this case, each graph assumes a unit length 1. As shown in Figure 3.13, there are 13 nodes and 12 graphs passing from the foot or hand (START in the figure) to the other hand or foot (END in the figure). In the structure, the number of nodes passing through the path from START to END is 13-2 = 11, excluding the first and last nodes. The graphs connected to these nodes represent the respective dance motion (vibration) elements in the link structure. This structure is composed of the IMFs shown in the Hilbert-spectrum in Figure 3.9 (a).

The average of the angular velocities of IMFs is averaged in time, so that dance moves are long enough to cancel the phase difference between the rotational angular velocities of each decomposed joint IMF. This is the reason why the Fibonacci sequence in average angular frequencies is observed in human dance motions. If this relationship is correct, then the further away a joint is from the observation point, the greater the angular velocity will be from the base position obtained due to the above Fibonacci sequence relationship. Then, when a dancer performs different choreographies, the average of the angular velocities of IMFs will change into different pseudo-Fibonacci sequences. Thus, the different pseudo-Fibonacci sequences can be considered as different choreographies. For example, in the Perfume Hilbert spectrum of Figure 3.9(a), the Fibonacci sequence relationship was established approximately from IMF1-3 and IMF4-6. This indicates that Perfume performs two different choreographies, and each joint rotation speed is relatively constant and synchronized. Because the choreography composed of IMF7-11 and the trend was the basic motion of step and position from the hip joint, the human body of high-frequency IMFs (choreographies) can be considered as a motion blended into the low-frequency IMFs (basic motion) and the trend, as shown in Figure 3.14. The further an IMF is from the origin, the higher the average frequency. Figure 3.15 shows the flow of motion blending using the Fibonacci sequence. For example, Figure 3.14 (a) shows choreographies extracted from the Perfume dance as an example. Here, IMF4-6 is the shoulder and knee motions, and IMF1-3 is the wrist and ankle motions. Figure 3.14 (b) is the basic motion extracted from the target dance. We can exchange, blend, and add decomposed choreographies into the basic motions of the

target dance of (b).

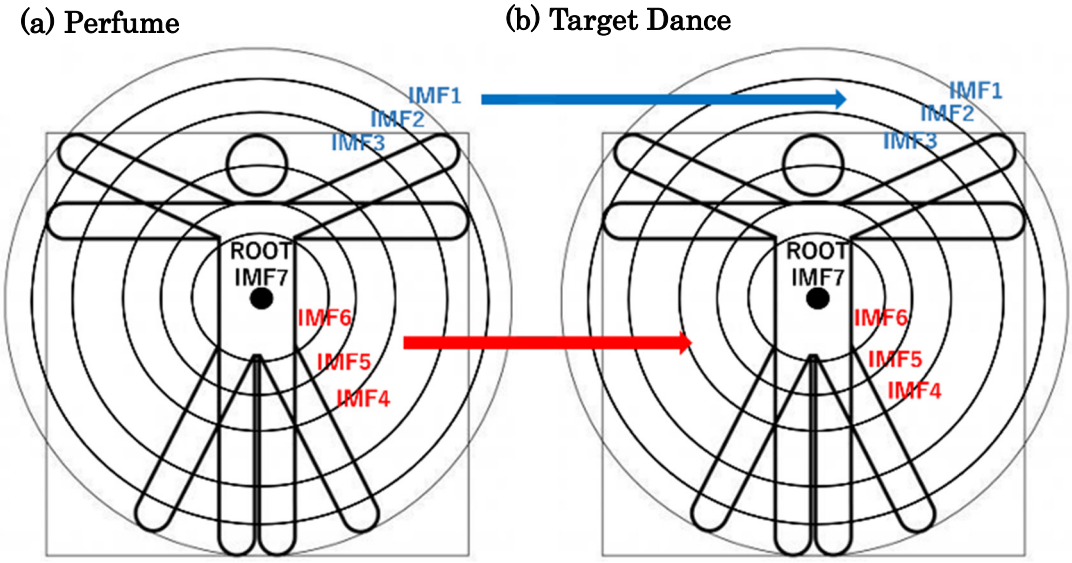

Figure 3.14 (a) Two choreographies decomposed by NA-MEMD (b) Blending two choreographies by Perfume into the basic motion target dance

Thus, if we assume the dancer is synchronized with the music and moving all joints with almost the same frequency (rhythm), the average IMF instantaneous frequencies must follow the Fibonacci sequence corresponding to different choreographies. Therefore, we can distinguish which frequency the specific IMF should belong to and associate the specific IMF with its corresponding joints that play a part in the choreographies. Once all the IMFs are classified, we can reconstruct and combine the corresponding IMFs into one dancer's choreography.

3.5.3 Dance choreographic reconstruction using the Fibonacci sequence

Using IMFs associated with each Fibonacci sequence corresponding to the link hierarchical structure showing in Figure 3.13, we can reconstruct and combine each dance choreography that consists of a few joint motions and their corresponding IMFs. For example, the human body skeleton of Perfume motion data has (23 joints) × (3 Euler angles) channels. We chose only one of the Euler angles of the root (hip) joint IMF to explain the way the dance choreographic motions are segmented. The same procedure can be applied to all other channels to decompose the motion. Thus, the decomposed IMFs from the original dance motion can be defined as follows:

$$\sum_{k} IMF_k(t) = \sum_{n} C_n(t) + B(t) \qquad (12)$$

Here, $IMF_k(t)$ is the IMF set decomposed from the original dance motion using NA-MEMD, and $C_n(t)$ is a dance choreography set reconstructed from several IMFs whose average frequencies $F_k(t)$ are approximately the Fibonacci sequence frequency. Here, $B(t)$ is the basic dance motion including dance positions and postures, reconstructed and combined by several IMFs whose average frequencies $F_k(t)$ are lower than the root joint (hip joint) frequency. $F_k(t)$ can be calculated using WAFA [28].

In the present research, we focused on one of the Euler angles of the root joint (hip joint) to reconstruct the choreographies. The decomposed average frequency of one IMF, the Euler angles of the root joint (hip), is denoted by $F_i(t)$. A flowchart of our algorithm is outlined in Figure 3.15 and our algorithm is summarized as follows:

1. Obtain the $IMF_k(t)$ set as the input from NA-MEMD;

2. Extract the basic motion $B(t)$ by combining the IMFs with an average frequency lower than the root (hip) (frequency threshold);

3. Reconstruct and combine choreographies
   a) Set $IMF_i(t)$ as the root joint and $IMF_{i+1}(t)$ as a Fibonacci sequence $IMF(t)$ set.
   b) If the average frequency of the IMFs satisfies $F_{i+2} \sim F_i + F_{i+1}$, add F_{i+2} to the $IMF(t)$ set, and check the next IMF average frequency F_{i+3}.
   c) Combine the $IMF(t)$ set $(A_i > \varepsilon)$ as one choreography $C_j(t)$ by adding into the choreographies set $C_n(t)$ for all joints. Here, A_i denotes the average amplitude of $IMF_i(t)$, and ε is the amplitude threshold that can be considered as a parameter to control the sparsity of the choreography; and

4. Repeat step 3 until all dance choreographies are composed.

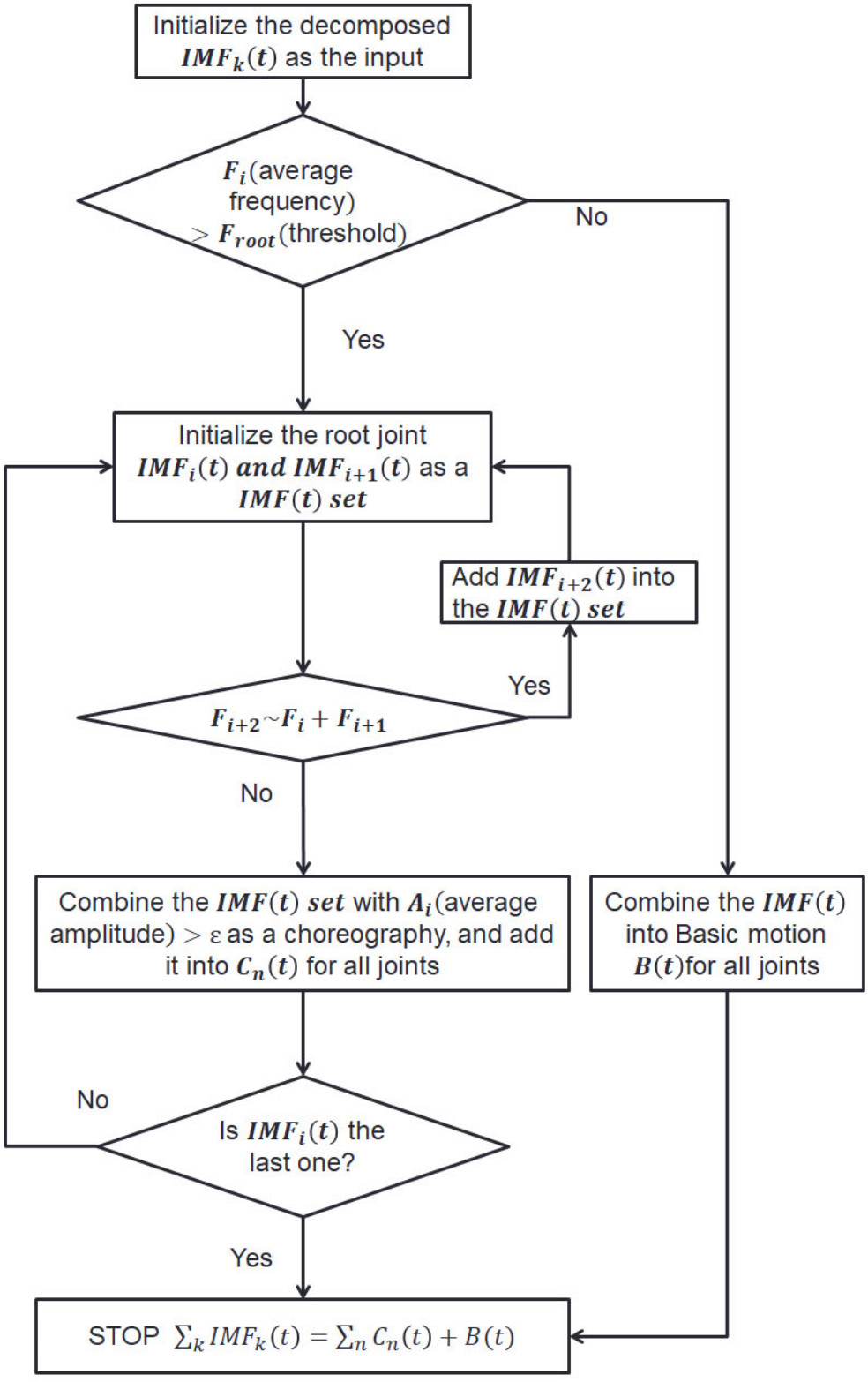

Figure 3.15 Flowchart of the proposed algorithm using the Fibonacci sequence

As we can see from the proposed algorithm, the extracted high-frequency IMFs can be reconstructed and combined into several choreographies $C_n(t)$ for blending and editing. The extracted low-frequency IMFs are the basic motion $B(t)$ including dance positions and postures, which can be considered as a basis for adding high-frequency choreography with the trend.

3.5.4 Dance beat adjustment and rescaling, and skeleton rescaling

The extracted choreographies need to be rescaled before they are blended into another dance motion. Because the dance motion is synchronized with a dance music tempo or beat, the extracted choreographic speed needs to be adjusted and rescaled to the targeted music.

We use beat tracking to calculate BPM in dance music, proposed by Dan Ellis [36] introduced in section 3.2.2. Using BPM, we can adjust the extracted choreography's speed to the targeted dance motion speed.

In addition, because motion-captured data often include inputs from different dancers using different capture systems, the motion data need to be rescaled to the targeted dancer's skeleton structure. After these adjustments, dance choreographies can be blended into another dance motion consistently.

3.6 Dance motion editing using the proposed framework

In this section, we show the results of the dance motion editing method proposed in Section 3.5.

3.6.1 Dance choreographic IMF reconstruction

In the Perfume dance, IMF1-3 and IMF4-6 consist of two sets of Fibonacci sequences, as shown in Figure 3.14 (a). First, we used the Fibonacci sequence in section 3.5.2 to reconstruct and combine IMF1-3 and IMF4-6 into two different dance choreographies, separately. Figure 3.16 shows the two Perfume dance choreographies extracted by our proposed framework. In the figure, two different perfume choreographies have been extracted, separately. The IMF4-6 choreographies consisted mainly of both the shoulder and knee motions, and the IMF1-3 choreographies consisted mainly of both the wrist and ankle motions. After beat adjustment and rescaling, and bone rescaling, these choreographies were

blended into the basic motion of a target dance, as shown in Figure 3.14 (b).

Figure 3.16 Perfume dance choreographies extraction

3.6.2 Target dance basic motion extraction

Figure 3.17 shows the salsa basic motion extracted by using $F_i > F_{root}$ (the hip average frequency threshold that was the lowest frequency). Salsa basic dance motions (dance step) have a root position synchronized with the beat frame. This indicates that IMF 7-11 and trend was a basic dance motion that is not exchangeable. This basic motion was considered as a target for blending other high-frequency

choreographies, as in Figure 3.14 (a). Then, except for the salsa dance step, the two different high-frequency choreographies of the Perfume dance in Figure 3.16 were blended separately into the extracted Salsa basic dance motion shown in Figure 3.17, to create two novel dances.

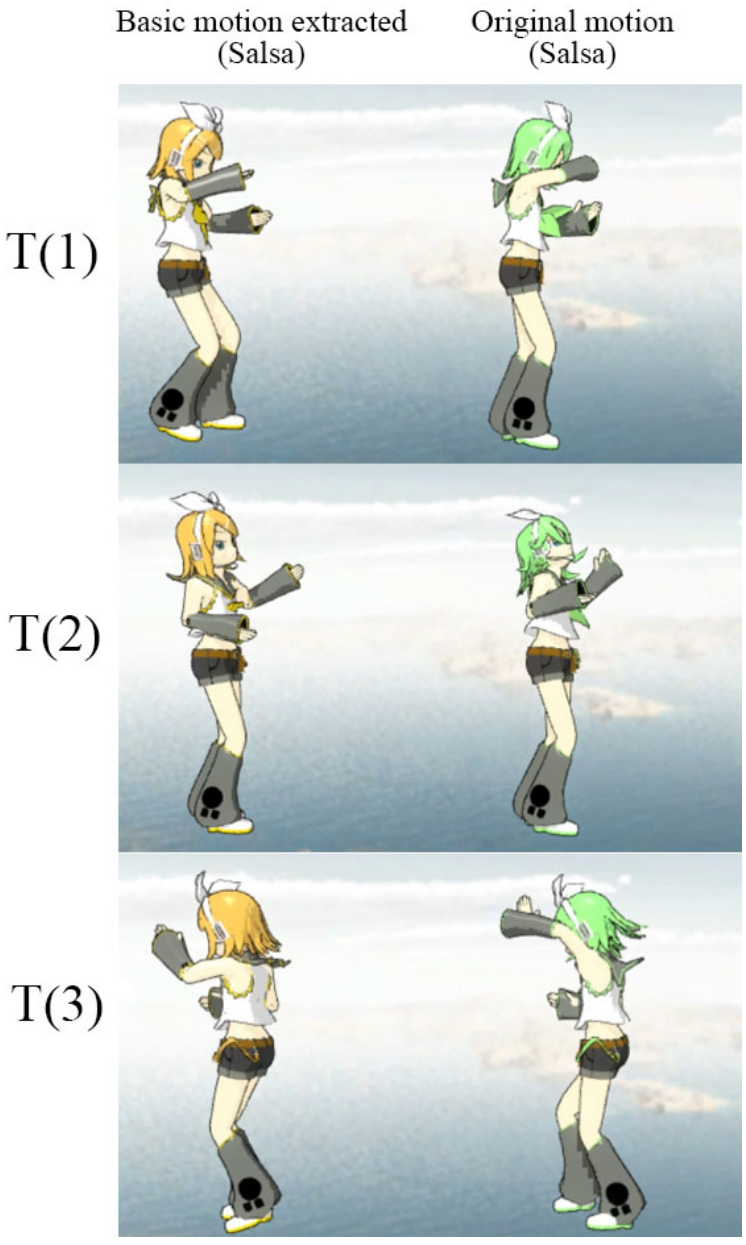

Figure 3.17 Salsa dance basic motion extraction

3.6.3 Blending extracted dance choreography

In Figure 3.18, after adjusting the beat from 2.0Hz (Perfume beat frequency) to

2.5Hz (salsa beat frequency), a piece of Perfume dance choreography was blended into the salsa basic motion (Figure 3.18). As shown in the figure, the different Perfume choreographies were blended into the salsa basic motion. This can create a new salsa dance with a different style. The dance choreography 1 (IMF4-6) is the set of Perfume's shoulder and knee motions that were blended into the salsa basic motion (IMF7-11 + trend). The dance choreography 2 (IMF1-3) is the set of Perfume's wrist and ankle motions, which were blended into the salsa basic motion. Using this proposed framework, animators can easily extract and edit dance choreographies using different IMFs reconstructed from the IMF sets that consist of Fibonacci sequences.

Figure 3.18 Perfume dance choreographies blended into the salsa dance

3.7 Discussion of dance motion editing

Human motions such as dance motions are very noisy and extremely difficult to edit. It is known that using NA-MEMD can clearly decompose noisy dance motions into distinct pseudo "monochromatic" IMFs. The human joint or link-dynamical system in dance often results in a Fibonacci sequence frequency relation. Using this Fibonacci sequence relation, the dance choreographies can be clearly reconstructed. Our framework is a powerful and useful tool to decompose, blend, and edit dance motions for animators.

In section 3.6, we use our proposed a framework for dance motion editing and blending. The main results of our research can be summarized as follows. First, we proposed a framework for dance choreographic editing using HHT and the Fibonacci sequence. Second, the low-frequency IMFs can be reconstructed and combined with the trend as a basic motion. Third, the high-frequency IMFs can be reconstructed and combined with several different choreographies. These choreographies can be exchanged, blended and edited into the basic motion of different dance styles. Fourth, beat adjustment and bone rescaling must be done before choreographies blending. Fifth, human dance motion editing and blending using the decomposed nonlinear signals as the instantaneous frequency in time-frequency space can give new insights into motion editing and blending techniques.

3.8 Limitations of the proposed framework

The following points are the limitations of the proposed framework.

- Singular IMF

The EMD used in our framework usually decomposes a signal from highest to lowest frequency [8]. However, frequency is not necessarily decomposed from highest to lowest. Singular IMF exists that is specifically decomposed at a higher frequency than that of the previous IMF [33]. Previous research has shown that a singular IMF may occur while the decomposing human motion data [49, 50]. As a result, the average frequencies of the IMFs cannot be calculated correctly. Therefore, our proposed algorithm based on Fibonacci frequency could not be used to extract the choreographies directly. To resolve this problem, we applied basis pursuit denoising (BPD) [52] as a pre-processing step to remove singular IMFs that can cause mode mixing and IMF singularities.

- Body collision

Our proposed dance choreography editing framework use NA-MEMD to decompose motion data into several choreographies (high-frequency IMFs) and a basic motion (low-frequency IMFs) with a trend. The trend can be considered as the posture of the whole dance. Thus, body collision will not occur in blending only one dance choreographies into another basic motion (low-frequency IMFs + trend). However, if we blend multiple choreographies extracted from two or more different dance motion into other basic motions into another basic motion in the same frame, body collision may occur due to the motions from different joints. For example, if we want to blend Perfume choreography extracted from the right-hand joint, and hip-hop choreography extracted from the left-hand joint into the salsa basic motion, body collision detection such as [53, 54] should be applied to verify whether two different dance choreographies can be blended at the same time.

3.9 Summary and conclusions

In this chapter, we explained the process of our proposed method from the input motion capture data to the output generated motions. First, we demonstrated the proposed framework of dance motion analysis using HHT. Second, as an example, we analyzed Japanese techno-pop unit Perfumes' dance motions, salsa, waltz, and hip-hop motions by our proposed framework. Third, we introduced the proposed framework of dance motion editing using HHT. Fourth, we used Perfume dance and salsa dance to show the results of our proposed dance motion editing framework. Fifth, we discussed the performance and limitations of our framework. Then, the conclusions of this chapter are as follows:

1. We proposed a framework for dance analysis using HHT. Using this framework, we can decompose different dance motions into several nonlinear modes that are pseudo monochromatic waves so-called IMFs. After applying HT to each IMF, their instantaneous frequency and amplitude can be obtained. Thus, the dance motions can be analyzed in the instantaneous frequency domain with the Hilbert spectrum; and

2. We proposed a framework for dance editing using HHT with a Fibonacci sequence based on the structure of the human body. Using our proposed framework, we can decompose dance motions of different styles into several choreographies and one basic motion. The choreographies can be edited separately and can be also blended into the basic motion of other styles of dance. Thus, a new dance motion can be created without motion capturing again.

Overall, our method can reveal novel features from decomposed IMFs such as a Fibonacci sequence that can be observed in human motion. Also, these IMFs can be scaled, combined, subtracted, exchanged, and modified, and can be blended into new dance motions. Thus, our proposed framework helps choreographers to analyze dance motions in the primitive unit. And it also provides a method for animators to edit dance motions for creating a new dance from different dance styles.

Chapter 4

Bunraku puppet motion analysis using Hilbert-Huang transform

This chapter is organized as follows. Section 4.1 introduces the research background and the problem of the analysis of Bunraku puppets. Section 4.2 demonstrates the Bunraku puppet and its manipulation. Section 4.3 introduces the Jo-Ha-Kyū, a traditional Japanese art principle, and unique interaction techniques of Bunraku puppets. Section 4.4 shows the results of Bunraku puppet motion analysis based on Jo-Ha-Kyū. Section 4.5 discusses the results in section 4.4. Section 4.6 summarizes this chapter.

4.1 Research background

The affective motions of the Intangible Cultural Heritage “Ningyo Joruri Bunraku”, is evaluated by UNESCO as “the most beautiful motion in the world”. Bunraku is the oldest mechanical structure of puppets in Japan and began in the Heian period (AC 794-1185). Creating affective motions to interact with humans is extremely complicated because emotional expressions have different elements such as the story of the whole play, the chant of Tayu, the music of Shamisen, and the motion of Bunraku puppets. In this dissertation, to analyze these complicated emotional expressions in Bunraku play, we only focus on the mechanism of how puppeteers manipulate Bunraku puppets to express their affective motions. Manipulation of Bunraku puppets requires sophisticated techniques. These techniques create empathy for audiences using a traditional Japanese art principle called “Jo (beginning)-Ha(breaking)-Kyū(rapid)” [55-58]. There have been several studies on Bunraku puppets in different areas. Because three puppeteers manipulate one puppet, some studies have focused on puppeteers’ non-verbal communication, so-called “Zu” (“ず” in Japanese) [59]. Hattori et al. [60-63] analyzed the emotional

motion patterns of Bunraku puppets and investigated robot motion designs based on them. Chen et al. proposed a method to generate Jo-Ha-Kyū motions for robots using a neural network [64]. However, no research has analyzed the Jo-Ha-Kyū mechanism by focusing on the Bunraku performing art techniques themselves.

In this chapter, we present the Bunraku motions using the famous concept Jo-Ha-Kyū [55-58], which is used in expressing emotional changes synthesizing with stories. The concept of “Jo-Ha-Kyū” is widely used in Noh, Ningyo Joruri Bunraku, and Kabuki, which are all recognized by UNESCO as intangible cultural heritages. Previous researches showed that all of these art forms express emotions mainly based on Jo-Ha-Kyū [55-58].

In Japan, Jo-Ha-Kyū is a very broad concept that can apply to numerous fields and concepts [55-58]. Typically, Jo-Ha-Kyū changes the rhythmic speeds in three stages: "Jo (beginning)"; "Ha (breaking)"; and "Kyū (rapid)". However, sometimes, Jo-Ha-Kyū can be only “Jo” or only “Jo-Ha” depending on the play stories. This has been intensively studied in Noh. Jo-Ha-Kyū also differs slightly according to different Noh schools. Furthermore, Jo-Ha-Kyū in Bunraku is also somewhat different from that of Noh. Therefore, Jo-Ha-Kyū is considered to be an ambiguous concept that depends on its art form. Thus, in this dissertation, we simply limit or define the Jo-Ha-Kyū as changes and breakings of rhythmic speeds (i. e., pace or tempo) in music, which will be described in detail later.

4.2 Bunraku puppets

Ningyō jōruri Bunraku is a traditional Japanese performing art in a puppet theatre. It started at the beginning of the 17th century in Osaka. In the Bunraku theater, three different types of performers (puppeteers, narrator, and musician) play in one show. Three puppeteers manipulating one Bunraku puppet. The narrator (Tayū) tells the story, and the musician (Shamisen) plays the music. The motions of the Bunraku puppet are led by the chants of the Tayū, who controls the Jo-Ha-Kyū with breaking rhythms so-called “Ma” (“間” in Japanese).

As displayed in Figure 4.1, the puppet master manipulates only the head and right hand, the left sub-master manipulates only the left hand, and the other sub-master manipulates the two legs only. It is well-known that puppet gestures and motions are very affective and sophisticated. On the one hand, humans express their affections using facial expressions, eye movements, and gestures at the same time.

On the other hand, the puppet masters use a unique technique (Jo-Ha-Kyū) mainly involving gestures to express emotions without any facial expressions.

Figure 4.1 Manipulation of Bunraku puppet

As shown in Figure 4.2, the puppet masters use sophisticated link-dynamics of the head joint to control delicate head (eye) motions and express various affections. When you see the puppet in its holder, it is just a puppet. However, when the puppet master begins to control the head, it looks like a real human expressing his/her affections.

Figure 4.2 Structure of Bunraku puppet

It is known that the famous and beautiful puppet motions can be classified into about 50 types, “Kata” (“型” in Japanese) [24]. In this chapter, we introduce these unique techniques with Jo-Ha-Kyū and capture these sophisticated puppet motions using motion capture. Using our proposed framework of motion analysis introduced in chapter 3, we identified the mechanism of Jo-Ha-Kyū with these affective motions of the Bunraku puppets using HHT.

4.3 Jo-Ha-Kyū

Jo-Ha-Kyū is the concept of rhythmic or tempo modulations with special breaks in Japanese traditional art. Western classical music creates emotions mainly based on a change of key. However, Japanese traditional music has no pitch standard (i.e., “Do” is 440Hz). Thus, Japanese traditional music uses a change of tempo (Jo-Ha-Kyū) instead of a change of key to create emotions. Jo-Ha-Kyū is derived from Japanese “Gagaku” (“雅楽” in Japanese) and is a term used in Japanese traditional arts. It has been traditionally used in Japan since the middle ages. Jo-Ha-Kyū has been studied mainly in the Noh [55-58]. Thus, in this section, we first introduce Jo-Ha-Kyū in Noh and give the definitions of Jo-ha-kyu in this dissertation. Next, we introduce Jo-Ha-Kyū in Bunraku and Puppet Principles.

4.3.1 Jo-Ha-Kyū in Noh

Noh is a field of Japanese traditional performing arts. Until the Edo era, it was called “Sarugaku”, and after the Meiji restoration, it was called Noh. In the Edo era (1603-1868), the Tokugawa government had absolute power, and Noh became popular among the ruling military class. This cultural and social unrest stimulated the development of entertainment.

Usually, Jo-Ha-Kyū means only a series of dramatic changes of speed following the stories. We use this definition in this dissertation. From previous studies [55-58], it has been verified that the rhythms of the Noh performance gradually rise in time, as shown in Figure 4.3 for the schematic and typical pattern of Jo-Ha-Kyū [55-58].

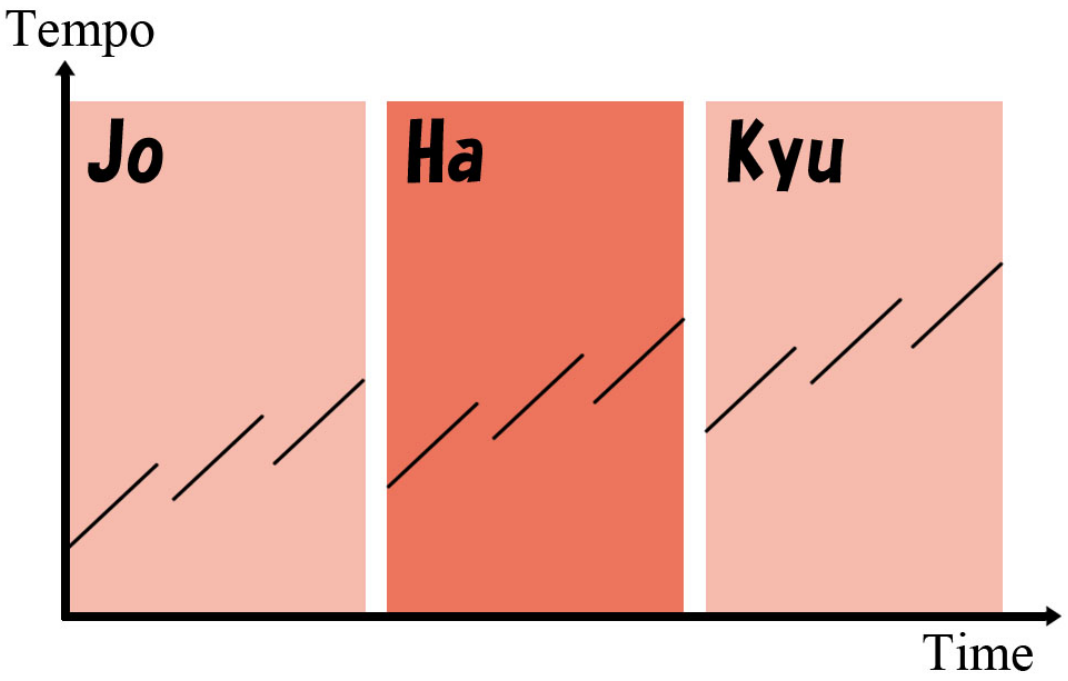

Figure 4.3 A schematic picture to explain Jo-Ha-Kyū

Please note that Figure 4.3 is only conceptual and is one of the Noh examples. It is only schematic, and the patterns differ according to the stories. In addition, "Jo (beginning)", "Ha (breaking)" and "Kyū (rapid) " can have “Jo", "Ha" and "Kyū" structures in themselves that are recursive, self-similar, or fractal. The small intervals between the rhythmic changes are Ma, breaking rhythms. The tempos or rhythmic changes and breaks are generated from the story. Thus, Jo, ha Ha and Kyū patterns vary according to the stories. For the real Jo-Ha-Kyū patterns, please see [57, 58].

4.3.2 Jo-Ha-Kyū in Bunraku puppets

Jo-Ha-Kyū in Noh plays is somewhat different from that of Bunraku. On the one hand, according to the previous researches [57, 58], Noh is composed of formal rhythms consisting of a nine-step structure (Figure 4.3). On the other hand, Bunraku play abandons this formal nine-step structure to establish in Noh. In so doing, Bunraku established modern dramatic structures that are more flexible and correspond to more complicated stories [57, 58].

Jo-ha-kyū in Bunraku puppets can be roughly translated as "beginning, break, rapid", as shown in Figure 4.3. It means that like in Noh, all actions or efforts of

Bunraku should begin slowly, speed up, and then end swiftly. In the Bunraku plays, the puppets display the emotional expressions in accordance with the narrator's (Tayū's) chants and Shamisen's music, etc. Bunraku puppets have unique changes and breaks of tempo (Jo-Ha-Kyū) in their motions to express emotions according to the Joruri storyline [65].

However, Bunraku does not have any fixed score, and the Syamisen(music), Tayū (narrator), Puppeteers (motion) perform continuously throughout the whole story. Thus, to perform a Bunraku play, Tayū uses his or her own tonal center and changes tempo based on Jo-Ha-Kyū according to the Jōruri (storyline). This leads the whole performance (Bunraku puppet and music), as shown in Figure 4.3. The master puppeteer creates the puppet affective motions that are synchronized and desynchronized with the Tayū's chant based on the Jo-Ha-Kyū tempo. It is a unique interaction technique that is not apparent in Western music.

4.3.3 Bunraku puppet principles in Jo-ha-kyu

To analyze the emotional expression techniques of the Bunraku puppet, we used 12 animation principles published by Disney animators Lasseter et al. [66] Based on Disney animations since the 1930s. Animation principles can make characters extremely lifelike in a 3D virtual world. Similarly, puppet principles can make puppets extremely lifelike in the real world. Although almost all 12 animation principles are used in the Bunraku puppet, in this dissertation, we show only the puppet principles most related to Jo-Ha-Kyū (Table 4.1). For the original 12 animation principles, please see [66-68].

Table 4.1 shows 4 Bunraku puppet principles in Jo-Ha-Kyū based on 12 animation principles. Bunraku puppets use these techniques to form the short-term Jo-Ha-Kyū, and finally form the long-term Jo-Ha-Kyū of the whole story. In the next section, we use motion capture to show the mechanism of Jo-Ha-Kyū and puppet principles in Bunraku play.

Table 4.1 Bunraku puppet principles in Jo-Ha-Kyū

Principles	Definitions
Arcs	In Bunraku, most “Kata” use arcs such as circles and lines to make motion expressions. Soft motions are represented by circles, and hard motions are represented by lines. Jo-Ha-Kyū (the change of motion strength) can be created by these arcs.
Anticipation	Anticipation is considered as "Ma"(Breaking rhythms) and is a link among Jo, Ha, and Kyū. Anticipation motions can also be performed by such as drawing a circle of head motions and standing up motions.
Pose-to-pose action	According to reference [24], these famous and beautiful puppet motions can be classified into about 50 kata. Pose-to-pose action refers to one Kata from another Kata in the case of Bunraku puppets. There are many cases where Kata is connected to another Kata by a circle, and the emphasis motion performed by Ma (Breaking rhythms).
Timing	Timing expresses different emotions using different lengths of "Ma" for the same motions.

4.4 Motion data capturing and analysis of Bunraku puppet motions using HHT

In order to analyze the Jo-Ha-Kyū in Bunraku and its emotional expression mechanisms, we collected Bunraku motions and music data (Tayū and shamisen) simultaneously at the Bunraku theater. We used our proposed framework in chapter 3, to analyze Bunraku motions related to Jo-Ha-Kyū using HHT.

4.4.1 Motion data capturing of Bunraku puppet motions

In this subsection, we focused on Bunraku puppet motions and analyzed the mechanisms of Jo-Ha-Kyū. Figure 4.4 shows one of the Bunraku play scenes performed by the Bunraku performers (Tayū: Tsukoma Takemoto, Shamisen: Sosuke Takezawa, Puppeteers: Kanjuro Kiritake and 2 others), for which we captured puppet motions and recorded the Bunraku music and chant. In order to obtain high accuracy motion data, we use both optical and magnetic mocap systems as shown in Figure 4.4.

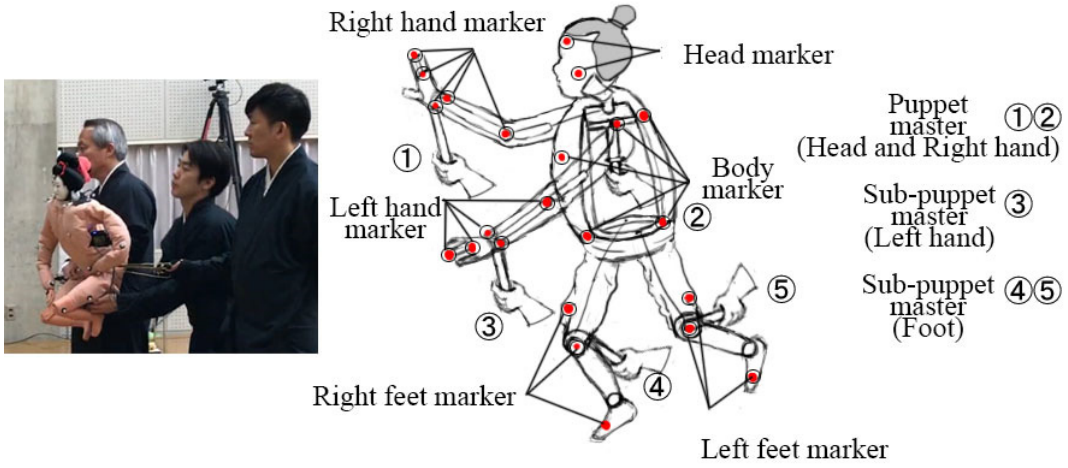

Figure 4.4 Bunraku motion capturing using both optical and magnetic mocap measurements

We selected a scene named “Sugisakaya” ("杉酒屋の段" in Japanese) from a famous Johruri story, "Imoseyama Onna Teikin” ("妹背山婦女庭訓" in Japanese). This scene is composed of the Jo-Ha-Kyū stories, and, thus, the changes of motion speed should follow Jo-Ha-Kyū.

4.4.2 Analysis of Bunraku puppet motions focusing on Jo-Ha-Kyū in long-term

To show the Jo-Ha-Kyū of Bunraku puppets motions in the “Sugisakaya” scene, we plotted angular velocities of four main joints of the puppet as shown in Figure 4.5 (a). The joint angle velocities were averaged using a 4-second window for smoothing. As shown in Figure 4.5 (a), the tempos of the angular velocities of the neck and hip joints, which should represent the motion speed of the puppet, can be divided into three Jo-Ha-

Kyū parts in time, correspondingly. To analyze Jo-Ha-Kyū in detail, we also used our proposed framework in chapter 3 to plot a Hilbert spectrum of head joint as shown in Figure 4.5 (b). Head motions are most relevant with emotion expression, such as drawing a circle of head motion as an anticipation in Ma (breaking rhythms), which introduced in table 4.1. As shown in Figure 4.5 (b), the three Jo-Ha-Kyū parts are clearer, and small Ma can be detected and decomposed in the Kyū part by our proposed framework using HHT.

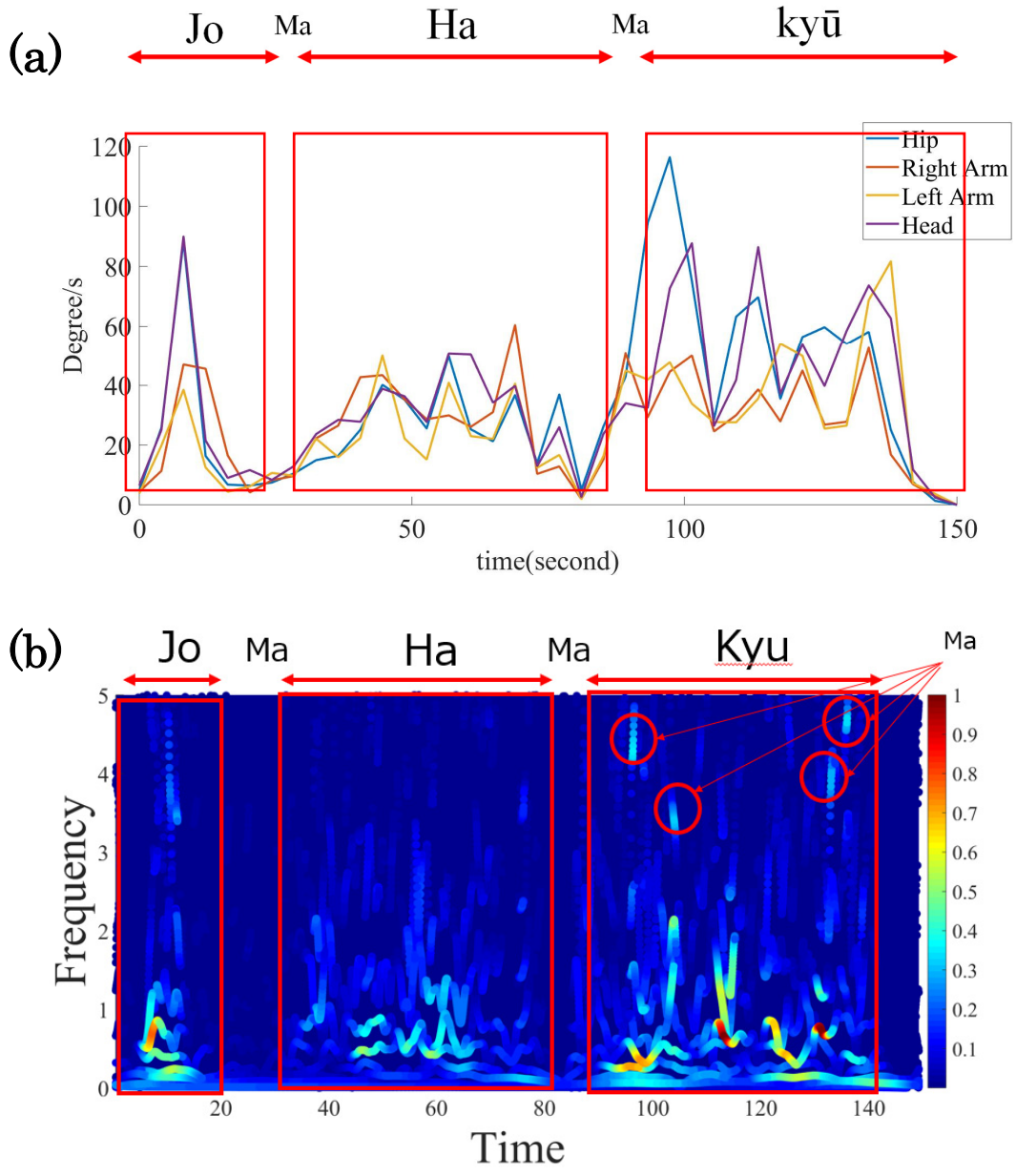

Figure 4.5 Data from the Sugisakaya (a) Time sequence of the angular velocities of upper body joints. (b) The Hilbert spectrum of the dead θ_z

Figure 4.5 shows real Jo-Ha-Kyū changes of motion speed according to the Tayu's chant. Please note that Figure 4.3 is only a schematic, and we assumed Jo-Ha-Kyū if the motion speeds changed significantly according to the chants or stories, as shown in Figure 4.5. The question here is whether such significant and complex changes of motion speed exist in our modern dances. We tried to answer this question by showing one modern dance example. Therefore, Figure 4.6 compares that one of the most famous modern dances by Perfume dance data used in chapter 3 before. We use Perfume dance to compare to the changes of joint speeds of Bunraku puppet motions. As shown in the Figure 4.6, the Perfume dance is a typical western dance and has no significant changes of speed in this example. Motion data in Figure 4.6 (a) were averaged by a window of 0.5 seconds (beat interval). The strong sharp peaks are the dancer's rotational motions, which were insignificant on average in the changes of speed.

As can be seen from Figure 4.6, the Perfume dance is a modern pop dance. Therefore, the performers dance at a constant rhythmic speed with no significant changes of tempo. Thus, it is apparent that the Figure 4.6 cannot be decomposed into three "Jo-Ha-Kyū" parts. By comparing Figure 4.5 with Figure 4.6, it can be seen that "Perfume" does not use Jo-Ha-Kyū or changes of motion speeds in this dance example. In a Bunraku play, the tempos of Tayū's chants are always changing because Japanese plays and music express their emotions mainly by changes and breaks of tempo or speed of rhythm [65]. It is expected these puppet motions will follow the changes and breaks of tempo or speed of Tayū's chants. Thus, it can be safely assumed that Bunraku puppet motions use the Jo-Ha-Kyū principles, which are the changes and breaks of the motions speed according to the Tayū's chant speed. In contrast, we know that the "Perfume" dancers' motions are kept almost at a constant tempo.

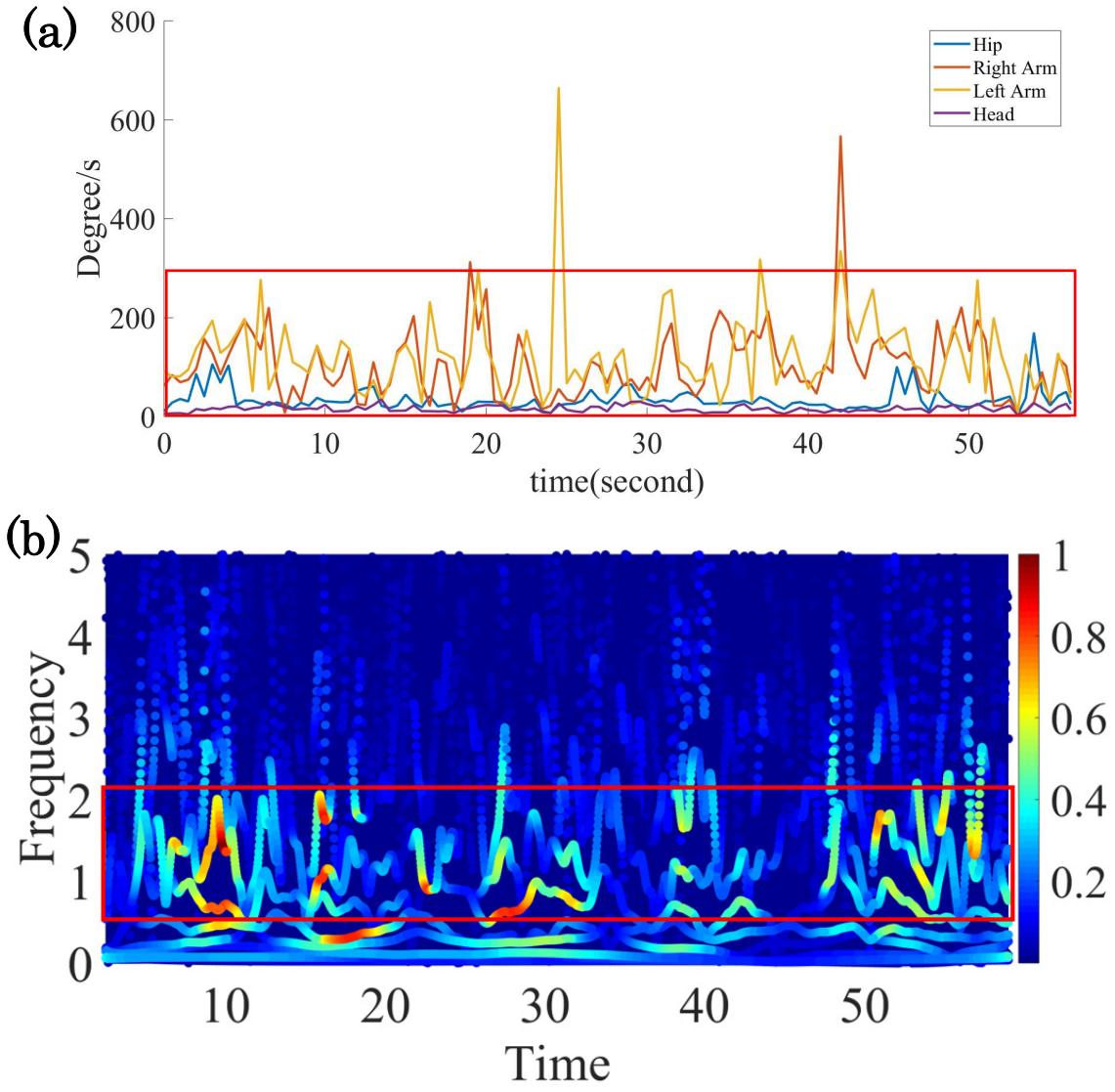

Figure 4.6 Perfume dance entitled "Enter the Sphere." (a) The time sequence of angular velocities of the neck, right, left arm, and hip joints. (b) The Hilbert spectrum of head θ_z using HHT

To confirm bunraku motions are led by Tayū to form the Jo-Ha-Kyū, we convert the narration of Tayū and the music of shamisen to a music score. Because Japanese traditional music has different principles from western music, so algorithms for western music like beat tracking cannot be applied. Therefore, in our study, we ask the musician Dr. Tomonari Higaki, a guest professor in Osaka university of arts, to perform a qualitative musical analysis (scoring by staff notation) of the scene where motion capture was performed.

Figure 4.7 shows an example of the music notation, for which we were able to score

the narration (Tayū) and the music(shamisen). The point to note here is the part where the tempo of Tayū constantly changes. As the score shows, the tempo changes rapidly even from one Japanese character to another Japanese character. This kind of musical expression is totally different from Western music [65]. In the Japanese performing arts, emotions are created by changing the tempo in this way [65].

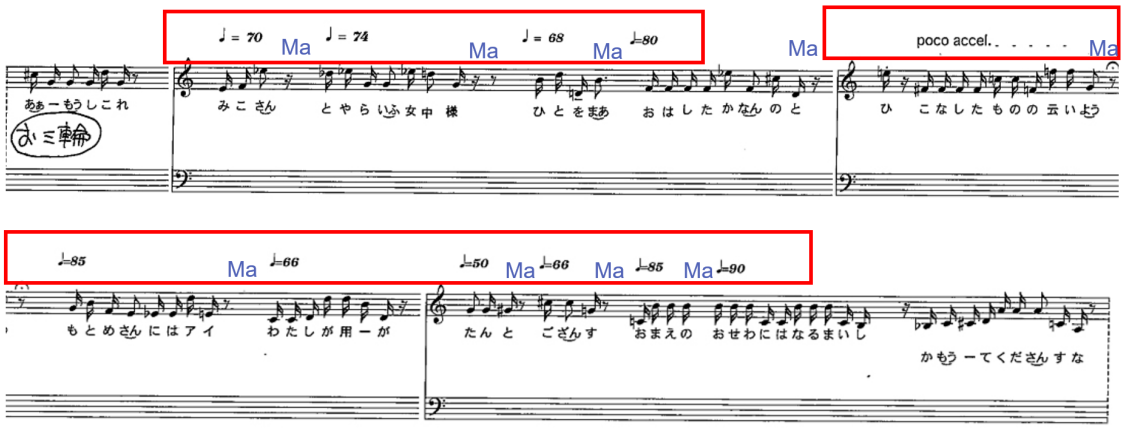

Figure 4.7 An example (Sugisakaya) music score by musician

Figure 4.8 (a) shows the musical score of the Sugisakaya scene we analyzed in Figure 4.5. Figure 4.8 (b) is the same plot as Figure 4.5(a).

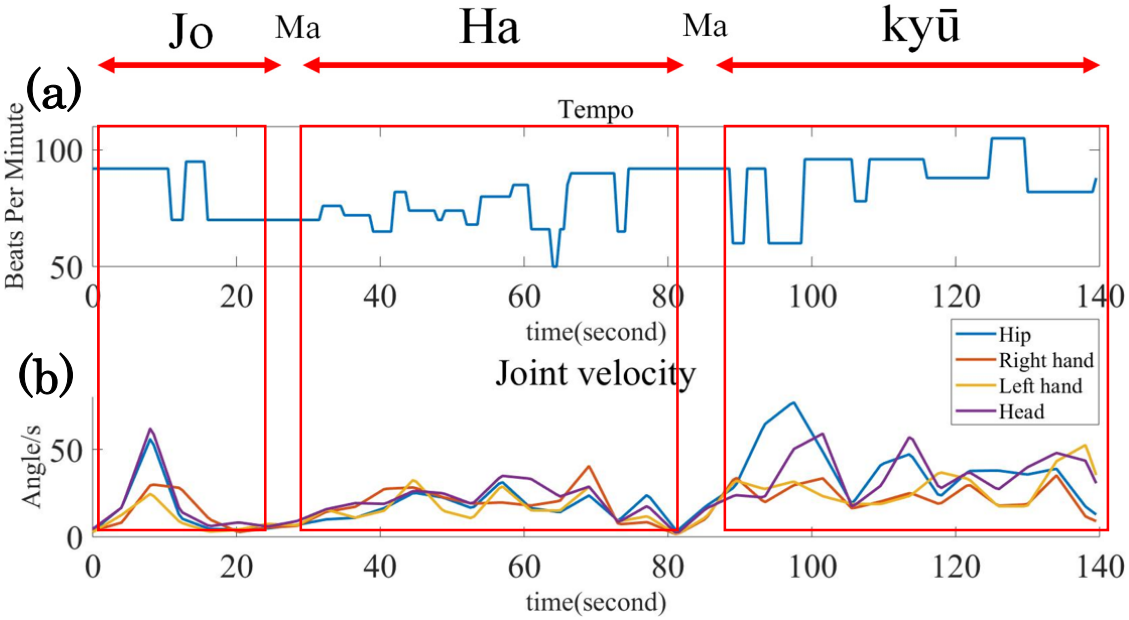

Figure 4.8 Sugisakaya scene from Imoseyama Onna Teikin. (a) The change of tempo of Tayū chant. (b)The average speed change of the puppet head motion smoothing by measure average unit (4s)

As shown in Figure 4.8, using the musical score, we can get the same Jo-Ha-Kyū with motion data. Thus, the Bunraku puppet motions are following the Tayū narration to form the Jo-Ha-Kyū that follows the storyline.

4.4.3 Analysis of Bunraku puppet motions focusing on Ma with puppet principles in short-term

Anticipation and Timing are described in table 4.1, and they play an important role to make "Ma" (Breaking rhythm) among each Jo, Ha and Kyū in the short-term. The content expressed by the word "Ma" is very wide. Originally, Ma indicates a spatial separation. However, when Ma is expanded to the concept of time, it is forming a metaphysical world that transcends time and space in an empty space. In this subsection, Ma we discuss here is during the Jo-Ha-Kyū. The result of the Anticipation and Timing with Ma is as Figure 4.9.

The "Anticipation" in "Animation principles [66]" is confirmed in Bunraku by our research. As an example, Figure 4.9 shows an example of "Anticipation" (ma) selected from the Kyū part shown in Figure 4.5(b), using our proposed framework with HHT. The "Anticipation" is completely decomposed as a nonlinear mode at the turning point where the narrative tempo of Tayū is changing faster (BPM 82 → 88).

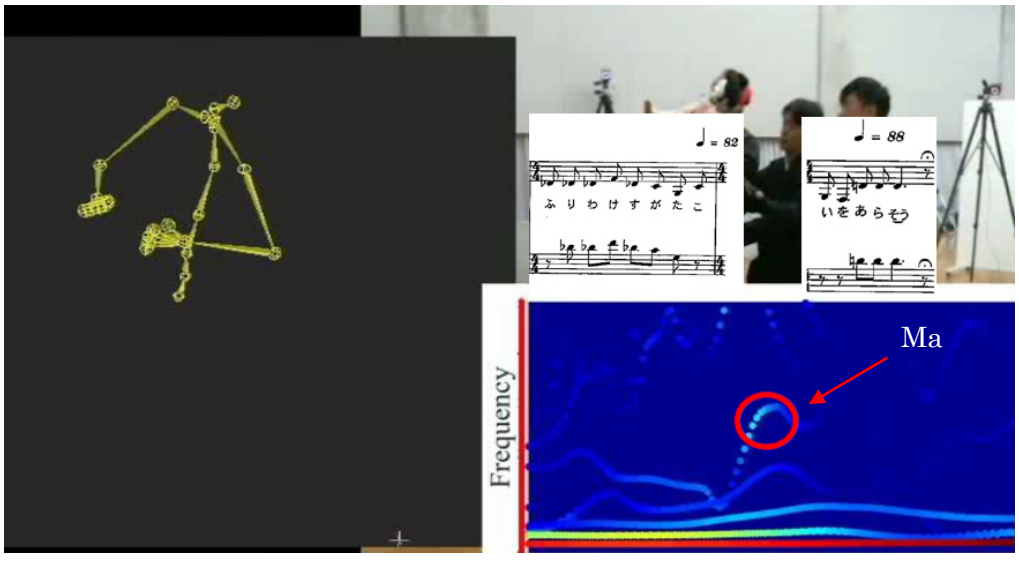

Figure 4.9 Extraction of Ma (breaking rhythm) using HHT

As shown in the Figure 4.9, Bunraku puppeteers cooperate with the shamisen and Tayū, and give the audience a chance to recognize the next action. In this way, by

making full use of puppet principles to create different short-term Jo-Ha-Kyū, bunraku play can be finally formed in the long-term Jo-Ha-Kyū as one performing art. Therefore, it became clear that Bunraku has a stage composition based on completely different principles from Western dance.

4.5 Discussions

Using our proposed framework, the Jo-Ha-Kyū and Ma of the Bunraku puppet can be decomposed into nonlinear signals (IMFs) by HHT. Then, by extracting and analyzing the motion capture data of Bunraku based on Jo-Ha-Kyū, we show the mechanism of affective motion in both long term Jo-Ha-Kyū and shot term Jo-Ha-Kyū.

Our research reveals the results as follows. First, Section 4.2 shows a Jo-Ha-Kyū (change of tempo) in Tayū's narrative and Bunraku puppet motions. Our research reveals that there was a correlation between the narrative tempos and the puppet's motions. This means that Bunraku puppet is led by Tayu's narrative to make Jo-ha-kyu with the story. Second, the motion of Ma (breaking rhythm) can be decomposed and detected from the puppet head joint as movements of 0.2 to 0.3 seconds (3-5 Hz). These motions among the Jo, ha, and kyu where the puppet and the shamisen are synchronized. It reveals a relationship with Jo-Ha-Kyū and Ma that is the key to creating Jo-Ha-Kyū. Third, the Jo-Ha-Kyū pattern like Bunraku is not confirmed in Perfume, a Western dance. By summarizing the above, we can use the Jo-ha-kyu mechanism for motion design in preforming arts.

4.6 Summary and conclusions

In this chapter, we showed the results of Bunraku puppet motion analysis using HHT. First, we used both the optical and magnetic mocap system to capture the Bunraku puppet emotional motions from the “Sugisakaya” scene selected from the famous Johruri “Imoseyama Onna Teikin”. Second, using our proposed method in chapter 3, our research revealed the anticipation occurs in “ma”, which plays an important role in a Japanese traditional art principle, Jo-Ha-Kyū. Using the proposed method in chapter 3, we observed the Jo-Ha-Kyū and Ma in the spectrum from emotional expression motions of a Bunraku puppet. We compared Bunraku

puppet spectrum with those of the typical western modern dance motions of “Perfume” to show the mechanism of Bunraku affective motion design techniques, which can be used in performing arts design. Thus, in this chapter, we analyzed the Jo-Ha-Kyū mechanism of Japanese traditional art (Bunraku) and the conclusions are as follows:

1. Unlike Western dance motions (Perfume), Bunraku puppets use Jo-Ha-Kyū (Change of speed) to express emotions.

2. Bunraku puppets change their motion speed in accordance with the Jo-Ha-Kyū (change of tempo) of the Tayū (story) and Shamisen (music).

3. Bunraku expresses the emotional expression of Bunraku's unique puppets using puppet principles, arcs, anticipation, timing, and pose to pose actions similar to the animation principles.

4. Bunraku puppeteers make long term Jo-Ha-Kyū while making short term Jo-Ha-Kyū using the puppet principles.

Overall, our research showed that the Jo-Ha-Kyū mechanism of Bunraku motions is completely different from the Western. It can be used in performing arts like actors, robots and CG characters to create emotional expression motions interacting with humans. In the next chapter, as an example, we use this Jo-ha-kyu mechanism to develop a framework that can generate emotional expression motion using Jo-ha-kyu in robot motion design.

Chapter 5

Applying Hilbert-Huang transform to robot motion synthesis using Bunraku puppets with Jo-Ha-Kyū

This chapter is organized as follows. Section 5.1 introduces the research background and the problem of human-robot interactions. Section 5.2 is the proposed framework using Jo-Ha-Kyū principle extracted from Bunraku in chapter 4. Section 5.3 shows the results of the proposed framework using Jo-Ha-Kyū principle of Bunraku puppets. Section 5.4 discusses the results shown in section 5.3. Section 5.5 summarizes this chapter.

5.1 Research background

Home robots are expected to spread in the future due to the development of Artificial Intelligence technology. However, the movement of robot eyes that cannot communicate with humans causes an "Uncanny Valley" phenomenon [25], which leads to human discomfort. To solve this interaction problem between humans and robots, numerous researches on robot verbal interaction have been performed [69]. For instance, Valin et al. [70] used the microphone to develop a sound localization and tracking system. Nakamura et al. [71] conducted a research on localization proposing a system for robot interactive sound. However, very few studies focusing on human senses affected by the sounds and motions in human-robot interaction. The mechanisms that control the human senses affected by the sounds and motions by multiple robots need to be understood. Nakagawa [72] proposed that, in the design of humanoid service robots, the appearances, motions, and voices of robots produce discomfort or a negative feeling if they do not satisfy people's expectations. Especially, it is essential that autonomous robot designs have to be comprehensive and

integrated designs, which integrates all sounds, motions, stories in a balanced and harmonic manner.

As we discussed in chapter 4, Ningyo Joruri Bunraku has no eye movement and expresses emotions with motions of the head and shoulders. Its beautiful emotional expression behavior attracts audiences deeply. By using Bunraku's affective motion, it is considered to reveal an interactive motion design technique that makes a robot to overcome the "Uncanny Valley" phenomenon.

In this chapter, we use the results of traditional Bunraku plays obtained in chapter 4 and apply them to robot motion designs. Our hypotheses are shown in Figure 5.1. As we introduced in chapter 4, Multiple Bunraku puppets play interactively on the stage following the chant by Tayu (Chanter and Narrator), and music by Shamisen players. Each puppet is operated by three puppeteers in the Bunraku theaters. Thus, our proposed concept is that multiple robots can create an attractive "Bunraku stages," and coexist in human life with appropriate designed sounds, motions, and stories by robots. In this dissertation, we mainly focus on using Jo-Ha-Kyū mechanism analyzed in chapter 4 and use it to robot motion design.

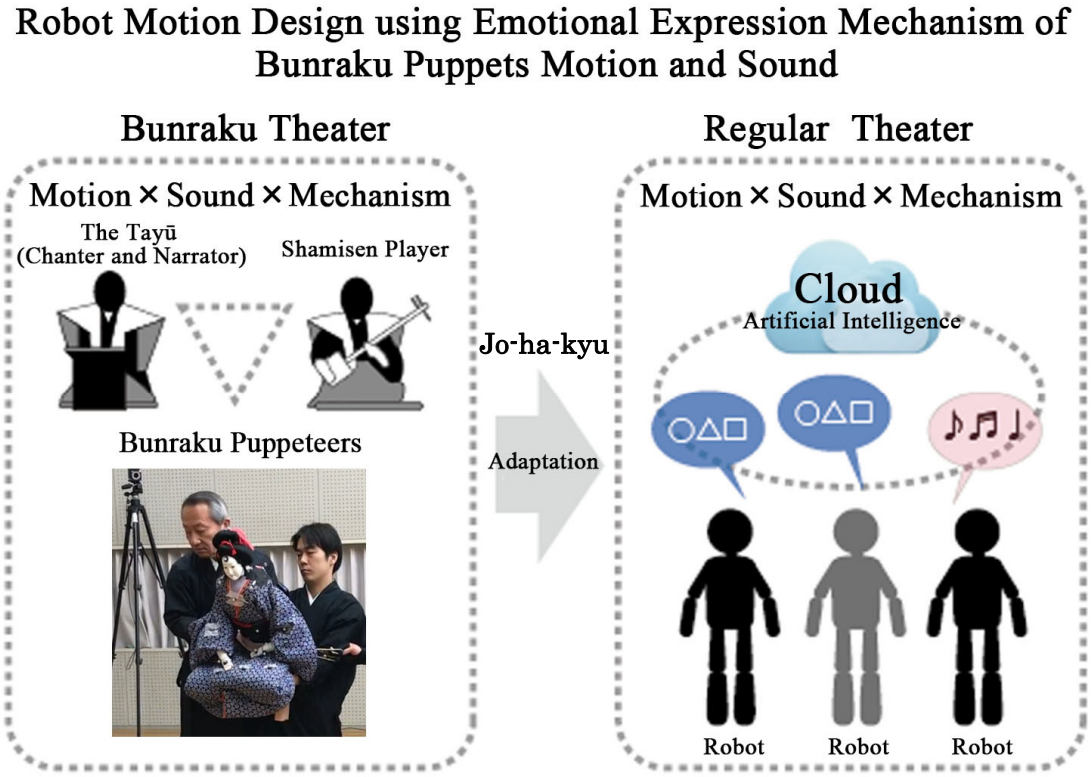

Figure 5.1 Schematics views of Bunraku plays with motions and sounds applied to "robot plays" in real lives

In our study, we propose a new robot motion design framework based on Bunraku puppet motions and the concept of Jo-Ha-Kyū. Using our framework, it is possible to generate affective robot motions using the beat tracking, deep neural network, and HHT from Bunraku puppets.

5.2 Proposed framework

In this dissertation, we present a robot motion design framework using Bunraku affective motions that are based on "Jo-Ha-Kyū," and convert a few simple Bunraku motions into a robot motions using one of deep learning methods. Our primitive experiments show that Jo-Ha-Kyū can be incorporated into robot motion design smoothly, and some simple affective robot motions can be designed using our proposed framework.

5.2.1 Related works and proposed robot motion design framework

Numerous motion synthesis and modifications studies on robotics have been conducted. Somani et al. [73] proposed a framework for cup grasping in 6 to 7 DOF robot arms using the prioritized nonlinear inequality constraints. To ensure the stability of a robot arm, Karami et al. [74] presented a new algorithm in redundant robot manipulators for control of multiple tasks. For humanoid robots, Posa et al. [75] conducted research on whole-body robot dynamic motions. They reported an approach to generate robot motion for completing tasks of walking and climbing based on constrained dynamical systems. Kanehiro et al. [76] also proposed a reaching motion planning method for a 30 DOF humanoid robot using the inverse kinematics. These researches are mainly aimed at the tasks that are analyzed in advance, while the Bunraku motions are interactive and different in every play [57, 58].

Thus, we propose a motion design framework illustrated in Figure 5.2 to develop communication robot motions using Jo-Ha-Kyū. This framework does not necessarily reproduce the motion of Bunraku puppet itself. The robots only follow the Jo-Ha-Kyū principle in their motions just to be "natural." The rhythms and tempos of Jo-Ha-Kyū can only be determined by the stories in Bunraku chants and narrations. These musical factors are out of scope in this dissertation, and we assume Jo-Ha-Kyū of the plays are already known in the proposed framework.

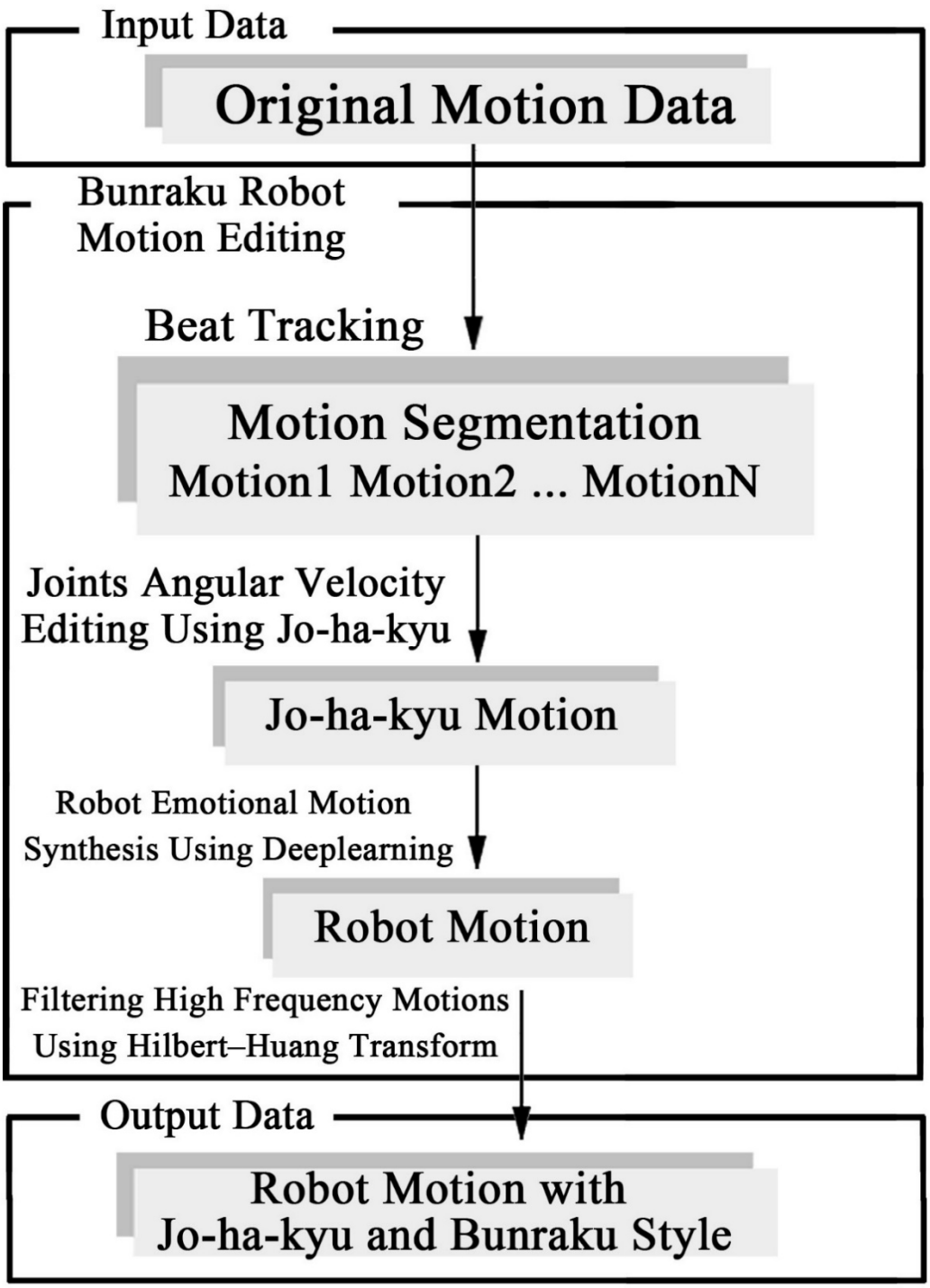

Figure 5.2 General schemes of Bunraku robot motion design using the Jo-Ha-Kyū principle, retargeting deep learning and HHT

5.2.2 Motion editing using Jo-Ha-Kyū principle and beats tracking

As we discussed in chapter 4, Jo-Ha-Kyū motion speeds of Bunraku increase and break the tempos following the play stories. Therefore, it is necessary to edit and change the motion speed following the play stories. As mentioned in the previous chapter, the Jo-Ha-Kyū structure is recursive. It is also necessary to edit and change the motion's speed in a unit length based on their Jo-Ha-Kyū recursive levels. The smallest unit length in the lowest recursive level is "motion primitives." The motion primitive is the minimum unit of motions and it can be segmented following the beats or tempos of music [35]. Using the beat tracking method we discussed in the previous chapter [36], as shown in Figure 5.3 (a), we propose a motion segmentation method to create Jo-ha-kyu using the beat positions and extracts the minimum action or motion primitives [35]. Indeed, the Bunraku puppeteers change and break the motion speeds in each motion primitive. Thus, motion segmentation is essential in our proposed framework. The real Jo-Ha-Kyū patterns have to follow the stories. The way how the motion speeds or tempos increase has to be decided manually according to the stories in our proposed motion design framework.

After the input motions have been segmented, the speed of motions can be edited and adjusted manually according to the stories. Figure 5.3 (b) shows edited sample motions using a simple Jo-Ha-Kyū principle. The increasing red lines represent the slope of peak joint velocities. The tempos or speeds both increase and break 9 times in total in time. Using this method, Jo-Ha-Kyū motions can be created by gradually increasing the angular velocities of each joint as shown in Figure 5.3 (b). Please note Figure 5.3 is only schematic. The real Jo-Ha-Kyū has to follow the stories. Careful manual designs according to the stories are essential.

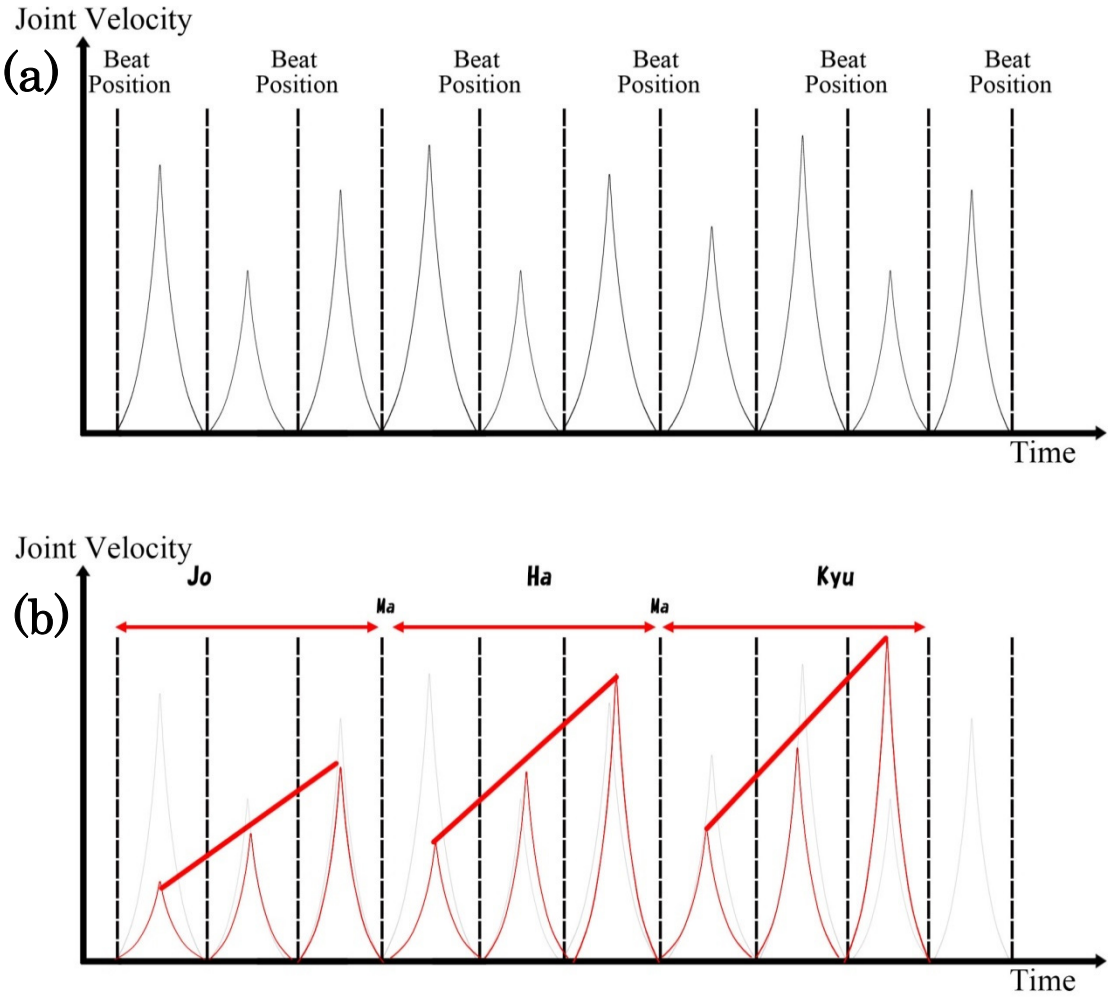

Figure 5.3 (a) Our motion segmentation using beat tracking method (b) Edited segmented motions following the Jo-Ha-Kyū principles

5.2.3 Jo-Ha-Kyū design principle based on Weber-Fechner Law

To design a Jo-Ha-Kyū robot motion, the designer has to decide the number of the Jo-Ha-Kyū recursive levels and its Jo, Ha, Kyū cell structures in each recursive level based on the stories. In this dissertation, for simplicity, we only focus on the two-level recursive structure, and the minimum unit is the motion primitives.

After we determine the time structure or the Jo, Ha, and Kyū cell structures of a robot motion as shown in Figure 5.3 (b), the question here is how we increase and break the angular velocities of each joint motions. For this purpose, we use the Weber-Fechner law to determine the changes in the motion speed in each Jo, Ha, and Kyū cell structures [57, 58]. The Weber-Fechner law is first proposed in 1860

[77] and relates to human perception, more specifically, the relationship between the actual change in a physical stimulus and the perceived change. The Weber–Fechner law can be written as below [78]:

$$S = c\ log_e I + K \qquad (13)$$

Here, S is the magnitude of sensation, I is the intensity of the exciting cause, K is the former stimulus, c is a constant [78].

Using Equation (13), once the motion speed R_i at the time i is obtained, the next motion speed R_{i+1} can be determined following Weber–Fechner law as [57, 58]:

$$R_{i+1} = K\ log 2 R_i + R_i \qquad (14)$$

Here, R_{i+1} are the magnitudes of the sensation of the Jo, Ha, and Kyū at i, R_i is the magnitude of the sensation before R_{i+1}, the number i corresponds to time, and K is a constant that is defined for the play [78]. Some typical values of K are introduced in [57, 58] for Noh play. If K is larger, the audiences feel the plays or motions are moving faster.

After we set the average "Ha" motion speed in "Jo" as the original average motion to be edited, we use the Weber–Fechner law to determine the increases or decreases of the motion speeds.

5.2.4 Robot motion retargeting deep neural learning

After segmenting, editing, modifying the motions following the Jo-Ha-Kyū principle, the motion-captured data must be retargeted to the real robot motion data because the robots usually have less degree of freedom in kinematics. In order to retarget and implement motions to real robots, we have to rewrite the motion data changing and reducing the degree of freedom (DOF) of the motions.

This is simply because the DOFs of the captured motion data from human beings, or in our case, Bunraku puppets are not the same as those of the robots. In reality, this retargeting process needs special skills, thus, is very difficult to do this. Table 5.1 shows the difference between the degrees of freedoms of the Bunraku puppet [79] and the dancing robot named "PremaidAI ™" developed and sold by DMM [80], which is used in this dissertation.

Table 5.1 The Degrees of freedoms of each joint of the Bunraku puppet and the dancing robot PremaidAI™ developed by DMM.

	Bunraku Puppet	PremaidAI
Head	1	3
Chest	3	0
Left Shoulder	3	2
Left Elbow	3	2
Left Hand	1	1
Right Shoulder	3	2
Right Elbow	3	2
Right Hand	1	1
Hip	3	0
Left Hip	3	3
Left Knee	1	1
Left feet	0	2
Right Hip	3	3
Right Knee	1	1
Right feet	0	2
Total	29	25

In order to reduce the DOFs of Bunraku puppet to those of PremaidAI™, in this dissertation, we use a retargeting neural network with deep learnings [1, 48, 64]. Poses of robots are usually described by the joint angles of the motor. This representation is suitable for data processing, but valid retargeted motion only exists in a small subspace of this whole representation set. Motion manifold is a subspace of valid motions [48]. It can be considered as the probability distribution of all valid robot motion data here. Using the retargeting neural network shown in Figure 5.4, we can extract an affective robot motion manifold of the Bunraku puppet [1, 48, 64]. After inputting a set of motion capture data into this neural network, a Bunraku-style robot motion can be retargeted. Here, the term B and R in Figure 5.4 refer to the DOFs of the Bunraku puppet (m = 29) and those of the robot (n = 25), and m> n in Figure 5.4.

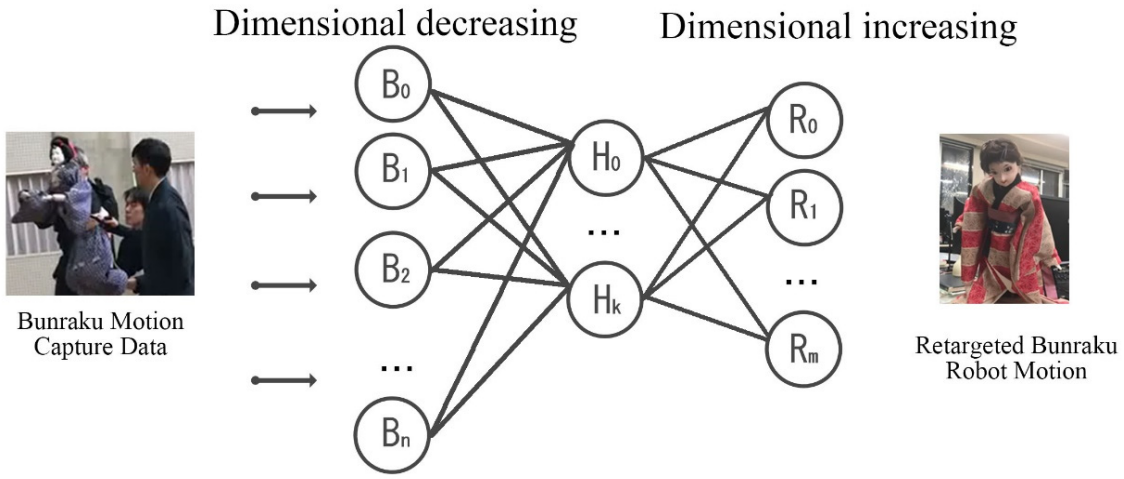

Figure 5.4 Robot motion retargeting using a deep learning network

The convolutional neural network performs a one-dimensional convolution over the temporal domain, independently, for each filter. The dimensional decreasing filters consist of three one-dimensional convolution layers. The dimensional increasing filters also consist of three one-dimensional convolution layers. This network provides a forward (dimensional decreasing) operation Φ and a backward (dimensional increasing) operation $\Phi^{\dagger}$. The forward operation is:

$$\Phi(B) = ReLU(B * W + b) \qquad (15)$$

The backward operation is:

$$\Phi^{\dagger}(R) = (R - b) * \widetilde{W} \qquad (16)$$

Here, B represents Bunraku motion capture data as the input, R represents Robot affective motions as the output, W is the weights of dimensional decreasing network filter, $\widetilde{W}$ is the weights of dimensional increasing network filter, b is a bias for each network, Φ and $\Phi^{\dagger}$ use the max-pooling and up-sampling operation. The values of W and $\widetilde{W}$ are usually initialized to small and random values. The values of b are usually initialized as zero.

The input layer is 29 DOFs (Bunraku puppet) and the output layer is 25 DOFs (PremaidAI™). The learning rate is set to 0.001 and we choose the adaptive moment estimation (ADAM) as the optimizer. The cost function of dimensional increasing is

given and is minimized with respect to the corresponding network parameters as follows:

$$Cost(R) = \|R - \Phi(B)\|_2^2 \qquad (17)$$

After appropriate pieces of training, a robot affective motion manifold is found. By minimizing the cost function Equation (17), an output of Bunraku style robot retargeting is obtained.

In addition, in order to train the neural network in Figure 5.4, the motion capture data of the Bunraku puppet, and the retargeted robot motion data are required as a set. We capture 40 minutes Bunraku puppet data set as shown in chapter 4 and prepared the same 40 minutes robot motion created manually. These are used for direct teachings as shown in Figure 5.5.

Figure 5.5 (a) Capturing Bunraku motion data. (b) Retargeting robot motion data by direct teachings

The training data include 3 scene (20 minutes) selected from "Imoseyama Onna Teikin" introduced in chapter 4, 2 scene (20 minutes) selected from "Keisei Awa no Naruto" ("傾城阿波の鳴門" in Japanese). Their frame rate is 60Hz including 141550 motion data.

5.2.5 Filtering high-frequency motions using HHT

First, the speeds of Bunraku motions are changing rapidly based on Jo-Ha-Kyū. Second, the robot motions generated from the neural network also contain high-frequency noises that exceed the robot motor maximum speed. Thus, motions whose

frequencies are higher than the robot motor speed has to be filtered.

We use the HHT to decompose the motions from high to low angular speeds [8, 49, 50]. We delete all decomposed motions that have higher angular speeds than the motor speed as shown in Figure 5.6 [81]. In the figure, all 5 higher angular velocity IMF motions are filtered or deleted. Thus, the generated motions from our deep learning neural network can well be retargeted to the robot motions.

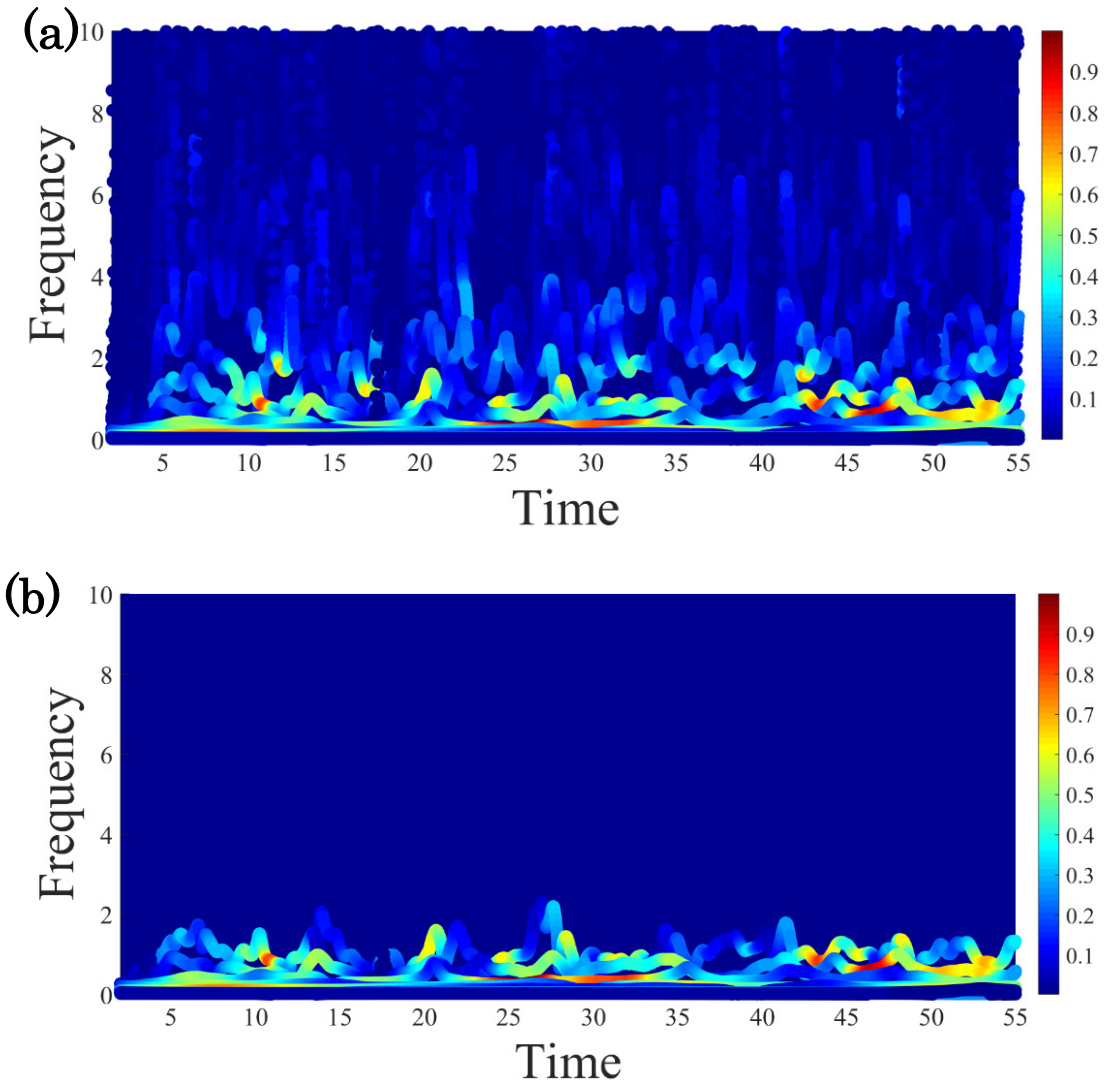

Figure 5.6 (a) The head joint Hilbert spectrum of motions generated from deep learning neural network. (b) The Hilbert spectrum deleted the 5 higher frequency exceed the motor speed

5.3 Result

In this dissertation, our framework uses a deep learning method proposed by Holden et al. [1, 48, 64]. They used 10 hours of motion training data to achieve high-quality complicated motion syntheses or stylizations. However, the purpose of our research is only to retarget simple filtered and edited Bunraku puppet motions to robot motions. As we discussed in chapter 4, Bunraku puppets have approximately 50 beautiful basic motion primitives so-called “Kata” [24]. In the present research, we focus mainly on several basic motions from “Omiwa” (“三輪” in Japanese) of "Imoseyama Onna Teikin”, for example, “walk”, “kneel up”, “stand”, and “sit” [24]. Thus, we use only 40-minute training data that are enough to retarget a few basic motions in the present research. Please note that, in our deep learning network, we do not aim for the deep neural network to obtain the ability to synthesize or stylize the motions.

5.3.1 Basic motion example

In this subsection, we use the retargeting method in section 5.2 to retarget a short basic Bunraku puppet motion to robot motion. Because the original Bunraku puppet motions have Jo-Ha-Kyū in themselves, we do not need to incorporate the Jo-Ha-Kyū rhythms into the original motions using the Jo-Ha-Kyū design method in section 5.2.

Figure 5.7 shows the result of the Bunraku robot's basic motions using our proposed framework. We use about 20-second motion capture data of bunraku puppet selected from “Sugisakaya” part in the lowest recursive level of Jo-Ha-Kyū, which is composed of roughly 17 motion primitives. In the figure, the red lines are the slope of the average joint angular velocities. As evidenced in the figure, the original input data have Jo-Ha-Kyū in themselves. The Bunraku puppet motions are retargeted successfully to the robot motions, smoothly. The robot motions are synchronizing with Tayu’s chant that is one of the significant keys to creating affective motions.

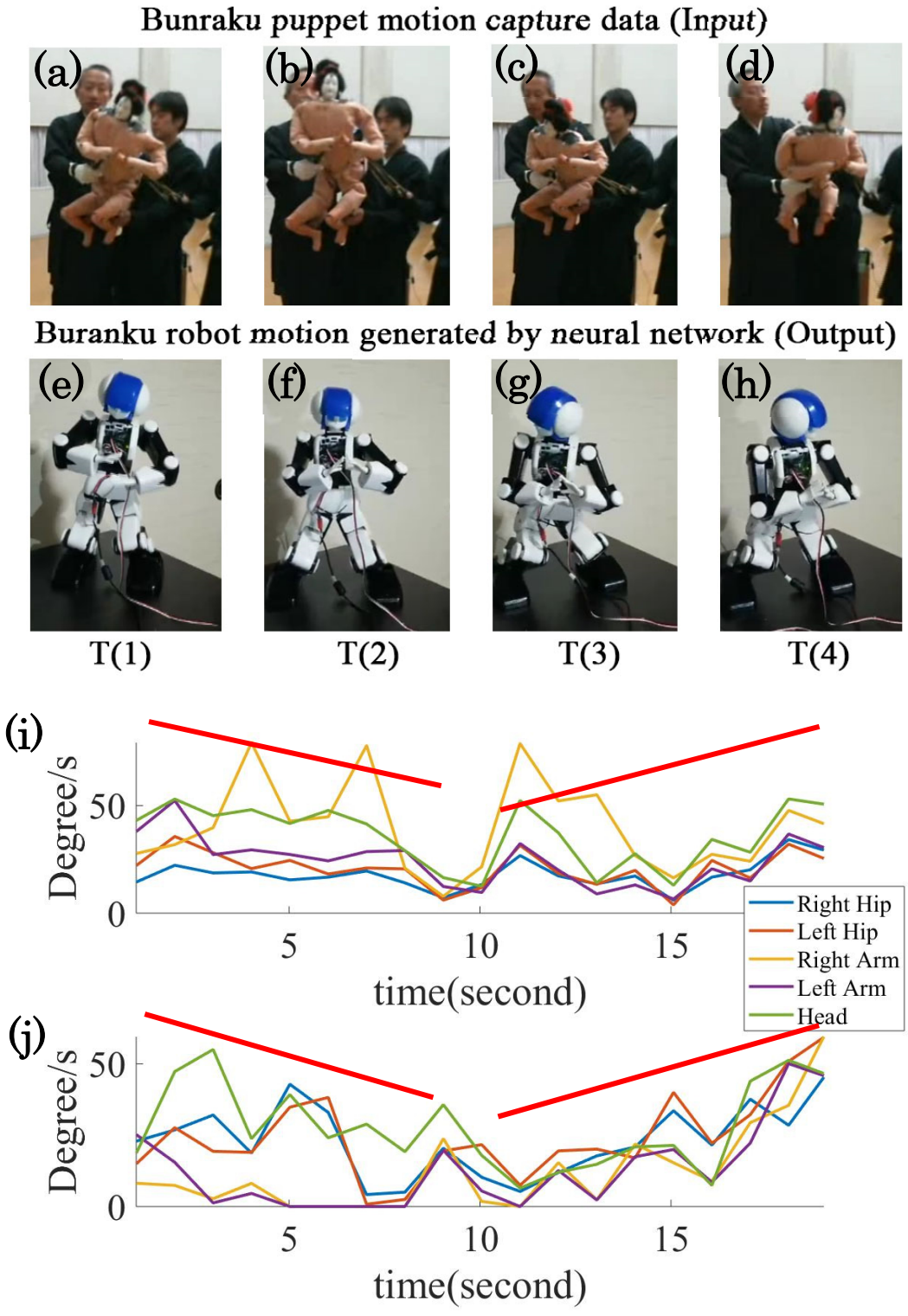

Figure 5.7 (a)-(d) Capturing Bunraku puppet motion data. (e)-(h) Retargeted robot motion generated from deep learning neural network. (i) The upper body joints speeds of Bunraku puppet motion. (j) The retargeted upper body joints speeds of the robot using the deep learning neural network

Here, Figure 5.7 (i) shows the upper body joints speeds of (a)-(d), which have Jo-Ha-Kyū in themselves, and (j) shows the retargeted robot upper body joints speeds of (e)-(h). Thus, Bunraku and robot motion have similar Jo-Ha-Kyū patterns. As can be seen from Figure 5.7, the 40-minute training is enough to retarget the basic bunraku motion.

Using Equation (14), we obtain the coefficient K that controls the impression of plays or motion speeds in Figure 5.7 (i-j). For the original head Bunraku motion of Figure 5.7 (i), we obtain K=2.6±0.3. For the retargeted head Bunraku motion of Figure 5.7(j), we obtain K=2.2±0.3. The retargeted K value is 0.4±0.3 smaller. This is because the high-frequency modes are filtered using the HHT due to the limit of motor speed. However, the changes of motions (Jo-Ha-Kyū) are preserved.

5.3.2 Advanced motion example

Next, we use 1-minute "Perfume" dance data introduced in section 3.3 as the input data for retargeting. Here, Figure 5.8 (a) shows the upper body angular speeds of the original Perfume motion data. The time variations of data are relatively flat. In the figure, the speeds are smoothed over 2 seconds. First, we edit manually Perfume dance motion data into the "Jo-Ha-Kyū" style using the beat tracking method as introduced in section 5.2.2. Figure 5.8 (b) shows the edited or modified joint angular speed using the Jo-Ha-Kyū principle mentioned in chapter 4. Perfume motions are divided into 3-part or Jo-Ha-Kyū. These results show that the speeds of joint motions change (increase) and break (down) three times. Second, we input the edited motion shown in Figure 5.8 (b) to retarget robot motion using the deep learning neural network of Figure 5.4. Figure 5.8 (c) shows the joint angular velocity of robot motions. Thus, the top speeds of robot motion increase following Jo-Ha-Kyū in chapter 4.

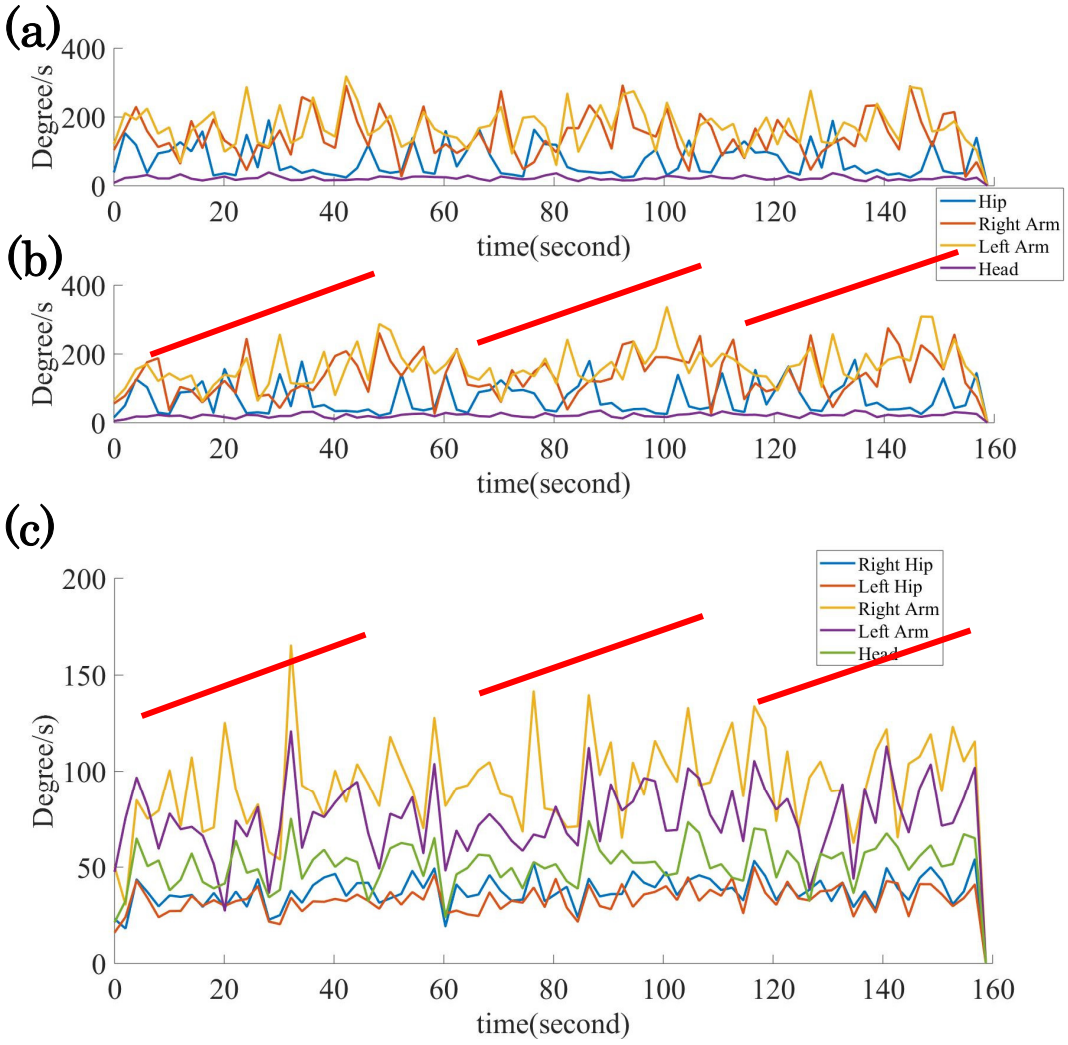

Figure 5.8 (a) The upper body joints speeds of "Perfume" original dance motion data. (b) The modified upper body joints speeds of Perfume dance motion modified manually following Jo-Ha-Kyū principle. (c) The retargeted upper body joints speeds of the robot using the deep learning neural network

Here, Figure 5.9 (a-d) shows the original Perfume dance motion and Figure 5.9 (e-h) shows the retargeted robot motion corresponding to the input data of "Perfume" in Figure 5.8 (b). As shown in Figure 5.8, the input motions are edited into three-part at different speeds. Thus, the output robot motion data has different speed variations, so-called Jo-Ha-Kyū as shown in Figure 5.9. As a result, thus, the modern "Perfume" dance motions are retargeted into the motions that follow Jo-Ha-Kyū principle, which can hopefully express the robot emotion more naturally and comfortably overcoming "Uncanny Valley."

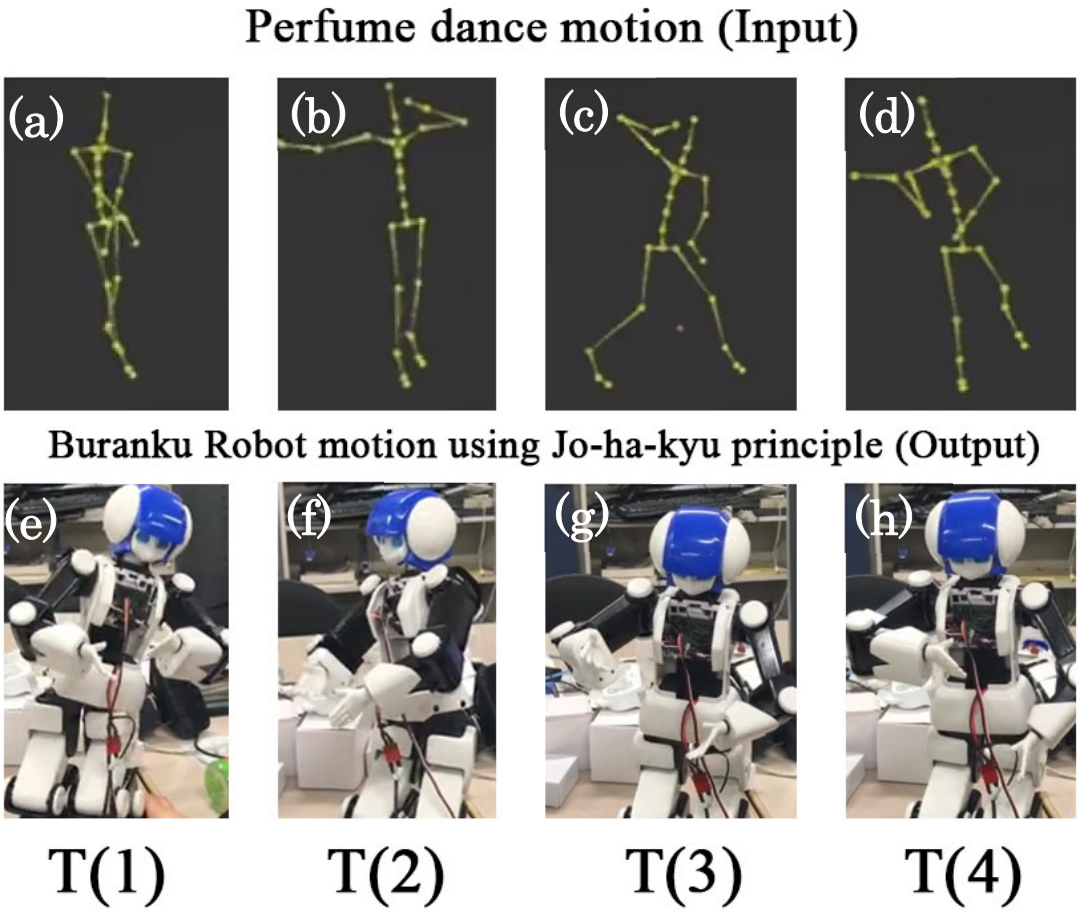

Figure 5.9 (a)-(d) Input motions (Perfume original dance motion). (e)-(h) Retargeted motions (Bunraku robot motion using Jo-Ha-Kyū principle)

Using Equation (14), we can obtain the coefficient K that controls the impression of dances or motion speeds in Figure 5.8(b, c). For the original hand Perfume motions of Figure 5.8(b), we get $K = 1.6 \pm 0.2$. For the retargeted head Bunraku motion of Fig. 5.8(c), we get $K = 1.4 \pm 0.2$. The retargeted K value is slightly smaller. This is because the high-frequency modes are filtered using the HHT due to the limitations of the motor speed. However, the characteristics of Jo-Ha-Kyū is preserved.

5.4 Discussions

In this chapter, we try to apply sound and motion mechanisms of Bunraku plays, Jo-Ha-Kyū, to affective robot motion designs. In order to do so, first, we propose a general scheme of Bunraku robot motion design using the Jo-Ha-Kyū principle and

retargeting deep learning method. Second, we detail a motion editing and morphing method using both the Jo-Ha-Kyū principle and the beat tracking method to segment the motions. Third, we propose to segment a story into three "Jo," "Ha", and "Kyū" cell structures, recursively. We propose to use the Weber–Fechner law to control the increase of motion speeds in each cell structure [57, 58] shown in chapter 4. Fourth, we introduce a motion retargeting deep learning neural network that reduces the Bunraku motion DOFs to the robot motion DOFs. Fifth, after the robot motions are retargeted, we propose to use HHT to filter the high-frequency motions to obtain smoothed motions. Sixth, to evaluate our proposed framework, we give two motion examples to edit and retarget from both the puppet and human motions into robot motions. Then, the main contributions of our works can be summarized as follows.

First, using both Jo-Ha-Kyū principles and Weber-Fechner law, we proposed a unique and new affective robot motion design method. The new method can be summarized as follows: (i) The motions are segmented into the motion primitives using the beat tracking method; (ii) Using the Weber–Fechner law, the segmented motion speeds are modified to Jo-Ha-Kyū cell structures following the stories or can be adopted from some Bunraku puppet motions that intrinsically have Jo-Ha-Kyū in themselves. (iii) Using the retargeting deep learning neural network, the Jo-Ha-Kyū motions including the detailed emotional expressions are retargeted to the robot motions; (iv) Using HHT, the motions are decomposed into different modes or IMFs from fast to slow motions, and those IMF motions faster than the robot motor speeds are canceled or filtered; (v) The final output robot motions are adjusted manually to balance robot inertia forces. Thus, the robot emotional motions based on Jo-Ha-Kyū can be generated, easily. Robot motion designers can create affective Jo-Ha-Kyū motions without special knowledge and techniques.

Second, we use only 40-minute training data to retarget a few basic motions, while Holden et al. [1, 48] used 10 hours of motion training data to achieve high-quality complicated motion syntheses or stylizations. We evidence the 40-minute training is enough to retarget the basic bunraku motion. And using our proposed neural network, detailed emotional expressions of Bunraku motion can be retargeted automatically, while a 3 minutes Bunraku motion needs 30 hours to retarget manually.

5.5 Summary and conclusions

As discussed in chapter 4, the affective robot motions have to be accompanied by sounds and narrations and cannot be generated only by the motion generation system. After analyzing the mechanism of Jo-Ha-Kyū in the instantaneous frequency domain in chapter 4, in this chapter, we try to convert these emotional motions into robot affective motions so that robots can interact with human beings more comfortable. As shown in the results, we can automatically design the Jo-Ha-Kyū tempos or rhythms using the Weber-Fechner law after the play motions are organized to be Jo-Ha-Kyū cell structures. To show the results, we convert a few simple Bunraku motions into a robot motion. Our experiments reveal that Jo-Ha-Kyū can be incorporated into robot motion design smoothly, and some simple affective robot motions can be designed using our proposed framework. Thus, in this chapter, our conclusions can be summarized as follows:

1. Because Jo-ha-kyu is used in both long term and short-term with Bunraku puppet principles, in our proposed framework, we used beat tracking to segment motion for Jo-Ha-Kyū editing. As shown in the results, Jo-Ha-Kyū can be created by changing the motion speed of each segmented motions by beat tracking;

2. In our present framework, we use the Weber-Fechner law to create affective robot motion based on previous research of Jo-Ha-Kyū [57, 58]. By editing the motion speed of each segmented motion by beat racking, affective robot motions can be synthesized automatically;

3. We use deep learning to retarget Bunraku affective motion for a robot. By using the motion synthesis framework of previous researches [57, 58], retargeted robot motion can be generated from input Bunraku puppet motions. The generated motion can also represent the Jo-Ha-Kyū pattern; and

4. To fit affective Jo-Ha-Kyū motion into a robot, we use HHT to sparsify motion to fit the motor speed of the robot. By using our proposed framework, the generated affective Jo-Ha-Kyū motion can be expressed correctly by the robot.

In this chapter, we only focused on the affective motion designs, and we manually

edited the input motions to be Jo-Ha-Kyū. However, once the robot sounds and narrative AI systems are developed, our proposed scheme can easily be incorporated into them, and the affective motions can easily be generated automatically using our method.

Chapter 6

Conclusions and future works

In this chapter, we summarize the main contributions of this dissertation and discuss possible future works.

6.1 Conclusions

Traditional methods of data analysis are based on linear and smooth hypotheses. Recently, new methods have been introduced to analyze non-stationary and non-linear data. In addition, some nonlinear time series analysis methods have been designed for nonlinear but stationary deterministic systems. However, in most practical systems, whether natural or artificial, the data are likely to be nonlinear and unstable. To solve this problem, Hilbert-Huang Transform (HHT) uses Empirical Mode Decomposition to decompose a real-world signal into a set of Intrinsic Mode Functions (IMFs) and a residual (trend). The IMF is a decomposed pseudo monochromatic signal. Thus, we can apply the Hilbert transform (HT) to each IMF to get instantaneous frequency and amplitude of the original signal correctly. This transform is suitable for the analysis of non-linear and non-stationary data in an empirical way. In all cases, the results of HHT in time-frequency energy representation are much clearer than those of any traditional analysis method.

Although HHT has been applied in a wide range of researches, there are few relevant researches on motion analysis and editing for signals collected from the real world. Therefore, we carried out the experiments and analyses around this issue, proposed an analysis and the editing framework and applied them to CG character and robot motion synthesis. To show the results of our proposed framework, we applied our framework to three different problems: (i) Dance motion analysis and editing; (ii) Bunraku puppet motion analysis; and (iii) Robot motion synthesis using Bunraku puppets. The conclusions of this dissertation can be summarized in three parts.

6.1.1 Dance motion analysis and editing

We proposed a framework for dance motion analysis and editing using HHT. As an example, Japanese techno-pop unit Perfumes' dance motions, salsa, waltz, and hip-hop motions were analyzed and edited. In addition, some of their interesting and unique features and editing results were discussed with regard to Hilbert spectrums. The conclusions of this research can be summarized as follows:

1. Hilbert-Huang Transform can be considered as a powerful tool for dance motion analysis and editing. By using NA-MEMD to decompose dance motions into a set of IMFs, we can extract dance choreographies from dance motions in the instantaneous frequency domain. Thus, our proposed framework helps choreographers to analyze dance motions in the choreography units. Thus, animators can edit dance motion not just for editing whole dance motion, but also can edit each choreography separately in the instantaneous frequency domain for creating novel different dances; and

2. Hilbert-Huang Transform can reveal new insights in motion analysis. For example, our research reveals the Fibonacci sequence in human dance motions is corresponding to each joint. Also, the Fibonacci sequence is associated with the golden principle, which is one of the important ruling factors for aesthetics. Thus, there is a possibility that this relationship may have some connections with the aesthetic elements of dances.

6.1.2 Bunraku puppet motion analysis

We first collected Bunraku motion using motion capture. Next, we analyzed the Jo-Ha-Kyū mechanism of Bunraku puppet motions with puppet principles using HHT in both long term and short term. The conclusions of this part of the research can be summarized as follows:

1. Hilbert-Huang Transform can decompose Bunraku puppet affective motions into several IMFs. These decomposed motions can be analyzed in the instantaneous frequency domain. By analyzing Bunraku motion with their Hilbert spectrum, the Jo-Ha-Kyū of Bunraku motion was confirmed in our research. We also showed the difference in motion features between Bunraku and Western dance (Perfume dance) using HHT; and

2. Hilbert-Huang Transform can extract delicate and complicated motion of Bunraku puppet motions, like Ma between the Jo, Ha, and Kyū. The extracted Ma can be considered as “Anticipation”, one of the puppet principles, that plays an important role in interaction with humans. These complicated motions extracted by HHT can be used in affective motion synthesis.

6.1.3 Robot motion synthesis using Bunraku puppets

We developed a new robot motion design method that retargets the beautiful affective Bunraku motions to the android robot motions using Jo-Ha-Kyū. Then, we presented a framework using deep learning to generate emotional motions. These emotional motions were synchronized with the concept of Jo-Ha-Kyū (序破急). By our proposed framework, a robot can express affective motion synchronized with narration (story) based on Jo-Ha-Kyū (序破急). The conclusions of this research can be summarized as follows:

1. Hilbert-Huang Transform can be considered as a great tool for motion sparsity. We used HHT to remove complicated high-frequency motions. These motions cannot be represented by the robot because of the motor speed limitation; and

2. These decomposed motions can be considered for use in deep learning methods for interaction motion synthesis between humans and robots. By using HHT, unnecessary high-frequency IMFs can be removed. Thus, it is possible to easily sparsify motion data by removing the IMFs with t small spectrum energy in each channel and remove the unnecessary IMFs. These can be used for sparse coding to generate efficient training data in deep learning.

6.2 Future works

Our research reveals that HHT can be a useful tool for motion analysis and editing for different areas like CG characters, Bunraku puppets, and robots. However, for the three different problems we explored in this dissertation, there are also three future works regarding our works regarding motion analysis and editing using HHT.

6.2.1 Dance motion analysis and editing

We used HHT for dance motion analysis and editing. Our research revealed the Fibonacci sequence in human dance motions corresponding to each joint. The Fibonacci sequence is associated with the golden principle, which is one of the important ruling factors for aesthetics. Thus, there is a possibility that this relationship may have some connections with the aesthetic elements of dances. However, the relationship between the Fibonacci sequence and the beauty of dance motions is unclear. Also, other motions, like motions in a normal scene or battle scene, are also considered can be used with our proposed framework. We would like to leave these topics for future work.

6.2.2 Bunraku puppet motion analysis

Jo-Ha-Kyū motion patterns are one of the effective ways to express emotions. We used HHT to analyze and extract Bunraku puppets Jo-Ha-Kyū motion like "Anticipation", one of the Bunraku puppet principles, by our proposed framework. There are also other puppet principles like "slow in slow out" and "squash and stretch". These principles are also considered to be very effective for expressing emotions to overcome the "Uncanny Valley." However, Jo-Ha-Kyū principles can also be easily incorporated into these other methods. The Bunraku masters only teach their students by direct teaching. They never use textbooks and many other details of Jo-Ha-Kyū remain unknown to us. It is essential to reveal these unknown details for the deep neural network training to generate emotional motions in the future.

6.2.3 Robot motion synthesis using Bunraku puppets

Using our proposed framework using deep learning and HHT with Jo-ha-kyu can make robots interact with humans automatically. We believe that our proposed framework can contribute to the automatic generation of robot emotional motions using Jo-Ha-Kyū in the future. However, our research is still at an early stage, and a large number of training datasets, with Tayū's narrations that should contain natural emotions following the Joruri stories, are essential. This is because the Jo-Ha-Kyū motion patterns should be created following the stories automatically by the deep learning neural network. However, large data acquisition from Bunraku puppeteers, Tayū and Shamisen players is not easy. In addition, some objective emotional evaluation metrics in accordance with the emotional motions following the

stories are required to train the deep neural network. The generated emotional robot motions also have to be objectively evaluated. We would like to address these issues in future work.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Professor Dongsheng Cai for his continuous supports of my doctoral study with his patience and immense knowledge. Prof. Dongsheng Cai is a great source of inspiration and motivation for my doctoral research. I have to admit that without the help of my supervisor, I could not finish my studies.

Besides my advisor, I would like to thank the following committee members of my doctoral dissertation: Professor Akihisa Ohya, Professor Takashi Tsubouchi, Professor Jun Mitani, and Professor Hiroyuki Kudo for their valuable comments and encouragement. Their questions and comments greatly encourage me to improve the quality of this doctoral dissertation.

Next, I would like to express my deepest gratitude to my co-researcher, Professor Shinobu Nakagawa, Professor Tomonari Higaki, Professor Shingo Hayano, Professor Nobuyoshi Asai, for making this research possible. Their excellent caring and patience provide me an excellent environment for my doctoral research. I would like to thank the performers of Bunraku, Tsukoma Takemoto (Tayu), Sosuke Takezawa (Shamisen), and the Puppeteers team Kanjuro Kiritake and 2 other team members, for providing the sound and motion data for this research. I would also like to thank all members of Ningyo Joruri Club at Tokushima Bunri University for providing the puppet motion as the training data. My sincere thanks also go to Kuryu Nishikawa and all members of Hachioji Kuruma Ningyo, who give me profession advises on the Japanese traditional art concept, Jo-ha-kyu in Bunraku puppet, which has bought tremendous help to my research and to this dissertation. I also would like to thank all members, especially Mrs. Obana in CAVE Lab, for their kindness helping in my daily life of study in Japan.

I would like to thank the Takaku-foundation and Rotary Yoneyama Memorial Foundation for the scholarship support. Especially I would like to thank my instructor, Mr. Kazuyuki Itou, for all his kindness to a foreign student like me. I would also like to thank the Japan Science and Technology Agency and the Telecommunications Advancement Foundation for funding this research.

Finally, I would like to thank my families for their unconditional support. I would also like to thank my dear wife, Ni Shaowen who is a doctoral student of Policy and Planning Sciences. She is my primary motivation for finishing my doctoral degree.

Bibliography

[1] Holden, D., Saito, J. and Komura, T. A deep learning framework for character motion synthesis and editing. *ACM Transactions on Graphics (TOG)*, 35, 4 (2016), 138.

[2] Choi, B., Lewis, J., Seol, Y., Hong, S., Eom, H., Jung, S. and Noh, J. SketchiMo: sketch-based motion editing for articulated characters. *ACM Transactions on Graphics (TOG)*, 35, 4 (2016), 146.

[3] Aberman, K., Wu, R., Lischinski, D., Chen, B. and Cohen-Or, D. Learning character-agnostic motion for motion retargeting in 2D. *ACM Transactions on Graphics (TOG)*, 38, 4 (2019), 75.

[4] Kim, M., Hyun, K., Kim, J. and Lee, J. Synchronized multi-character motion editing. *ACM transactions on graphics (TOG)*, 28, 3 (2009), 79.

[5] Kim, T. H., Park, S. I., and Shin, S. Y. . Rhythmic-motion synthesis based on motion-beat analysis. In *ACM Transactions on Graphics (TOG),* 22, 3(2003), 392-401.

[6] Pons-Moll, G., Romero, J., Mahmood, N. and Black, M. J. Dyna: A model of dynamic human shape in motion. *ACM Transactions on Graphics (TOG)*, 34, 4 (2015), 120.

[7] Kwon, T., Cho, Y.-S., Park, S. I. and Shin, S. Y. Two-character motion analysis and synthesis. *IEEE Transactions on Visualization and Computer Graphics*, 14, 3 (2008), 707-720.

[8] Huang, N. E. *Hilbert-Huang transform and its applications*. World Scientific, 2014.

[9] Bracewell, R. N. and Bracewell, R. N. *The Fourier transform and its applications*. McGraw-Hill New York, 1986.

[10] Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C. and Liu, H. H. The empirical mode decomposition and the Hilbert

spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454, 1971 (1998), 903-995.

[11] Wiener, N. Generalized harmonic analysis. *Acta mathematica*, 55 (1930), 117-258.

[12] Yamazaki, Y. Time-frequency analysis of acoustic signals (<Special issue> Sound field / acoustic signal model and its analysis). *Journal of the Acoustical Society of Japan*, 53, 2 (1997), 147-153. Japanese.

[13] Rehman, N. and Mandic, D. P. Multivariate empirical mode decomposition. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 466, 2117 (2009), 1291-1302.

[14] Rilling, G., Flandrin, P., Gonçalves, P. and Lilly, J. M. Bivariate empirical mode decomposition. *IEEE signal processing letters*, 14, 12 (2007), 936-939.

[15] ur Rehman, N. and Mandic, D. P. Empirical mode decomposition for trivariate signals. *IEEE Transactions on signal processing*, 58, 3 (2009), 1059-1068.

[16] ur Rehman, N., Park, C., Huang, N. E. and Mandic, D. P. EMD via MEMD: multivariate noise-aided computation of standard EMD. *Advances in Adaptive Data Analysis*, 5, 02 (2013), 1350007.

[17] Guo, M.-F., Yang, N.-C. and Chen, W.-F. Deep-Learning-based Fault Classification using Hilbert-Huang Transform and Convolutional Neural Network in Power Distribution Systems. *IEEE Sensors Journal* (2019).

[18] Kaleem, M., Guergachi, A., and Krishnan, S. Empirical mode decomposition based sparse dictionary learning with application to signal classification. In *2013 IEEE Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)* (2013) , pp. 18-23.

[19] Qiu, X., Ren, Y., Suganthan, P. N. and Amaratunga, G. A. Empirical mode decomposition based ensemble deep learning for load demand time series forecasting. *Applied Soft Computing*, 54 (2017), 246-255.

[20] Rodriguez, M. A., Sotomonte, J. F., Cifuentes, J., and Bueno-López, M. (2019, September). Classification of Power Quality Disturbances using Hilbert Huang Transform and a Multilayer Perceptron Neural Network Model. In *2019 International Conference on Smart Energy Systems and Technologies (SEST)* (2019) pp. 1-6.

[21] Vazirizade, S. M., Bakhshi, A. and Bahar, O. Online nonlinear structural damage detection using Hilbert Huang transform and artificial neural networks. *Scientia Iranica*, 26, 3 (2019), 1266-1279.

[22] Zhang, G., Wang, Z., Zhao, L., Qi, Y. and Wang, J. Coal-rock recognition in top coal caving using bimodal deep learning and hilbert-huang transform. *Shock and Vibration*, 2017 (2017).

[23] VoiceLabs. The 2017 Voice Report by VoiceLabs. 2017. http://voicelabs.co/2017/01/15/the-2017-voice-report

[24] Culture Digital Library. Invitation to UNESCO Intangible Cultural Heritage Bunraku. http://www2.ntj.jac.go.jp/unesco/bunraku/jp/introduction/index.html

[25] Mori, M., MacDorman, K. F. and Kageki, N. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19, 2 (2012), 98-100.

[26] Chen, Q., Huang, N., Riemenschneider, S. and Xu, Y. A B-spline approach for empirical mode decompositions. *Advances in computational mathematics*, 24, 1-4 (2006), 171-195.

[27] Boashash, B. Estimating and interpreting the instantaneous frequency of a signal. I. Fundamentals. *Proceedings of the IEEE*, 80, 4 (1992), 520-538.

[28] Niu, J., Liu, Y., Jiang, W., Li, X. and Kuang, G. Weighted average frequency algorithm for Hilbert–Huang spectrum and its application to micro-Doppler estimation. *IET Radar, Sonar & Navigation*, 6, 7 (2012), 595-602.

[29] Li, Y., Wang, T., and Shum, H. Y. Motion texture: a two-level statistical model for character motion synthesis. In *ACM transactions on graphics (ToG)* ,21,3 (2002), 465-472.

[30] Shiratori, T., Nakazawa, A., and Ikeuchi, K. Detecting dance motion structure through music analysis. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings* (2004). pp. 857-862.

[31] Shiratori, T. and Ikeuchi, K. Synthesis of dance performance based on analyses of human motion and music. *Information and Media Technologies*, 3, 4 (2008), 834-847.

[32] Chan, J. C., Leung, H., Tang, J. K. and Komura, T. A virtual reality dance training system using motion capture technology. *IEEE Transactions on Learning Technologies*, 4, 2 (2010), 187-195.

[33] Huang, J., Xie, J., Li, F. and Li, L. A threshold denoising method based on EMD. *Journal of Theoretical and Applied Information Technology*, 47, 1 (2013), 419-424.

[34] Shiratori, T., Nakazawa, A. and Ikeuchi, K. The Structure Analysis of Dance Motions using Motion Capture and Musical Information. *IEICE transactions on information and systems (Japanese edition)*, 88, 8 (2005/08/01 2005), 1662-1671. Japanese.

[35] Yamada, Y. and Ueda, A. 2P2-Q05 Analysis of classical dance for extracting motion primitives (Informative Motion & Motion Media). *Proceedings of Robotics and Mechatronics Lectures*, 2011 (2011 2011), _2P2-Q05_01-_02P02-Q05_02. Japanese.

[36] Ellis, D. P. Beat tracking by dynamic programming. *Journal of New Music Research*, 36, 1 (2007), 51-60.

[37] Morishita, H. Biomechanism of dance: Japanese movements / Western movement. *Journal of the Society of Biomechanisms*, 5 (1980), 5-16.

[38] Perfume global site. Perfume dance motion. http://www.perfume-global.com/

[39] songbpm Perfume Enter The Sphere BPM. https://songbpm.com/perfume/enter-the-sphere

[40] Motion Capture Database HDM05. Waltz motion data. http://resources.mpi-inf.mpg.de/HDM05/03-01/index.html

[41] tf3dm. Hip-hop motion data. http://tf3dm.com/3d-model/bvh-1-74368.html

[42] CMU's Motion Capture Database. Salsa motion data. https://sites.google.com/a/cgspeed.com/cgspeed/motion-capture/cmu-bvh-conversion

[43] Hip-hop BPM. http://hiphop.ldblog.jp/archives/28050799.html

[44] Waltz BPM. http://www14.plala.or.jp/nekokirin/02aboutbpm/01aboutbpm.html

[45] Salsa BPM. http://strings6.blog14.fc2.com/blog-entry-543.html

[46] Yamane, K. and Nakamura, Y. Natural motion animation through constraining and deconstraining at will. *IEEE Transactions on visualization and computer graphics*, 9, 3 (2003), 352-360.

[47] Fragkiadaki, K., Levine, S., Felsen, P., and Malik, J. Recurrent network models for human dynamics. In *Proceedings of the IEEE International Conference on Computer Vision* (2015), pp. 4346-4354.

[48] Holden, D., Saito, J., Komura, T., and Joyce, T. Learning motion manifolds with convolutional autoencoders. In *SIGGRAPH Asia 2015 Technical Briefs* (2015), p. 18.

[49] Dong, R., Cai, D., and Asai, N. Dance motion analysis and editing using hilbert-huang transform. In *ACM SIGGRAPH 2017 Talks* (2017), p. 75.

[50] Dong, R., Cai, D., & Asai, N. Nonlinear dance motion analysis and motion editing using Hilbert-Huang transform. In *Proceedings of the Computer Graphics International Conference* (2017), p. 35.

[51] Albert, R. and Barabási, A.-L. Statistical mechanics of complex networks. *Reviews of modern physics*, 74, 1 (2002), 47.

[52] Gill, P. R., Wang, A. and Molnar, A. The in-crowd algorithm for fast basis pursuit denoising. *IEEE Transactions on Signal Processing*, 59, 10 (2011), 4595-4605.

[53] Macagon, V., and Wünsche, B. Efficient collision detection for skeletally animated models in interactive environments. In *Proceedings of IVCNZ* (2003), Vol. 3, pp. 378-383.

[54] Moore, M., and Wilhelms, J. Collision detection and response for computer animation. In *ACM Siggraph Computer Graphics* , 22, 4(1998), pp. 289-298.

[55] Hui, Y.-c. *Japanese Noh theatre: the aesthetic principleof Jo-ha-kyu in the play Matsukaze*. HKU Theses Online (HKUTO) , 1999.

[56] Konparu, *K. Invitation to Noh - Science between Jo-Ha-Kyū and Ma.* Tanko magazine, 1980. Japanese.

[57] Tamba, *A. The musical structure of Nō.* Tokai University Press, 1981.

[58] Tamba, A. *The Beauty of "Jo-Ha-Kyū": Japanese music thinking style revives today*. Ongaku no Tomo magazine, 2004. Japanese.

[59] Tsuji, T., Yuki, N., Kazuya, S., Toshio, M. and Ueda, K. Nonverbal Information that Enables Bunraku Puppeteers to Cooperatively Manipulate a Bunraku Puppet. *The IEICE transactions on information and systems (Japanese edition)*, 96, 1 (2013/01/01 2013), 195-208. Japanese.

[60] Hattori, M. *A study on the analysis of emotionality in acting performance of Bunraku puppet.* Kobe University, 2000. Japanese.

[61] Hattori, M., Tsuji, S., Tadokoro, S., Toshi, T. and yamada, K. A Method to Emphasize the Emotional Factor of the Humanoid Robots' Actions. An Analysis of the Bunraku Puppet's Actions using KM2O-Langevin Equations with Initial Times. *Transactions of the Japan Society of Mechanical Engineers*, 66, 644 (2000 2000), 1236-1242. Japanese.

[62] Hattori, H., Nakabo, Y., Tadokoro, S., Takamori, T., and Yamada, K. An analysis of the Bunraku puppet's motions based on the phase correspondence of the puppet's motions axis-for the generation of humanoid robots motions with fertile emotions. In *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)* (1999), Vol. 2, pp. 1041-1046.

[63] Hattori, M., Tsuji, M., Nakabou, Y., Tadokoro, S., Takamori, T., and Yamada, K. A motion analysis of the Bunraku puppet for generation of emotional robot actions based on KM/sub 2/O-Langevin model. In *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation* (1997),

Vol. 4, pp. 3267-3272.

[64] Chen, Y., Dong, R., Cai, D., Nakagawa, S., Higaki, T., and Asai, N. The beauty of breaking rhythms: affective robot motion design using Jo-Ha-Kyū of bunraku puppet. In *ACM SIGGRAPH 2019 Talks* (2019), pp. 1-2.

[65] Yamada, C. *Standards and transformations in narration of Tayu.* Kyoto City University of the Arts, 2017. Japanese.

[66] Lasseter, J. (1987, August). Principles of traditional animation applied to 3D computer animation. In *ACM Siggraph Computer Graphics* (1987), Vol. 21, No. 4, pp. 35-44.

[67] Kazi, R. H., Grossman, T., Umetani, N., and Fitzmaurice, G. Motion amplifiers: sketching dynamic illustrations using the principles of 2D animation. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), pp. 4599-4609.

[68] Johnston, O., and Thomas, F. *The illusion of life: Disney animation* (p. 576). New York: Disney Editions. 1981.

[69] Mavridis, N. A review of verbal and non-verbal human–robot interactive communication. *Robotics and Autonomous Systems*, 63 (2015), 22-35.

[70] Valin, J.-M., Michaud, F. and Rouat, J. Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering. *Robotics and Autonomous Systems*, 55, 3 (2007), 216-228.

[71] Nakamura, K., Sinapayen, L., and Nakadai, K. Interactive sound source localization using robot audition for tablet devices. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2015), pp. 6137-6142.

[72] Nakagawa, S. *The Study of Robotics Design to Express Robot Emotion: Effect of a New Emotional Robot Therapy for Demented Elderly Patients -Using the Automatic Generation of Face Movie* -. Mie University, 2016. Japanese.

[73] Somani, N., Rickert, M., Gaschler, A., Cai, C., Perzylo, A., and Knoll, A. Task level robot programming using prioritized non-linear inequality constraints. In *2016*

IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2016), pp. 430-437.

[74] Karami, A., Sadeghian, H. and Keshmiri, M. Novel approaches to control multiple tasks in redundant manipulators: stability analysis and performance evaluation. *Advanced Robotics*, 32, 10 (2018), 535-546.

[75] Posa, M., Kuindersma, S., and Tedrake, R. Optimization and stabilization of trajectories for constrained dynamical systems. In *2016 IEEE International Conference on Robotics and Automation (ICRA)* (2016), pp. 1366-1373.

[76] Kanehiro, F., Yoshida, E. and Yokoi, K. Efficient reaching motion planning method for low-level autonomy of teleoperated humanoid robots. *Advanced Robotics*, 28, 7 (2014), 433-439.

[77] TF, G. *Elemente der psychophysik*. Leipzig: Breitkopfund Hartel, 1860. German.

[78] Hecht, S. The visual discrimination of intensity and the Weber-Fechner law. *The Journal of general physiology*, 7, 2 (1924), 235-267.

[79] Miyao, S. *The Book of Shigeo Miyao <1> Bunraku puppet Chart*. Kanosyobo, 1984. Japanese.

[80] DMM.com PremiadAI - World-class dance communication robot - http://robots.dmm.com/robot/premaidai/spec

[81] KONDO Robot. KRS-2552RHV ICS. https://kondo-robot.com/product/03067e

List of Publications

Refereed International Journal

Ran Dong, Yang Chen, Dongsheng Cai, Shinobu Nakagawa,Tomonari Higaki, Nobuyoshi Asai, Robot Motion Design Using Bunraku Emotional Expressions - Focusing on Jo-ha-kyū in Sounds and Movements. Advanced Robotics, Special Issue - Full paper.

Refereed Japanese Domestic Journal

董然, 蔡東生, ヒルベルトーファン変換を用いたダンスモーション解析, 電子情報通信学会和文論文誌(D) Vol.J102-D,No.12,pp. 843-853, Dec. 2019.

Refereed International Conference Papers

Ran Dong, Dongsheng Cai, and Nobuyoshi Asai. 2017. Dance Motion Analysis and Editing using Hilbert-Huang Transform. In Proceedings of SIGGRAPH '17 Talks, Los Angeles, CA, USA, July 30 - August 03, 2017, 2 pages.

Ran Dong, Dongsheng Cai, and Nobuyoshi Asai. 2017. Nonlinear Dance Motion Analysis and Motion Editing using Hilbert-Huang Transform. In Proceedings of CGI '17, Yokohama, Japan, June 27-30, 2017, 6 pages.

Non-refereed Domestic Conference Papers

董然, 蔡東生 , 中川志信 , 檜垣智也, 文楽における序破急のメカニズム解析, 2019 年電子情報通信学会総合大会, 2019-03-20

董然, 蔡東生, 中川志信, 檜垣智也, "ヒルベルトーファン変換を用いた文楽人形動作解析とロボットモーションデザイン", 情報処理学会研究報告. 研究報告コンピュータグラフィックスとビジュアル情報学, 2018-CG-171(18),1-2 (2018-9-23) , 2188-894

董然, 蔡東生,浅井信吉, "ヒルベルトーファン変換を用いた日本舞踊と欧米舞踊の動作

特徴分析 －AyaBambi の舞踊動作解析を中心に－”, 情報処理学会研究報告. 人文科学とコンピュータ研究会報告（CH）,2018-CH-117(4),1-4 (2018-05-05) , 2188-8957