THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA VOL.24, NO.2E 2005. 6 pp. 67-70

Determining Key Features of Recognition Korean Traditional Music Using Spectrogram

Jae Chun Kim* Kyung Sup Kwak**
*Department of Electronics Engineering, Inha University,
**Graduate School of Information Technology and Telecommunications, Inha University.
(Received May 24 2005; revised May 31 2005; accepted Jun 8 2005)

Abstract

To realize a traditional music recognition system, some characteristics pertinent to Far East Asian music should be found. Using Spectrogram, some distinct attributes of Korean traditional music are surveyed. Frequency distribution, beat cycle and frequency energy intensity within samples have distinct characteristics of their own. Experiment is done for pre-experimentation to realize Korean traditional music recognition system. Using characteristics of Korean traditional music, 94.5% of classification accuracy is acquired. As Korea, Japan and China have the same musical roots, both in instruments and playing style, analyzing Korean traditional music can be helpful in the understanding of Far East Asian traditional music.

Keywords: *Music, Spectral analysis, Classification algorithm, Auditory model, Feature extraction*

I. Introduction

Music recognition is a very complex procedure. Though accomplishments have been made in the western music genre recognition, Asian traditional music has not yet been explored. As a direct result, this paper confronts the unexplored Korean traditional music recognition using Spectrogram[1]. Fundamentally western music and Asian traditional music are similar. So various tools used in western music recognition can be applied to Korean traditional music analysis.

Music is a result of human perception. Though the physical attributes of sound are well known, the human's processing methods leading to sound perception are not. Hence, there are many problems facing the realization of fully human perception simulated systems.

Corresponding author: Jae chun Kim (milgaroo@naver.com)
Hitech center 428, Telecommunication Lab, Department of Electronic Engineering, Inha Univ., 253 YongHyun-Dong, Nam-Gu, Inheon. 402-751, Korea

II. Music Recognition Possibility in Short Time Sample

The astounding ability of human's music perception abilities are revealed by D. Perrot and R. O. Gjerdige[2]. Classical, blues, country, dance, jazz, latin, pop, r&b, rap and rock, 10 genres used for experiment and each genre is composed of 8 songs. 3000ms, 475ms, 400ms, 325ms, 250ms length samples picked from each of the songs. 52 college students participated in the experiments. They are heard arbitrarily chosen samples, subsequently choosing one genre within the above 10 genres that best described it. The results compared to the CD vendor's classification.

Accuracy of success was 70% using 3000ms samples. Considering the obscurity of the music genre in some cases, the success rate is relatively high. While the success rate of the experiment is 40% when using 250ms samples. This research showed that any high level inferences are not necessary to genre

classification. As implied in the short time sample where it was difficult to grasp rhythm, melody and structure, the students in the experiment totally depended on spectral, timbral attributes for their judgments.

III. Music Signal

To classify musical categories, it is indispensable to extract some characteristics from music that represents a short time frame within the music sample. Characteristics can be called features. Feature values are acquired from raw data. As the dimension of raw data is very high, some converting process is required. The human auditory system transforms the sound into frequency components. Fourier analysis is a powerful tool used in music recognition systems as it can decompose sound into frequency.

3.1. Signal Processing

Though the time domain of a sound signal reveals relatively little information, only the amplitude of signal as the time passes can be determined. However, in the frequency domain of a sound signal, Fourier analysis discloses information regarding amplitude, frequency and phase, which can be integrated into important features through mathematical manipulations.

3.2. Physical Feature & Perceptual Feature

Usually, the physical and perceptual features are employed when classifying music. Physical features are based on the analysis of mathematical and statistical values, while perceptual features are based on the pitch, timbre, rhythm and so on. The later being the way a human recognizes music.

As a matter of fact, physical and perceptual features are related to each other. As an example, the amplitude of a signal, which is one of the physical features, is directly connected to loudness which is one of the perceptual features. Vice versa, perceptual loudness is closely related to the position of the physical frequency components. As perceptual features incorporate more physical features, it grows more difficult in expressing in mathematical form. Thus, strict discrimination between physical and perceptual features dos not make sense.

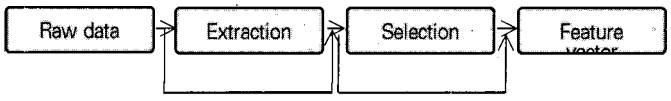


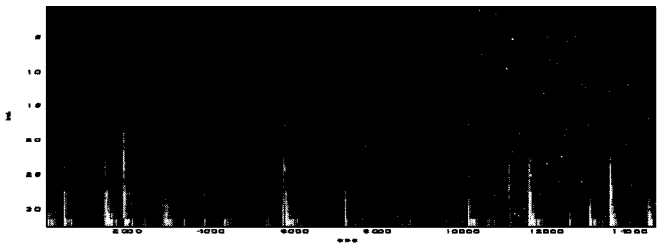
Fig. 1. Generating a feature vector from input dataset.

In order to acquire the perceptual features, physical features are recruited. As an example, pitch can be acquired from physical features that influence the pitch attributes. To fully realize the perceptual system, it is necessary in the feature extraction step to obtain the perceptual features. This can be the ultimate model in music recognition. Though the scope of understanding regarding the human recognition process is not so broad to embody such system.

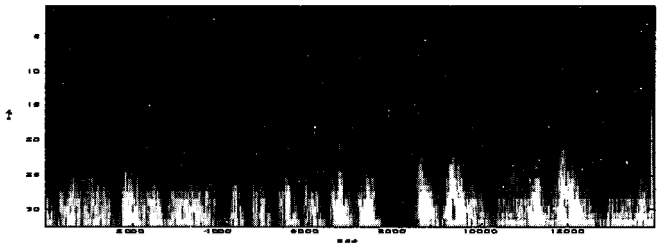
3.3. Distribution of Frequency

Fourier analysis decomposes frequencies in a similar manner as the human auditory system. This implies that there is important information transferred through the frequency. Frequency distribution one of the main factors in determining the characteristics of music.

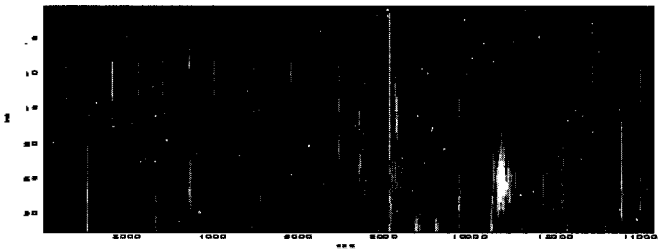
Spectrogram approaches music recognition by way of visualization. Researchers at Stanford university realize systems which transform the wave signal into the visual domain using an



(a) Korean traditional music (Sanjo)



(b) Classical



(c) Jazz

Fig. 2. Spectrograms for music samples from three genres. Time is shown on the x-axis(increasing from the top), grey values indicate power.

edge detection algorithm[3]. Spectrogram an intuitional presentation of FFT results[4]. Spectrogram visualizes data at the time-frequency domain. X-axis is time and y-axis is frequency. Point (t, f) represents signal intensity. In grey scale, brighter region indicates strong intensity.

Figure 2 visualizes music samples of 10s in length. Every sample is sampled at 44100 Hz. Each sample is selected to be considered as representing a typical genre style. Figure 2 (a) is a fraction of Kyong bokkung T'aryong played by Kayagum Trio accompanying Jango. (b) is a fraction of Mahler Symphony No. 2, 5th movement. (c) is a fraction of Jacques Loussier Trio organized of Piano, Base, Drums.

Korean traditional music (Sanjo) shows an irregular rhythm pattern with each beat intensity different from each other. Classical music shows high-energy frequencies and no rhythm pattern. The lower portion of the classical spectrogram is brighter than others. Jazz music shows regular rhythm pattern and Frequencies distributed all over the region. Each genre has it's own distinct attributes.

IV. Experiment

Several traditional Korean music genres are categorized according to their playing style and composition of instruments. To test the slight difference of Korean traditional music, Several Korean traditional music are integrated into 4 genres.

Using feature extraction function, 8 features are acquired from each 800 sample data. Features used in this experiment are RMS of frequency amplitude, percentage of low frequency, spectral flux, zero crossings, average of signal, variance of signal, average peak level and class. RMS coordinates loudness of music. Percentage of low frequency is useful to classification of musical instruments. Zero crossings useful to telling the continuous sound and intermittent sound. Average peak level discriminates frequency of beats. Other features such as average of signal, variance of signal and class discriminates overall discrepancies in music files. As an example, equations are presented.

Table 1. Classification of Korean traditional music genre by way of playing style.

	contents
genre-1	Nongak/Samulnori/Muak
genre-2	Pansori/Byungchang/Sijo/Minyo
genre-3	solo/duo
genre-4	Jereyak/Orchestral/Julpungryu

Table 2. Experiment result of genre classification.

	genre-1 output	genre-2 output	genre-3 output	genre-4 output
genre-1 input	189	4	4	3
genre-2 input	7	184	5	4
genre-3 input	3	7	189	1
genre-4 input	0	6	0	194

Table 3. Entire accuracy of parameters.

Parameters	k-NN
Correctly Classified Instances	94.5%
Incorrectly Classified Instances	5.5%
Kappa statistic	0.9267
Mean absolute error	0.0294
Root mean squared error	0.1654
Relative absolute error	7.8415%
Root relative squared error	38.1951%
Total Number of Instances	800

- Spectral flux

$$SF_t = \sum_{n=1}^{N} (N_t[n] - N_{t-1}[N-1])^2$$

N[ ] is normalized value at frame t.

- Average peak level

$$AVpeaklevel = \sum \frac{x(t)}{f_p}$$

$x(t)$ is measured peak value and f_p is frequency of peaks.

Although the music is same, These feature extraction functions are designed to estimate rhythm pattern, frequency energy distribution, spectral intensity[5].

V. Conclusion

The music recognition field has many areas to explore due to it's youth. A voice recognition tool can be applied to this area because they have a lot in common. It is possible to discern music in very short time frames (within several seconds) as proved by D. Perrot and R. O. Gjerdige's experiment. But there are some problems, as an example, it is very difficult to detect rhythm and melody structure from free, and avant-garde jazz music. Although there are deviations which avert typical form of musical genre, it's proportion is very small. It will be able to be

considered only after more research has been accumulated. As the Korean traditional music has clear characteristics, a qualified recognition system can be expected. From the results of experiment, Frequency distribution, frequency intensity and frequency intensity intervals are effective to categorize music samples. Japan, China and Korea have the same root in musical instruments and their playing style. Study on Korean traditional music can be helpful to Far East Asian music understanding.

References

1. Malcom Slaney, *Auditory toolbox version2*, (Technical Report 1998-010, Interval Research Corporation, 1998).
2. D. Perrot, and R. O. Gjerdigen, *Scanning the dial : An exploration of factors in the identification of musical style*, (In Proceedings of the 1999 Society for Music Perception and Cognition, 1999).
3. Hrishikesh Deshpande, Unjung Nam, and Rohit Singh, *Mugec : Automatic music genre classification*, (Technical report, Stanford University, June 2001).
4. Z. Liu, Y. Wang, and T. Chen, "Audio feature extraction and analysis for scene segmentation and classification", Journal of VLSI Signal Processing System, 20, 61-79, 1998.
5. G. Tzanetakis and P. Cook, Automatic musical genre classification of audio signals, in Proc. Int. Symp. Music Information Retrieval ISMIR, Oct 2001.

[Profile]

•Jae Chun Kim

2002 Complete Ph.D Course of Electronics Engineering
1999 Master of Advanced Precision Engineering
1996 Bachelor of Electronics Engineering

•Kyung Sup Kwak

1977: Inha Univ./Electrical Engineering (BS)
1981: Univ. of So. Calif./Electrical Engineering (MS)
1988: Univ. of Calif., San Diego/Comm. Theory/ (Ph.D.)
1998~1989: Hughes Network Systems (USA)/ Research staff
1989~1990: IBM Network Analysis (USA)/Research staff
2000~2002: Dean, Graduate School of IT Telecom., Inha Univ.
2002~present: Director, Advanced IT Research Center
2003~present: Director, UWB Wireless Comm. Research center