Study of Feature Values for Subjective Classification of Music

Masashi MURAKAMI*, Kazuo KAWAGUCHI*, Tetsuya OMURA**, Toshikazu KATO*

**Chuo University, ** Shobi Gakuen University*

Abstract: In this research, we analyze how the sound and music relate to humans from the aspect of Kansei engineering. We analyze what features of the sound humans pay attention and how humans interpret sound.

Therefore, we divide the signal processing of sound that humans do into four levels. At the physiological level, processing is done by the auditory characteristic. In this level, humans don't interpret the image of the sound yet. There is no subjectivity for the sound. By using auditory characteristic, we investigate the features which help in the case that sound and music is analyzed.

We consider that the processing at early stage of auditory nervous system is to extract the change in power, which is obtained from the segmentation of the sound-signals which is divided by band of the frequency and time interval, and its contrast. We also consider the features obtained by that extractation.

Moreover, in the cognitive level, we analyze the correlation of that features with the word of interpretation that humans do subjectively.

By these modelings, we develop the method of retrieving the sound and music that having the similarity, or having the image that is expressed by any subjective words.

Key words**:** *Music, Hierarchical modeling of Kansei, Auditory characteristic*

1. Introduction

Various studies have been made of music retrieval due to the increase in multimedia technology and its contents of recent years (1) (2). However, these retrievals are based on the metadata of music and there are few on the classification of music itself. As the similarity of music cannot be obtained by these studies, it can't be said that the listeners have been always satisfied with the result of the retrievals. Therefore, in this study, we attempted the analysis of music by the attributes of auditory perception from the viewpoint of the audiological psychology which recognizes the sound at a physiological level.

With this approach, it is possible to establish the feature values of music by the attribution of auditory perception at a physiological level, without using a complicated model.

2. Hierarchical modeling of Kansei (sensitivity)

We notice the individual difference of the standard of interpretation which appears through the process of human perception and we are seeking to conduct a further study of the engineering modeling of this phenomenon (3) to (7). By the application of this way of thinking, a model of human beings and the sound is shown hierarchically (Fig.1.)

2.1 Kansei model at a physical level

The level based on the features (the intrinsic features as a frequency-based physical signal) of sound itself before a human being can perceive the sound.

2.2 Kansei model at a physiological level

1-13-27, Kasuga, Bunkyo, Tokyo, Japan 〒112-8551, Tel. 03-3817-1943, E-mail :{masa_m, kato}@indsys.chuo-u.ac.jp

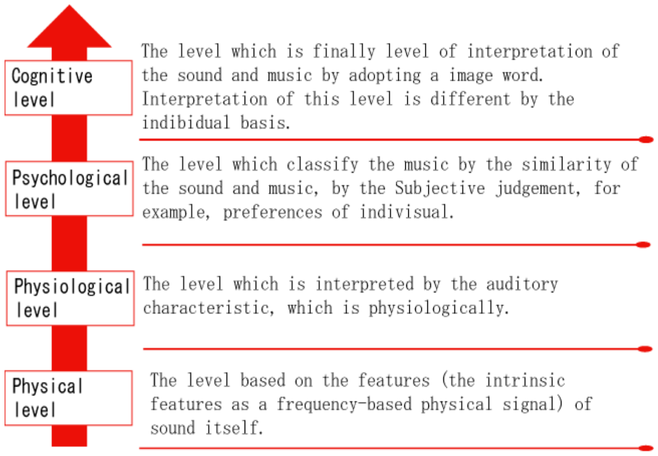
The level which perceives sound based on the results of the extraction of various features, which are implemented in the physiological response feature by the attribution of auditory perception and the nerve pathways, through the human senses.

2.3 Kansei model at a psychological level

The level which expresses or interprets the features of sound or music by classifying the weight of the features. Features are obtained at the physiological level and by its subsequent grouping.

2.4 Kansei model at a cognitive level

The level that interprets music by adopting a generic word (image word) for each of the groups classified at a psychological level.



The level which is finally level of interpretation of the sound and music by adopting a image word. Interpretation of this level is different by the indibidual basis.

The level which classify the music by the similarity of the sound and music, by the Subjective judgement, for example, preferences of indivisual.

The level which is interpreted by the auditory characteristic, which is physiologically.

The level based on the features (the intrinsic features as a frequency-based physical signal) of sound itself.

Fig. 1: Hierarchical modeling of Kansei

3. Sound as a physical signal

When conducting experiments on music, it is very important to consider the source of sound to be investigated. Taking recent automatic playing systems for examples, many of them re-create faithfully the contents of music written on scores (8) and (9). However, we receive a different impression of musical performances given by world-class musicians from that of the automatic playing system.

This difference can be explained by tempo fluctuations in musical performance (10).

4. Feature values at a physiological level (11)

It can be considered that human beings take notice of some features of sound at a physiological level and can evaluate those features.

We take notice of the following three points as the above features and seek the amount of features for the purpose of the human classification of the sound.

4.1 Pitch of sound

Pitch of sound is not measured directly and it is given by the attributes of auditory perception which can perceive relatively the change of sound. Therefore, we consider it important to use the pitch of sound as feature values at a physiological level.

4.2 Intensity of sound

It is difficult to measure directly the intensity of sound. Therefore, such a method is employed that measures relatively the intensity of sound by logarithmics, indicating a comparison of the intensity of two sounds.

As the intensity of sound is one of the attributions of auditory perception, we consider it important to use it as feature values for the classification of the intensity of sound.

4.3 Time variation of sound

Most sounds have a time variation and it is important to perceive this time variation for the study of the auditory perception. As music is composed of the time variation of sound, we consider it important to use the time variation of sound as a feature value.

5. Experiment at a physiological level

For each feature value at a physiological level as mentioned in Section 4, for this study, analysis was made by changing the frequency data into the short-lasting Fourier transform.

Additionally, we aim to model on a physiological level by conducting the retrieval experiment of similar music through the evaluation experiment of feature values.

5.1 Short-lasting Fourier transform of music data

A power spectrum can be obtained by the use of the short-lasting Fourier transform. In this study, we divided the result of the short-lasting Fourier transform into plural ranges by the following process in order to study the relationship between the time variation of sound and the attributes of auditory perception. The result of the division is shown in Fig. 2.

1-13-27, Kasuga, Bunkyo, Tokyo, Japan 〒112-8551. Tel. 03-3817-1943, E-mail :{masa_m, kato}@indsys.chuo-u.ac.jp

(i) The frequency, which represents the vertical axis of the power spectrum, is divided by the six ranges shown in Table. 1.

Table. 1 Frequency spectrum per range

Treble	5000～10000Hz
Middle Treble	2600～5000Hz
Mediant	320～2600Hz
Middle Bass	160～320Hz
Bass	40～160Hz
Bassy	20～40Hz

(ii) The time period, which represents the horizontal axis of the power spectrum, is divided by “S” seconds.

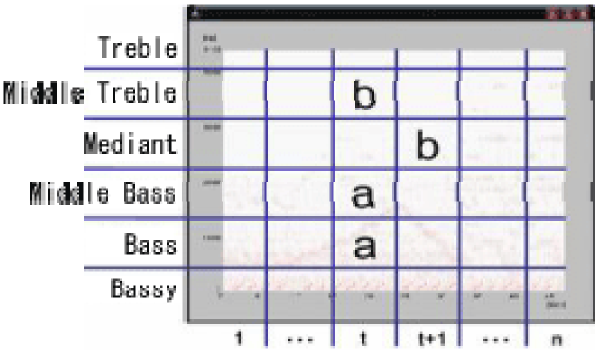


Fig. 2: Division of spectrum by range and time

Now, the value of “S” is established by the formula (1) with the following conditions;

(A) Sampling frequency of the data used is 22050Hz.

(B) Sampling frequency is halved in the process of the short-lasting Fourier transform.

(C) In order to implement the short-lasting Fourier transform, the number of data should be 2 multiplied by n.

$$s = \frac{2^n}{22050/2} = \frac{2^{12}}{11025} \fallingdotseq 0.37(n=12)6 \quad (1)$$

As it is set as “n=12”, the time period per divided range is equal to the time per note of a tune with tempo of BPM 160. It is considered that the value of “n= 12” is the most appropriate for the time period per divided range.

By performing the following processing of the divided ranges, the amount statistics is sought from feature values. For the next step, evaluation is made of this statistics quantity.

(a) According to the Weber-Fechner Law, “The amount of perception is proportional to the logarithmic in the amount of stimulus,” a logarithmic value (logPS) is sought of the power spectrum

(b) LogPS is sought of each of the six (6) ranges divided by (i).

(c) The contrast of logPS is sought between the different ranges (Fig. 2.a) at the same time, “t.”

(d) The contrast of logPS is sought between the different ranges (Fig. 2.b) at different times, “t, t+1.”

5.2 Evaluation of the proposed feature values

In order to examine the significance of feature values and its statistics quantity, a retrieval experiment of similar music was performed by the use of the statistics quantity which was sought, based on the proposed feature values.

5.3 Summary of the experiment

The music data were collected from the 576 tunes of the Sound Material Collection (12), excluding recordings of sound effects and human voices.

Fifty (50) tunes were selected at random among them, and the retrieval experiment of similar music was performed using these tunes as the key tunes.

An evaluation was made of the proposed feature values by seeking the average relevance ratio of the top 5, 10, 15 and 20 similar music of the examination result.

Further, in order to compare with the proposed feature values, the same experiment was performed using a power spectrum obtained from the fast Fourier transform as feature values.

5.4 Result of the experiment

The relevance ratios are shown in Table. 2 of the result of the retrieval experiment of similar music, by the fast Fourier transform and by the proposed feature values.

Table.2 Relevance ratio of the top 20 similar music, of the examination result

	～5th	～10th	～15th	～20th
Proposed feature values	72.8%	64.2%	56.4%	48.7%
fast Fourier transform	59.1%	53.2%	47.5%	42.3%

The Wilcoxon rank-sum test was made using the results shown in Table. 2 and also an examination was performed of the significance of the difference of relevance ratios by the proposed feature values and by the use of the fast Fourier transform. As a result, at a significance level of 1%, the difference was recognized as significant between the two relevance ratios of the feature values.

6. Experiment at a cognitive level

In order to establish a Kansei model at a cognitive level, an experiment was conducted by the use of an adjective called an "image word" to combine the human Kansei and the physical features of music.

6. 1 Selection of image words and the summary of the experiment

A questionnaire was carried out in advance on the subjects of "Image words frequently used for the evaluation of music." The following words were selected as the image words for use in the experiment (Table. 3.).

Table. 3 Image words used in the experiment

vivid	fresh	tranquil
pleasant	monotonous	swinging

As the next step, a study was made on a five (5) grade evaluation to determine whether each image word was applicable or not for the 288 tunes which were selected at random among the music data base of 576 tunes, and then a model was formed of Kansei of auditory perception. Further, the two ways of analysis were performed using physical feature values at the dimension of 216 without any modification, and by the use of the stepwise selection.

6.2 Result of the experiment

Table 4 and 5 show the Multiple Correlation Coefficients adjusted for degrees of freedom by the Multiple Regression Equation, which were sought as the result of the experiment.

Table. 4 Multiple correlation coefficients when variable selection was not performed

Image terms	Multiple correlation coefficient
vivid	0.5024
fresh	0.6217
tranquil	0.7997
pleasant	0.5956
monotonous	0.6661
swinging	0.6817

Table. 5 Multiple correlation coefficients when variable selection was performed

Image terms	Multiple correlation coefficient
vivid	0.6534
fresh	0.6516
tranquil	0.8513
pleasant	0.6842
monotonous	0.5548
swinging	0.7594

In the case where variable selection was applied only for the image word "monotonous," it resulted in a low multiple correlation coefficient.

Of the other image words, higher multiple correlation coefficients resulted when variable selection was performed.

However, of the difference of the value of multiple correlation coefficients by variable selection, a Wilcoxon rank-sum test was conducted.

As a result, no significance was recognized in the difference.

6.3 Evaluation experiment of the Kansei model of auditory perception

A retrieval experiment was performed about the Kansei of music by the use of the established Kansei model of auditory perception, and then by its retrieval precision, an evaluation was made of the established Kansei model of auditory perception. Retrieval experiment of the music was performed by selecting 288 tunes out of the music data base of 576 tunes. An evaluation was performed about the established Kansei model of auditory perception, by seeking the relevance ratio of the top 20 estimated values per image word.

Further, in the experiment in the Section 5.1, an experiment was carried out about the difference of the multiple correlation coefficients by variable selection.

6.4 Result of the experiment

A retrieval experiment of Kansei was performed about music by the use of the Kansei model of auditory perception. As a result of the experiment, the relevance ratios of the top 20 tunes were found as shown in Table. 6 and 7, which were retrieved per image word.

Table. 6 Relevance ratios when variable selection was not performed

Image terms	Relevance ratio of the top 20 tunes
vivid	70.0%
fresh	75.0%
tranquil	80.0%
pleasant	70.0%
monotonous	80.0%
swinging	80.0%

Table. 7 Relevance ratios when variable selection was not performed

Image terms	Relevance ratio of the top 20 tunes
vivid	80.0%
fresh	75.0%
tranquil	95.0%
pleasant	80.0%
monotonous	85.0%
swinging	90.0%

Table. 6 and 7 show that the relevant ratio was higher in all the image words when variable selection was performed.

A Wilcoxon rank-sum test was made using results of the difference of the relevance ratio. As a result, the difference of relevance ratio was recognized at a significant level of 1%.

Additionally, a retrieval experiment of Kansei was performed about the difference by variable selection. As a result, the difference of the relevant ratio was recognized.

Table 7 illustrates that extremely high accuracy can be obtained of each image word as the result of a retrieval experiment of the Kansei of music, using a hearing Kansei model, which was established based on the proposed feature values and which underwent variable selection.

7. Improvement of feature values

7.1 Selection of feature values

As the number of the dimensions of the proposed feature values decreased by 1/10 when variable selection was performed, it is considered that the proposed feature values are redundant. Additionally, it is necessary to seek the feature values necessary for each image word.

7.2 Relationship between feature values and image words

In order to seek the feature values necessary for each image word, the authors sought the standard partial regression coefficient by the multiple regression of the Kansei model of auditory perception.

The feature values were sought of the top five (5) most important. Table 8 shows one example of the variety of the said feature values.

Table. 8 Top 5 of the level of importance of "vivid"

Image terms	Vivid
1st	Whole of midrange
2nd	Contrast between middle treble and midrange
3rd	Contrast between treble and bass
4th	Contrast between middle bass and treble
5th	Contrast between middle treble and treble

The following can be considered from the result of Table. 8.

"Vivid" - As the contrasts between various ranges of sound are ranked high on the axis of the "whole of midrange," it can be considered that both features are important in a wide range, of the whole of tune and of a short time.

8. Challenge and perspective

In this study, we proposed the feature values of sound from the viewpoint of human attributes of auditory perception at a physiological level.

With the results of the evaluation experiment of statistics quantity which was obtained from the proposed feature values, we have obtained the possibility of classification not depending on the name of tunes or composers, but by the use of the feature values which we proposed for this study,

As a future policy, we will seek the improvement of statistics quantity, the most appropriate division of sound range, the increase in experiment data and others. Furthermore, by the use of the established auditory perception model, we are attempting the quantification / classification, by audio comparison experiment, of musical performances given by master musicians and by machines.

Acknowledgement

This research is partially supported by Grant-in-Aid for Scientific Research (S), No. 19100004, "Robotics modeling of diversity of multiple KANSEI and situation understanding in real space", Japan Society for the Promotion of Science.

References

(1) Yuya Hashimoto, Kenichi Fukui, Kouichi Morimoto, Satosi Kurihara, Masayuki Numao;
“Acquisition Mechanism of Personal Sensitivity in Musical Composition Mechanism”
National Convention of Japanese Society for Artificial Intelligence (20th)

(2) Yuya Ichikawa, Tetsuji Tamura, Satoru Hayami,
“Recommendation System of Musical Composition by the Use of the Grouping of Impression Words”
National Convention of Japanese Society for Artificial Intelligence (20th)

(3) Syunichi Kato, Katsumasa Sakai
“*Research and Development of Sensitivity Agent and Human Media Data Base—Sensitivity Work Shop, System/Information/Control ”* V*ol.42，No.5，pp.253-259(Jun 1998)*

(4) Kato，T.:Trans-category Retrieval Based on Subjective Pereception Process Models，*Proc.IEEE Multimedia and Expo ICME2004*，TP9-5(2004).(CD-ROM)

(5) Kouta Hagino, Tosikazu Kato
“*Design Approach of the Retrieval System of Sensitivity; Sensitivity System Modeling”*
Journal of the Information Processing Society of Japan; Data Base; Vol.46，No.SIG19(TOD 29)(2005)

(6) Masahiro Tada, Tosikazu Kato,
“Modeling of the Sensitivity of Auditory Perception by the Use of Hierarchical Classification and Application for the Retrieval of Similar Image”
Journal of the Information Processing Society of Japan;
Data Vol44，No.SIG8(TOD 18)，pp.37-45(2003)

(7) Yasuhiko Tada, Tosikazu Kato,
“Feature Analysis of the Range of Similar Images and the Modeling of the Sensitivity of Auditory Perception”
Journal of the Institute of Electronics, Information and Communication Engineers;
D-Ⅱ，Vol.J87-D-Ⅱ，No.10，pp.1983-1995(2004)

(8) Yasuhiko Oda, Kenichi Sirakawa, Yutaka Murakami, Yosinobu Kajikawa, Yasuo Nomura,
“Establishment of an Automatic Playing System for Piano, adding the Information of Musicians/Extraction of the Features of Performance in Topical Parts? by Neutral Network//”.
Study Report, Information Processing Society of Japan

(9) Takayuki Hoshishiba, Susumu Horiguchi,
“Automatic Playing of Music for Piano by use of Standard Performance Data”
Study Report, Information Processing Society of Japan

(10) Tetsuya Omura
“Base of the Style of Musical Performance”
Shunjusha Co., Ltd. pp158-165(1998)

(11) B.C.J.Moor, Kengo Ookusi (translation supervisor)
“Survey of Audiological Psychology”
Seisinshobo Co., Ltd (1994)

(12) *“On-Man-Tan DX”*
E-Frontier Co., Ltd