Study of Music Effects on Human Factors and Its Application

2012

Ryosuke Yamanishi

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Music is one of arts that have been loved all around the world since ancient times, and presently we do not spent daily time without listening to music. And most of us experienced to play some musical instruments, we enjoy music not only passively but also actively. This study focuses on music as one of arts, and researched music effects on human factors and develops applications based on the revealed facts to receive more pleasure of music. To achieve this study, varied and many computer science technology is applied: bio-signal analysis, audio-signal analysis, and varied data-mining techniques.

The following sections in this chapter describe the relation between computer science and human factors, how music is penetrated in human life, computer science for music informatics, and the detail composition of this paper.

1.1 Computer Science and Human Factors

In the field of computer science and engineering, so many studies recently focus on *Human-factors* [1] which deals psychology, physiology, engineering, industrial design, statistics, and so on. Human factors engineering focuses on "How people interact with objects considering characteristics of human," and "How people feel the product affectively and emotionally." Evaluations of the end-user should be respected to construct all products in engineering, as long as the products are eventually used by public users. Even though the products demonstrate mathematically superior performance, if the product does not fulfill the users' demand, the product comes down [2]. This is the reason why human-factors recently attracts attention as with product's performance. That is to say, it has changed from "Physical period" to "Heart period."

Emotional/affective engineering (i.e., *Kansei* engineering in Japanese) has been expressly focused, that intends to construct user-centric and user-satisfactory products. To achieve this, it has to clarify the relationships between the feelings and the product's properties, and design the products reflecting user sensory preference. And computer science technologies are used as the key tools, for example, data-mining, signal processing, modelization, machine learning and so on. With emotional/affective engineering, there is a lot of research about Human-Computer/Robot interaction, analysis on human communication and arts such as painting and **music**, because the emotional/affective engineering provides knowledge on how to design more attractive and available products and make user-satisfy.

1.2 Music and Human Life

Nowadays, we can not spend any day without listening to music; we can hear so many promotion music from store when we walk the streets of the city, we can listen to some music that enhances the mood from most commercial messages and TV programs. Even in airport and station which are public institutions, music is used to call attention to passengers. Regardless of being conscious or unconscious, music has always been in human lives and provides us varied effects.

Most music we daily listen to is referable to western music. In western music, music has to have the following three requirements, A) using sound as the component, B) combine sounds with sound property, C) construct sounds in **time variation**. That is to say, music is temporal art that is constructed in time variation whereas caving and painting are special arts.

1.2.1 Three modes of music

Generally there are three modes to enjoy music, "composing," "playing," and "listening." Musicians express their feelings by composing music, and the composed music is played and reproduced by players. And the musicians' feelings come across to listener with the played music.

To put it plainly, composing is "to make melody," which is not only for vocal but also for some instruments. Briefly, making plural melodies considering the harmony among the many instruments and combination of the harmonious plural melodies become a song, this is composing. To compose music, gifted musicians can compose brand new music based on their own senses with the knowledge of the musical theory. Rather, it can be thought that great music based on brilliant musicians' senses has been formalized and become the musical theory in a long time history of music. However, what music most people feel great has been not clear, assumedly these may be determined by human natural senses. I believe that the emotional/affective engineering can be of service to solve this ancient question.

As long as the composed music is scored, it can be played by anyone throughout the ages. With the changes over time, the musical instruments has changed and the tone color also has changed. So the music played in ancient may be different from the music now we can listen to. However, the popular classical music is now still loved and listened by so many people around the world. This means that the composition of music may be more important for human feelings than tone colors. Meanwhile, the music sometimes can be daringly played by the musical instruments that differs from the original, and some parts and accompaniments of the music are arranged without composer's intent. These are also ways of enjoying to play music, the differences among arrangements, players and musical instruments let us to find other appeal of the music and to receive varied impressions. Especially in Jazz music, most music is improvised and the players and even listener enjoy the differences among the playing.

The most accessible way to enjoy music is listening. In any way, almost all people have enjoyed listening to music. Although people could listen to music only live in ancient times, people became to be able to listen to music whenever they want because record medium was developed by the breakthrough of computer science. As shown in Figure 1.1, the record medium has changed with the changes over time: record, cassette tape, and compact disk, and most recently almost all people store music in their own hard disks and carry out the portable music player with the hard disks. As the record medium changes, the capacity of storage has increased and the acoustic quality has also improved. Then the listening style has also changed, people regardlessly store many and varied songs in the hard disk and listen to the chosen songs without playback for each album. Thus I believe that the song selection method complying the new styles of listening to music has been needed.

1.2.2 Application of Music on Human Life

In several ways, music is applied in human life. It is useful to play the musical instruments and to move/sway/rock to the music to improve human abilities [3]: exercise capacity, hearing sense and so on. As an illustration, it has known that rhythmical auditory stimulation enables higher cerebral dysfunctions to improve sensibility and induce them to achieve functional movements, and singing and playing the wind can be a rehabilitation for physical functions: breathing, cardiopulmonary function,

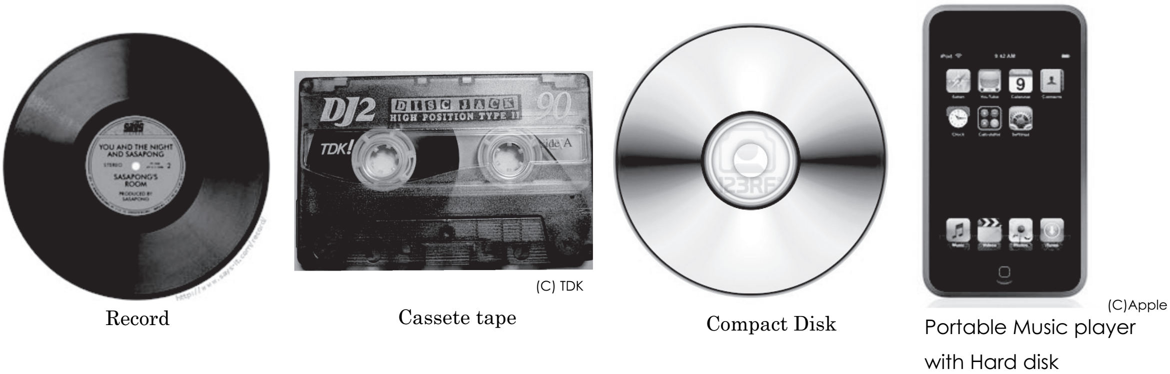


Figure 1.1: Record medium changes with changes over time

pronunciation and so on.

Music is applied also in psychological field, even in medical field [4, 5, 6]. It has been thought that the acoustic vibration and the harmonic overtone provide people with relaxation: alleviation of anxiousness and stress, hypnotic action and so on. And attentiveness and power of concentration are improved by listening to music. Moreover the music is used in shopping stores to enhance the costumers' buying inclination [7].

As these facts have shown, music fits in human life and we perceive the varied effects of music in whichever consciousness or unconsciousness.

1.3 Computer Science and Music

As we can see some electrical musical instruments, music has a lot of connection with computer science. Nowadays some computation technology (for example, a large capacity of storage and cloud computing) has been developed, and it becomes to seem that not only computer scientist and developer but also the public interested in music focuses on the computer technology.

For example, time-series data processing and frequency analysis are used in the acoustic analysis, and Musical Instrument Digital Interface (MIDI) [8] are used as the universal unified standard for the computer musical score. In the appendixa, the Fast Fourier Transform (FFT) that is continually used in this paper as the time-series processing and frequency analysis techniques and basic knowledge for MIDI will be explained.

1.4 Composition of this paper

The purpose of this study is to reveal the important factor of music

Music is composed by the sound variation with time passage, and the variation may influence effects on psychology and physiology when people listen to music. Thus the sound variation with time passage is focused as a key of this study. This study has the following three main contents.

In Chapter 2, this study will focus on bio-signals as physiology, and attempt to verify the importance of the sound variation through the analyze of brain waves while listening to harmony. Then, this study originally propose a presentation sound and analytical time zone to reveal the influence on EEGs of the combination of previous sound and following sound. Moreover, the relaxation effects of music are also studied.

In Chapter 3, this study focuses on the fluctuation properties as features that can explain the temporal variation of music/sound, and reveals the relationships between acoustic features and emotional/affective evaluation of music, and develops the song selection system complying with emotional/effective requests. And focusing on also the musical performance, emotional/affective evaluations of musical performance will be quantified with acoustic features, which is played with the violin. And the relationships between acoustic features and emotional/affective evaluation of musical performance will be studied.

And, in Chapter 4, musical score will be modeled for data mining and information processing, then the sound variation on musical score will be taken in consideration. This study will develop two types of the automated accompaniment composing system, and these system will be subjectively compared. As the mechanism of learning and expressing, Hidden Markov Model and Bayesian Mining of Score context will be each applied.

At last, this study will be concluded in Chapter 5.

Chapter 2

Music Effect on Human Factors

2.1 Introduction

It has reported that people in contemporary society have a lot of stress from various sources. About one million people around the world commit suicide every year. Also the number of people suffering from depression in Japan based on psychological evaluations is approximately 3.84 million people (3% of the population) according to the WHO. These facts lead us to conclude that mental disorders are spreading among people in contemporary society. In addition to their usual treatment, patients need relaxation to recover from their depression. Many services and products that help people to relax have become popular over the years, and they have been developed and offered as ways to overcome various forms of stress in contemporary society. Music has also been used in medical treatment [9, 10].

Since ancient times, people have spent a lot of time listening and playing music as a means to relax. Music is now media friendly and is easy to acquire. It can be inferred from that relaxation music has been an established musical genre, modern people seek to relax by listening to music. However many songs have this effect on people, but we do not have any clear idea of this; which elements and parts are important for these effects on psychology and physiology. Many songs that cause relaxation have the same harmony, and these harmonies have been frequently used in music since ancient times. Moreover, they say "The preceding sounds enable us to predict the ones that follow in music, and the following sounds enable us to confirm the preceding ones. The correlation between the preceding sounds and the following ones is the point of music" according to Gestalt cognitive psychology [11, 12].

This study focuses particular attention on harmony (i.e., variation of multiple sounds), and describes an original measurement and analysis method of brain waves when listening to harmony. Clarifying the difference of brain waves between "listening to single chord" and "listening to harmony," the potential of the variation of sounds on physiology are verified. And the correlation between relaxation and harmony are also the focus of this study.

2.2 Music and Significance of Harmony

2.2.1 Modern Music and The Musical Theories

The songs that people listen to daily are generally classified into genres in western music; when we hear the word "music," we are reminded of these classifications. Musical theories are general ideas having unified rules regarding songs, and these rules have provided a good estimation since the beginning of recorded history. In contemporary and in classical times, most songs that have been classified in western music have been composed using these theories. It is believed that the musical theories have some factors that affect people in various ways, because quite a few songs following these theories have been cherished for many years by music enthusiasts all over the world. Thus, this study considers that the musical elements in the musical theory influence human feelings: relax or excite.

2.2.2 The Main Three Elements of Modern Music

One musical theory states that music is made up of three main elements: melody, rhythm, and harmony.

Melody is an array of notes on a musical scale, and important factor for music as a type of $face$ for the music. However, it differs from song to song, and determining the common aspects among many songs is quite difficult, because there are no rules that must be obeyed with the exception of major and minor scales.

The songs tend to have repetitive elements time. This is called rhythm, and some specific aspects have rules. The rules of rhythm depend on the genre of music, the country, and the climate.

Harmony is the variation of multiple sounds which establishes a certain mood to a song. and has common rules used among many songs. Harmony is easier than melody to detect the common aspects among many songs, thus some standard rules apply. Also, some standard rules of harmony are often used in many songs, and they are not confined to one musical genre.

Harmony Harmony can be another word for cadence which is the sequence of chords. Because harmony does not depend on musical genre and common rules of harmony are often used in many songs, this study considers that the study of harmony, variation of multiple sounds, can obtain a general factor encapsulating the various effects that music has on people. We can also more easily determine the specific aspects of harmony than the specifics of other elements of modern music. This study may help to establish the factors in the effects of "relaxation music." This study focus on harmony for these reasons, and we herein present our experiment and the results of our analysis. Few papers have been written on harmony and relaxation, but some papers about chords and relaxation are in the literature [13].

2.2.3 Conversion Score for Expressions

Almost all songs we listen to daily are made by musical scale n^a. That can be written as Equation 2.1. Incidentally, a refers to a high level octave.

$$n^a \in \{A^a, A\#^a, B^a, C^a, C\#^a, D^a, D\#^a, E^a, F^a, F\#^a, G^a, G\#^a\} \tag{2.1}$$

Plural musical scales that are pronounced at the same time are called chords. Chord c_j can be written as Equation 2.2 with musical scale n^a.

$$c_j = \{n_1^a, n_2^a, n_3^a, \ldots, n_i^a \mid n_1^a \neq n_2^a \neq n_3^a, \ldots, \neq n_i^a\}. \tag{2.2}$$

The procession made up of chords is called harmony. Harmony h_k can be written as Equation 2.3 with chord c_j.

$$h_k = c_1, c_2, c_3, \ldots, c_j. \tag{2.3}$$

For example, Figure 2.1-(a) the musical scale can be written as C^0 with Equation 2.1. Also, Figure 2.1-(b) can be written as $c_1 = \{C^0, E^0, G^0\}$ with Equation 2.2. When it corresponds to the second chord in Figure 2.1-(c) as $c_2 = \{G^0, B^0, D^1\}$, this harmony can be written as $h_1 = c_1, c_2$ with Equation 2.3.

This study assumes the harmony made by two chords as the minimum unit of harmony, and hereafter this minimum unit is assumed as the *harmony*.

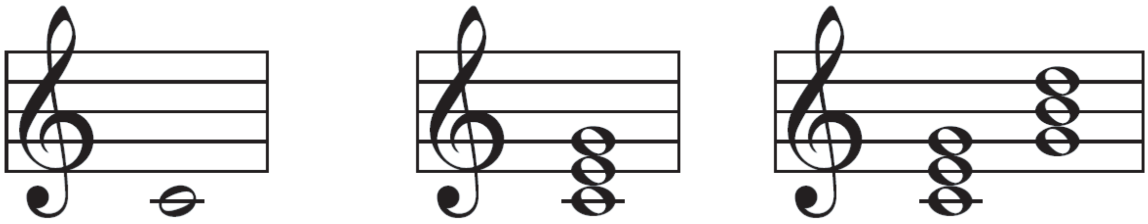


Figure 2.1: Expression on the musical score

2.3 Brain Waves and The Relaxation Index

This study especially focus on relaxation effects as the evaluation index for musical influence on physiology. The effects of relaxation in people can be thought of in psychological and physical terms. This study conducted an experiment and analyzed the effects on only the body through brain waves for showing relaxation, because we were interested in the involuntary effect on people. Thus this study evaluated brain waves, especially α wave bands. Many papers focused on α wave bands in brain waves to determine the effects of listening to music, sounds, or chords [14, 15, 16, 17, 18, 19]. This study also extracts α wave bands and β wave bands in brain waves that caused by listening to harmony, and analyzes brain waves especially focusing on α wave bands.

2.3.1 Brain waves

Brain waves can be considered as a time series signal, and can be written as $x_i(t)$, where i shows the electrode part. Then, frequency response $\tilde{x_i}(f)$ can be written as Equation 2.4 with a Fourier transform, where f shows frequency.

$$\tilde{x_i}(f) = \int_{-\infty}^{\infty} x_i(t)e^{-j\omega t}dt, \tag{2.4}$$

α waves bands are $(7 - 13Hz)$, and β waves bands are $(13 - 30Hz)$. Thus, power spectrum values of α and β waves can be calculated by Equation 2.5 and Equation 2.6, respectively.

$$W_\alpha(i) = \int_{7}^{13} |\tilde{x}_i(f)|^2 df, \tag{2.5}$$

$$W_\beta(i) = \int_{13}^{30} |\tilde{x}_i(f)|^2 df. \tag{2.6}$$

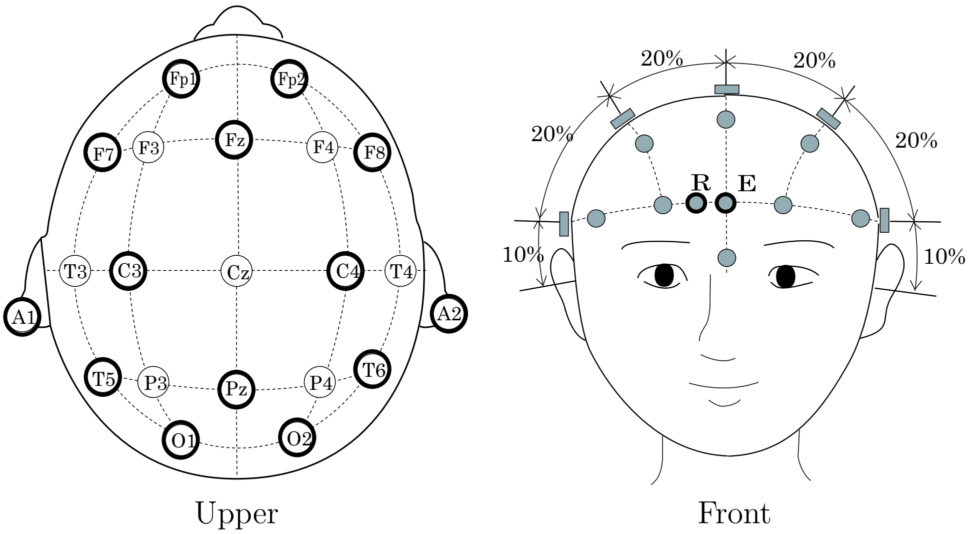


Figure 2.2: Electrode arrangement

2.3.2 Relaxation index

This study used the α waves content rate to β waves as the evaluation standard for relaxation, because the range of power spectrum values differed in the brain waves of each subject and this is used as index for activation level of brain waves (for example , in [20]). The α waves content rate to β waves is defined as $D_\alpha(i)$ and calculated by the following equation.

$$D_\alpha(i) = \frac{W_\alpha(i)}{W_\alpha(i) + W_\beta(i)}. \tag{2.7}$$

This study measured the average $D_\alpha(i)$ for all 12–channel electrodes (see Figure 2.2) through Electoro-Cap by Electro-Cap International, Inc. [21]. The average $D_\alpha(i)$ when subjects listen to the sound stimulation y is defined in the following equation.

$$Relax(y) = \frac{1}{12}\sum_{i=1}^{12} D_\alpha(i). \tag{2.8}$$

2.4 Problem in Measuring Brain Waves While Listening to Harmony and The Solution

The brain waves of subjects listening to chords were already measured and analyzed in related works. Figure 2.3 shows the time chart and the corresponding brainwaves

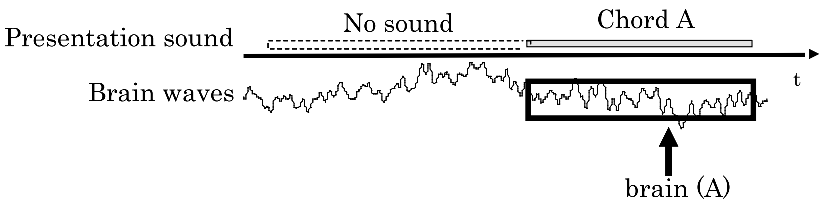


Figure 2.3: Measurement of brain waves while listening to a chord in related works

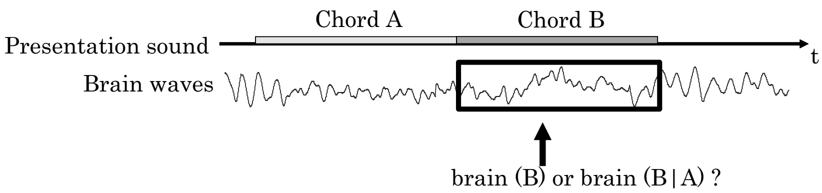


Figure 2.4: Problem on measuring brain waves while listening to harmony

in related works, where $brain(N)$ shows the effect on brain waves while listening to sound N.

It is indispensable to measure brain waves in subjects while they listen to harmony for obtaining the relationships between music and brain waves. However, when we only measure the brain waves while they listen to harmony, serious problems occur. With brain waves while they listen to the following chord (see Figure 2.4), we can not know either "the effect caused by listening to only the following chord ($brain(B)$)" or "that caused by listening to the both preceding and following chords ($brain(B|A)$)."

Therefore, this study proposes an original presentation sounds to solve this problem. The results of an experiment will allow us to determine the effects on brain waves caused by listening to only the following chord or the preceding and following chords when listening to two consecutive chords: influence of temporal variation of music on brain waves.

2.4.1 Presentation sounds

In this research, this study used nine kinds of chords for the $first$ chords (c_j); we selected these nine chords because they are commonly used in modern music [22]. We also used a chord that was fixed for the *second* chord (c_s), which is major chord.

Table 2.1: Chord sets that compose the standard consonance harmonies

Harmony	*Chord name*		Expression	
	First chord	*Second* chord	*First* chord	*Second* chord
h_1	G6		$\{G^0, B^1, D^1, E^1\}$	
h_2	Em		$\{E^0, G^0, B^0\}$	
h_3	Am		$\{A^0, C^1, E^1\}$	
h_4	C	G	$\{E^0, G^0, C^1\}$	$\{G^0, B^1, D^1\}$
h_5	C6		$\{E^0, G^0, A^0, C^1\}$	
h_6	D		$\{F\#^0, A^0, D^1\}$	
h_7	F#dim		$\{F\#^0, A^0, C^1, D\#^1\}$	
h_8	D7		$\{D^0, F\#^0, A^0, C^1\}$	
h_9	Bm		$\{D^0, F^0, B^0\}$	

Figure 2.5: Time chart of presentation sound

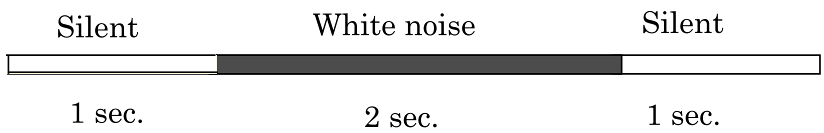


Figure 2.6: Time chart of white noise

Then the nine first chords were selected considering the law of harmony and frequent chords in popular music. That is to say, nine harmonies were made of one of the first chords and the second chord, and the harmonies differed. Table 2.1 shows the chord sets that compose harmonies. All sounds were made with MIDI, and all the tones for them were made on a piano.

Both the first chord and the second chord are presented for 2 sec. in a harmony; a harmony was 4 sec. in all. Chord progression in music is generally occurred at bar progression, and chord presentation for 2 sec. corresponds to a time of an bar in song whose beat per minutes is 120. And this study also prepared only a single chord as presentation sound, that was made of 2 sec. silent which corresponds to c_j and the second chord. Figure 2.5 shows the presentation of each harmony.

To repeatedly measure influences on brain waves while listening to each harmony, each harmony was presented for four times, for each participant. White noise (see

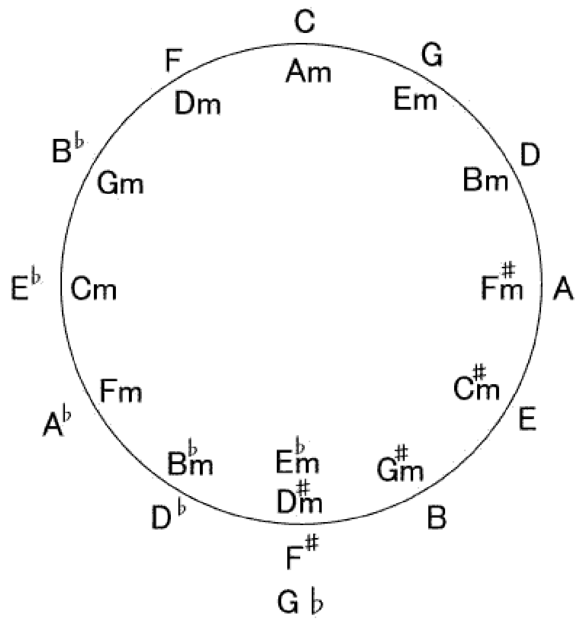


Figure 2.7: Circle of Fifth

Figure 2.6) were presented in interval between presentation sounds, the influence of listening to previous presentation sound seemed to be vanishingly small. Each participant listened to presentation sounds in random order, and the influence of presentation order of presentation sound also seemed to be vanishingly small.

Circle of Fifths

Circle of fifths, which is a musical terminology, has been known as one of the modulatory space, and can be shown as Figure 2.7. In the figure, the perfect fifths are arranged next to each other, this idea was devised by Pythagoras in ante-Christum period. Most music, whether consciously or not, adhere fundamentally to circle of fifths, and the temporal tonality is maintained.

This study prepared nine harmonies considering the circle of fifth as the standard presentation sounds. Considering musical theory, the prepared harmonies are often used in music.

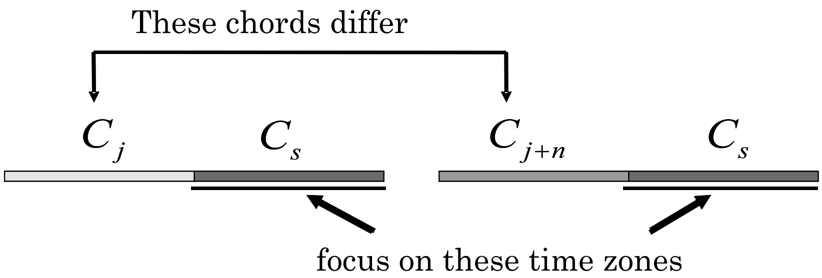


Figure 2.8: Proposed analytical objects

2.4.2 Analytical Object

For the experiment of measuring brain waves, many studies focused on when the brain wave fired such the P300 wave. The P300 wave means the brain wave that fires at about 300 [ms] after stimulus, has been known as an event related potential [23]. And it has known that the P300 is useful to detect the subject's focus and intend (i.e., yes/o, experienced/unexperienced), and applied in some brain-computer interface [24]. However, this study dare raises questions; the P300 wave seems to be inadequate as the index that shows activation on brain waves for contiguous stimulus such music. The audio-signals continuously changes in music, hence it may be not clear which time zone should be focused to analyze the brain waves while listening music.

As a first step of measuring true activation on brain waves while listening to music, this paper proposed the original presentation sounds, that is harmonies with a little ingenuity which have the nature of music: continuously changes of plural sounds, detailed in section 2.4.1. Figure 2.11 shows the proposed analytical object, brain waves while listening to the following chord is set as the analytical objects. The following chord is same and fixed in all harmonies, thus it is expected that the differences depending on the previous stimulus will be determined as measuring the brain waves while listening to the following chord.

2.4.3 Experimental environment

The experimental environment is shown in Figure 2.9. The low-pass and high-pass filter in the brain waves analysis system was adjusted to 13 Hz and 8 Hz, respectively. The output signal from a MIDI device was transferred to the amplifier through a fiber optical cable. The sound volume was fixed at a comfortable level for the participants.

The method of measuring brain waves conformed with the standards of the International 10-20 method. Brain waves were measured with an Electro-Cap, which has

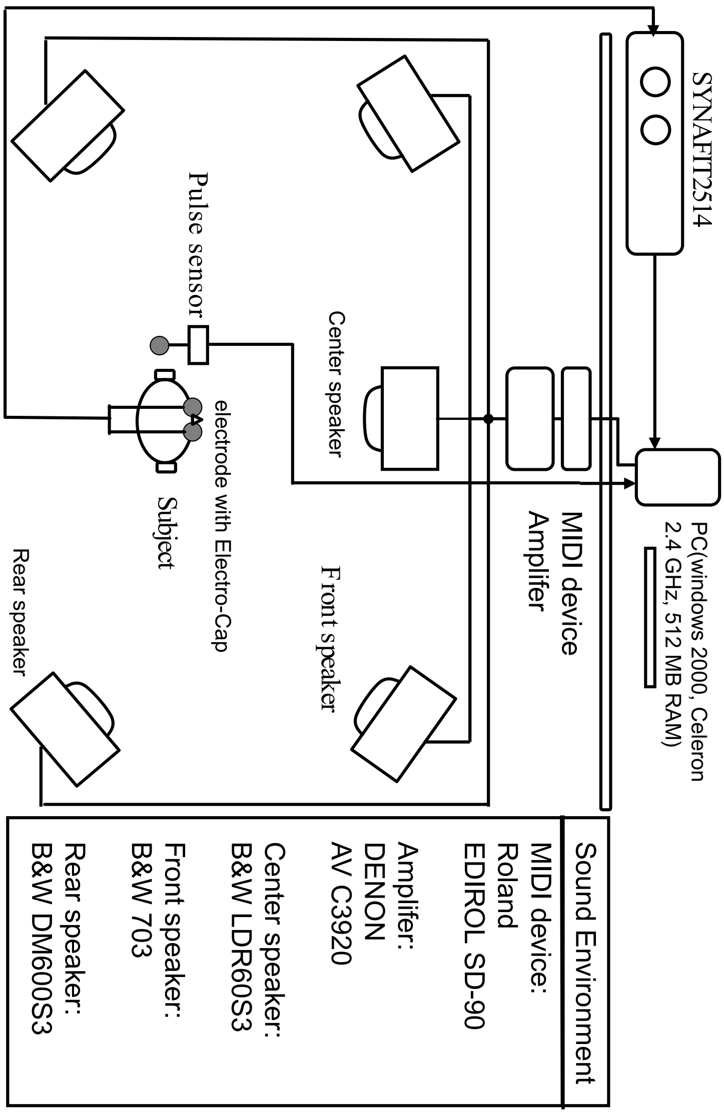


Figure 2.9: Experimental environment

twelve Electrodes (Fp1, Fp2, F7, F8, Fz, C3,C4, Pz, T5, T6, O1, O2): circles of bold-faced type except ears and R, E in Figure 2.2 were used for the measurements.

In brain waves analysis, sampling frequency, resolution time, and frequency resolution were 500 Hz, 1 sec., and 512Hz, respectively.

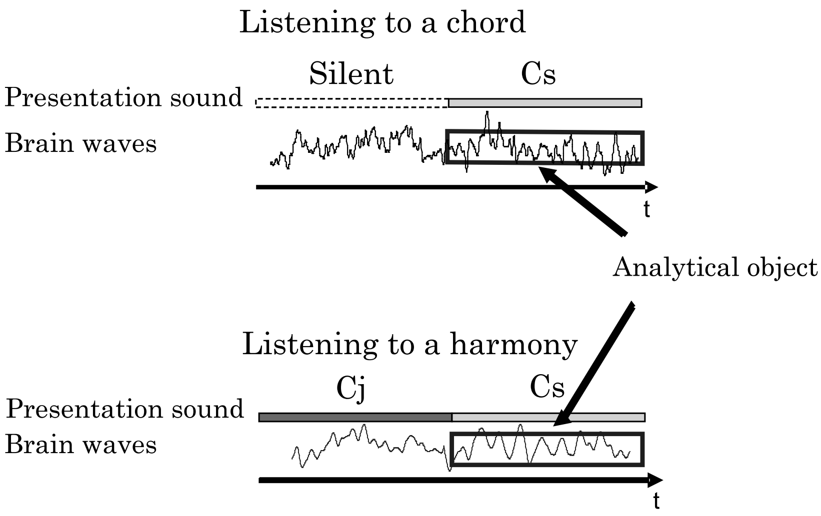


Figure 2.10: Analytical objects on experiment 1

2.5 Influence of Temporal Variation of Music

Eight healthy men and women in their twenties participated. In the experiment, the participants laid quietly on a chair for 90 seconds with their eyes closed and listened to presentation sounds, and then their brain waves were measured. In this study, brain waves were measured in two types of procedure.

First, in experiment 1, this study verify the difference of brain waves between listening to only a single chord and listening to the same chord after other given chord (in fact, listening to the chord as the second chord in harmony) using chi-square test. And a source of brain waves while listening to the second chord in harmony is examined.

And in experiment 2, the differences of relaxation index when listening to each harmony are verified using Kruskal-Wallis test, and the differences of influence on brain waves with respect to each listening harmony is examined.

2.5.1 Experiment 1: a source of brain waves while listening to the second chord in harmony

This study should clarify that brain waves while listening to the second chord in harmony is influenced by whether only the second chord or the combination of the second chord and the precious chord. Thus this study examine a source of brain waves while listening to the second chord by analysing the analytical object which is shown in section 2.5.1.

Table 2.2: Result of Chi-square test

Group A	Group B
h_3, h_4, h_5, h_7, h_8, h_9	h_1, h_2, h_6
show significant differences	no significant differences

Analytical object on Experiment 1

Figure 2.10 shows analytical objects on experiment 1. This study focuses on average relaxation index while listening to the second chord c_s in harmony $Relax(c_s)$ and average relaxation index while listening to c_s as just a single chord, and used them as analytical objects. In both case, presented c_s are same chord. If relaxation index while listening to the second chord in harmony is influenced by only c_s, the activation common to listening to c_s as just a single chord should be observed. And if the brain waves activation differs between listening to a chord as component of harmony and listening to the chord as just a single chord, it is suggested that the first chord also influences brain waves while listening to the second chord in harmony.

Result of experiment 1

Using relaxation index while listening to just a single chord on analytical object as a control, Chi-square test (significance level = 5%) was conducted to verify the difference of the distribution of each harmony. The result of the test is shown in Table 2.2. Group A shows harmonies that have significant difference between relaxation index while listening to a given chord as a component of harmony and listening to the chord as a single chord. Group B shows harmonies that have no significant difference.

For many harmonies, it was confirmed that relaxation index while listening to a single chord was different from one while listening to the chord as a component of harmony. From this, it was suggested that relaxation index on analytical object (i.e., while listening to the second chord in harmony) was not $Relax(c_s)$ but $Relax(c_s|c_j|)$.

2.5.2 Experiment 2: relaxation effects caused by listening to harmony

In this experiment, relaxation effects caused by listening to harmony are studied with a focus on analytical object shown in section 2.5.1.

From the result of the experiment 1, it was suggested that brain waves while listening

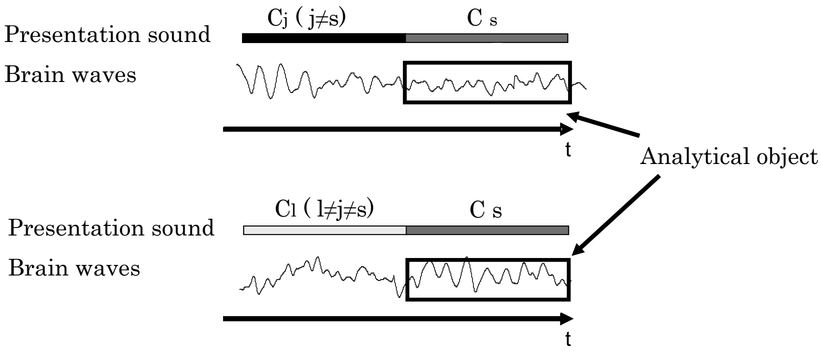


Figure 2.11: Analytical objects on experiment 2

to the second chord in harmony were influenced by listening to the first chord. So, relaxation index on analytical object shown in section 2.5.1 can be redefined as the following equation.

$$Relax(h_k) = Relax(c_s|c_j), \tag{2.9}$$

hereafter, this study assumes $Relax(h_k)$ as the relaxation index caused by listening to *harmony*.

Figure 2.11 shows analytical objects on experiment 2. Here, this experiment focuses only the chords that are included in cadence. This study measures differences on $Relax(h_k)$ due to kind of harmony; the first chord that participants previously listened influences relaxation index on analytical object or not.

Result of experiment 2

Average relaxation index for each harmony of all participants, $\overline{Relax(h_k)}$ is defined as the following equation.

$$\overline{Relax(h_k)} = \frac{1}{p}\sum_{p=1}^{P} Relax_p(h_k), \tag{2.10}$$

where, P shows number of the participants; in this study $P = 8$. Figure 2.12 shows $\overline{Relax(h_k)}$ of each harmony, and error bars in the figure show standard errors (the number of samples is 32 for each harmony). Nevertheless the chords on the analytical object were same among all harmonies, the high-low difference was confirmed. And referring the circle of fifth (see Fig. 2.7), it was suggested that the harmonies whose the first chord closes to the second chord G on the circle of fifth tend to show the relatively high relaxation index.

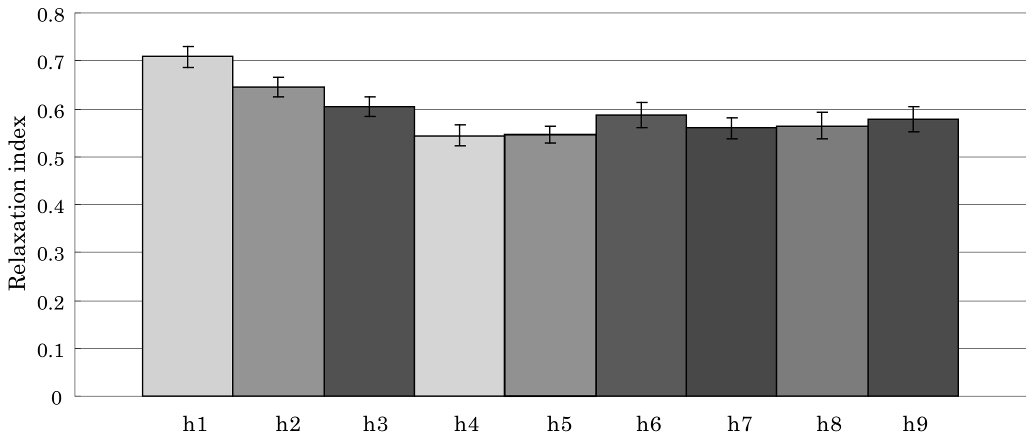


Figure 2.12: Relaxation index on analytical object of each harmony

To statistically analyze this difference, Kruskal-Wallis test was conducted. Table 2.3 shows the test result. The significant difference could be not confirmed among most harmonies, however h_1 shows significantly higher relaxation effects than every harmonies except h_2. So it was suggested that listening to h_1 gave relatively high relaxation effects to human.

From the result, it was confirmed that kinds of listening harmony influeced hi-low difference on relaxation index.

2.6 Conclusions

This study proposed original presentation sounds and analytical time zone on brain waves, and verified the relation between listening to harmony and relaxation effect through the experiments.

In experiment 1, the source of brain waves while listening to the second chord in harmony was examined. As the result, it was suggested that not only the second chord in a harmony but also the first chord in the harmony influences brain waves while listening to the second chord in the harmony. In the experiment and analysis, the result provided support for the Gestalt cognitive concept about music, "The preceding sounds enable us to predict the ones that follow in music, and the following sounds enable us to confirm the preceding ones. The correlation between the preceding sounds and the following ones is the point of music."

And in experiment 2 based on the result of experiment 1, this study verified whether the difference of harmony, that is the difference of the first chord in harmony, influenced

Table 2.3: Result of Kruskal-Wallis test

	h_1	h_2	h_3	h_4	h_5	h_6	h_7	h_8	h_9
h_1			$*$	$*$	$**$	$**$	$*$	$**$	$**$
h_2					$*$	$*$			
h_3									
h_4									
h_5									
h_6									
h_7									
h_8									
h_9									

$**$ 1% significance
$*$ 5% significance

relaxation index or not. Analyzing relaxation index while listening to the second chord, the relaxation index while listening to the second chord in each harmony differed each other nevertheless the second harmonies are common and fixed among all harmonies. Moreover, it could be confirmed that harmony such h_1 gives higher relaxation effect. This results suggested that there were some harmony that gives higher relaxation, and the relaxation effect by listening to music seemed to be caused by harmony.

In future of this work, it will analyze in detail the relationship between the second chord (c_s) that was fixed to one in each harmony and the first chord (c_j) that was different in all of the *harmonies*, and propose the solvable method for this problems. Also, it will test our discoveries on the relationship between harmonies and the effects of high relaxation found in our analyses.

In this chapter, it was confirmed that the temporal variation of music influences not only psychology but also physiology. In the following chapters, the temporal variation of music is especially focused and introduced in analysis and system.

Chapter 3

Acoustic Features and Emotional Evaluation

3.1 Introduction

This chapter proposes three studies: relationships between acoustic fluctuation features and emotional evaluation of music as "Acoustic–Emotion" model in section 3.2, song selection system complying with emotional requests which is based on Acoustic–Emotion model in section 3.3, and quantification of musical performance impressions based on acoustic analysis for violin playing in section 3.4.

When people listen to music they feel some affection/emotion from the song, in the other words, music has variety of impressions and provides uplifting feelings or sentimental feelings with people [11]. Focusing on this knowledge, many and several studies about music and human emotion have been reported: the musical genre estimation based on human emotion [25], and the correlations between acoustic features and emotional evaluation [26, 27, 28, 29].

This study focuses on the temporal variation of music based on the Gestalt psychology and the effect on physiology caused by harmony described in Chapter 2. In section 3.2, this paper reveals the relationships between musical fluctuation features and emotional evaluation of music. Musical fluctuation features can cover temporal variation of musical rhythm, pitch and volume. Moreover in section 3.3, the song selection system is constructed using the relationships revealed in section 3.2 as the "Acoustic–Emotion" model. This study proposes two types of song selection algorithm for the proposed song selection system considering characteristics of human instinct. Through song select demonstration and subjective evaluation experiment, the availability and usability of the proposed system is verified.

And it has known that slight differences in musical performances influence human emotional evaluation. In section 3.4, this study quantifies the musical performance impressions based on acoustic analysis focusing on the acoustic features that cover the temporal variation and harmonics.

3.2 Relationships Between Acoustic Features And Emotional Evaluation of Music

Music has recently attracted attention as the one of the media that enriches human cultural lives. It has recently become reasonable to suppose that music is regarded as one of the most important factor in human lives considering the widespread use of portable music players. The development of the data processing technology has enabled us to have and carry the portable music player that has the large capacity of database.

We select the song we would like to listen to using bibliographic data, for example, the name of a song, an artist, an album, a genre and so on. However as capacity of the database increases, it is difficult to select the song with the bibliographic data. Thus, some music retrieval techniques based on human instinct have been studied, for example, music retrieval by hamming [30, 31], and music classification by audio signals [32]. Additionally, we have known that most people selects and listens to music with respect to their feelings at the given time, and also the mood of the music influences their feelings [33, 34] and behaviors [7]. So, the song selection method based on the given instincts enables us to not only select the song from a large-capacity database based on our feelings at the time, but also enjoy the psychological and biological effectiveness.

To develop the song selection system based on the instincts, we need to know "what emotional evaluation song has in the database" as the first step. Then, this paper studies the relationships between the emotional evaluation of music and the acoustic fluctuation features. This study extracts the acoustic fluctuation features that can cover the musical features by conducting an acoustic analysis, which will be described in section 3.2.1 . On the oter hand, this study obtains the results from an emotional evaluation of music for several adjectives by conducting subjective evaluation experiment, which will be described in section 3.2.2. Then thhe relationships between the extracted acoustic features and the emotional evaluations of music can be resulted through a multiple classification analysis, which is described in section 3.2.3, and the acoustic features that influence the emotional evaluation can be determined in section 3.2.4.

3.2.1 Fluctuation features

Some existing studies have tried to relate the audio signals with the emotional evaluation of the music [28, 29, 35, 36], however these studies do not cover the acoustic features that can explain the temporal variation of music. The primary consideration in music should be the temporal variation of sounds. The temporal concept of music has been regarded as the one of the more important factors for music in the field of the cognitive psychology [11], and it was also inferred from the suggestion in Chapter 2. Thus, we use the **fluctuation features** that can explain the temporal variation of music as the acoustic features of music.

Melody, rhythm, and harmony are the three major factors for composing music [37]. They are signed and sealed as the time structures of both the pitch and volume, and are different among songs. We have believed that the emotional evaluations of songs can be influenced by their differences. Melody and harmony are signed and sealed by the temporal regularity of both the pitch and volume. And, rhythm is signed and sealed by the temporal structure of the pronunciations. In this paper, we focused on and prepared the fluctuation features about volume, pitch, and rhythm.

Feature extraction

In this study, we used songs stored as WAV format data in order to extract the acoustic features that people can precisely perceive. An acoustic analysis of WAV formatted song data enables us to visualize the structures about both the time and frequency of music. All the song data used in an analysis are taken from Compact Discs and stored as stereo data, the sampling frequency is 44100 Hz, and they were linear quantized to 16 bits in order to standardize the acoustic quality. Using a short-time Fourier transform (Hanning window, length of window 2048 points, resolution time is 25 ms), frequency spectrum on each time, $freq(\omega, t)$ are calculated. Then ω and t each shows the frequency and the given time, respectively.

Volume features We believe that not only the acoustic features of music for an entire frequency band, but also the ones on several frequency bands influence the emotional evaluations. So, we prepared five kinds of frequency bands: Low, Middle, High, Ultrahigh, and All. The acoustic features of music were extracted from these five frequency bands. In this study, the Low frequency band, Middle frequency band, High

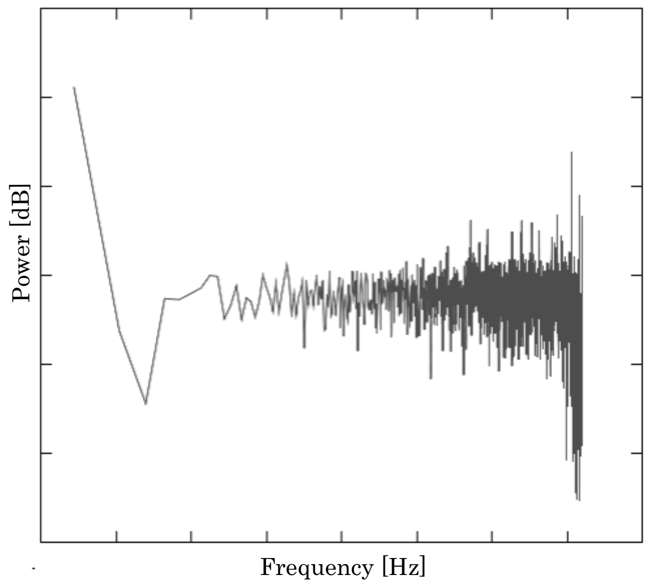


Figure 3.1: An example of the volume fluctuation spectrum

frequency band, Ultrahigh frequency band, and All frequency band are each defined as a band lower than 200 Hz, 200 - 800 Hz, 800 - 2000 Hz, more than 2000 Hz, and a band that is as long as we can calculate, respectively.

For each time t, the value of the integral of each power spectrum is calculated by following equation, with respect to each frequency band.

$$S_\omega(t) = \int^{\omega} freq(\omega, t) d\omega. \tag{3.1}$$

The value of the integral is regarded as the volume on each frequency band. The volume features are generated by regarding the volume at the given each time as the time-series data with respect to each frequency band. Then, the fluctuation spectrum concerning the volume on each frequency band can be generated from the volume feature on each frequency band by using the Fast Fourier Transform.

Pitch features We prepared following five acoustic features that are calculated from the frequency spectrum on each given time.

1. Centroid frequency $f_{centroid}(t)$:

$$f_{centroid}(t) = \frac{\int \omega \cdot freq(\omega, t) d\omega}{S_{all}(t)}. \tag{3.2}$$

2. The ratio of the low frequency band $ratio_{low}(t)$:

$$ratio_{low}(t) = \frac{S_{low}(t)}{S_{all}(t)}. \tag{3.3}$$

3. The slope of the regression line $\alpha(t)$:
   The slope of the regression line for $freq(\omega, t)$ is calculated by using the least-squares method,

$$\alpha(t) = \frac{n \sum^{n} \omega \cdot freq(\omega, t) - \sum^{n} \omega \sum^{n} freq(\omega, t)}{n \sum^{n} \omega^2 - \left(\sum^{n} \omega \right)^2}. \tag{3.4}$$

4. The intercept of the regression line $\beta(t)$:
   The intercept of the regression line for $freq(\omega, t)$ is calculated by using the least-squares method,

$$\beta(t) = \frac{\sum^{n} \omega^2 \sum^{n} freq(\omega, t) - \sum^{n} \omega \cdot freq(\omega, t) \sum^{n} \omega}{n \sum^{n} \omega^2 - \left(\sum^{n} \omega \right)^2}. \tag{3.5}$$

5. The peak frequency of the spectrum $f_{max}(t)$:

$$f_{max}(t) = \arg \max_{\omega > 0.1} \left(freq\left(\omega, t\right) \right). \tag{3.6}$$

Temporal sequence of the each acoustic features about pitch are regarded as the pitch features. Then, the each fluctuation spectrum concerning the pitch is generated from the each pitch feature by using the Fast Fourier Transform, respectively.

Rhythm features In this study, the beat spectrum [38] is regarded as the fluctuation about rhythm. The cosine similarity $D_C(i, j)$ between vector v_i and v_j is dynami-

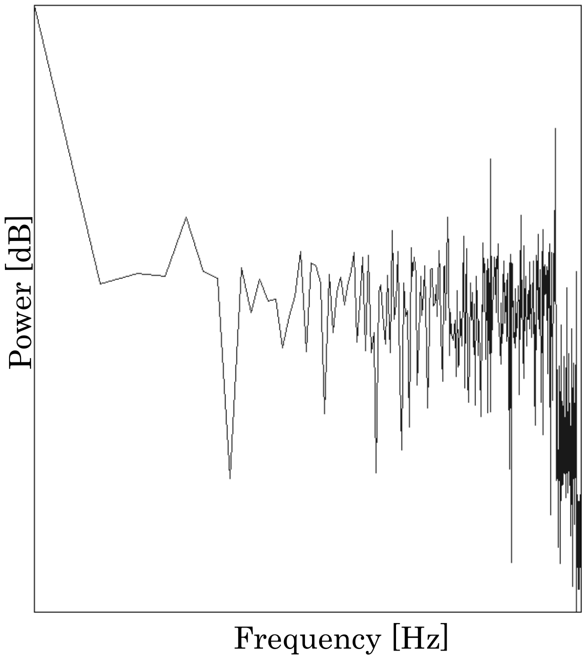


Figure 3.2: An example of the pitch fluctuation spectrum

cally calculated by the following equation, and then the similarity table D_C (reference Figure 3.3) is generated. Then, v_i and v_j each shows $freq(\omega, \boldsymbol{i})$ and $freq(\omega, \boldsymbol{j})$, respectively.

$$D_C(i, j) = \frac{v_i \cdot v_j}{|v_i||v_j|}. \tag{3.7}$$

The fluctuation about rhythm, that is the beat spectrum, on the given time t, $B(t)$ is generated by calculating cumulative similarity on the given time t for each time k, that is shown in the following equation.

$$B(t) = \sum_k D_C(k, k + t), \tag{3.8}$$

Figure 3.4 shows an example of the fluctuation about rhythm. This example was generated from the 120 Beat Per Minutes tempo song which contain drums patterns. From the figure, we can confirmed that the peaks of the spectrum are happened by each 0.5 seconds, and the beat spectrum seem to be appropriate as the fluctuation of rhythm.

The rhythm fluctuation spectrums can be generated from the rhythm features by

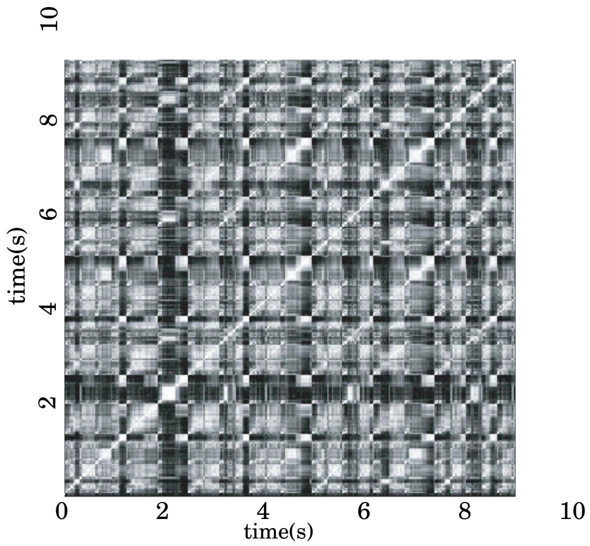


Figure 3.3: An example of the similarity table D_C

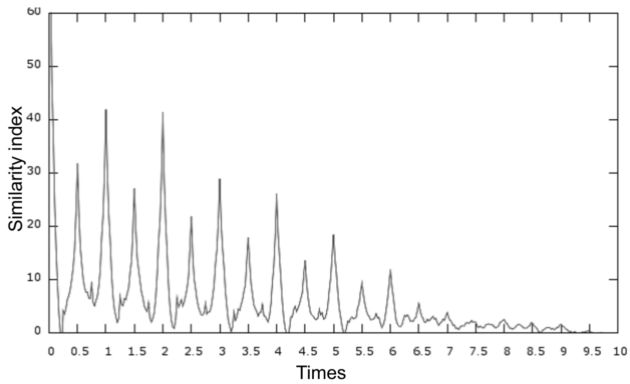


Figure 3.4: An example of the fluctuation features about rhythm

using the Fast Fourier Transform.

Fluctuation spectrum features

Fluctuation spectrum features are calculated from each generated fluctuation spectrum. In this paper, we prepared the following features to explain the structure of the each fluctuation spectrum.

(1) Value of the integral S

(2) Centroid frequency f_c

(3) Max value of the fluctuation spectrum $fluc^{max}$

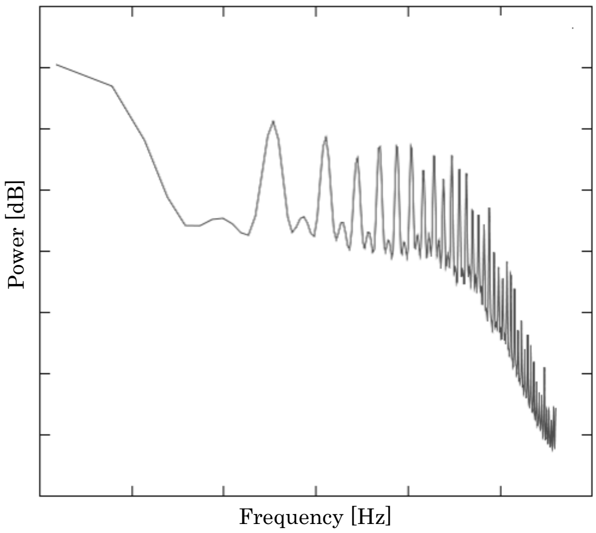


Figure 3.5: An example of the rhythm fluctuation spectrum

(4) Max frequency of the spectrum f_{max}

(5) The slop of the regression line

(6) The intercept of the regression line

Take an example of the pitch fluctuation, the fluctuation spectrum features are shown in Figure 3.6.

This study used the sixty six fluctuation spectrum features listed in Table 3.1 as features of a song. The features of songs are each normalized, and then the average property values and standard deviations of all songs reach each zero and one, respectively.

3.2.2 Emotional Evaluation

It has been proposed that the emotional evaluation is important for considering the significance and effectiveness of music, and has been focused and studied in some related works [7, 11, 39]. This study obtained the emotional evaluations of music by conducting the subjective evaluation experiments based on the Semantic Differential method whose criteria were varied [40].

This study prepared one hundred and fifty songs that were arbitrarily selected from several genres, e.g., Pop, Classical, Jazz, Rock, House and so on. And these prepared songs were selected throughout the generation and nations, thus it is believed that these

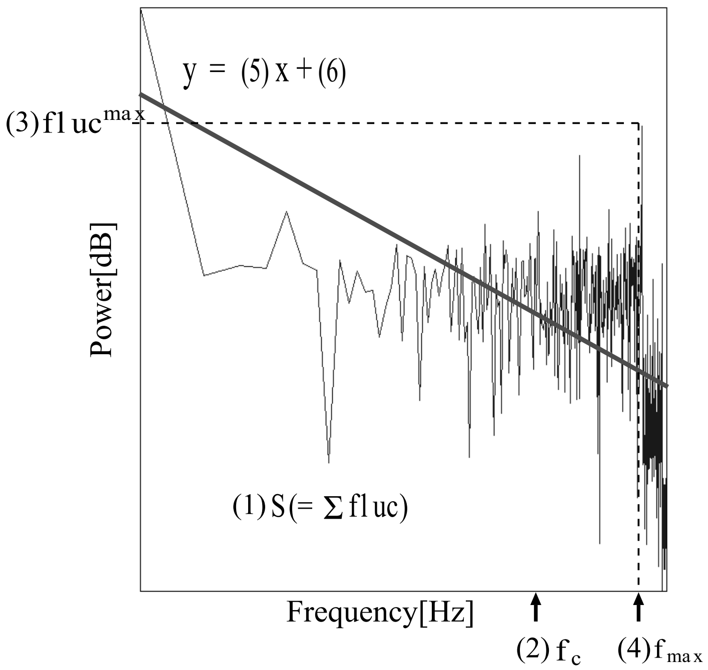


Figure 3.6: An example of the pitch fluctuation spectrum

songs are sufficient as a learning database for "Acoustic - Emotion" model. Almost all of the songs provide people with different emotional impressions throughout a song. Then this study used an arbitrarily-prescribed phrase that was extracted from a song, and it was about ten seconds long and used as the object to be subjectively evaluated. Hereafter evaluation of this one phrase are used as the emotional evaluation of a song.

This study prepared eight pairs of adjectives for evaluating the songs. Table 3.2 lists the eight pairs of adjectives. All the prepared songs were evaluated on a scale of one to seven for the eight adjectives based on the Semantic Differential Method. Eighteen healthy Japanese males and females (fifteen males, and three females) in their twenties participated in our evaluation. The subjects were asked to listen to and evaluate all the songs in random order through a prepared interface on the given web browser. With this interface, the songs were played by using the Windows Media Player, and the subjects could rewind and play back the songs over and over again.

3.2.3 Acoustic - Emotion model

With the fluctuation spectrum features of the one hundred and fifty songs used in the subjective evaluation experiment (detailed in section 3.2.1) and the emotional evaluations for the songs, the estimate spaces for emotional evaluations are constructed

Table 3.1: Acoustic features

Feature index	Fluctuation spectrum features
$\boxed{1}$ - $\boxed{6}$	Features (1) - (6) about volume on the frequency band All
$\boxed{7}$ - $\boxed{12}$	Features (1) - (6) about volume on the frequency band Low
$\boxed{13}$ - $\boxed{18}$	Features (1) - (6) about volume on the frequency band Middle
$\boxed{19}$ - $\boxed{24}$	Features (1) - (6) about volume on the frequency band High
$\boxed{25}$ - $\boxed{30}$	Features (1) - (6) about volume on the frequency band Ultrahigh
$\boxed{31}$ - $\boxed{36}$	Features (1) - (6) about pitch concerning $f_{centroid}(t)$
$\boxed{37}$ - $\boxed{42}$	Features (1) - (6) about pitch concerning $ratio_{low}(t)$
$\boxed{43}$ - $\boxed{48}$	Features (1) - (6) about pitch concerning $\alpha(t)$
$\boxed{49}$ - $\boxed{54}$	Features (1) - (6) about pitch concerning $\beta(t)$
$\boxed{55}$ - $\boxed{60}$	Features (1) - (6) about pitch concerning $f_{max}(t)$
$\boxed{61}$ - $\boxed{66}$	Features (1) - (6) about rhythm (beat spectrum)

Table 3.2: Adjectives used in the subjective evaluation experiments

Adjective index	Evaluation on a one-to-seven scale
1	Bright - Dark
2	Mellifluence - Harsh
3	Rich - Poor
4	Clear - Murky
5	Agitato - Peaceful
6	Heavy - Light
7	Soft - Hard
8	Thick - Thin

and used as a "Acoustic - Emotion" model. Table 3.3 shows the number of the songs on each degree for each adjective, and these songs were used for analysis. There is no song on +3 for adjective 8 (i.e., highly thin), this study conducted 6 classes discriminant analysis for this adjective.

This study compares the following four discriminant methods.

- Canonical Discriminant Analysis (CDA)
  CDA is a learning model that maximizes separation metrics and estimate the objective variable under the hypothesis that the objective variable can be linearly estimated by the learning data.

- k-Nearest Neighbor algorithm (k-NN)
  k-NN estimates objective variable by majority voting. First, this method represents class distribution using all learning data And for the given test data, it

Table 3.3: Number of songs on each degree for each adjective

Adjective index	On a one to seven scale						
	-3	-2	-1	0	+1	+2	+3
1	7	41	29	1	32	17	4
2	5	26	39	4	21	22	5
3	2	22	29	11	13	4	1
4	9	26	44	2	27	12	5
5	9	25	40	3	22	21	13
6	5	6	20	5	28	22	5
7	9	18	29	5	29	14	4
8	4	14	41	14	16	2	0

Table 3.4: Selected Musical Fluctuation Features for Each Adjective

Adjective index	Selected Musical fluctuation features
1	x_5,x_{15},x_{27},x_{36},x_{54}
2	x_6,x_{38},x_{39},x_{51},x_{53},x_{66}
3	x_{20},x_{34},x_{42},x_{63},x_{66}
4	x_{11},x_{32},x_{35},x_{42},x_{56}
5	x_2,x_3,x_{12},x_{13},x_{14},x_{15},x_{16},x_{19},x_{51}
6	x_{11},x_{47},x_{56},x_{60}
7	x_{23},x_{42},x_{59}
8	x_4,x_{10},x_{15},x_{16},x_{33},x_{51},x_{56},x_{60}

estimates the objective variable based on majority voting of k-Nearest Neighbors around the test data. This study set optimal k by leave-one-out cross validation.

- Support Vector Machine (SVM)
  SVM is a nonlinear pattern recognition method. It classifies a lot of data expressed as vector for multi classes by maximizing the margin based on Kernel function.

- ADA-BOOST
  ADA-BOOST is one of machine learning meta algorithm for supervised learning [41], and achieves 2 classes discriminant This study improves it to cover multi-classes discriminant applying ECOC method [42].

Table 3.5: Accuracy rate (%) by each discriminant method

Adjective index	CDA	k-NN	SVM	ADA-BOOST
1	52	44	42	45
2	48	29	35	33
3	52	37	44	49
4	53	42	34	42
5	56	37	26	41
6	47	36	8	36
7	44	32	44	36
8	58	45	47	37

Table 3.6: Absolute average error by each discriminant method

Adjective index	CDA	k-NN	SVM	ADA-BOOST
1	$\pm$ 0.94	$\pm$ 1.08	$\pm$ 1.16	$\pm$ 1.08
2	$\pm$ 0.94	$\pm$ 1.34	$\pm$ 1.15	$\pm$ 1.25
3	$\pm$ 0.80	$\pm$ 1.02	$\pm$ 0.94	$\pm$ 0.89
4	$\pm$ 0.97	$\pm$ 1.20	$\pm$ 1.30	$\pm$ 1.22
5	$\pm$ 0.74	$\pm$ 1.18	$\pm$ 1.11	$\pm$ 1.05
6	$\pm$ 1.02	$\pm$ 1.16	$\pm$ 1.43	$\pm$ 1.18
7	$\pm$ 1.08	$\pm$ 1.29	$\pm$ 1.06	$\pm$ 1.30
8	$\pm$ 0.59	$\pm$ 0.81	$\pm$ 0.74	$\pm$ 1.01

Future selection

The contributory features for estimating degrees of the each adjective were selected from the prepared sixty six fluctuation spectrum features listed in Table 3.1 using a step wise method, and the Fisher's linear discriminant function was used as the evaluation function [43]. Table 3.4 lists the selected fluctuation spectrum features for estimating degrees of the each adjective.

Analysis results by each discriminant method

With sample songs listed in Table 3.3, each discriminant analysis was conducted, where the explanatory and dependent variables were the selected contribute musical fluctuation features and the degrees of each adjective, respectively. For generalization, leave-one-out cross validation was used in each discriminant analysis. Table 3.5 and Table 3.6 each shows the average accuracy rate and the average error of each discriminant method, respectively. Here, the average error means the absolute diffrence between the correct degree of the subjective evaluation and the degree estimated by the each discriminant analysis on each adjective.

From Table 3.5 and 3.6, CDA shows the best performance on both the average accuracy rate and the average error. For adjective 1, 2, 4, 5, and 8, CDA achieves upper than 50% accuracy rate, it has more than tripled accuracy rate to theoretical one. And for all adjectives, the average error by CDA are less than 1.1, and it is suggested that most songs are estimated in ± 1 classes to the actual class even though misclassification occurs. The reason why CDA showed the best performance is suggested that CDA restructures the canonical discriminant spaces with the explanatory variables, that are appropriate for discriminant, while other methods (k-NN, SVM, and ADA-BOOST) do not restructure the estimation spaces.

3.2.4 Relationships between evaluation and acoustic features

This study focuses and considers the result by CDA in more detail below, which shows the best estimation performance. CDA generates canonical variates with the selected contribute features with respect to each adjective. And then, the estimate space for each adjective is generated by the each corresponding canonical variates, respectively. Table 3.7 lists the selected features and the coefficients of each canonical variate (CV) which discriminant was conducted using all samples as learning data, and shows that the first and second canonical variates contribute the estimation. Figure 3.7 and Figure 3.8 show point diagrams of songs for each adjective expressed by the first and second canonical variates.

The emotional evaluations of songs for these adjectives are discriminated well by CV 1 and CV 2, which are composed of prepared fluctuation features. It seems reasonable to suppose that the coefficients of CV for each fluctuation features can be regarded as the influence rate of each fluctuation features for each CV. This study would like to focus our attention on the adjectives whose indices are one, two, and five detailed in below.

Adjective 1: Bright–Dark Figure 3.7 shows that Bright songs and Dark songs are each in the minus area and the plus area of CV 1, respectively. The coefficient of x_{54} for CV 1 that represents the intercept of the regression line for the fluctuation spectrum about pitch concerning $\beta(t)$ is relatively big in positive. And it was suggested that the fluctuation features about pitch highly influences the emotional evaluation, Bright–Dark.

Adjective 2: Mellifluence–Harsh Figure 3.7 shows that Mellifluence and Harsh songs are each in the minus and plus areas of CV 1, respectively. The coefficients of x_6 for CV 1 that represents the intercept of the regression line for the fluctuation spectrum about volume on the all frequency band is relatively big in positive, and it seems to influence the emotional evaluation Harsh. And the coefficients of x_{51} for CV 1 that represents the max value of the fluctuation spectrum about pitch concerning $\beta(t)$ is relatively big in negative, and it seems to influence the emotional evaluation Mellifluence.

Adjective 5: Agitato–Peaceful Figure 3.8 shows that Agitato and Peaceful songs are each in the minus and the plus area of CV 1, respectively. In addition, each evaluation is distributed in a continuous way along CV 1. The coefficients of x_3 for CV 1 that represents the max value of the fluctuation spectrum about volume on the all frequency band is relatively big in positive, and it seems to influence the emotional evaluation Agitato. And the coefficients of x_{15} for CV 1 that represents the intercept of the regression line for the fluctuation spectrum about volume on the middle frequency band is relatively big in negative, it seems to influence the emotional evaluation Peaceful.

Regarding the adjectives 3 and 8, the classes of evaluation are mixed in each discriminant space (see Figure 3.7 and Figure 3.8) even though accuracy rate are sufficient. It was suggested that this result was caused by the facts that the cumulative contribution ratio by CV 2 for adjectives 3 and 8 are relatively low.

3.2.5 Discussion

In this section, the acoustic fluctuation features were extracted from music and emotional evaluation of the music were obtained through subjective evaluation experiment. Then the relationships between the acoustic fluctuation features and the emotional evaluation of music were analyzed by conducting a multiple discriminant analysis. As a result, it was suggested that the acoustic fluctuation features influenced the results from an emotional evaluation of music, and CDA could obtain high average accuracy rate and low average error. Moreover, it was confirmed that the acoustic fluctuation features that influence the discrimination for the degrees of adjectives: Bright–Dark, Mellifluence–Harsh, and Agitato–Peaceful.

This study focused our attention on and took into consideration the adjectives that are easily discriminated by using only CV1 and CV2. However, the discriminant of some of the adjectives might depend on not only CV1 and CV2, but also more CVs (see Table 3.7). So taking more CVs into consideration by using some more techniques that can be used to describe multiple dimensions such as the Varimax rotation [44], other findings about relationships between acoustic features and emotional evaluation of music might be provided.

Table 3.7: Selected features and the coefficients of each canonical variate

Adjective index	Future index	CV1	CV2	CV3	CV4	CV5	CV6
1	x_5	0.20	0.62	0.64	-0.44	0.39	
	x_{15}	-0.26	0.61	0.45	1.29	-0.03	
	x_{27}	-0.50	-1.03	0.46	-1.37	-0.63	
	x_{36}	-0.36	1.04	-0.74	0.01	0.05	
	x_{54}	0.82	0.11	0.41	-0.34	-0.74	
	Cumulative contribution ratio	0.717	0.863	0.960	0.993	1.000	
2	x_6	0.80	0.45	0.41	0.49	0.42	0.99
	x_{38}	-0.68	0.61	0.49	0.77	0.14	0.15
	x_{39}	0.32	0.39	-1.08	0.14	-0.22	0.15
	x_{51}	-0.91	-0.14	-0.18	0.24	-0.06	0.83
	x_{53}	0.07	0.12	0.12	0.31	-0.97	0.24
	x_{66}	-0.03	1.04	-0.04	-0.35	0.31	0.88
	Cumulative contribution ratio	0.670	0.870	0.938	0.967	0.989	1.000
3	x_{20}	-0.64	0.88	0.28	0.45	-0.46	
	x_{34}	0.86	0.53	-0.22	-0.95	-0.37	
	x_{42}	-1.45	-0.21	-0.42	-0.57	-0.13	
	x_{63}	0.56	-1.06	0.20	-1.72	-1.85	
	x_{66}	-0.80	0.65	-1.07	1.91	1.19	
	Cumulative contribution ratio	0.569	0.796	0.917	0.999	1.000	
4	x_{11}	-0.22	-0.32	-0.51	-0.47	0.68	
	x_{32}	1.18	1.06	-1.24	1.02	0.60	
	x_{35}	1.58	0.39	-1.40	0.64	-0.06	
	x_{42}	-0.72	0.27	-0.26	-0.79	-0.64	
	x_{56}	-0.53	-0.67	-0.28	0.70	0.13	
	Cumulative contribution ratio	0.553	0.827	0.937	0.981	1.000	
5	x_2	-0.73	-0.56	0.54	0.80	0.39	-0.58
	x_3	1.67	1.40	1.35	1.36	-0.01	-0.01
	x_{51}	-0.23	-0.10	0.58	0.75	-1.11	0.58
	x_{12}	0.75	-0.91	-0.70	0.19	0.22	0.78
	x_{13}	-0.06	0.94	1.60	-0.44	0.58	0.24
	x_{14}	0.61	0.14	-0.26	-1.29	-0.03	0.69
	x_{15}	-1.23	-1.07	-1.13	-1.03	-1.15	0.40
	x_{16}	-0.02	0.10	-0.42	0.05	0.59	0.34
	x_{19}	0.32	-1.13	-0.86	-0.05	-0.71	-1.16
	Cumulative contribution ratio	0.731	0.901	0.960	0.984	0.994	1.000
6	x_{11}	0.18	-0.79	0.31	0.62		
	x_{47}	-0.45	0.54	0.34	0.85		
	x_{56}	-0.08	-0.32	-1.12	0.47		
	x_{60}	1.00	0.50	0.39	0.37		
	Cumulative contribution ratio	0.627	0.904	0.983	1.000		
7	x_{23}	0.53	-1.00	-0.21			
	x_{42}	0.58	0.08	-1.12			
	x_{59}	0.94	0.53	0.35			
	Cumulative contribution ratio	0.727	0.960	1.000			
8	x_4	0.20	-0.79	0.27	0.23	-0.26	
	x_{10}	-1.54	-0.04	0.75	-0.36	-0.23	
	x_{15}	-0.42	0.50	-0.48	-0.01	-0.38	
	x_{16}	0.57	0.90	-0.01	-0.01	0.62	
	x_{33}	-0.44	0.39	-0.94	1.01	-0.44	
	x_{51}	0.80	0.72	0.36	-0.20	-0.81	
	x_{56}	-0.26	0.03	-0.41	-0.57	-0.69	
	x_{60}	0.88	0.24	1.30	-0.41	-0.10	
	Cumulative contribution ratio	0.561	0.772	0.911	0.965	1.000	

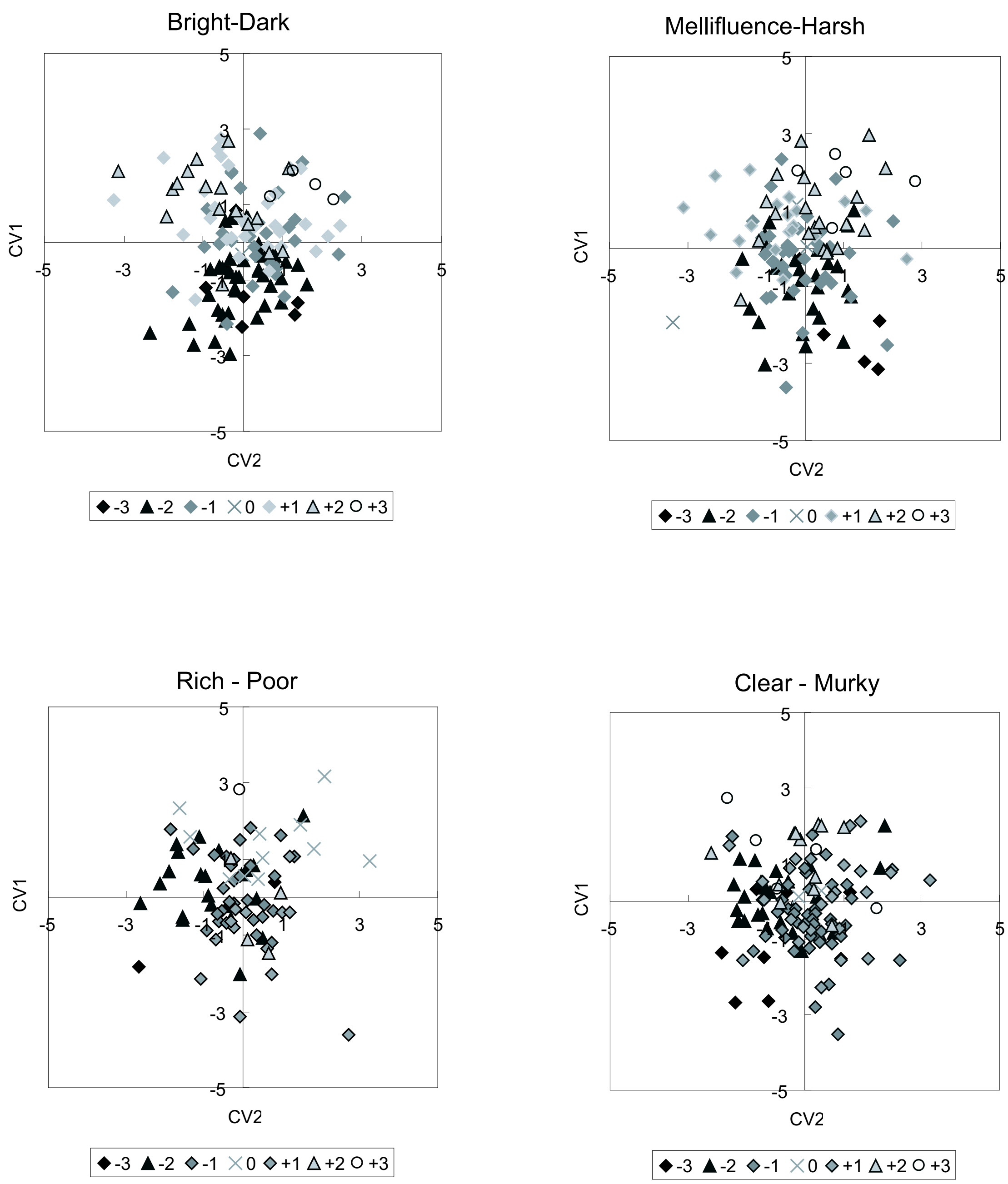


Figure 3.7: Point diagrams of songs for each adjective expressed in 2-dimensions (1)

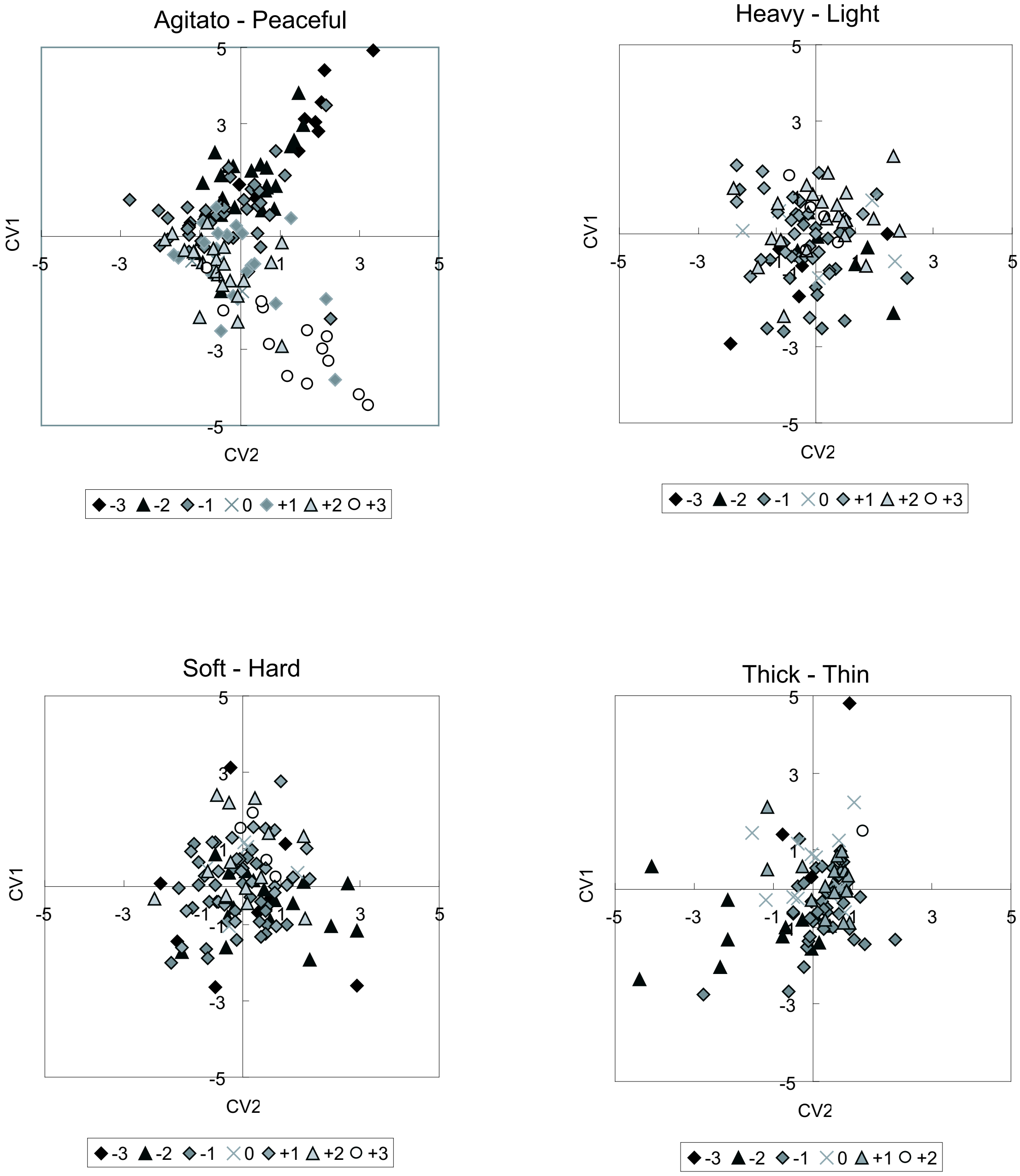


Figure 3.8: Point diagrams of songs for each adjective expressed in 2-dimensions (2)

3.3 Song Selection System Complying With Emotional Requests

Based on the relationships between the musical fluctuation features and the emotional evaluations of music, this study proposes the song selection system complying with emotional requests. The proposed system stores the each canonical variates that compose the estimate space with respect to each adjective and its corresponding coefficients as **"Acoustic–Emotion" model**.

Figure 3.9 shows the general idea of the proposed system. The system is constructed with Acoustic–Emotion model that estimates each emotional evaluations of a song in the given music database from its acoustic features and labeled them with the music.

As users input their requests for a song with the adjective words and their degrees, then the system selects appropriate songs for the requests. The proposed system can be thought user-intuitive, because it needs only the degrees of the adjectives for requesting a song.

Figure 3.10 shows the interface of the proposed system. Users can select a song from their own music database with their emotions as the following steps.

1. Fill the check box of the requested adjective word on the left side of the interface.
2. Select its degree as sliding the bar on the center of the interface, then the degrees of the requested adjective is shown in the *Degree* box on the right side of the sliding bars.
3. All requested adjectives and their degrees are shown in the *Requested* box on the right side of the interface.
4. As clicking the *Enter* button on the bottom of the interface, the system searches suitable song for the requested and the selected song then is appeared on the right side of the *Enter* button

In the system, the emotional evaluations of a song s in the given "Music Database" were labeled based on the Acoustic–Emotion model. The adjective evaluations of the song s, Emo_s can be shown as following equation.

$$Emo_s = (Ad_1, Ad_2, Ad_3, Ad_4, Ad_5, Ad_6, Ad_7, Ad_8). \tag{3.9}$$

And user's emotional request, REQ can be shown as following equation

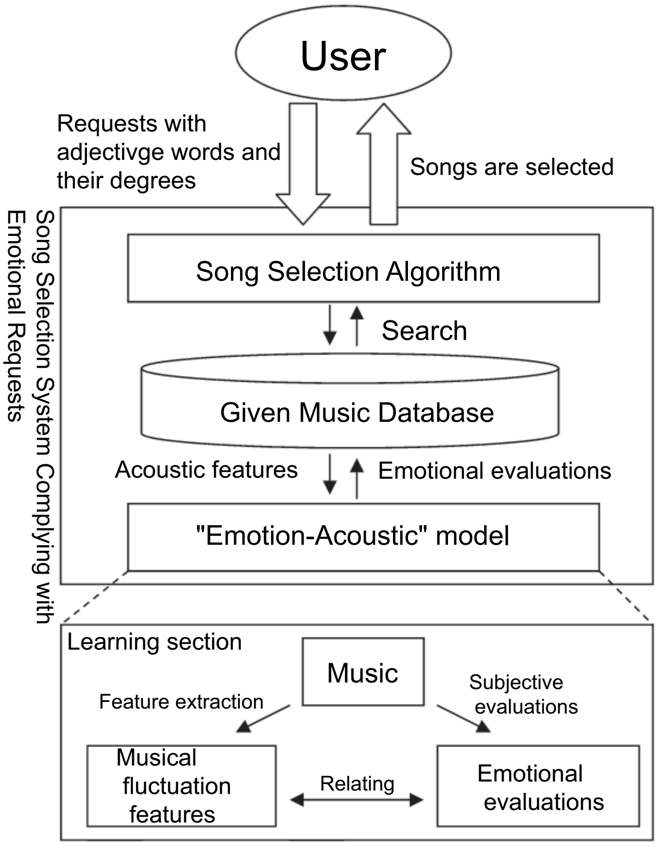


Figure 3.9: General of the proposed system

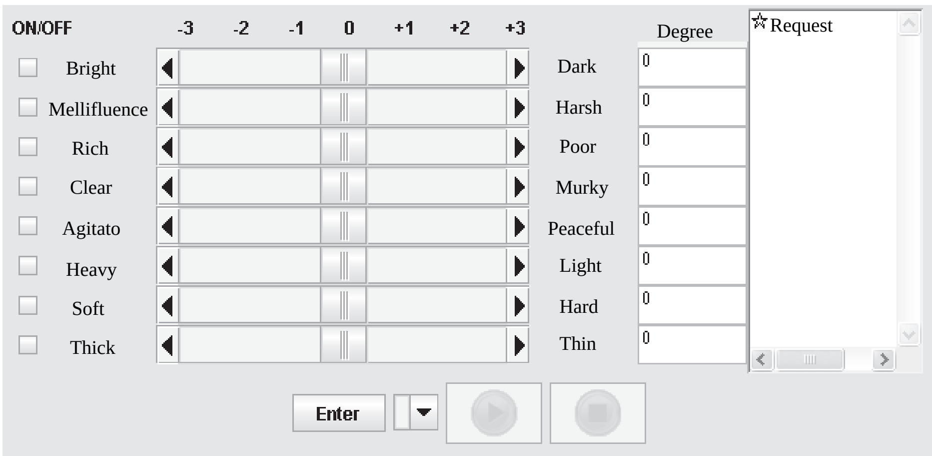


Figure 3.10: The interface of the proposed system

$$REQ = \{Req|Req = (j, Deg), 1 \leq j \leq 8, -3 \leq Deg \leq 3\}, \tag{3.10}$$

where j and Deg are both integer numbers, and each shows adjective index (see Tab. 3.2) and their degree, respectively. The element count of REQ is only less than eight, and the adjective index is not duplicated among the elements. REQ is used as a retrieval key, the songs that can satisfy the user's request are selected from the given "Music Database."

3.3.1 Song Selection Algorithm

In this paper, we prepared two types of song selection algorithm. With each algorithm, two types of song selection system are composed.

Exactly-matching

"Exactly-matching" (EM) system is one of the general song selection system, which selects songs that exactly matches user's request. The EM system selects songs that **exactly** satisfying user's request REQ as $Requested\_songs$ from the "Music Database" $MusicDB$ as follows.

SELECT $Requested\_song$ FROM $MusicDB$

WHERE FORALL $Req \in REQ$

Tolerant-list

People can be satisfied with the songs which have a few difference from their requests, hence this study prepared the song selection system considering tolerance of affection, "Tolerant-list" system. The TL system selects songs that **tolerantly** satisfying user's request REQ as $Requested\_songs$ from the given "Music Database" $MusicDB$ as follows.

SELECT $Requested\_songs$ MIN (SDL_s)

FROM $MusicDB$

where SDL_s shows user's satisfaction level for the song s, and is calculated by the following equations.

$$SDL_s = \sum_{Req \in REQ} DL_{s,Req}, \tag{3.11}$$

$$DL_{s,Req} = \begin{cases} |emo - req| : & (emo * req \geq 0) \\ \infty : & (emo * req < 0), \end{cases} \tag{3.12}$$

$$emo = \arg(\arg(1, Req), Emo_s), \tag{3.13}$$

$$req = \arg(2, Req), \tag{3.14}$$

where $\arg(i, Tuple)$ shows the function that gets i th element from n-tuple. And ∞ in Eq.3.12 means that the TL algorithm don't allow opposite sign between Req and Aff_s.

We believed that song selection system considering the prior adjective of user's request can get more usability, especially in the case that there are many songs having same SDL_s values. Thus this study defined magnitude relation between song k and l ($k \neq l$) that the both have same SDL_s values, $Compere(k, l, REQ)$ as follows.

SELECT $Compare$

CASE($\text{argmax}_{Req \in REQ}(|\arg(2, Req)|)$),

WHEN $DL_{k,Req} > DL_{l,Req}$ THEN $k > l$,

WHEN $DL_{k,Req} < DL_{l,Req}$ THEN $k < l$,

WHEN $DL_{k,Req} = DL_{l,Req}$

THEN $Compare$ $(k, l, REQ \backslash \{Req\})$.

If magnitude relation can not be determined by max absolute value of user's requested adjective degree, the magnitude relation is recursively determined by the next big absolute value of one. And if the magnitude relation can not be determined though all of user's requested adjective degree are used, the songs are not discriminated and listed at random order.

By above steps, the TL system selects and lists 10 songs as *Requested_songs* that have emotional evaluation approaching user's request.

3.3.2 Subjective Evaluation Experiment

The subjective evaluation experiment is conducted for comparing the two types of algorithm. In this experiment, it was evaluated that correctness of the selected songs for user's emotional requests and the usability of the song selection system considering the selected songs.

One hundred and fifty songs, which were not used in construction of Acoustic–Emotion model, were prepared for Music Database. Twenty three males and females in their twenties participated. The EM system and the TL system are each shown as "system A" and "system B" to participants. They were asked to use the both two systems freely, and evaluate them for prepared two pair adjectives listed in follows on a scale of one to seven.

- **Correct – Incorrect**
  highly-correct (+3), correct (+2), mildly-correct (+1), neither (0), mildly-incorrect (-1), incorrect (-2), highly-incorrect (-3)
- **Satisfiable – Unsatisfiable**
  highly-satisfiable (+3), satisfiable (+2), mildly-satisfiable (+1), neither (0), mildly-unsatisfiable (-1), unsatisfiable (-2), highly-unsatisfiable (-3)

Fig. 3.11 shows average and standard deviation of all subjects' evaluation of each system for the adjectives. For "Correct – Incorrect", it was confirmed that both EM system and TL System were evaluated as more than 1.0 on a scale of one to seven. This suggested that the proposed system can select "corrective" songs for users' requests without difference of song selection algorithm, and the proposed Acoustic–Emotion model has sufficient usability. For "Satisfiable - Unsatisfiable", we prepared the evaluation of the Shuffle system which is widely used in present as the control sample. And compare the evaluations each other with Kruskal Wallis test by Steel-Dwass (1% significance). Then it was confirmed that both the evaluations of TL system and EM system show significantly high evaluations than the one of the Shuffle system, especially the TL system shows the most high satisfiable evaluation.

For more detailed analysis, we focused on the subjective evaluation for numbers of the requested adjectives. Figure 3.12 shows the average and the standard deviation of all subjects' evaluation of "Correct – Incorrect" for the numbers of the requested adjectives. Though the evaluation of the EM system for "Correct" became low as the number of the requested adjectives increased, the evaluation of the both system were maintained at positive. And Figure 3.13 shows the average and the standard deviation of all subjects' evaluation of "Satisfiable – Unsatisfiable" for the numbers

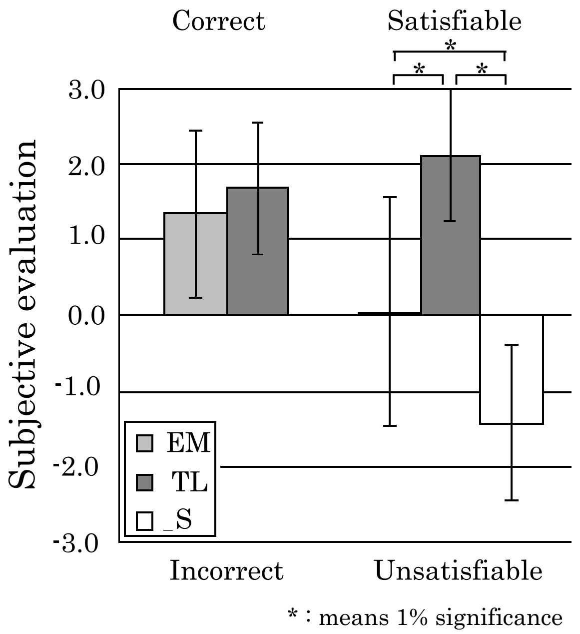


Figure 3.11: Subjective evaluation results

of the requested adjectives. From the figure, it was confirmed that the evaluation of the EM system for “Satisfiable” became especially low to negative as the number of the requested adjectives increased, though the one of the TL system was maintaned at the high level in positive. The TL system provided some nearest songs to the users’ requests, and it seemed to cause the high satisfiable evaluation. And this result seemed to be caused by that the tolerance of the human instincts was taken in the consideration in the TL system. It was suggested that the EM system provided only the songs exactly matched to the requests, the EM system could not comply the complex requests from the users; the EM system seems to become more satisfiable by adding more music in the database.

From these, it was suggested that the proposed systems with **both algorithm** enable users to select **correct** songs for their requests, and the TL system showed higher **satisfiable** evaluation on song selection.

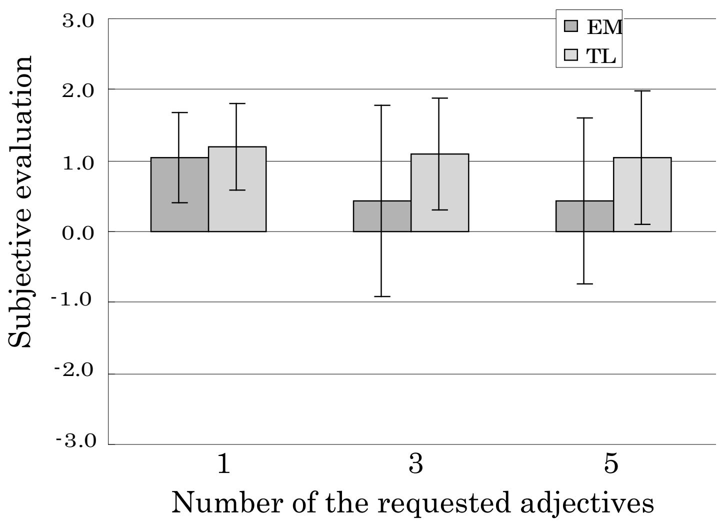


Figure 3.12: Subjective evalution of “Correct-Incorrect” for the numbers of the requested adjectives

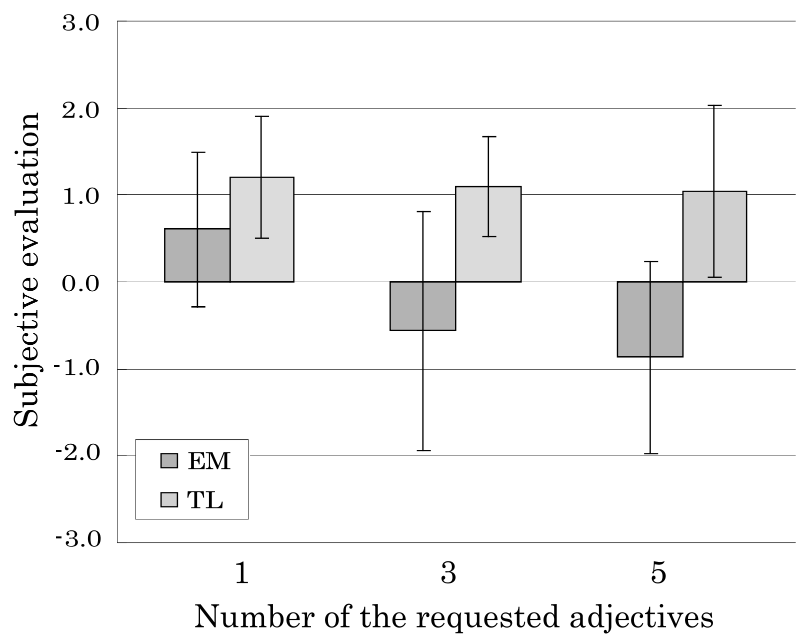


Figure 3.13: Subjective evalution of “Satisfiable-Unsatisfiable” for the numbers of the requested adjectives

3.3.3 Discussion

In this section, this study proposed the song selection system complying with emotional requests, that is based on the Emotion–Acoustic model. Two types of song selection algorithm were prepared for the system: EM algorithm and TL algorithm. Through subjective evaluation experiments, it was confirmed that the proposed select appropriate songs for the requests without difference of the song selection algorithm. And the proposed system with TL algorithm was more satisfiable than the one with EM algorithm and Shuffle system which is used widely in present.

In future, this study will test the usability of the proposed system for searching back ground music under the assumption that scenes the proposed system will be used, particularly in the game sound field where game sound designers/developers have to search through tens of thousands of sound clips. Moreover concurrently using dialogue mood estimation system [45], an automated suitable song selection system for dialogue mood will be developed.

3.4 Quantification of Musical Performance Impressions Based on Acoustic Analysis

It can be thought that most people can perceive each impression from not only the mood and melody of music, but also the emotional performance by the player. Even a single sustained tone can provide many and several impressions with people as the instruments are played creatively and emotionally. The musical maestros can move the listener with their playing, because they have a lot of and exquisite expressive skills. However the public interested in music do not have sufficient expressive skills, therefore they can not give some impressions to their playing and move listener. It is important for them to objectively evaluate their own musical performance for improving their expressive skills. Therefore, the system that can output objective evaluation of their musical performance is useful to educate the expressive skills of the public. This study proposes the quantification method for the impressions about musical performances focusing on the single sustained tone based on its acoustic property.

This study especially focused on the violin as the one of the instruments whose musical performance can be reflected by the minute changes of the expressions. Violin is far from keyboard instruments, percussion instruments, and plucked string instruments, and can almost permanently sustain its sound. And it also can give several expressions, e.g., vibrato, accent, and dynamics in the single tone. Moreover vibration of strings is leaded to their body through the bridge, and several and many harmonic tones are generated in playing the violin. It has been known that the harmonic tone is one of the important factors for generating tone color in acoustic instruments [46].

This study is different from the existing studies, discrimination or classification of musical instrument by their acoustic features [27, 47] and genre estimation by acoustic features [25] in that this study focuses on the impressions about musical instrument's performance. And the study covers not only general acoustic features, such a composition ratio of harmonic component, a distribution of non-harmonic component, and amplitude, but also fluctuation properties that can explain the temporal variation of a sound.

3.4.1 Presentation Sounds of Violin Performance

A professional violin performer played and recorded a single sustained tones with several playing techniques [48], and the pitches of the single sustained tone are discussed in detail below. All sound data used in an analysis are taken as WAV format, the sampling frequency is 96000 Hz, and they were linear quantized to 24 bits in order to standardize the acoustic quality. In order to study the minute difference of sounds, this study used the higher quality than Compact Disk in the analysis.

It is supposed to prepare plural pitch data to cover several sound inputs. However it is difficult to record all pitches that can be played with the violin, therefore the eight pitches were used considering the circle of fifths in musicology [49]; $A3$ (220 Hz), $C4$ (261 Hz), $E \flat 4$ (331 Hz), $F \sharp 4$ (370 Hz), $A4$ (440 Hz), $C5$ (522 Hz), $E \flat 5$ (622 Hz), $F \sharp 5$ (740 Hz); All pitches can be reasonably played with violin. Then each pitch is played and recorded twelve times, and used for the presentation sounds.

3.4.2 Acoustic properties

This study used the one hundred and twelve acoustic properties that can be available for quantifying the impressions of the sounds: the frequency properties, amplitude properties, and fluctuation properties of the single tone. Before the extracting properties, the power spectrum on each frequency and each time are calculated by conducting Short-time Fast Fourier Transform (Hamming window, length of window 2048 points, resolution time is 25 [msec]).

Frequency spectrum properties

This study extracts the power values of the harmonic component as the steady property that can explain the structure of the frequency structure about tone color. The harmonic component is the high order frequency component that is whole-number multiple of the fundamental frequency, and it is generated with the sound that people recognize as the pitch (fundamental pitch). Most people can not perceive the harmonic component by their ears, however it has been known that the composition ratio of the harmonic component structure influences the impressions of the sound in their unconsciousness.

There are two types of the harmonic components, harmonic component that is harmonious for the fundamental pitch and non-harmonic component that is not harmonious for the one. As dimension number of the harmonic component increases, it is often non-harmonic for the fundamental pitch. This study extracts thirty one properties shown in follows, then parenthetic number means the index of property.

(1) The ratio about the power of the fundamental pitch to the sum of the power of all harmonic components.

(2)-(28) The ratio about the power of the nth harmonic component to the sum of the power of all harmonic components ($n = 2, 3, \ldots, 28$).

(29) The sum of the power of the odd harmonic components (This include the fundamental pitch).

(30) The sum of the power of the even harmonic components.

(31) The ratio about the sum of the power of the odd harmonic components to the sum of the power of the even harmonic components.

Amplitude power variation properties

In order to explain the amplitude structure relating the volume, following five properties are extracted. Then parenthetic number means the index of property.

(32) The slope of collinear approximation by the linear least-squares method for the power envelope curve.

(33) The intercept of collinear approximation by the linear least-squares method for the power envelope curve.

(34) The average of the amplitude power.

(35) The maximum of the amplitude power.

(36) The minimum of the amplitude power.

Fluctuation properties

The fluctuation shows the temporal variation of physical something, and can be found through nature, such a murmur of a brook. In the musical sound, there are fluctuations about pitch and volume, which probably influence the impressions of the sound. Thirty six properties about fluctuations of both pitch and volume are extracted.

I. Fluctuation properties of pitch

   The average pitch on the each given time is calculated by conducting Short-time Fast Fourier Transform detailed in Section 3.4.2. And the average pitch in the time series is used as the fluctuation of pitch, in which the minute changes of pitch can be found. Then this study conducts Fast Fourier Transform to the given fluctuation of pitch, and the fluctuation spectrum of pitch is calculated.

II. Fluctuation properties of volume

In a similar way to the fluctuation properties of pitch, the power spectrum on each time is calculated by conducting Short-time Fast Fourier Transform. In this study four relative frequency bands are prepared: Low, Middle, High, Ultrahigh, and All. And each fluctuation of volume on the each frequency band is then calculated. Here, the each relative frequency band is prepared as follows according to the fundamental frequency of the eight pitch prepared as the presentation tones; Low band is lower than fundamental frequency, Middle band is from fundamental frequency to the eighth harmonic component, High band is from from the eighth harmonic component to the sixteenth harmonic component, Ultrahigh band is over than the sixteenth harmonic component, and All means the frequency as long as we can calculate. This study calculates the integral of the each power spectrum on each frequency band, each time, and used as the volume on the each frequency band. The volume in the time series is used as the fluctuation of volume. Then this study conducts Fast Fourier Transform to the given fluctuation of volume, and the fluctuation spectrum of volume is calculated.

From the fluctuation spectrum defined above, the following each fluctuation property on each frequency is extracted.

1. The value of the integral: S

$$S_\omega(t) = \int^{\omega} freq(\omega, t) d\omega.$$

2. Centroid frequency: $f_{centroid}(t)$

$$f_{centroid}(t) = \frac{\int \omega \cdot freq(\omega, t) d\omega}{S_{all}(t)}.$$

3. Max value of the fluctuation spectrum: $fluc^{max}$

$$fluc^{max} = \max_{\omega > 0.1} \left(fluc\left(\omega\right)\right).$$

4. Max frequency of the spectrum: $freq^{max}$

$$freq^{max} = \arg\left(\max_{\omega>0.1}\left(fluc\left(\omega\right)\right)\right).$$

5. The slope of the regression line: $\alpha(t)$
   The slope of the regression line for $fluq(\omega, t)$ is calculated by using the least-squares method,

$$\alpha(t) = \frac{n\sum^{n}\omega \cdot freq(\omega, t) - \sum^{n}\omega\sum^{n}freq(\omega, t)}{n\sum^{n}\omega^2 - \left(\sum^{n}\omega\right)^2}.$$

6. The intercept of the regression line: $\beta(t)$
   The intercept of the regression line for $fluq(\omega, t)$ is calculated by using the least-squares method,

$$\beta(t) = \frac{\sum^{n}\omega^2\sum^{n}freq(\omega, t) - \sum^{n}\omega \cdot freq(\omega, t)\sum^{n}\omega}{n\sum^{n}\omega^2 - \left(\sum^{n}\omega\right)^2}.$$

Where n, $fluc(\omega, t)$, and ω each shows number of the spectrum, fluctuation spectrum on the given time t, and frequency, respectively. With the properties explained above, the indices of the fluctuation spectrum properties were defined as follows. Then parenthetic number means the index of property.

(37)-(42) Fluctuation spectrum property of volume on All band 1. $\sim$ 6.

(43)-(48) Fluctuation spectrum property of volume on Low band 1. $\sim$ 6.

(49)-(54) Fluctuation spectrum property of volume on Middle band 1. $\sim$ 6.

(55)-(60) Fluctuation spectrum property of volume on High band 1. $\sim$ 6.

(61)-(66) Fluctuation spectrum property of volume on Ultrahigh band 1. $\sim$ 6.

(67)-(72) Fluctuation spectrum property of pitch 1. $\sim$ 6.

Harmonic component kurtosis properties

Harmonic component is the component that is whole-number multiple of the fundamental frequency, which influences the tone color. On another front, non-harmonic component is not whole-number multiple of the fundamental frequency, and that is often in the noise and explosion, which does not have any pitch. This study calculates the matter of the non-harmonic component around the each harmonic component peak from each harmonic component kurtosis. As extracting the data around the harmonic component peak that the length is from the fundamental pitch to the twentieth component, and the harmonic component kurtosis is calculated from fourth-order moment. Harmonic component kurtosis K_w can be shown as the following equation.

$$K_w = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^4}{(nV^2)}, \tag{3.15}$$

where, X_i $(i = 1, 2, \ldots, n)$, i, $\bar{X}$, and V each shows the measured value of the power spectrum in the extracted data, the number of the sampling, the average of the power of the spectrum, and the dispersion of the power of the spectrum, respectively.

As the harmonic components are clear and pointed (see Figure 3.14), the spectrum has high kurtosis and a little non-harmonic components. However, as harmonic component is not clear and pointed (see Figure 3.15), the spectrum has low kurtosis and much non-harmonic components.

In the period between beginning of the sound and 150 [msec], the following harmonic component kurtosis properties. Incidentally, the 0th harmonic component means the fundamental pitch.

(72)-(91) The average of K_w for each time on the nth harmonic component ($n = 0, 2, 3, \ldots, 20$).

(92)-(111) The amplitude of K_w for each time on the nth harmonic component ($n = 0, 2, 3, \ldots, 20$).

3.4.3 Impressions of Musical Performance

We conduct the subjective evaluation experiment to obtain the impression values of musical performances. In the experiment, we prepared twelve patterns for eight pitches (ninety six musical performances) as the presentation sounds.

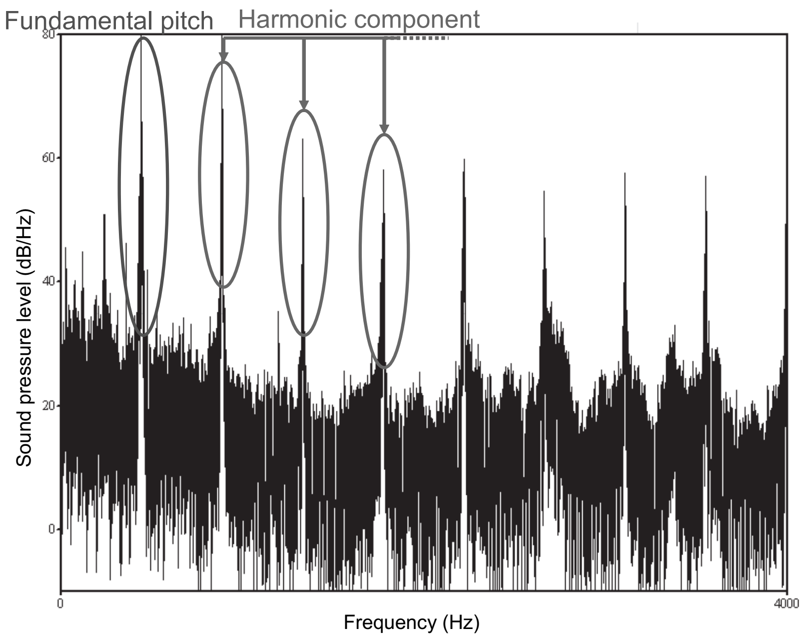


Figure 3.14: An example of harmonic components and high Kurtosis

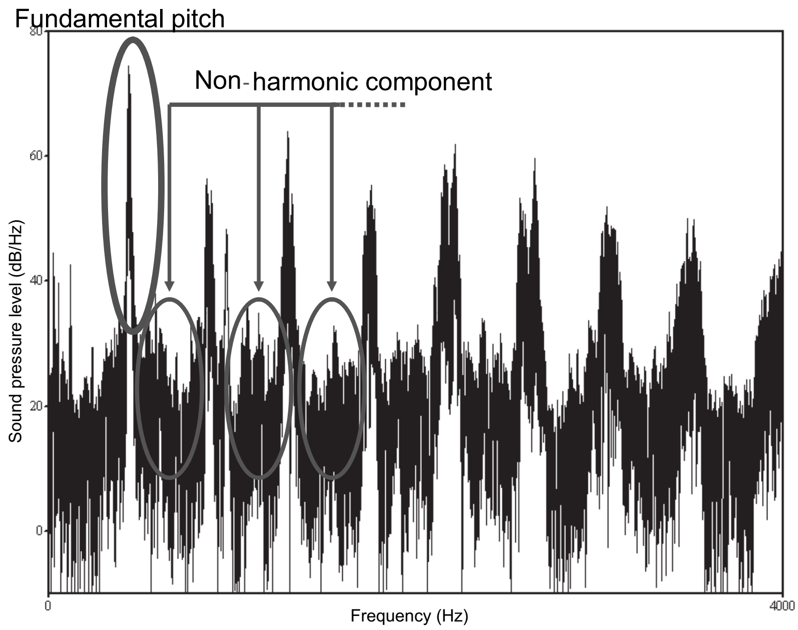


Figure 3.15: An example of non-harmonic components and low kurtosis

Table 3.8: Adjective pairs

Adjective index	Evaluation point (0 - 5) ← 5 4 3 2 1 0 1 2 3 4 5 →
1	Bright - Dark
2	Clear - Murky
3	Mellifluence - Harsh
4	Rich - Poor
5	Agitato - Peaceful
6	Heavy - Light
7	Cold - Warm
8	Soft - Hard

Subjective evaluation experiment

We prepared eight pairs of the adjective shown in Table 3.8, which can be considered suitable for evaluating the tone color. In the experiment, thirty subjects (twenty three males, seven females) are asked to listen to presentation tones and evaluate the degree of the each prepared adjectives on a scale of one to five. The pitches are presented in random order for subjects, and the performance patterns of the pitch are presented in random order through a prepared interface on the given web browser and the 5.1 ch surround speakers. With this interface, the subjects could rewind and play back the presentation sounds over and over again. To alleviate the subjects' tiredness by the experiment, the subjects listen to the presentation sounds not only once, in two periods that each period has forty eight presentation sounds (four pitch times twelve performance patterns). And in an interval between the two periods, subjects have sufficient time to rest.

Impression value

We calculate the seven hundred and sixty eight impression values (ninety six musical performance patterns times eight pairs of adjective) from the evaluation obtained through the experiment detailed in the preceding section. We use the average of the evaluation of all subjects as the impression value of the presentation sound. Then the evaluations that disobey normal distribution are removed from the analytical objects.

Table 3.9: Selected acoustic properties for each adjective

Adjective index	Properties
1	q_{32}, q_{33}, q_{47}, q_{51}, q_{55}, q_{60}, q_{71}, q_{81}, q_{86}, q_{102}
2	q_{12}, q_{13}, q_{51}, q_{75}, q_{78}, q_{83}, q_{84}, q_{85}, q_{86}, q_{88}
3	q_{5}, q_{6}, q_{35}, q_{44}, q_{51}, q_{57}, q_{71}, q_{73}, q_{82}, q_{87}
4	q_{47}, q_{53}, q_{56}, q_{57}, q_{83}
5	q_{33}, q_{34}, q_{36}, q_{56}, q_{60}, q_{68}, q_{73}
6	q_{4}, q_{11}, q_{12}
7	q_{23}, q_{28}, q_{36}, q_{47}, q_{71}, q_{80}, q_{86}
8	q_{23}, q_{25}, q_{34}, q_{59}, q_{60}, q_{73}, q_{82}

3.4.4 Quantification based on the acoustic properties

We conduct multiple linear regression analysis, and study the correlation between the acoustic properties detailed in section 3.4.2 and the impression values detailed in section 3.4.3. The contributory properties for multiple linear regression analysis were selected from the one hundred twelve different acoustic properties with respect to each adjective. Then the selected properties and the impression values are each used as the explanatory variable and the objective variable, respectively. The appropriate features properties were selected by using a stepwise method.

In this paper, we would like to reveal the correlation between the acoustic properties and the impression values not depended on the pitch. Therefore, we use all ninety six data used in subjective evaluation experiment as the analytical object.

Correlation between acoustic properties and the impression values

Table 3.9 lists the selected properties q_i (i shows the index of the acoustic property) for each adjective. For each adjective, the multiple linear regression analysis was conducted, and each correlation between measured value and the predicted value by the multiple regression equation are listed in Table 3.10.

For adjective 2, 3, 5, and 8, we can confirm the high correlations, as their adjusted multiple correlation coefficients are over than 0.60. And especially for adjective 3

Table 3.10: Adjusted multiple correlation between the acoustic properties and the impression value

Adjective index	Adjusted multiple correlation coefficient R^2	Standardized residual
1	0.49	1.08
2	**0.66**	1.39
3	**0.76**	1.28
4	0.25	1.71
5	**0.68**	1.29
6	0.29	0.95
7	0.40	1.19
8	**0.65**	1.39

(Mellifluence - Harsh) and adjective 5 (Agitato - Peaceful), the adjusted multiple correlation coefficients are relatively high, it seems that the highly accurate multiple regression equation are obtained. For example, Figure 3.16 shows the correlation between measured value and the predicted value by the multiple regression equation for adjective 3 (Mellifluence - Harsh). The red line in Figure 3.16 shows the regression line, and the coefficient and the intercept of the regression line are each 0.82 and -0.22, respectively. As we can see the regression line in Figure 3.16, the impression values are effectively predicted by using the multiple regression equation. For other adjectives that have high adjusted multiple correlation coefficients, it is no different from this example. Thus we suggest the high correlation between the impression value and the acoustic properties which does not depend on the pitch, and the availability of the proposed method for quantification of the violin performance impressions.

Effective properties for quantification of musical performance impressions

Table 3.11 lists the selected properties and the standardized partial regression coefficient of the multiple regression equation for each adjective. Then the standardized partial regression coefficient shows the influence rate of each selected properties for quantifying the degree of each adjective. In following sections, we detail the properties that influences the quantification of adjective 2, 3, 5, and 8, because the degrees of their adjectives are quantified in relatively high accurate rates.

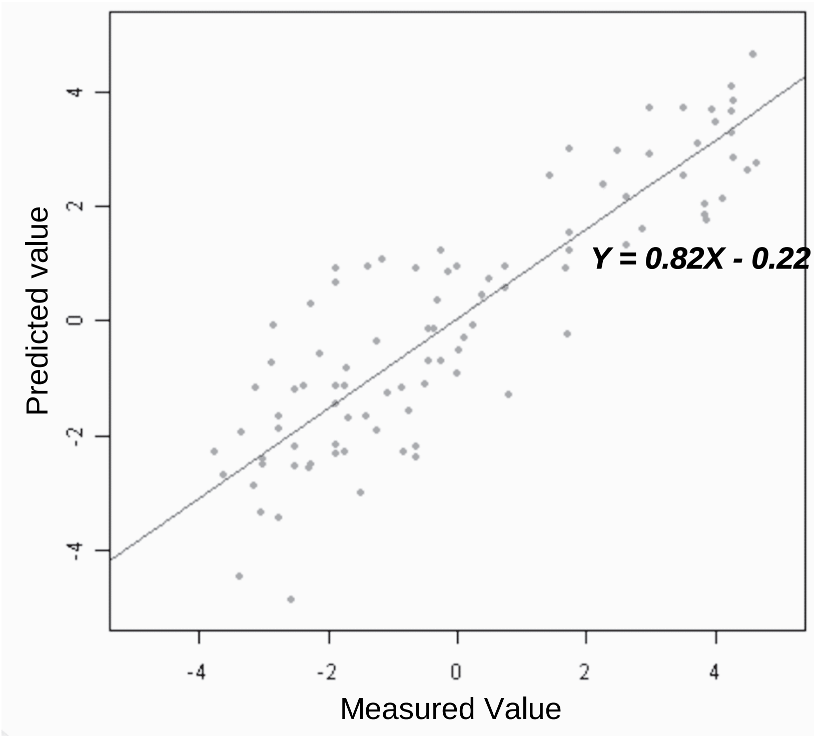


Figure 3.16: Correlation between the measured value and the predicted value for adjective 3 (Mellifuluence - Harsh)

Adjective 2: Clear - Murky For adjective 2, the standardized partial regression coefficients of q_{12} and q_{13} are both big numbers in positive. Then q_{12} and q_{13} each shows the ratio about the power of the 12 th and 13 th harmonic component, respectively. This results suggest that as the powers of the 12 th and 13 th harmonic component increase, the impression value for "Clear" is high. The 12 th harmonic component is three octaves and perfect fifth high for fundamental pitch, and the 13 th harmonic component is three octaves and major sixth high for fundamental pitch, which both of them are harmonious for fundamental pitch. Hence it suggests that the harmonious frequencies' power influence the impression "Clear - Murky".

Adjective 3: Mellifluence - Harsh For adjective 3, the standardized partial regression coefficients of q_5 and q_6 are both awfully big numbers in positive. The property q_{12} and q_{13} each shows the ratio about the power of the 5 th and 6 th harmonic component to the sum of the power of all harmonic component, respectively. Therefore it is suggested that as the 5 th and 6 th harmonic component increase, the musical performance enables people to feel "Mellifluence". The 5 th harmonic component is

Adjective index	Properties	Standardized partial regression coefficient	Adjective index	Properties	Standardized partial regression coefficient
2	$\mathbf{q_{12}}$	**52.47**	3	$\mathbf{q_5}$	**42.46**
	$\mathbf{q_{13}}$	**51.55**		$\mathbf{q_6}$	**35.04**
	q_{51}	-0.01		q_{35}	0.37
	q_{75}	-0.36		q_{44}	-0.001
	q_{78}	0.37		q_{51}	-0.003
	q_{83}	0.45		q_{57}	0.001
	q_{84}	-0.47		q_{71}	0.85
	q_{85}	0.44		q_{73}	-0.17
	q_{86}	0.42		q_{82}	0.27
	q_{88}	-0.34		q_{87}	-0.22
5	q_{33}	-0.34	8	$\mathbf{q_{23}}$	**-28.78**
	q_{34}	-0.05		$\mathbf{q_{25}}$	**-26.81**
	$\mathbf{q_{36}}$	**2.79**		q_{34}	0.04
	q_{56}	-0.005		q_{59}	0.91
	q_{60}	0.64		q_{60}	0.31
	q_{68}	0.005		q_{73}	-0.19
	q_{73}	0.17		q_{82}	0.26

Table 3.11: Standardized partial regression coefficient of each properties

two octaves and major three high for fundamental pitch, and the 6 th harmonic component is two octaves and perfect fifth high for fundamental pitch. The high powers of these properties mean that the major harmony made of fundamental pitch, the third, and the fifth are formed. From this, it seems that the harmony of fundamental pitch, the third, and the fifth influence the impression “Mellifluence - Harsh”.

Adjective 5: Agitato - Peaceful For adjective 5, the standardized partial regression coefficient of q_{36} which shows the minimum of the amplitude power is big number in positive. This result shows that as the minimum of the amplitude power increases, the performance impresses people “Agitato”,

Adjective 8: Soft - Hard For adjective 8, the standardized partial regression coefficients of q_{23} and q_{25} are both big numbers in negative. The property q_{23} and q_{25} each shows the ratio about the power of the 23 th and 25 th harmonic component to the sum of the power of all harmonic component, respectively. The 23 th harmonic component and the 25 th harmonic component each means four octaves and augmented fourth high pitch for fundamental pitch, and four octaves and augmented fifth high pitch for fundamental pitch, respectively. As these unharmonious pitch for fundamental pitch decrease, the music performance impresses people "Soft".

3.4.5 Discussion

In this section, this study proposed the quantification method for the musical performance impressions based on the acoustic analysis, which especially focused on the violin playing. Then the frequency spectrum, amplitude power variation, fluctuation of pitch and volume, and harmonic component are extracted and used as the acoustic properties. Through the experiment, it was confirmed that the proposed method enables to quantify the impressions of musical performance in high accuracy rates. Moreover, the effective acoustic properties for quantifying the degrees of each adjective were also suggested through the discussion of the selected properties.

Based on this study, the automated evaluation system for musical performance will be constructed in future. And this study will cover other musical performance by other instruments to get more usability.

3.5 Conclusion

In this chapter, this study focused on the human emotion/affection as one of the human factors, and intended to reveal the relationships between the acoustic features and the emotional/affective evaluation. Applying the result, the intuitively song selection system was developed.

In section 3.2, this study clarified the relationships between the acoustic features and the emotional evaluation of music especially focusing on the musical fluctuation properties, which can show the temporal variation of music. As musical fluctuation properties, the temporal variation of volume, pitch, and rhythm were extracted from music, on the other hand the emotional evaluations of the music were obtained through subjective evaluation experiments. Using features selection and the multiple discriminant analysis, the musical fluctuation properties were related with the emotional evaluation of the music. The emotional evaluation of music can be estimated by the musical fluctuation properties, and the high accuracy rates and low errors were confirmed in cross validation. The selected properties for estimating an emotional evaluation composes the estimation space, the coefficients of the selected properties for estimation space were considered, and noted the effective properties for emotional evaluation of music.

The generated estimation spaces for the emotional evaluation were composed and assumed as the *Acoustic-Emotion model*. The discriminant results show high accuracy rates and low discriminant errors for all adjectives. Using the acoustic-emotion model, the song selection system complying with emotional request was constructed. For the system, two types of song selection algorithm were prepared: the exactly-matching and the tolerant algorithm. Through the subjective evaluation experiment, the high usability was confirmed, and the selected songs were evaluated as the corrective songs for the user's request.

And in section 3.4, the quantifying method for impressions of musical performances was proposed. The musical performance is also important to provide emotional/affective effects with human beings. By the acoustic analysis, the musical performance were successfully quantified. And this study also suggested the effective acoustic properties for the emotional evaluation of the musical performance. This facts can be applied to the automated musical performance evaluation system to improve the musical performance to have more emotional/affective colors, which is especially for the public interested in music.

Overall, the high accuracy rates were confirmed, and the effective acoustic features that identify the emotional/affective evaluations were determined. As the application of this result, the emotional song selection system was developed. In the future, user-intuitively system for music will be developed, such the automated back ground music selection system. This study in this chapter was based on the fact that the temporal variation of music should be taken in the consideration, and the acoustic fluctuation features were used; I believe that the acoustic fluctuation features lead these pleasing results.

Chapter 4

Accompanying System for Composing Support

4.1 Introduction

Recently, there are diverse and numerous forms of entertainment to enjoy, which has been known as *Multimedia entertainment*, and almost all of them are composed with music because music provides people with more affective effects. To create more affective multimedia contents, an original dramatic music should be enclosed. Many composers are guided by their "senses" when writing music, and they have often moved people with their expressive compositions. Most music is composed by taking many elements into consideration, e.g., rhythm, harmony, and melody [50]. However it is not clear that what elements in a song can move people, and only the gifted composer can write music that will be loved by many people for a long time. The question we have to ask here is whether people who do not have much music experience can compose songs.

In the last several years, in the field of Information Science, sound and music recognition [51, 52, 53, 54] and generation [55, 56] systems have been focused and studied. And some products have also been published [57, 58, 59], hence composing music has been accessible for people. However, it is still difficult for people who have not sufficient musical experiment to compose music, because high level knowledge and experience are needed for composing. The music composing support system should be useful to present pleasure of music, and the music accompaniment composing is paid particular attention in this study.

Accompaniment is the one of the most effective factors that influences emotions. Nevertheless accompaniment is the most difficult part for composing music, because

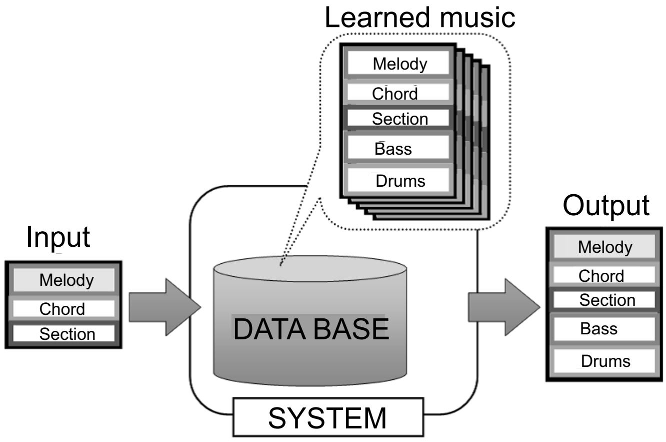


Figure 4.1: General flowchart of the proposed systems

it is necessary for knowledge of mutual interaction among plural musical instruments to compose them.

Figure 4.1 shows the general idea of the system proposed in this section. The system needs some information: inputs, melody, chord, and section. And the melody with accompaniment will be outputted, where the system composes music considering the correlation between melody and each instrument in the learning database. Users do not need any knowledge and experience for composing, but have to only previously make the system learn target plural music.

4.2 Accompanying System Using HMM

In the field of voice recognition, a lot of studies have used Hidden Markov Models (HMMs) for learning [60, 61]. The HMM is known as a special model for learning time-series signals, and has been confirmed to be an effective model for recognizing voice, which has a frequency fluctuation with a time series. In this study, the temporal variation of music on one bar is leaned and the musical score can be generated and output by HMM for every bar.

Music can also be regarded as a frequency fluctuation with a time series, so it is expected that HMMs would be used in music recognition studies. Additionally HMMs is a probability based model, hence the contingency would be made in composing. From this fact, it would be expected that highly-diverse music is composed by using HMMs. For these reasons, it can be thought that using an HMM is a suitable model for recognizing and composing music. Some studies have already suggested the application of HMMs for music recognition and composing systems, and the feasibility of this has been reported, for example, automatic harmonization [62] and automatic estimation of chords [63].

HMM is composed of the four elements: $s_{A_s} \in S$ as the *States*, $o_{A_o} \in O$ as the *Output symbols*, $a_{i,j} \in A$ as the Transition probability from s_i to s_j, and $b_{s_i}(o_l) \in B$ as the Probability that state s_i outputs symbol o_l. Where each A_s and A_o shows the number of states and output symbols, respectively. The Baum-Welch algorithm is used to learn the HMM. When a series data is given as input data In, then HMM λ that satisfies the following equation can be generated.

$$\lambda(In) = (A_{In}, B_{In}), \tag{4.1}$$

where, A_{In} and B_{In} represent A and B that make likelihood $P(In|\lambda(In))$ maximum.

4.2.1 Learning Section

In this study, rhythm, chords, and pitch were each modeled and learned from songs that had melodies and bass part. Score information of a song $sn \in SN$ was modeled according to the following procedure, where SN indicates the database of songs which are 4×4 and without modulation. Each bar t in sn were learned respectively using Left-to-Right HMM where states change in a single direction. Then the eighth note and rest were assumed to be one beat, and used as the minimum unit for learning. The term p shows the number of beats in a bar.

Modeling and learning rhythm series

Rhythm has been modeled in some related studies, for example, [64]. Rhythm is composed of notes and rests, and is also known as a series having the following three states.

RS1: a note that produces a sound

RS2: a rest

RS3: continuance of the previous state

State of rhythm on the beat p, bar t is indicated as $r_{(t,p)}$ for these three states. In bar t of song sn, the rhythm series of melody mr_t^{sn} and the rhythm series of bass br_t^{sn} can be expressed as the following equations.

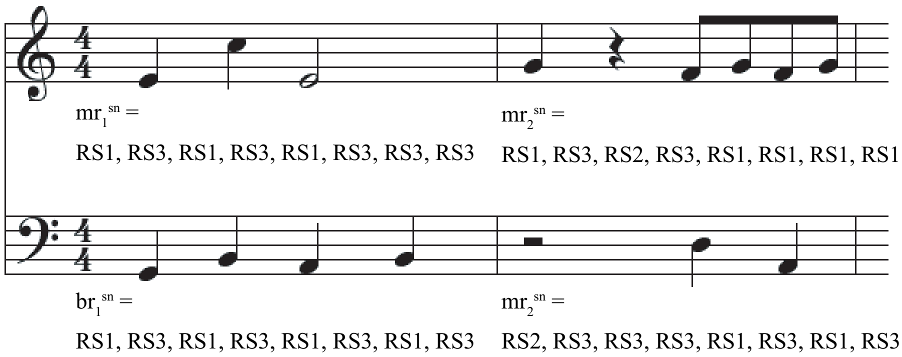


Figure 4.2: An example of the rhythm model

$$mr_t^{sn} = r_{(t,1)}, \ldots, r_{(t,P)}, \tag{4.2}$$

$$br_t^{sn} = r_{(t,1)}, \ldots, r_{(t,P)}, \tag{4.3}$$

$$r_{(t,h)} \in \{\mathrm{RS1}, \mathrm{RS2}, \mathrm{RS3}\},$$

where P shows the number of all beats in a bar. In this paper $P = 8$, because the eighth note were used as the minimum unit for learning. Figure 4.2 shows an example of the rhythm model.

If at least one whole rest is observed in a rhythm series of melody or bass in a bar t, song sn, then mr_t^{sn} and sr_t^{sn} are both removed from the learning. In this condition, database for rhythm series of melody and bass are shown as follows.

$$MR = \bigcup_{sn \in SN} \bigcup_{t} \{mr_t^{sn}\}, \tag{4.4}$$

$$BR = \bigcup_{sn \in SN} \bigcup_{t} \{br_t^{sn}\}, \tag{4.5}$$

here, MR and BR each shows the database of melody and bass, respectively.

Conditional probabilities for rhythm are calculated from appearance probabilities for each series in the database SN. That is to say, these show the relationships between the rhythm series of the melody and the one of bass. In the case of that HMM outputs the rhythm series of bass br when the rhythm series of melody mr is observed, the conditional probability $P(br|mr)$ can be shown as calculated from the following equation.

Table 4.1: Number of tonic

Tonic	C	D♭	D	E♭	E	F	G♭	G	A♭	A	B♭	B
Number of keys	1	2	3	4	5	6	7	8	9	10	11	12

Table 4.2: Number of chord

Name of chords	C	D♭	D	E♭	E	F	G♭	G	A♭	A	B♭	B
Number of chords	1	2	3	4	5	6	7	8	9	10	11	12

$$P(br|mr) = \frac{P(br, mr)}{P(mr)}, \tag{4.6}$$

where $P(mr)$ is the appearance probability of mr in the database SN, and $P(br, mr)$ is the joint probability of br and mr.

HMM set Λ_r is generated through this procedure. Λ_r can be shown as the following equation with Equation 4.1 whose input $In = mr$.

$$\Lambda_r = \{\lambda(mr)|\forall mr \in MR\}. \tag{4.7}$$

Modeling and learning chord progressions

In this paper, each bar in a song $sn \in SN$ is defined to have one chord. The pitch used in the melody and bass on the bar is generally based on the chord of the bar. Taken in the light of tonality, on bar t, song sn, relative degrees for tonic $Degree_t^{sn}$ were used to represent the name of chord [37]. Table 4.1 and 4.2 each lists the numbers of tonic and chords in this study, respectively. $Degree_t^{sn}$ can be calculated with the numbers of tonic and chords listed in Table 4.1 and 4.2, and shown as the following equation.

$$Degree_t^{sn} = \Big\{\{12 + (chord - key)\} \bmod 12\Big\} + 1. \tag{4.8}$$

Chord progression has been known as one of the most important factors that can make a certain impression on people, and has been researched in some related studies [65, 66]. Therefore, taking chord progression into consideration, three bars before and after centering on bar t in a song sn were assumed in a couple of $(t-1, t, t+1)$; three chords over the three bars $(t-1, t, t+1)$ were used as a chord progression. Chord

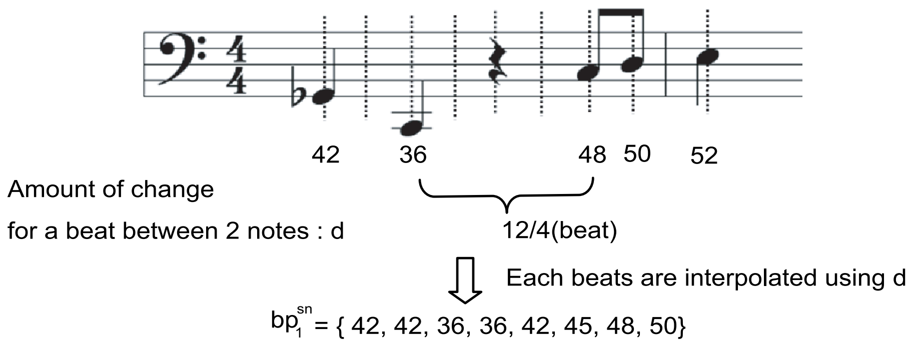


Figure 4.3: An example of pitch series model

progression c_t^{sn} among $(t-1, t, t+1)$ can be shown as the following equation with $Degree_t^{sn}$.

$$c_t^{sn} = (Degree_{t-1}^{sn}, Degree_t^{sn}, Degree_{t+1}^{sn}). \tag{4.9}$$

Database for chord progressions can be shown as the following equation. Here, if there are not any notes on a bar t in song sn, then c_t^{sn} is removed from C.

$$C = \bigcup_{sn \in SN} \bigcup_{t} \{c_t^{sn}\}. \tag{4.10}$$

Modeling and learning pitch series

The pitch on a beat p, bar t can be shown as $pitch_{(t,p)}$. A pitch series of bass on each bar in song sn, bp_t^{sn} can be shown as the following equation with $pitch_{(t,p)}$.

$$\begin{aligned} bp_t^{sn} &= pitch_{(t,p)}, \ldots, pitch_{(t,P)}, \\ pitch_{(t,p)} &\in 1, \ldots, 128. \end{aligned} \tag{4.11}$$

Pitch of not only pronounced beat but all pitch should be learned, because the same length of series are needed for learning HMM. Thus, the pitch for each beat were fixed using linear interpolation for subsequent notes to cover the pitch on unpronounced beat in the learned song. Only notes based on the key of the song were used as substitute notes while fixing the note of each beat. Figure 4.3 shows an example of pitch series model. The database for the pitch series of bass, BP can be shown as the following equation.

$$BP = \bigcup_{sn \in SN} \bigcup_{t} \{bp_t^{sn}\}. \tag{4.12}$$

Conditional probabilities concerning the relationships between pitch series of bass and chord progressions can be calculated from appearance probabilities for each series in the song set SN. In the case of that HMM outputs as bp when c is observed, the conditional probability $P(bp|c)$ can be calculated from the following equation.

$$P(bp|c) = \frac{P(bp, c)}{P(c)}, \tag{4.13}$$

where $P(c)$ shows the appearance probability for c in the song set SN, and $P(bp, c)$ shows the joint probability of sp and c.

Through this procedure, HMM set Λ_c is generated with Equation 4.1 whose input $In = c$.

$$\Lambda_c = \{\lambda(c) | \forall c \in C\}. \tag{4.14}$$

4.2.2 Composing Section

In this study, a song without bass part was used as the input melody. Then a bass part corresponding to the input melody was generated considering the rhythm series and chord progressions of the input melody.

Input rhythm series and chord progressions

The rhythm series on bar t of the input melody imr can be shown as the following equation.

$$imr_t = r_{(t,1)}, \ldots, r_{(t,P)}. \tag{4.15}$$

The rhythm series of the input melody IR can be shown as the following equation with equation 4.15.

$$IR = ir_1, \ldots, ir_T. \tag{4.16}$$

Chord progressions of the input melody IC can be shown as the following equation with equation 4.9.

$$IC = ic_1, \ldots, ic_T, \tag{4.17}$$

where T shows the number of bars of the input melody.

Output the rhythm series of bass part

The rhythm series of bass part of the input melody OBR can be generated from the following equation.

$$OBR = obr_1, \ldots, obr_T, \tag{4.18}$$

where, osr_t is the output rhythm series corresponding to ir_t, and can be calculated from the following equation.

$$obr_t = \operatorname*{argmax}_{br \in BR} P(br|mr)P(ir_t|\lambda(mr))_{\mathrm{for}\forall mr \in MR}. \tag{4.19}$$

Like-hoods $P(ir_t|\lambda(mr))$ for each $\lambda(mr) \in \Lambda_r$ are calculated using the Viterbi algorithm, where input $In = mr$. Then the products of $P(br|mr)$ and $P(ir_t|\lambda(mr))$ are calculated. The br that has the maximum product of $P(vr|mr)$ and $P(ir_t|\lambda(mr))$ is used as obr_t, that is the output rhythm series of bass part corresponding to ir_t.

Output the pitch series of bass part

The pitch series of bass part of the input melody OBP can be generated from the following equation.

$$OBP = obp_1, \ldots, obp_T, \tag{4.20}$$

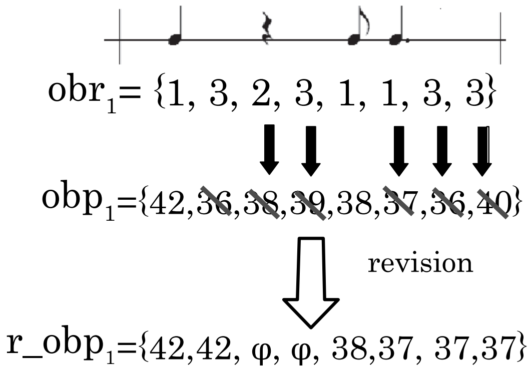


Figure 4.4: An example of pitch revision

where, obp_t is the output rhythm series corresponding to ir_t, and can be calculated from the following equation.

$$obp_t = \operatorname*{argmax}_{bp \in BP} P(bp|c)P(ic_t|\lambda(c))_{\text{for}\forall c \in C}. \tag{4.21}$$

Like-hoods $P(ic_t|\lambda(c))$ for each $\lambda(c) \in \Lambda_c$ are calculated using the Viterbi algorithm, where input $In = c$. Then, the products of $P(bp|c)$ and $P(ic_t|\lambda(c))$ are calculated. The bp that has the maximum product of $P(bp|c)$ and $P(ic_t|\lambda(c))$ is used as obp_t, which is the output pitch series for ic_t.

Revision of the output pitch series From above procedures, output rhythm series and output pitch series of each beat can be obtained from equation 4.18 and equation 4.20, respectively. To allocate the generated pitch to each beat, the output pitch series should be revised considering the state of the output rhythm series on the beat. Specifically, referring to the states of obr_t and obp_t on beat p in the case of that $r_{(t,h)}$ = RS3 in obr_t is observed, $pitch_{(t,p-1)}$ is used as $pitch_{(t,p)}$ (i.e., r_obp$_t$ in Figure 4.4) while increasing p by when $r_{(t,h)} = \mathrm{RS1}, \mathrm{RS2}$ can be observed. Figure 4.4 shows an example of the pitch revision. Then pitch is pronounced only when $r_{(t,p)}$ = RS1 is observed. If $r_{(t,p)}$ = RS3 is observed after $r_{(t,p-1)}$ = RS1, $pitch_{(t,p-1)}$ is pronounced. If $r_{(t,p)}$ = RS3 is observed after $r_{(t,p-1)}$ = RS2, no pitch is pronounced.

4.2.3 Composing demonstration

Using the system proposed in this section, melody accompanied with bass part is composed. Considering musical genre, this study prepared three types of database to learn: five songs from the "RWC Music Database: Jazz Music Database [67]" as the database for jazz music, five songs from dance songs as the database for Dance music, and five rock songs as the database for Rock music. This study set $A_s = 4$ to learn the HMM for rhythm. For learning chord progressions, this study set $A_s = 3$ to learn the HMM for chords.

Composing bass part with various databases

Through the composing demonstration, the various bass part was each composed according to the three different databases though only one common melody was used as the input. Figure 4.5, 4.6, and 4.7 shows the example of the accompanied music by the proposed system. Although the input melodies are the same over the three demonstrations, the bass part composed by the proposed system in this section were different due to the used different HMMs; in the other words, the used databases were different. Additionally, it seems that each composed bass part matches the input melody, and each accompanied music has the nuance of the used database.

Figure 4.5: Output example composed by the proposed system in this section with Jazz database

Figure 4.6: Output example composed by the proposed system in this section with Dance database

Figure 4.7: Output example composed by the proposed system in this section with Rock database

4.3 Accompanying System Using Bayesian Mining of Score Context

Not only by using the musical score directly but also coding the musical score, the time-variation on musical score can be represented. In this section, this study shows the temporal variation on musical score as the given one note shall have the previous and the following note information as "score context." Then the relationships among musical elements of plural musical instruments are studied by using Bayesian network, and Bayesian network can not only study the relation among plural things but also detect the missing value based on the studied relation. This system focuses on musical score information on every beat and the previous and following beats as the temporal variation of music. This proposed system study the structure of accompaniments of stored music, and the accompaniments corresponding the user-input melody can be output based on the studied relations between melody and the accompaniments.

4.3.1 Bayesian network

This study analyzes score information as a contexts, and learn the causal connection among components constructing music such as rhythm and pitch. Thus, it is expected that the knowledge that are needed for composing music would be obtained as a mathematical model by using Bayesian mining of score context.

Knowledge representation of Bayesian network

Bayesian network is a one of the learning and detecting method by modeling the dependence relationship among plural things, which is shown as a connected directed acyclic graph (DAG), and shows the strength of the relation as a conditional probability [68].

As the dependence relationship and stochastic correlation are assumed as one, the joint probability distribution of the N nodes stochastic variables $(X_1, \ldots, X_n)$: P can be shown as the following equation.

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} (X_i | Pa(X_i)), \tag{4.22}$$

where, $Pa(X_i) \in \{X_1 \ldots, X_{n-1}\}$ shows a set of parent nodes of X_i. The equation 4.22

shows that each node X_i depends on only $Pa(X_i)$ and are conditionally independent from the nodes that can be not followed from X_i.

The strength of the dependence relationship among each stochastic variable can be shown as Conditional Probability Table (CPT). CPT is composed of nodes that are related with the dependence relationships and the conditional probability distribution. Conditional probability distribution of X_i that has n possible values($y_1, \ldots, y_n$) can be shown as $p(X_i = y_1|x), \ldots, p(X_i = y_n|x)$ under that the node has parent nodes: $Pa(X_i) = x$, where x shows a vector composed of each value of parent nodes. The topology of Bayesian network and the CPT are given, the dependence relationships among stochastic variables can be modeled. From Bayesian network which is graphical model, the presence or absence of the dependence relationships can be overviewed and intuitively considered.

Bayes' theorem

Bayes' theorem is derived from the following two formulas.

$$P(A, B) = P(A|B)P(B). \tag{4.23}$$

$$P(A, B) = P(B|A)P(A). \tag{4.24}$$

In the two equations the left sides are same, and these can be represented as following equation as connecting the right sides and dividing the both side by $P(A)$.

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}. \tag{4.25}$$

The relational expression 4.25 has been known as the Bayes' theorem. In present, The relational expression provides the fundamentals for stochastic reasoning in AI system. For multiple variables, this equation can be generally represented as follows in P notation.

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}. \tag{4.26}$$

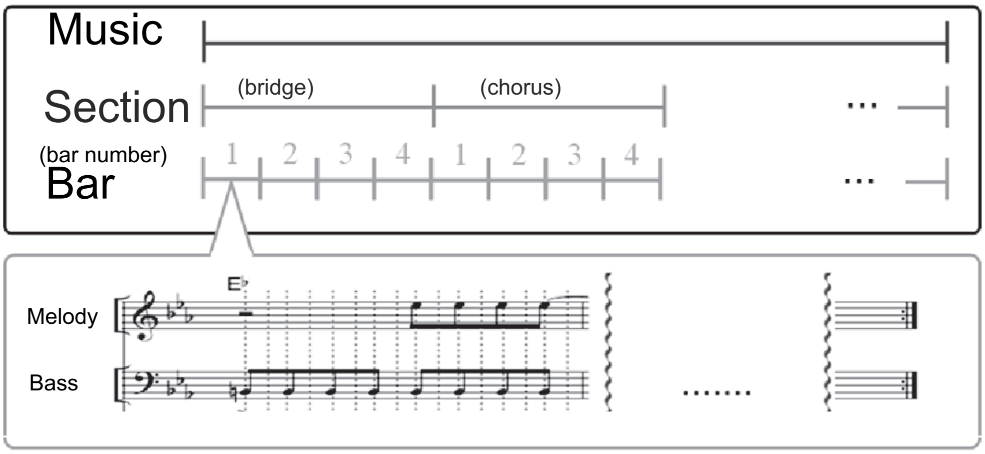


Figure 4.8: Structure of music defined in this paper

4.3.2 Score context

Fig. 4.8 shows the structure of the music defined in this paper. Section is made of the group of the bars, that is defined and classified considering similarities of both the melodies and the chord progression patterns on each bar [69, 70], Section is generally well known as *bridge*, *motif* or *chorus*, and believed as the important concept for music composition. There are no universal definition of section, thus we assume the group made of the multiple of four bars as the section in this paper. Music for learning are all four by four, and composed by melody, bass, drums, chord and information of sections. Sixteenth note and rest are the minimum unit, and are assumed as the one beat.

In this study, the score context is made from the six-tuple set: chord, rhythm of the melody, pitch of the melody, rhythm of the bass, pitch of the bass, and bar counter. Score context on beat p included in bar t, sc_p^t is shown by following equation, and which elements are detailed in following sections.

$$sc_p^t = (bch_p^t, bmr_p^t, bmp_p^t, bbr_p^t, bbp_p^t, bbc_p^t). \tag{4.27}$$

Modeling Chord

Chord on beat p included in bar t, bch_p^t is shown by chord name. In this study, the chord name is described by root of the chord and sign that discriminates the tonality of the chord: major or miner [22]. The pitch of the melody and bass are both based on the chord.

Table 4.3: The reference tone of each chord

Chord	C	C ♯	D	D ♯	E	F	F ♯	G	G ♯	A	A ♯	B
The reference tone	60	61	62	63	64	65	66	67	68	69	70	71

Modeling melody and bass

Melody and bass are each modeled with dividing into two parts, rhythm and pitch [71].

Modeling rhythm The rhythm of the melody and the bass on beat p included in bar t are each shown as bmr_p^t and bbr_p^t, respectively. Rhythm is composed of notes and rests, and known as a series having the following three states: RS1, RS2, and RS3 as noted earlier in section 4.2.1.

Modeling pitch Pitch of the melody and the bass are described by numbers 0 to 127, which conform to the note number of Musical Instrument Digital Interface (MIDI). If the rhythm state in the beat is RS2, then the pitch of the melody and the bass are ϕ. The pitch of the melody and the bass on beat p included in bar t are each shown as bmp_p^t and bbp_p^t, respectively. bmp_p^t and bbp_p^t are shown by relative pitch to bch_p^t. Then, relative pitch is calculated by the difference between the pith and the reference tone of the corresponding chord in the beat (see Tab. 4.3).

Bar counter

The index of the bar in a section is assumed as the bar counter, and the bar counter on beat p included in bar t is shown as bbc_p^t. When it transitions to the following section, bar counter is reset. Because bar counter is decided with respect to bar, all bbc_p^t in a bar have same number.

Fig. 4.9 shows an example of the score context in a bar. When whole note rest is observed on bass or drums in a bar, the score context in the bar is removed from learning.

p	1	2	3	4	5	6	7	8	9	...
bch_p^t	D maj	D maj	D maj	D maj	D maj	D maj	D maj	D maj	D maj	...
bmr_p^t	RS1	RS3	RS3	RS3	RS1	RS3	RS3	RS3	RS2	...
bmp_p^t	11	11	11	11	11	11	11	11	φ	...
bbr_p^t	RS1	RS3	RS1	RS3	RS1	RS3	RS1	RS3	RS1	...
bbp_p^t	81	00	10	00	12	00	10	02	10	...
bbc_p^t	1	1	1	1	1	1	1	1	1	...

Figure 4.9: Example of score context

4.3.3 Accompany model

Storing user's target songs into the learning database, the proposed system composes the accompany models for rhythm and pitch of bass. The accompany models are each constructed by the parameters calculated from score context using Bayesian network.

Accompany model for rhythm of bass

Accompany model for rhythm of bass is composed by constructing the correlation between the rhythm of bass and the following three parameters.

mr : the rhythm of the melody
$partition$: the number of the beat in the bar
bc : the bar counter

Accompany model for pitch of bass

Accompany model for pitch of bass is composed by constructing the correlation between the pitch of bass and following seven parameters.

$length$: the length referring rhythm of the bass
$cadence$: the relative difference between the corresponding chord route and the key of the music

Table 4.4: Adjectives used in the subjective evaluation experiments

Evaluation on a one-to-five scale
Like - Dislike
Natural - Unnatural
Creative - Uncreative
Comfortable - Uncomfortable

pre_cadence : the cadence of the previous note

next_cadence : the cadence of the next note

chord_key : distinction between major and miner for the key of the chord

partition : the number of the beat in the bar

bc : the bar counter

When the generated rhythm of bass is RS1, the pitch of bass on the beat is generated.

4.3.4 Composing demonstration

The availability of the proposed system is verified through the composing demonstration Three types of databases are prepared, which each learned five Jazz, Dance, and Rock music. In composing demonstration, three different accompaniments with respect to each database are generated with common input information: melody, chords, and the information of the section. Fig. 4.10, 4.11, and 4.12 each shows the example of accompaniment generated by the proposed system with each database, respectively. Despite only one input information is used, it is confirmed that three generated accompaniments are different by the used learning database, and it seems that they have each nuance of the genre of learned database.

4.4 Comparative experiments for impressions between the two proposal systems

Through the subjective evaluation experiments, the impressions of the composed music by the system proposed in section 4.2 (named *HMM* system) and the one by the system proposed in section 4.3 (named *BMSC* system) are compared. In this experiment, twenty five subjects (twelve subjects have some musical experiences) are

Figure 4.10: Output example composed by the BMSC system using Jazz database

Figure 4.11: Output example composed by BMSC system using Dance database

Figure 4.12: Output example composed by BMSC system using Rock database

asked to listen to each composed music by *HMM* and *BMSC*, and evaluate each music in one to five scales with the four adjectives listed in Table 4.4.

Table 4.5, 4.6, and 4.7 each lists the average of the subjective evaluation of the composed music by the BMSC and HMM system using each database for each adjective, respectively. For "Like – Dislike," "Natural – Unnatural," and "Comfortable – Uncomfortable" in Table 4.5 and 4.7, and for "Like – Dislike," "Creative – Uncreative," and "Comfortable – Uncomfortable" in Table 4.6, it was confirmed that the evaluation of the music composed by the BMSC system were the significantly better than the one by HMM through the significance test using Scheffe Paired Comparison (Significance level is 1%).

From these results, it was comprehensively suggested that the BMSC system composes more musical and preferred accompanied music than the HMM system.

Table 4.5: The result of the comparative subjective evaluation experiment for the music using Jazz database

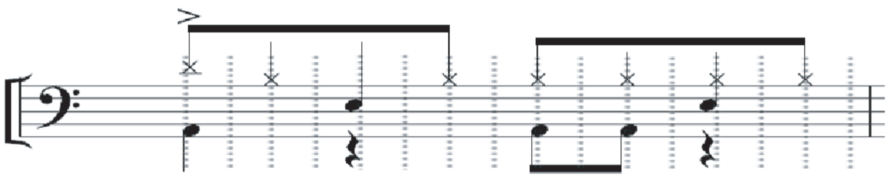
	HMM	BMSC	Significance
Like - Dislike	2.0	**3.5**	$\surd$
Natural - Unnatural	1.4	**3.4**	$\surd$
Creative - Uncreative	3.5	3.2	
Comfortable - Uncomfortable	1.7	**3.5**	$\surd$

Table 4.6: The result of the comparative subjective evaluation experiment for the music using Dance database

	HMM	BMSC	Significance
Like - Dislike	2.8	**3.6**	$\surd$
Natural - Unnatural	2.5	2.6	
Creative - Uncreative	2.7	**4.1**	$\surd$
Comfortable - Uncomfortable	2.3	**3.3**	$\surd$

Table 4.7: The result of the comparative subjective evaluation experiment for the music using Rock database

	HMM	BMSC	Significance
Like - Dislike	2.4	**3.28**	$\surd$
Natural - Unnatural	1.72	**3.32**	$\surd$
Creative - Uncreative	3.4	2.96	
Comfortable - Uncomfortable	2.16	**3.16**	$\surd$



Crash-cymbal	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ride-cymbal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Open-hihat	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Closed-hihat	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
Tom	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Floor-tom	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Snare	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Bass-drum	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
p	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
dr_p^t	81	00	10	00	12	00	10	00	11	00	11	00	12	00	10	00

Figure 4.13: Example of drums pattern and the corresponding drums state

4.5 Additional Accompany Part

In the previous section, the BMSC system showed the higher performance in accompaniment of bass-part for several adjectives. Thus this study expands the BMSC system to deal with drums-part in this section, drums-part is also important for music to make the mood of music.

4.5.1 Modeling Drums

Drums on beat p included in bar t, dr_p^t is described by the combination of the pronounce states of the percussion instruments that compose drums. In this paper, eight percussion instruments are used as drums: Crash-cymbal, Ride-cymbal, Open-hi hat, Closed-hi hat, Tom, Floor-tom, Snare, and Bass-drum. Percussion instruments do not have any idea of length, then each instrument has only two states, RS1 and RS2 (reference section 4.3.2). Fig. 4.13 shows an example of drums score and the corresponding drums state dr_p^t.

4.5.2 Accompany model for drums

Accompany model for drums is composed by constructing the correlation between the drums in the learning database and the following four parameters.

mr : the rhythm of the melody

$$\begin{aligned} br &: \text{ the rhythm of the bass} \\ partition &: \text{ the number of the beat in the bar} \\ bc &: \text{ the bar counter} \end{aligned}$$

where, the **composed** rhythm of bass is referred in the composing phase.

Fig. 4.14, 4.15, and 4.16 each shows the drums added the output example composed by the BMSC system; melody and bass are same as the one shown in Fig. 4.10, 4.11, and 4.12. The each drums part is different due to the each used learning database, and these seems to have the nuance of the genre of the learned songs in each database and well collaborate with each bass and melody.

4.6 Evaluation experiment for whether the composed music by the proposed system has the nuance of the learned songs

Through the subjective evaluation experiment, it will be verified whether the composed music by the BMSC system, which includes drums-part, has the nuance of the learned songs. In this experiment, twenty five subjects (twelve subjects have some musical experiences) are asked to listen to and evaluate the composed music by the BMSC system. Aside from the composed music by the BMSC system, music examples of each genre are prepared for a purpose of reference, which were not contained in the learning music databases. And this study then did not show the genre names of the music examples, "Jazz," "Rock," and "Dance," but shows each "Genre A," "Genre B," and "Genre C." Subjects choose the most appropriate genre for the each composed music by the BMSC system from four options: "Genre A," "Genre B," "Genre C," and "Other." Table 4.8 shows the result of the experiment. And the kappa coefficients is calculated from the results listed in Table 4.8, and the coincidence between the genres of the learning music database and the subjects' answer are evaluated. Generally, kappa coefficient $\kappa \geq 0.60$ means significantly high concordance rate, and $\kappa = 0.65$ was confirmed for the results. Therefore, it is suggested that the genre of the composed music substantially conform to the each genre of the learning music database. That is to say, the BMSC system with drums enables user to compose the accompanied music that have the nuance of user's target.

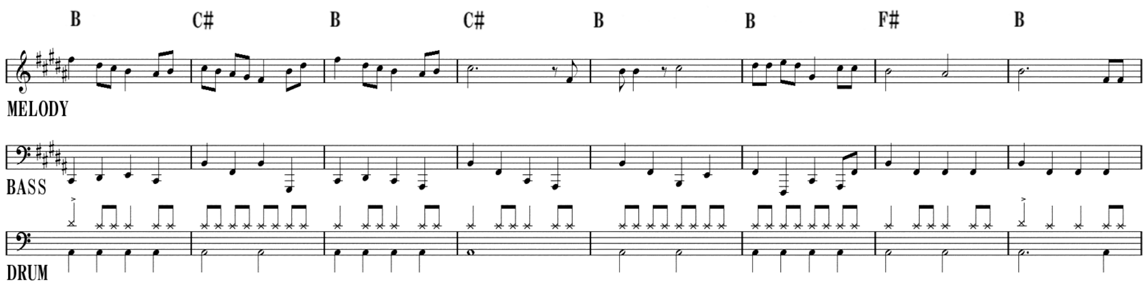


Figure 4.14: Output example composed by the BMSC system using Jazz database with Drums

Figure 4.15: Output example composed by BMSC system using Dance database with Drums

Figure 4.16: Output example composed by BMSC system using Rock database with Drums

Table 4.8: The result of the concordance rate

		Subjects' answers				
		A	B	C	Other	Concordance rate
The genres of the used database for learning	A (Jazz)	**16**	1	5	3	0.64
	B (Rock)	0	**22**	2	1	0.88
	C (Dance)	0	5	**19**	1	0.76

4.7 Conclusion

In this chapter, this study proposed two types of accompaniment system for composing support: HMM system using HMMs and the BMSC system using Bayesian mining of score context. The musical score were modeled to the form suitable for each system, then the time-variation, which is important factor of music on human factor, was taken in consideration.

In HMM system, the pitch variation and the rhythm variation on the given bar were modeled and studied by using HMM. In the composing phase, the accompaniments that has the highest likelihood for input-melody were output. Meanwhile in BMSC system, the pitch variation and the rhythm variation were modeled as context to represent the time-variation on musical score, and the relation between accompaniments and the melody were studied by using Bayesian network. In the composing phase, the accompaniments that are appropriate for the user-input melody were detected and output by Bayesian network.

Three types of learning music database were prepared for composing experiment, in which Jazz, Dance and Rock music were each stored. For each system, despite only one common input information was used, it was confirmed that three composed accompaniments corresponding to the each database are different. And it was suggested that composed accompaniment had each nuance of the genre of the learned songs in each database.

The impressions of the composed music were evaluated through the subjective evaluation experiment, and the proposed two systems were compared. The results of comparative experiment showed that the music composed by the BMSC system have better impressions than the one by the HMM system. Since this study make the BMSC system cover the drums-part as an additional accompaniment part. And through the subjective evaluation, it was confirmed that the music composed by the BMSC system, which includes drums-part, had the nuance that the user's target.

The future directions of this study are as follows.

- Advancement of the usability
  The proposed system needs the melody, chord, and the information of section, the system will automatically obtain the information of the chord and section with pattern detection [72, 73], then the system will need only melody.

- Dealing with other musical instruments
  More instruments will be covered, e.g., strings, piano and guitars.

- Considering the fingering
  The composed accompaniment is now for DAW, so the composed music is not appropriate for playing by the real musical instruments. Considering the fingering, it is expected that the composed accompaniment will be playable by the real musical instruments.

Chapter 5

Conclusion and Future Works

In this paper, three subjects were studied: music effects on physiology, the relationships between acoustic features and emotional/affective evaluation of the music, and automated accompaniment system for composing support.

In Chapter 2, this study focused on bio-signals as physiological index, and attempted to verify the importance of the sound variation through the analysis of brain waves while listening to harmony. Harmony can show the important features of music: plural sounds are produced at the same time, and the plural sounds variates with time axis. Then there was the problem to solve "how musical effects on brain waves should be measured and analyzed," because music and brain waves are both time-series data and the effects of music on brain waves can be not determined in general measuring and analyzing way. Thus this study proposed the original presentation sounds and analytical object on brain waves, and overcame the problem. As the result of experiments, this study confirmed the significance difference between the brain waves while listening to a single chord and harmony even though the subject listened to the same chord, and that the relaxation effects on brain waves were different depending on the listening harmony. To sum it up, the physiological importance of temporal variation of music was experimentally confirmed. From the result, this study especially focused and considered the temporal variation of music in acoustic analysis and modeling of musical score in the following chapters.

Next, in Chapter 3, reflecting the fact obtained in Chapter 2: the importance of temporal variation of music on physiological index, the relationships between acoustic features and emotional/affective evaluation of music were studied. First, the relationships between the acoustic features and the emotional/affective evaluation of music were studied. Then, the musical fluctuation features were used as the acoustic features to cover the temporal variation of music because it could show the variation of pitch, volume and rhythm as fluctuation. Using feature selection and some discriminant analysis, the musical fluctuation features and the emotional/affective evaluation of music were related; this study compared four kinds of discriminate method and confirmed that the canonical discriminant analysis showed the best performance on both the average accuracy rate and the average error. Then this study accepted the discriminate spaces of canonical discriminant analysis for emotional/affective evaluations of music, which were constructed by the contribute musical fluctuation features selected by using features selection, as the appropriate relationship diagrams of the acoustic features and the emotional/affective evaluations of music, and the effective features for discriminant of emotional/affective evaluations were determined from the diagram. Next, the discriminate spaces were assumed as the Acoustic-Emotion model, the song selection system complying with emotional/effective requests was developed. In this system, the emotional/affective evaluations of a song was labeled from the acoustic features of the song using the Acoustic-Emotion model. Moreover, this study constructed the emotional/affective song selection algorithm to cover the tolerance of human instinct. And the system selected the songs complying with the user's emotional/affective requests which were expressed by adjectives and the degrees. This system showed the high practicality and validity of the selected songs through the subjective evaluation experiments. And, this study also focused on the musical performance and quantified emotional/affective evaluations of musical performances with acoustic features, which were played with the violin. Then the acoustic fluctuation features that showed the slight sound variation of musical performances and harmonic component kurtosis properties that show the harmonic contents and other regular acoustic features were used in acoustic analysis. Using features selection, the contribute features were selected and used for multiple linear regression analysis. As the result, the high coefficient of regression was confirmed and the quantification of musical performance using acoustic features was suggested to be usable. And from the regression, the relationships between acoustic features and emotional/affective evaluations of musical performances were determined.

In Chapter 4, musical scores were modeled for data mining and information processing, then the temporal variation on musical score were taken in the consideration reflecting on the result of study in Chapter 2. This study developed and compared two types of the automated accompaniment composing system in which each Hidden Markov Model (HMM) and Bayesian Mining of Score context (BMSC) was applied as the mechanism of learning and expressing. In both systems bass part that was important for making a mood of music was adopted as the accompaniment. In demonstrations of both systems, the nuance of the each genre of learned songs which was in database seemed to be realized. The accompanied music composed by the each system was compared, the BMSC accompany support system showed higher performance on the composed music than the HMM accompany support system. Therefore drums part was additionally dealt in the BMSC system, the music accompanied by BMSC system was composed of melody, bass and drums. Through the subjective evaluation experiment, the emotional/affective evaluations of the music accompanied by the BMSC system were verified. Then the nuance that listeners perceived from the music accompanied by the BMSC system coincided with the nuance of the learned songs; the genres that listeners felt and the one of learned songs were coincided at a high level. This result suggested that the proposed system could help user to compose music that had the intended nuance, and the utility of the system was confirmed

With this study, at least the following applications are anticipated.

- The automated background music selection system
  Using the song selection system proposed in section 3.3 and the estimation system for dialogue mood [74], the suitable songs for dialogue mood would be able to be automatically selected; this system would be useful at many scene, for example, restaurant, cafeteria, and in driving. And studying the relationships between emotion/affection and brain waves, brain waves would be able to be used as the retrieval key for the song selection system. Then user would be able to select the songs that user would like to listen in the unconsciousness.

- The accompaniment system complying with emotional/affective requests
  Using the song selection system proposed in section 3.3 and the accompanying system proposed in Chapter 4, the accompaniment complying with emotional/affective requests would be able to be composed. As user would request the emotional/affective nuances, the song selection system automatically would select the suitable songs for the requests and the songs would be stored in the learning database for the accompanying system. Then, the accompanied song that has the requested emotional/affective nuance would be automatically composed.

Overall, this study revealed the influence of temporal variation of music on brain waves, and developed the systems to enjoy more pleasures of music based on the facts. The systems were constructed by combination of computer science and music, that is to say, combination of the civilization and culture. In the foreseeable future, it is expected that man-machine interaction and emotional/affective interaction will be more focused and active. I believe that the proposed systems in this paper will then support man-machine and man-man interaction as the emotional/affective and artistic interaction tools.

Appendix

Fast Fourier Transform

Fast Fourier Transform (FFT) is discrete Fourier transform based on the fast-acting algorithm on computer, and it is used for audio analysis and bio-signal analysis in this study. Discrete Fourier transform, hann and hamming windows, Cooley-Tukey algorithm that is generally used in FFT, and Short-time FFT will be explained in below.

Discrete Fourier transform

The discrete Fourier transform samples continuous period signals, and translates them to discrete period signals on Fourier series expansion [75]. The function $x^*(t)$ which cycle is NT can be shown as the following equation.

$$x^*(t) = \sum_{n=0}^{N-1} x(nT)\delta(t - nT), \tag{1}$$

where, N, T, $\delta(t)$ each shows integral, sampling period, and delta function, respectively.

Then, $x^*(t)$ is the function of discrete period signals (x_0, x_1, $\ldots$, x_{N-1}). And to one period of this function, the complex Fourier series expansion is conducted. With small positive number ϵ, the complex Fourier coefficient can be shown as the following equation.

$$\begin{aligned} c_k &= \frac{1}{NT}\int_{-\epsilon}^{NT-\epsilon} x^*(t)e^{-j2\pi k f_0 t}dt, \\ &= \frac{1}{N}\sum_{n=0}^{N-1} x(nT)e^{-j\frac{2\pi kn}{N}}, \end{aligned} \tag{2}$$

incidentally, the area of $\delta(t)$ is T.

And the c_{k+N} can be calculated as follows.

$$\begin{aligned} c_{k+N} &= \frac{1}{N}\sum_{n=0}^{N-1} x(nT)e^{-j\frac{2\pi kn}{N}}e^{-j2\pi n}. \\ &= c_k. \end{aligned} \tag{3}$$

from this equation, it is confirmed that the complex Fourier coefficient c_k also has the N cycle periodicity. Meanwhile, discrete period signals can be shown as the following equation with c_k.

$$x(nT) = \sum_{k=0}^{N-1} c_k e^{j\frac{2\pi kn}{N}}. \tag{4}$$

From these, the definitional identity of discrete Fourier transform and the inverse transform can be each derived as follows.

$$X(k) = \sum_{n=0}^{N-1} x_n e^{-j\frac{2\pi kn}{N}}, \tag{5}$$

$$x(n) = \frac{1}{N}\sum_{k=0}^{N-1} X_k e^{j\frac{2\pi kn}{N}}. \tag{6}$$

Definition and characteristics

Variable W corresponding the N-divided point of unit circle is defined as follows.

$$\begin{aligned} W &= e^{-j\frac{2\pi}{N}} \\ &= cos\left(\frac{2\pi}{N}\right) - j\dot{s}in\left(\frac{2\pi}{N}\right). \end{aligned} \tag{7}$$

Using W, the definitional identity of discrete Fourier transform and the inverse transform can be each rewritten as follows.

$$
\begin{aligned}
X(k) &= \sum_{n=0}^{N-1} x_n W^{kn}, & (8)\\
x(n) &= \frac{1}{N}\sum_{k=0}^{N-1} X_k W^{-kn}, & (9)
\end{aligned}
$$

where, W is rotator and the following equation can be realized.

$$
\begin{aligned}
W^N &= e^{-j2\pi} = 1, & (10)\\
W^{k+N} &= W^{k+2N} = \cdots = W^k. & (11)
\end{aligned}
$$

Discrete Fourier transform defined above can be shown as the following matrix.

$$
\begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_{N-1} \end{bmatrix}
=
\begin{bmatrix}
W^0 & W^0 & W^0 & \cdots & W^0 \\
W^0 & W^1 & W^2 & \cdots & W^{N-1} \\
W^0 & W^2 & W^4 & \cdots & W^{2(N-1)} \\
W^0 & W^3 & W^6 & \cdots & W^{3(N-1)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
W^0 & W^{N-1} & W^{2(N-1)} & \cdots & W^{(N-1)^2}
\end{bmatrix}
\begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{N-1} \end{bmatrix}, \tag{12}
$$

where, as representing signals $x_n (n = 0, 1, \ldots, N-1)$, spectrum $X_n (n = 0, 1, \ldots, N-1)$ as a vector, and the transform relation equation as matrix F, the foregoing matrix can be shown as follows.

$$
\hat{X} = F\hat{x}. \tag{13}
$$

And, the inverse transform can be shown as follows.

$$\begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{N-1} \end{bmatrix} = \frac{1}{N} \begin{bmatrix} W^0 & W^0 & W^0 & \cdots & W^0 \\ W^0 & W^{-1} & W^{-2} & \cdots & W^{-(N-1)} \\ W^0 & W^{-2} & W^{-4} & \cdots & W^{-2(N-1)} \\ W^0 & W^{-3} & W^{-6} & \cdots & W^{-3(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ W^0 & W^{-(N-1)} & W^{-2(N-1)} & \cdots & W^{-(N-1)^2} \end{bmatrix} \begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_{N-1} \end{bmatrix}. \tag{14}$$

As representing it as vector and matrix, the above matrix can be shown as follows. Where m $\overline{F}$ shows the complex conjugate.

$$\hat{x} = \frac{1}{N}\overline{F}\hat{X}, \tag{15}$$

where, matrix F corresponding discrete Fourier transform is regular matrix and the inverse matrix can be shown as follows.

$$F^{-1} = \frac{1}{N}\overline{F}.$$

As this matrix is premultiplied to discrete Fourier transform, the inverse transform function is derived as follows.

$$F^{-1}\hat{X} = F^{-1}F\hat{x} = \frac{1}{N}\overline{F}\hat{X}. \tag{16}$$

Hann window and hamming window

In this study, hanning and hamming windows were used as convolution filters. Using window function, both ends in window are 0 and only the finite interval remains since it helps numerical analysis.

Hann window

Hanning window is one of the most-often-used window function. The hamming window function $hn(n)$ is as follows.

$$hn(n) = \begin{cases} 0.5 - 0.5cos\left(\frac{2\pi n}{N-1}\right) & (0 \leq n \leq N-1) \\ 0 & (else) \end{cases} \tag{17}$$

Hamming window

Hamming window have higher frequency resolution and narrower dynamic range than hanning window. The hamming window function $hm(n)$ is as follows.

$$hm(n) = \begin{cases} 0.54 - 0.46cos\left(\frac{2\pi n}{N-1}\right) & (0 \leq n \leq N-1) \\ 0 & (else) \end{cases} \tag{18}$$

The spectrum X_k^* that is discrete signals x_n windowing hamming window function can be shown as follows.

$$\begin{aligned} X_k^* &= \sum_{n=0}^{N-1} (x_n \dot{h}m(n)) W^{nk} \\ &= \sum_{n=0}^{N-1} \left(\frac{1}{N} \sum_{m=0}^{N-1} Hm_m W^{-nm} \right) x_n W^{nk} \\ &= \frac{1}{N} \sum_{m=0}^{N-1} Hm_m \sum_{n=0}^{N-1} x_n W^{(k-m)n} \\ &= \frac{1}{N} \sum_{m=0}^{N-1} Hm_m X_{k-m}. \end{aligned} \tag{19}$$

Cooley-Tukey algorithm

Cooley-Tukey algorithm [76] is one of the typical algorithm for fast Fourier transform proposed by J. W. Cooley and J. W. Tukey in 1965. This algorithm focuses on the symmetric property of discrete Fourier transform and reduces calculation amount, and realizes high-speed processing.

The determinant 12 is divided into two matrics considering whether the subscript is even or odd, and the calculation amount halves. An example of calculation is shown as follows, where $N = 4$.

$$\begin{aligned}
\begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ X_3 \end{bmatrix} &= \begin{bmatrix} W^0 & W^0 & W^0 & W^0 \\ W^0 & W^1 & W^2 & W^3 \\ W^0 & W^2 & W^4 & W^6 \\ W^0 & W^3 & W^6 & W^9 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}. \\
&= \begin{bmatrix} W^0 & W^0 & W^0 & W^0 \\ W^0 & W^2 & W^1 & W^3 \\ W^0 & W^4 & W^2 & W^6 \\ W^0 & W^6 & W^3 & W^9 \end{bmatrix} \begin{bmatrix} x_0 \\ x_2 \\ x_1 \\ x_3 \end{bmatrix}. \\
&= \begin{bmatrix} W^0 & W^0 & W^0 W^0 & W^0 W^0 \\ W^0 & W^2 & W^1 W^0 & W^1 W^2 \\ W^0 & W^0 & W^2 W^0 & W^2 W^0 \\ W^0 & W^2 & W^3 W^0 & W^3 W^2 \end{bmatrix} \begin{bmatrix} x_0 \\ x_2 \\ x_1 \\ x_3 \end{bmatrix}. \\
&= \begin{bmatrix} 1 & 0 & W^0 & 0 \\ 0 & 1 & 0 & W^1 \\ 1 & 0 & W^2 & 0 \\ 0 & 1 & 0 & W^3 \end{bmatrix} \begin{bmatrix} W_2^0 & W_2^0 & 0 & 0 \\ W_2^0 & W_2^1 & 0 & 0 \\ 0 & 0 & W_2^0 & W_2^0 \\ 0 & 0 & W_2^0 & W_2^1 \end{bmatrix} \begin{bmatrix} x_0 \\ x_2 \\ x_1 \\ x_3 \end{bmatrix}. \qquad (20)
\end{aligned}$$

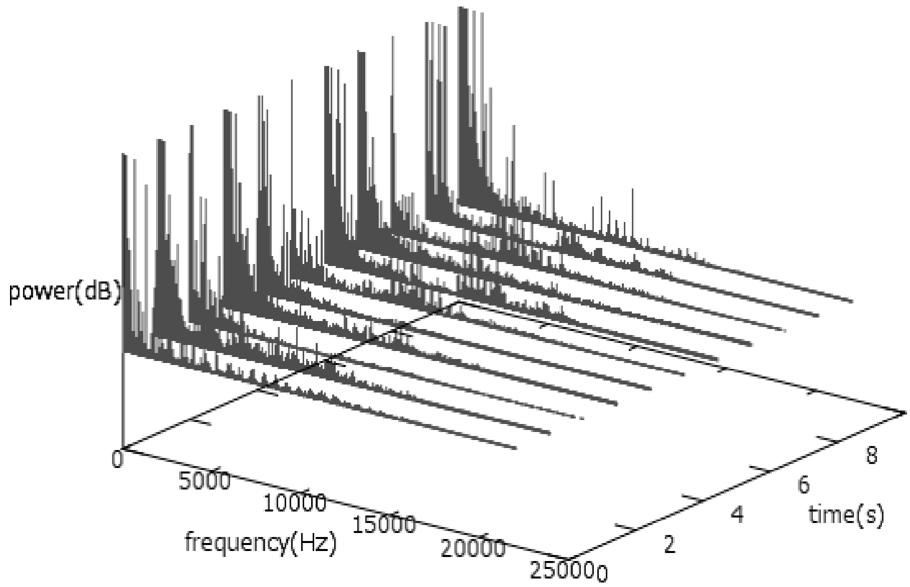


Figure 1: An example of the result of short-time fast Fourier transform

Short-time Fast Fourier Transform

Short-time fast Fourier transform has been often used to determine the frequency variation of signals that changes with time passages such sound and bio-signal. Short-time fast Fourier transform is simply to say, "fast Fourier transform is conducted taking as the window is slid along the time axis." And the Short-time fast Fourier transform (STFT) can be mathematically shown as follows.

$$\mathrm{STFT}\{x(t)\} = X(\tau, \omega) = \int_{-\infty}^{\infty} x(\tau) w(\tau - t) e^{-iwt}, \tag{21}$$

where, $w(t)$ and $x(t)$ are each the window function and the signal to be transformed, respectively.

Using Short-time fast Fourier transform, the power variation on each frequency along with the time axis can be determined. Figure 1 shows an example of the result of short-time fast Fourier transform, where the structure of the power variation on each frequency of a song was analyzed.

Musical Instrument Digital Interface

Musical Instrument Digital Interface (MIDI) has been generally applied in musical composing and playing. The data formatted to MIDI, that is MIDI data, has been dealt with sequence such Digital Audio Workstation (DAW) [57, 58, 59].

The information of MIDI data is not real tone but musical information for musical instruments and audio source: pronounce, pitch, and volume. So, the data size is much less than the digital real audio data recorded by microphone. When human plays the real musical instruments and records the sounds, the sounds have some difference of the used instruments and the player's nuance. However there is no difference between the sounds played with MIDI data, so universalistic sounds data can be prepared.

Presently, MIDI has been used in not only musical composing and playing. It is not too much to say that MIDI enables the public interested in music to enjoy composing and playing with their own computers.

MIDI data format

The all MIDI data are transmitted and received as MIDI massages, and all playing and sounds are controlled by MIDI massages. MIDI massage is composed of 8 bits data that have status bites and data bites. The status bites define "note on," "control change" and "system exclusive," and the data bites specify the contents and value of the defined status.

MIDI massage can be divided into two types according to the status bites: channel massage and system massage. Channel and system massage are MIDI massage specifying the channel and the one for the whole system regardless of channel, respectively. In the following paragraph, the channel voice massage that is the specifically important for playing will be briefly explained.

Channel voice massage

Channel voice massage specifies the necessary information for playing: pronounce, stop, change tone, and change pitch. The note number, which is specified in channel voice massage, shows the pitch of the musical instruments. In the note number, the lowest pitch and the highest pitch are each assigned to note number 0 and 127, respectively; incidentally, the centric C (261.63Hz) is assigned to note number 60. And the velocity, which is also specified in channel voice massage, means the strength of the sound (dynamic mark) which is shown as the number 0 to 127. In channel voice massage, the following information is specified.

- Note off
  "Note off" stops the sound that is ordered by "Note on."
- Note on
  "Note on" makes the system pronounce, and the sound keeps on being pronounced by the "Note off" is given.
- Polyphonic key pressure
  For keyboard instrument, "Polyphonic key pressure" is given and the key pressure is specified when the player presses the playing key harder. This is for only the playing key.
- Control change
  "Control change" is the command to control volume and sound quality.
- Program change
  "Program change" is the command to change tone color, the tone color can be selected from up to 128 kinds.
- Channel pressure
  For keyboard instrument, "Channel pressure" is given and the key pressure is specified when the player presses the pressed key harder. This is different from polyphonic key pressure, and for whole note numbers in the given channel.
- Pitch bend
  "Pitch bend" changes the pitch of the playing sound, the pitch can be minutely specified in 16384 steps.

Acknowledgements

From 2006 to 2009, at dept. of computer science and engineering, Nagoya Institute of Technology, I was affiliated with Itoh Hidenori laboratory as bachelor and master student, and I have been affiliated with Kato Shohei laboratory as doctor student since 2009. I have engaged in a research about "Music effects on psychology and physiology and its applications." For the research works, I have been taught and supported by so many respecters. I would like to show my acknowledgment to them in this section.

First, from the bottom of my heart, I would like to show my acknowledgement to emeritus professor Hidenori Itoh. When I was a high school student, I watched his research about relation between music and biosignals on TV program and was moved so deeply. That was my turning point to enter the institute and aim to research. He has given me the dreams and so many useful advises. I will be forever in his debt.

And I would like to show my best acknowledgement to associate professor Shohei Kato. He has told me a lot of things about research: motivation, research concept, detail techniques, how to write papers, and also about research life. Due to his support and teach, I have had a grateful and pleasant days. He is forever my example of researcher. I have definitely believed that my works could be not realized without him. From the bottom of my heart, I appreciate him deeply.

To respectful professor Tadashi Kitamura and Ichi Takumi, I would like show my deep acknowledgement. Due to their kind and critical comments and schooling for me improved the quality and significance of this thesis. I appriciate them for their affectionate lectures.

Additionally, I would like to appreciate my dear friend: Jangsik Cho, and dear my junior fellows: Yuya Ito, Ryogo Okamura, Takuya Naganawa, Atsushi Harada, Keigo

Tada, Kaoru Toyoda, Yasuhiro Suzuki, Yoshihiro Miyakoshi, Akinori Wakabayashi, and all of members in Shohei Kato laboratory.

And to my long-time friends: Eri Nishio and Kaori Kondo, and Kazuo Kato, Kenji Masegi and Tohru Ujihara, and all of my friends in Folk Song Society at Nagoya Institute of Technology; especially Nobuyuki Maezawa, Akihiro Asano, Terumasa Yuno, Yasuhisa Tsuchida, and the other members of *Heat-Pump* and its fans, and other all friends in varied fields, I would like to show my very special thanks for our friendships. Their trusts for me and the spent time with them have encouraged me, always.

Finally, I would like to thank my family: my father Nobuyoshi, my respectful mother Junko, my young sisters Saemi and Kaori, and my wise&pretty brother Hokuto always in my heart, and my grandfathers, grandmothers, uncles, and aunts for their supporting my research, life, and even heart. In particular, my mother Junko has always encouraged me kindly and critically. Everyday, every night, she has talked with me to cure my heart. Without her advises, I would not enter graduated school and write this thesis. I would like to appreciate her with my best love for her selfless love for me.

Bibliograghy

[1] N. A. Stanton, P. M. Salmon, G. H. Walker, C. Baber, and D. P. Jenlins, *Human Factors Methods: A Practical Guide for Engineering And Design.* Ashgate Publishing, 2005.

[2] Y. Shimizu, T. Sadoyama, M. Kamijo, S. Hosoya, and M. Hashimoto, "On-demand production system of apparel on the basis of kansei engineering," *International Journal of Clothing Science and Technology*, vol. 16, no. 1/2, pp. 32–42, 2004.

[3] S. K. Aung and M. H. Lee, "Music, sounds, medicine, and meditation: An integrative approach to the healing arts," *Alternative & Complementary Therapies*, vol. 10, no. 5, 2004.

[4] J. Alvin, *Music therapy for the autistic child.* Oxford University Press, 1978.

[5] C. E. Guzzetta, "Effects of relaxation and music therapy on patients in a coronary care unit with presumptive acute myocardial infarction," *Heart & Lung*, vol. 18, no. 6, 1989.

[6] S. Schneider, P. Schönle, E. Altenmüller, and T. Münte, "Using musical instruments to improve motor skill recovery following a stroke," *Neurol*, vol. 254, no. 10, 2007.

[7] G. Bruner, "Music, mood, and marketing," *Journal of Marketing*, vol. 54, pp. 94–104, 1990.

[8] MIDI manifactures association.
`http://www.midi.org/`.

[9] E. Hoshino, "What is music therapy? : The use of various powers inherent in the roots of music," *The Journal of the Acoustical Society of Japan*, vol. 64, pp. 468–474, 2008.

[10] M. Albert, R. Sparks, and M. Helm, "Melodic intonation therapy for aphasics," *Archives of Neurology*, vol. 29, pp. 130–131, 1973.

[11] L. B. Meyer, *Emotion and Meaning in music.* University of chicago Press, 1956.

[12] M. Suzuki, M. Kanamori, S. Nagasawa, T. Isowa, and T. Saruhara, "Music therapy-induced changes in behavioral evaluations, and saliva chromogranin a and immunoglobulin a concentrations in elderly patients with senile dementia," *Geriatrics & gerontology international*, vol. 7, pp. 61–71, 2007.

[13] G. Takeda, Y. Hirose, and S. Yamamoto, "A study of time varying eeg responses in the processes of chord cognition," *Journal of Acoustical Society of Japan*, vol. 2003, pp. 861–862, 2003. (in Japanese).

[14] M. Nuki, K. Nagata, and H. Kawakami, "The relations among eeg, mood, preference, personality, and spectrum power analysis in listening to healing music," *Information Processing Society of Japan (Special Interest Groups of Music) Technical Report*, vol. 2004, pp. 35–40, 2004. (in Japanese).

[15] Y. Okumura, K. Igawahara, and K. Matunami, "Emotional change in elderly persons induced by old japanese popular songs -a study with eeg and questionnaire method-," *Japanese Journal of Music Therapy*, vol. 5, pp. 177–187, 2005. (in Japanese).

[16] M. H. Thaut, *Rhythm, Music, And The Brain: Scientific Foundations and clinical applications.* Kyodoisho DTP station, 2006.

[17] S. Yamamoto, Y. Hirose, and Y. Sagawa, "Eeg power variation while listening to chords of synthetic piano tones," *Journal of Japan Society of Kansei Engineering*, vol. 6, pp. 51–60, 2005. (in Japanese).

[18] S. Chikahisa, M. Mitani, and H. Araki, "Effect of music constitution on alpha-wave changes," *Journal of physiological anthropology and applied human science*, vol. 21, p. 137, 2002.

[19] U. Yoshihiro, K. Hitoshi, I. Masaaki, Y. Toshio, K. Yuri, and S. Yoshikatsu, "Sound effect of a music box on eeg," *Journal of International Society of Life Information Science*, vol. 18, pp. 269–275, 2000.

[20] M. Nakamura, Q. Chen, T. Sugi, A. Ikeda, and H. Shibasaki, "Technical quality evaluation of eeg recording based on electroencephalographers' knowledge," *Medical engineering & Physics*, vol. 27, no. 1, pp. 93–100, 2005.

[21] Electro-Cap International Inc., “Electro-cap.” `http://www.electro-cap.com/`.

[22] G. Ewer, *Essential Chord Progressions*. Pantomime Music Publications, 2006.

[23] T. C. Handy, *Event-related potentials: a methods handbook*. The MIT Press, 2004.

[24] E. W. Sellers and E. Donchin, “A p300-based brain computer interface: Initial tests by als patients,” *Clinical Neurophysiology*, vol. 117, pp. 538–548, 2006.

[25] T. Kitahara, Y. Tsuchihashi, and H. Katayose, “Music genre classification and similarity calculation using bass-line features,” in *Proceesings of the 10th IEEE International Symposium on Multimedia, Workshop on Multimedia Audio and Speech Processing*, pp. 574–579, 2008.

[26] L. Lu, D. Liu, and H.-J. Zhang, “Automatic mood detection and tracking of music audio signals,” *IEEE Trance on Audio, SPEECH, And Lang Processing*, vol. 14, 2006.

[27] J. Marques and P. J. Moreno, “A study of musical instrument classification using gaussian mixture models and support vector machines,” tech. rep., Compaq Corporation, Cambridge Research laboratory, 1999.

[28] J. Skowronek, M. F. Mckinney, and S. V. D. Par, “Groundtruth for automatic music mood classification,” in *Proceedings of the 7th International Conference on Music Information Retrieval, Victoria BC*, 2006.

[29] P. Syank and A. Wieczorkowska, “Some issues on detecting emotions in music,” *Lecture Notes in Computer Science*, vol. 5097, pp. 164–174, 2008.

[30] A. Ghias, J. Logan, D. Chamberlin, and B. C. Smith, “Query by humming: musical information retrieval in an audio database,” *In ACM Multimedia*, pp. 231–236, 1995.

[31] N. Hu and R. B. Dannenberg, “A comparison of melodic database retrieval techniques using sung querie,” *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, 2002.

[32] C. Xu, N. Maddage, and X. Shao, “Automatic music classification and summarization,” *Speech and Audio Processing, IEEE Transactions on*, vol. 13, no. 3, pp. 441–450, 2005.

[33] M. E. Donald. and P. Joseph, *Music Therapy In Principle And Practice.* Charles C Thomas Pub Ltd, 2005.

[34] K. Momose, Y. Fujisawa, and A. Uchiyama, "Effects of music on physiological measurements and mood states, and their correlation with impression of the music," *Journal of International Society of Life Information Science*, vol. 22, no. 2, pp. 545–551, 2004.

[35] L. Lu, D. Liu, and H.-J. Zhang, "Automatic mood detection and tracking of music audio signals," *IEEE Transactions on Audio, Speech, and Langage Processing*, vol. 14, no. 1, pp. 5–18, 2006.

[36] Y. Feng, Y. Zhuang, and P. Yunhe, "Popular music retrieval by detecting mood," in *Proceedings of the 26th ACM SIGIR conference on Research and development in informaion retrieval*, pp. 375–376, 2003.

[37] S. Feldstein, *Practical Theory Complete: A Self-Instruction Music Theory Course.* Alfred Publishing Company, 1982.

[38] J. Foote and S. Uchibayashi, "The beat spectrums: A new approach to rhythm analysis," *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 881–884, 2001.

[39] P. N. Juslin and J. A. Sloboda, *Music and Emotion: Theory and Research.* Oxford University Press, 2001.

[40] C. E. Osgood, W. S. May, and M. S. Miron, *Cross-Cultural Universals of Affective Meaning.* University of Illinois Press, 1975.

[41] Y. Freund and R. E, "Schapire a decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[42] N. Yamaguchi and N. Ishii, "Combining classifiers in error correcting output coding," *Journal of The Institute of Electronics, Information and Communication Engineers*, vol. J86-D-II, no. 2, pp. 324–332, 2003. (In Japanese).

[43] H. Liu and H. Motoda, *FEATURE EXTRACTION CONSTRUCTION and SELECTION: A data Mining Perspective.* KLUWER ACADEMIC PUBLISHERS, 1998.

[44] H. F. Kaiser, "The varimax criterion for analytic rotation in factor analysis," *PSYCHOMETRIKA*, vol. 23, no. 3, pp. 187–200, 1958.

[45] K. Toyoda, Y. Miyakoshi, R. Yamanishi, and S. Kato, "Dialogue mood estimation focusing intervals of utterance state," *Transactions of the Japanese Society for Artificial Intelligence*, vol. 27, no. 2, 2012. (in Japanese) (to appear).

[46] C. M. sHutchins, "The future of violin research," *Catgut Acoustical Society Journal*, vol. 2, no. 1 (Series II), p. 3, 1992.

[47] S. M. S. McAdams, J. W. Beauchamp, "Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters," *Journal of Acoustic Society of America*, vol. 105, no. 2, pp. 882–897, 2006.

[48] C. Flesch, *Art of Violin Playing: Book One.* Carl Fischer Music Dist, 2000.

[49] C. R. Jensen, "A theoretical work of late seventeenth-century muscovy: Nikolai diletskii's "grammatika" and the earliest circle of fifths," *Journal of the American Musicological Society*, vol. 45, no. 2 (Summer), pp. 305–331, 1992.

[50] P. van der Merwe, *Origins of the Popular Style: The Antecedents of Twentieth-Century Popular Music.* Oxford University Press, 1992.

[51] M. Balaban, K. Ebciogle, and O. Laske, *Understanding Music with AI: Perspectives on Music Cognition.* AAAI Press, 1992.

[52] D. Cope, *Computers and musical style (Computer Music and Digital Audio Series).* A-R Editions, 1991.

[53] D. Hörnel, "A multi-scale neural-network model for learning and reproducing chorale variations," *Computing in Musicology*, vol. 11, pp. 141–158, 1998.

[54] N. Nettheim, "Melodic pattern-detection using musearch in schubert's die schöe mülerin," *Computing in Musicology*, vol. 11, pp. 159–168, 1998.

[55] M. Alfonseca, M. Cebrian, and A. Ortega, "A simple genetic algorithm for music generation by means of algorithmic information theory," *2007 IEEE Congress on Evolutionary Computation*, pp. 3035–3042, 2007.

[56] S. Fukayama, K. Nakatsuma, S. Sako, Y. ichiro Yonebayashi, T. H. Kim, Q. S. Wei, T. Nakano, T. Nishimoto, and S. Sagayama, "Orpheus: Automatic composition system considering prosody of japanese lyrics," *Entertainment Computing - ICEC 2009*, pp. 309–310, 2009.

[57] Steinberg Media Technologies GmbH, "Cubase." `http://www.steinberg.net/en/`.

[58] Cakewalk, "Sonar." `http://www.cakewalk.com`.

[59] Ableton, "Live." `http://www.ableton.com`.

[60] S. Kuroiwa, K. Kobayashi, T. Satoru, and F. Ren, "Packet loss concealment algorithm using speech recognition and synthesis," *The Journal of the Acoustical Society of Japan*, vol. 62, no. 1, pp. 3–11, 2005. (in Japanese).

[61] K.-F. Lee, "Context-dependent phonetic hidden markov models for continuous speech recognition," *Acoustics, Speech, and Signal Processing, IEEE Transactions on*, vol. 38, no. 4, pp. 599–609, 1990.

[62] K. Sugawara, T. Nishimoto, and S. Shigeki, "Automatic harmonization for melodies based on hmms including note-chain probability," *Information Processing Society of Japan SIG Notes*, vol. 2003, no. 127, pp. 49–54, 2003. (in Japanese).

[63] M. Khadkevich and M. Omologo, "Use of hidden markov models and factored language models for automatic chord recognition," in *Proceedings of the 10th International Society for Music Information Retrieval Conference 2009*, pp. 561–566, 2009.

[64] E. Cambouropoulos, "Musical rhythm: a formal model for determining local boudaries, accents and metre in a melodic surface," *Lecture notes in computer science*, vol. 1317, pp. 277–293, 1997.

[65] D. Werts, "'good,' 'fair,' and 'bad' chord progressions : A regression-analysis of some psychological chord progression data obtained in an experiment by j. bharucha and c. krumhansl," *Lecture notes in computer science*, vol. 1317, pp. 200–213, 1997.

[66] R. Parncutt, "A models of the perceptual root(s) of a chord accounting for voicing and prevailing tonality," *Lecture notes in computer science*, vol. 1317, pp. 181–199, 1997.

[67] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "Rwc music database: Music genre database and musical instrument sound database," *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*, 2003.

[68] G. F. Cooper and E. Herskovits, "A bayesian method for the induction of probabilistic networks from data," *MACHINE LEARNING*, vol. 9, no. 4, pp. 309–347, 1991.

[69] J.-J. Nattiez and C. Abbate, *Music and Discourse: Towarda Semiology of Music /Translated from French.* Princeton Univ Press, 1990.

[70] D. Stein, *Engaging Music: Essays in Music Analysis.* Oxford University Press, 2004.

[71] R. Yamanishi, K. Akita, and S. Kato, "Automated composing system for submelody using hmm: A support system for composing music," *Lecture notes in Computer Science*, vol. 6243, pp. 425–427, 2010.

[72] S. Gulati and P. Rao, "Rhythm pattern representations for tempo detection in music," *International Conference on Intelligent Interactive Technologies and Multimedia*, pp. 241–244, 2010.

[73] V. Zenz and A. Rauber, "Automatic chord detection incorporating beat and key detection," *2007 IEEE International Conference on Signal Processing and Communications*, pp. 1175–1178, 2007.

[74] K. Toyoda, Y. Miyakoshi, R. Yamanishi, and S. Kato, "Dialogue mood estimation focusing on utterance intervals by using tree-augmented naive bayes," *The 25th Annual Conference of the Japanese Society for Artificial Intelligence*, pp. 1l2–5, 2011. (in Japanese).

[75] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing.* Prentice Hall, 1975.

[76] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex fourier series," *Mathematics of Computation*, vol. 19, no. 90, 1965.

Related publication

- Ryosuke Yamanishi, Shohei Kato, Tsutomu Kunitachi, and Hidenori Itoh: "Effect on EEGs When Listening Harmony," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol.13, No.4, pp.366-372, 2009.
- Ryosuke Yamanishi, Shohei Kato, and Hidenori Itoh: "Study of Relation Effect on Brain Waves During Sound Listening: Difference Test Between Listening to Single Chord and Listening Harmony, and Relaxation Effect Caused by Listening to Harmony," *Journal of Japan Society of Kansei Engineering*, Vol.9, No.2, pp.293-299, 2010 (in Japanese).
- Ryosuke Yamanishi, Shohei Kato, Tsutomu Kunitachi, and Hidenori Itoh: "Effect on brain waves caused by listening to harmony," *Proceedings of the Joint 4th International Conference on Soft Computing and Intelligent Systems and 9th International Symposium on advanced Intelligent Systems*, pp.519-522, 2008.
- Ryosuke Yamanishi, Keisuke Akita, and Shohei Kato: "Automated composing system for sub-melody using HMM: a support system for composing," *Lecture notes in Computer Science*, Vol.6243, pp.425-427, 2010.
- Ryosuke Yamanishi, Yuya Ito, Shohei Kato: "Relationships Between Emotional Evaluation of Music and Acoustic Fluctuation Properties," *Proceedings of IEEE Symposium on Computers & Informatics 2011*, pp.721-726, 2011.
- Ryosuke Yamanishi, Yuya Ito, Shohei Kato: "Automated Song Selection System Complying with Emotional Requests," *Lecture notes in Computer Science*, Vol.6972, pp.367-370, 2011.
- Ryosuke Yamanishi, Ryogo Okamura, Shohei Kato: "Automated Accompaniment System Based on Bayesian Mining of Score Context," *Lecture notes in Information Technology*, 2012 (to appear).

Publication

- Yuya Ito, Ryosuke Yamanishi, Shohei Kato, and Hidenori Itoh: "A Method of Relating Sensibility Evaluation about Music with Acoustic Features Based on Fluctuation," *Journal of Japan Society of Kansei Engineering*, Vol.10, No.3, pp.341-348, 2011 (in Japanese).

- Yuya Ito, Ryosuke Yamanishi, and Shohei Kato: "Estimation of Music Impressions Using Musical Fluctuation Features," *Journal of The Acoustical Society of Japan*, Vol.68, No.1, pp.11-18, 2012 (in Japanese).

- Kaoru Toyoda, Yoshihiro Miyakoshi, Ryosuke Yamanishi, and Shohei Kato: "Dialogue Mood Estimation Focusing Intervals of Utterance State," *Transactions of the Japanese Society for Artificial Intelligence*, Vol.27, No.2, 2012 (in Japanese) (to appear).

- Kaoru Toyoda, Yuya Ito, Ryosuke Yamanishi, and Shohei Kato: "Kansei Song Selection System Based on Music Fluctuation Features and Comparison of Song Selection Algorithm," *Journal of Japan Society of Kansei Engineering*, 2012 (in Japanese) (to appear).

- Kaoru Toyoda, Ryosuke Yamanishi, and Shohei Kato: "Song Selection System with Affective Requests," *Proceedings of 12th International Symposium on Advanced Intelligent Systems*, pp.462-465, 2011.

- Ryosuke Yamanishi, Shohei Kato, Hidenori Itoh: "Analysis of EEGs during Listening to Harmony," The 11th Annual Conference of Japan Society of Kansei Engineering, Otsuma Women's University, 11B1-02(7 pages), Sep. 8-10, 2008 (in Japanese).

- Ryosuke Yamanishi, Shohei Kato, Hidenori Itoh: "Study of relaxation effect on brain waves during sound listening: Difference test between listening to single

chord and listening to harmony, and relaxation effect caused by listening to harmony," The 11th Annual Conference of Japan Society of Kansei Engineering, 3A1-3(6 pages), Sep. 08-10, 2009 (in Japanese).

- Yuya Ito, Ryosuke Yamanishi, Shohei Kato, and Hidenori Itoh: "Relationship on Music between Fluctuation-based parameters and Impressions," WiNF 2009, Nagoya Institute of Technology, Nov. 27-28, 2009 (in Japanese).

- Ryogo Okamura, Ryosuke Yamanishi, Shohei Kato, Hidenori Itoh: "Vibrato Performance Rendering Based on Music Score Context Clastering," *Special Interest Group on Musical Acoustic, Acoustical Society of Japan*, Vol.28, No.8, pp.43-48, Tsukuba University, Mar. 11 (2010) (in Japanese).

- Keisuke Akita, Ryosuke Yamanishi, Shohei Kato, Hidenori Itoh: "Automatically Composing Bass-line Using HMM: Modeling Based on Note And Rhythm in Jazz Music," *Special Interest Group on Musical Acoustic, Acoustical Society of Japan*, Vol.28, No.8, pp.49-54, Tsukuba University, Mar. 11 (2010) (in Japanese).

- Yuya Ito, Ryosuke Yamanishi, Shohei Kato, Hidenori Itoh: "A Multivariate Analytical Method of Extracting Fluctuation-based Parameters Related with Impressions on Music," *Special Interest Group on Musical Acoustic, Acoustical Society of Japan*, Vol.29, No.1, pp.1-6, The University of Electro-Communications, May 18, (2010) (in Japanese).

- Ryosuke Yamanishi and Shohei Kato: "Detection Method for Activated EEGs Components by Using Naive-Bayes Feature Selection, " Rengo2010, P3-3(1 page), Chubu University, Aug. 30-31, 2010 (in Japanese).

- Yuya Ito, Ryosuke Yamanishi, Shohei Kato, and Hidenori Itoh: "A method of relating sensibility evaluation about music with acoustic features based on fluctuation," The 12th Annual Conference of Japan Society of Kansei Engineering, 3A1-3(6 pages), Tokyo Institute of Technology, 3B1-5(6 pages), Sep. 11-13, 2010 (in Japanese).

- Ryosuke Yamanishi, Shohei Kato, and Hidenori Itoh: "A detection method for activated EEG parts when listening to harmonies giving different impressions," The 12th Annual Conference of Japan Society of Kansei Engineering, 3A1-3(6 pages), Tokyo Institute of Technology, Sep. 11-13, 2010 (in Japanese).

- Takuya Naganawa, Ryosuke Yamanishi, and Shohei Kato: "Impression quantification of violin performance sound with acoustic features," WiNF 2010, pp.23-28, Nagoya Institute of Technology, Dec. 10-11, 2010 (in Japanese).

- Ryogo Okamura, Ryosuke Yamanishi, and Shohei Kato: "Automated Composing System Based on Bayesian Mining of Score Context," The 73rd National Convention of IPSJ, Vol. 2, 247-248, Tokyo Institute of Technology, Mar. 2-4, 2011 (in Japanese).

- Kaoru Toyoda, Yoshihiro Miyakoshi, Ryosuke Yamanishi, Shohei Kato: "Dialogue mood estimation focusing on utterance intervals by using Tree-Augmented Naive Bayes," *The 25th Annual Conference of the Japanese Society for Artificial Intelligence*, Aiina center, Iwate, pp.1I2-5(4pages), Jun. 1-3, 2011 (in Japanese).

- Keigo Tada, Ryogo Okamura, Ryosuke Yamanishi, and Shohei Kato: "Accompanying System Based on Bayesian Mining of Musical Score Context: Ondemand Accompany for Affective Requests," *IPSJ Special Interest Group on MUSic and Computer*, 12 (6pages), Jul. 27-29, 2011 (in Japanese).

- Kaoru Toyoda, Yuya Ito, Ryosuke Yamanishi, and Shohei Kato: "Kansei Song Selection System Based on Music Fluctuation Features and Comparison of Song Selection Algorithm," *Japan Society of Kansei Engineering*, I44(6pages), Tokyo, Sep. 3-5, 2011 (in Japanese).