4C-2

Classification of Harmonic and Textural Keyboard Playing Style Using Acoustic Features

JooYoung Ahn, Akira Maezawa, Katsutoshi Itoyama, Toru Takahashi, Tetsuya Ogata, Hiroshi G. Okuno

Graduate School of Informatics, Kyoto University

1. Introduction

Capability to retrieve musical information based on the content of music is an important task both in music players and music stores. For example, a user might know neither the artist name nor the title of the song, but, say, the melody. The user might have a clear sense of chord progression or the timbre he/she wants to listen. Content-based music information retrieval (CBMIR) systems allow the user search for music based on different aspects of music, such as chord progression, timber rhythm or melody [1]. Yet, conventional CBMIR system is capable of dealing with only one musical aspect. Hence, the user must understand the aspect of music that a particular system accepts.

Hence, we believe that hybrid CBMIR system, that is, a system consisting of various CBMIR systems is essential to allow the user to search for a particular music without the hassle of understanding different musical aspects. In a hybrid CBMIR system identifying the kind of musical aspect that is present in the user query is mandatory. Without such identification, the system is incapable of properly choosing one of the constituent CBMIR systems to which the query is thrown. For example, a hybrid of query-by-tapping and query-by-humming system should use query-by-humming engine if the user input constituted of a melodic singing, whereas it should use query-by-tapping engine if the user input constituted of a rhythmic tapping.

As a first step towards hybrid CBMIR system, we seek to develop a hybrid CBMIR system that accepts audio input from a keyboard instrument such as a synthesizer.

Keyboard is chosen because it is capable of expressing wide variety of musical aspect. Keyboard played for query, can be broadly categorized into two categories: Harmonic style and Textural style. Harmonic style represents the keyboard playing style for describing the melody or the chord progression of the song, Textural style, on the other hand, expresses timbre.

Textural style typically contains segments where harmony is not clearly defined, but conveys, for example, the register of the query (e.g. strong low note suggesting strong bass), or the rhythm.

In this paper, we present a method to classify audio signal into harmonic playing or textural playing. The classified result may be used to choose between CBMIR systems dealing with timbral features and harmonic features. We employ features that are commonly used in timbre-based music retrieval and harmony-based music retrieval in order to separate out these two modes of query.

2. Classification Methods

To classify the user's keyboard playing into harmonic and textural styles, we extract audio features that convey both harmony and timbre, and use the k-Nearest Neighbor (k-NN) algorithm for classification. Monaural audio signal is recorded at 44100Hz sampling rate at a 32-bit resolution. Then, its short-time Fourier Transform (STFT) is computed, with window size of 32768 and 4410 frames of overlap to analyze the audio query in long-term unit.

From the obtained STFT, we compute the chroma vector and Mel-Frequency Cepstrum Coefficients (MFCCs), two features that convey harmony and texture, respectively.

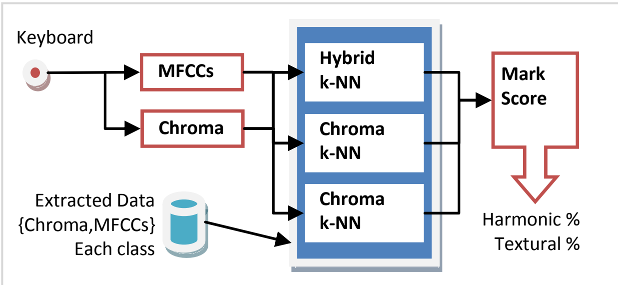

Figure 1. System Overview

2-1 Chroma Vector

Chroma vector is extracted from spectrogram of 22050Hz frequency band. We projected entire spectrum onto 128 bins corresponding to the MIDI note number. We then wrapped note numbers into an octave that is one 12-dimensional vector $N(f_k)$, where f_k is the kth frequency bin.

$$N(f_k) = \left(12\, log_2\left(\frac{fk}{440}\right) + 69\right) \text{ modulo } 12 \qquad (1)$$

Chroma vector conveys the pitch content, and is robust against timbral fluctuations [2]. So although the input audio is played by a beginner and so contains human error, we can still rely on the chroma vector as a key feature of observing the harmonic behavior.

In addition, we chose chroma vector not only because robust to human error, but also it is the most common feature on CBMIR system.

2-2 MFCCs

MFCCs is said to reflect timbre [3]. Therefore, we use seven bins of MFCCs as a feature that reflects the timbral factor of query. In our approach, we used window of 32768 frames in length, same as overall audio processing,

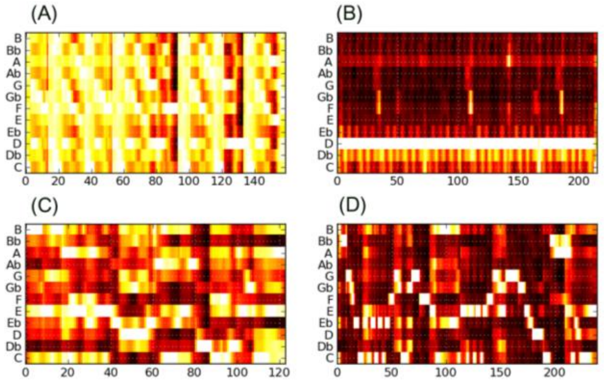

Figure 2. Various Chroma Vector of (a) textural playing, (b) textural playing with percussion, (c) harmonic chord playing, (d) harmonic playing with melody

then we transformed the log power spectrum into logarithmic $\mathrm{mel\_scale}(f) = 2595 * log\,(1 + f/700)$ and take each sum of uniformly quantized range, then normalize them in range-wise.

Similar to chroma vector, MFCCs is the most common feature to measure timbral difference of audio signal. So we choose MFCCs to have more adaptability to any of CBMIR system.

2-3 k-Nearest Neighbor

We use k-Nearest Neighbor (k-NN) classification algorithm to classify user's query because of its accuracy. k-NN algorithm has some strong consistency result. As the feature space, we use for every two classes of 12-dimensional chroma vector space, 7-dimensional MFCCs space, and 19-dimensional hybrid (using both chroma and MFCCs) space. We compute every frame of query to compare distance and choose 5 from the nearest.

3. Evaluation

We prepared about 50 minutes of audio data played by keyboard. Harmonic queries are basically chord playing from the popular song book. So Harmonic queries have human-understandable musical chord change shown as Fig.2(c). Also, Harmonic queries can have melodies above the chord playing shown as Fig.2(d). On the other hand, Textural queries are keyboard playing with more instrumental variations, but without many chords variation showed in Fig.2(a).

We performed 5-fold cross validation on each set {Textural set, Harmonic set}. Classification is performed by k-NN algorithm along every frame of query on three kinds of vector spaces: Chroma Vector space (Table.1), MFCCs space (Table.2), and Hybrid space (MFCCs and Chroma vector, Table.3)

Table 1. Result using Chroma Vector only

Recognized as Textural	Recognized as Harmonic	
55%	45%	Textural Query
26%	**74%**	Harmonic Query

Table 2. Result using MFCCs only

Recognized as Textural	Recognized as Harmonic	
55%	45%	Textural Query
35%	**65%**	Harmonic Query

Table 3. Result using both vectors

Recognized as Textural	Recognized as Harmonic	
53%	47%	Textural Query
24%	**76%**	Harmonic Query

Table 4. Accuracy of each / overall

	Accuracy (Textural)	Accuracy (Harmonic)	Accuracy (Overall)
MFCC	65%	54%	61%
Chroma	73%	55%	66%
Both	75%	53%	67%

4. Discussion and Conclusion

This paper presented a method for classifying audio segment of a keyboard playing into textural and harmonic playing style.

Evaluation showed that our method performs at 67% accuracy, suggesting the need for better features or modeling data. In detail, Table.1, 2 shows each feature's performance of when we throw {Textural, Harmonic} query to the each set. Our method, on Table.3, shows better result for Harmonic query then only using one feature. For the overall accuracy, our method shows 67% accuracy, while MFCCs shows 61%, and Chroma Vector shows 66%.

This time, the reason chroma vector showed good performance could be found from the test sets. Since the objective was to distinguish Textural style and Harmonic style, test sets had less variation of timbre. Rather, test sets had more variations of chord change.

Future work includes integration of out method into hybrid CBMIR system, and evaluation of hybrid CBMIR system using our method. Moreover, we would like to explore our method using non-keyboard audio signal as the input such as percussion.

References

[1] R. B. Dannenberg, et al. *A Comparative Evaluation of Search Techniques for Query-by-Humming Using the Musart Testbed*. Journal of the American Society for Information Science and Technology, vol. 58(3), 2007.

[2] E. Gomez. *Tonal Description of Music Audio Signals.* Ph.D thesis, 2006

[3] H. Terasawa, et al. *The thirteen colors of timbre*. IEEE WASPAA2005, 2005,

[4] M. Muller, et al., *Making Chroma Features More Robust to Timbre Changes*. ICASSP2009, 2009

[5] M.Muller, et al. *Audio matching via chroma-based statistical features.* ISMIR2005, 2005.

[6] S. Kim, et al. *Dynamic chroma feature vectors with applications to cover song identification*. MMSP, 2008