Auditory Feature Parameters for Music Based on Human Auditory Processes

Masashi Murakami and Toshikazu Kato

Chuo University,
1-13-27, Kasuga, Bunkyo-ku, Tokyo, Japan
{masa_m, kato}@indsys.chuo-u.ac.jp

Abstract. Authors aim to show the similarity and difference of sensibility effect when human listen to the music. For concrete example, we assume the music recommendation services. To propose the music that users really want, we use the image words that users feel for the music, and retrieve and propose the music based on the similarity of the image word. We intend to propose the music that users really want based on Kansei engineering. We make the Hierarchical model of Kansei. This model shows that human have the 4 processes when listen the music. Feature parameters are physical frequency data with physiological process. These parameters make us possible to represent how human listen the music. In next step, we represent how human feel for the music. These representations help users possible to retrieve the music really want. For representation of human feelings, we use image words and subjective estimation.

Keywords: Music, Hierarchical model of Kansei, Auditory characteristic.

1 Introduction

When human want to listen to the music, they have some needs of image. For example, when someone tired for working retrieve the music, he wants to freshen up, or concentrate. When someone dinking tea retrieve the music, he want the music that he feel “fresh” image for. We define the music that users really want is that the users feel image for. To propose the music that users really want, we use the image words that users feel for the music, and retrieve and propose the music based on the similarity of the image word. To realize this method, we make computers to trace how human listen the music, and how human feel for the music. Recently, the techniques intended for music retrieval mainly use the text-based metadata information that title, composers, and genres, as retrieval key. These services require a large amount of cost for giving metadata. Actually, quite a few systems are classified according to music itself. Therefore users retrieve the music by not the music itself but the music’s metadata information. The other techniques use the frequency-based data. In these techniques, users can retrieve the music by the similarity that presumed from the frequency of the music. For presuming the similarity of the music, various algorithms are used without how human feel for the music. Therefore the evaluation of these

techniques only intends to clarify the similarity of the frequency in music. These techniques are difficult to retrieve music users really want.

We intend to propose the music that users really want based on Kansei engineering. In Kansei engineering, we notice that the individualities of the basis of interpretation appear in the process of human perception and conduct a further study of the engineering modeling of the process of human perception. We make the Hierarchical model of Kansei. This model shows that human have the 4 processes when listen the music. This model, auditory processing process model shows human listen the music. In this model, we perceive the physical signals for physiological stimulus, and output the information weighted by similarity of images.

2 Hierarchical Model of Kansei

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 (1) to (5). By the application of this way of thinking, a model of human beings and the sound is shown hierarchically (Fig.1.). This Model is divided in several levels. There are physical level, physiological level, psychological level, and cognitive level.

Fig. 1. Hierarchical model of Kansei for Auditory Information Processing

At first, we can find physical level. This level is based on the features (the intrinsic features as a frequency-based physical signal) of sound itself. This level deals in signals, but signals are not received humans yet. So, this level deal before a human being can perceive the sound. At second, we can find physiological level. This level perceives sound based on the results of the extraction of various features. The sounds are implemented in the physiological response feature by the attribution of auditory perception and the nerve pathways, through the human senses. Next, we find psychological level. This level 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. At last, cognitive level that interprets music by adopting a generic word (image word) for each of the groups classified at a psychological level.

When conducting experiments on music, it is very important to consider the source of sound to be investigated. 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 (6).

3 Auditory Feature Values (7)

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.

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. And, Intensity of sound is difficult to measure directly the intensity of sound. Therefore, such a method is employed that measures relatively the intensity of sound by logarithmic, 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. Moreover, we take notice of the importance of 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.

In this section, we aim to mention the feature value at a physiological, and to analyze the data with short-lasting Fourier transform. In section 4, we aim to model on a physiological level by conducting the retrieval experiment of similar music through the evaluation experiment of feature values.

3.1 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. Fig. 2 shows the result of the division. In Fig. 2, vertical axis represents the frequency period. Frequency period is divided by six ranges (Table.1). Horizontal axis represents the time period, is divided by "S" seconds. 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.

Table. 1 Range Division per frequency

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

Fig. 2. Division of spectrum by range and time

$$s = \frac{2^n}{22050 / 2} = \frac{2^{12}}{11025} \approx 0.37 \ (n = 12) \ . \tag{1}$$

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 (log PS) is sought of the power spectrum
- b) Log PS is sought of each of the six (6) ranges divided.
- c) The contrast of log PS is sought between the different ranges at the same time, “t.”
- d) The contrast of log PS is sought between the different ranges at different times, “t, and t+1.”

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

4 Experiment of the Similarity

The music data were collected from the 200 tunes of the RWC Music Database Collection (8). 10 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.

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 similarity results and also an examination were 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.

5 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.

References

1. Kato, T., Sakai, K.: System/Information/Control. In: Research and Development of Sensitivity Agent and Human Media Data Base—Sensitivity Work Shop, vol. 42(5), pp. 253–259 (1998)
2. Kato, T.: Trans-category Retrieval Based on Subjective Perception Process Models. In: Proc. IEEE Multimedia and Expo ICME 2004, vol. TP9-5 (2004)
3. Ogino, A., Kato, T.: Design Approach of the Retrieval System of Sensitivity; Sensitivity System Modeling. Journal of the Information Processing Society of Japan; Data Base 46(SIG19(TOD 29)) (2005)
4. Tada, M., Kato, T.: 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 44(SIG8(TOD 18)), 37–45 (2003)
5. Tada, M., Kato, T.: 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, J87-D-II(10), 1983–1995 (2004)
6. Ohmura, T.: Base of the Style of Musical Performance, pp. 158–165. Shunjusha Co., Ltd. (1998)
7. Moor, B.C.J., Ohkushi, K. (eds.) (Translation supervisor): Survey of Audio logical Psychology. Seisinshobo Co., Ltd. (1994)