USING CHROMA HISTOGRAM TO MEASURE THE PERCEPTUAL SIMILARITY OF MUSIC

Linxing Xiao, Jie Zhou

State Key Laboratory on Intelligent Technology and Systems
Tsinghua National Laboratory for Information Science and Technology
Department of Automation, Tsinghua University, Beijing 100084, P.R. China
xiaolx02@mails.thu.edu.cn

ABSTRACT

Automatic evaluation of perceptual similarity is crucial for music retrieval. However, previous works mainly focused on the similarity of timbre and rhythm but not the musical pattern of a song, such as melody and chord. In this paper, we propose a new feature, chroma histogram, to summarize the musical pattern and use a transposition-invariant matching method to compare two chroma histograms. Experiment results demonstrate the efficiency of this method in measuring the similarity of musical pattern.

***Index Terms*—** Music, Acoustic signal analysis, Signal representations, Pattern matching

1. INTRODUCTION

With the increasing popularity of large databases of digital music, there is an growing need of intelligent systems which could quickly search desired song from such database. One crucial problem of building searching systems is how to automatically calculate the perceived similarity of music.

So far, several approaches of measuring similarity have been developed. The similarity factors they focus on can be roughly classified into two categories: timbre similarity [1] [2] [3]and rhythm similarity [4] [5]. Logan and Salomon [1] summarize a piece of music by using 16 typical spectral envelopes which are determined by k-means clustering, and use Earth Mover's Distance (EMD) to compute the distance between two pieces. In [2], a 3-center Gaussian Mixture Model (GMM) with diagonal covariance is used to summarize the spectral envelopes of a piece of music. Different GMMs are compared by computing the likelihood that samples from one distribution were generated by the other. The spectrum histograms (SH) [3] summarize the spectral shape by counting how many times a loudness level is reached or exceeded in the frequency bands. Different SHs are compared by calculating the Euclidean distance between them. Periodicity Histograms (PH) [4] and Fluctuation Patterns (FP) [5] are introduced to describe the rhythm pattern of music, and compared by using Euclidean distance.

Although the capacity of these approaches has been attested on some tasks like genre classification, they do not characterize the musical pattern such as melody and chord. Generally speaking, the musical pattern is the prior factor when people compare different pieces of music. One typical case is that different interpretations of the same piece of music are regarded very similar. Melody and chord are also efficient features for people to distinguish songs from different genres such as Blues, Jazz and Folk. Furthermore, musical pattern contains rich information related to human feelings. Thus, it can be used to measure the high-level music similarity such as emotion. Our goals are to design a feature which is able to capture the musical pattern of music, and to find a matching method to simulate the human's perception of musical patterns. To implement such a system, one has to account for the following fundamental issues. First, musical pattern is very robust to variations of parameters such as timbre and tempo. Second, the human's perception of musical pattern is irrelevant to the key transposition. For example, people usually think there is little difference between a song in C and its counterpart in G.

In this paper, we propose a new feature chroma histogram (CH), which is a long term statistics of chroma [6], to summarize the musical pattern of music. Chroma is the energy distribution on 12 pitch classes, and robust to the variation of timbre. The proposed feature absorbs this nice property of chroma and in addition, as a long term statistics of chroma, CH is robust to the variation in tempo. To eliminate the effects of key transposition, we present a transposition-invariant matching method to compare two CHs. The results of experiments show that our algorithm is efficient to evaluate the musical pattern similarity.

The rest of this paper is organized as follows. The procedure of feature extraction is presented in Section 2. The transposition-invariant matching method is introduced in Section 3. The experiment result is presented in Section 4 and Section 5 concludes the paper.

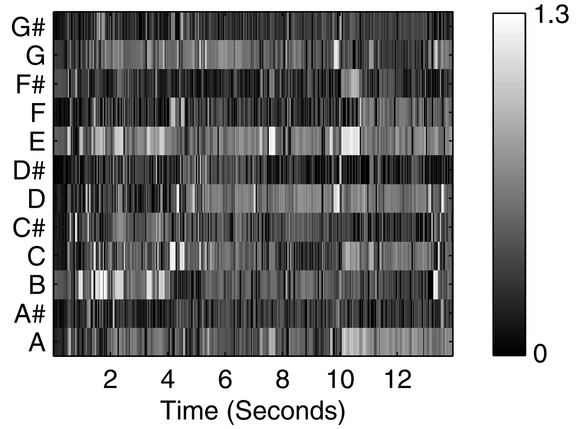

Fig. 1. Portion of the chroma features for Robbie Williams'song, "Better man"

2. AUDIO FEATURES

In this section, we will give a detailed description on the design of audio features. We proceed in two stages: first, we convert an audio music file into a sequence of chroma features. Then, we compute the large scale feature chroma histogram . We use MATLAB code provided by Dan Ellis ($www.ee.columbia.edu/\ dpwe/resources/matlab/chroma-ansyn$) to accomplish our first task. The algorithm is briefly reviewed in the following and for more details please see [7].

2.1. Chroma feature

First, a song is segmented into a series of 100ms frames with overlap of 75ms. For each frame, we calculate a 12-element chroma vector to capture the dominant note as well as the broad harmonic accompaniment. Chroma vector records the intensity associated with each of the 12 chroma bins within an overall octave, which is obtained by folding all octaves together.

One significant difference between [7] and other chroma calculating algorithm is that, it uses instantaneous frequency within each Fast Fourier Transformation (FFT) bin to locate strong tonal components in the spectrum and to get a higher resolution estimate of the frequency. Accounting for the situation when a piece is played slightly out of tune, [7] adjusts the mapping of frequencies to chroma bins for each piece of music by up to ± 0.5 semitones so that the strongest frequency peak from a long analysis window can line up accurately with a chroma bin center. Fig. 1 is a portion of chroma features of Robbie Williams' song, "Better man".

2.2. Chroma histogram

After the first stage, a piece of music is converted into a sequence of 12-element chroma features. The vectors are denoted as $CM = \{c(1), c(2), \ldots, c(l)\}$, where l is the total number of frames and $c(i)$ is a chroma feature. A CM uniquely characterizes a piece of music. However it is too specific for individual songs and cannot represent the common properties of a group of music. Therefore, we introduce a large scale feature, chroma histogram, to capture the common pattern of similar music. The process of extracting feature is given below:

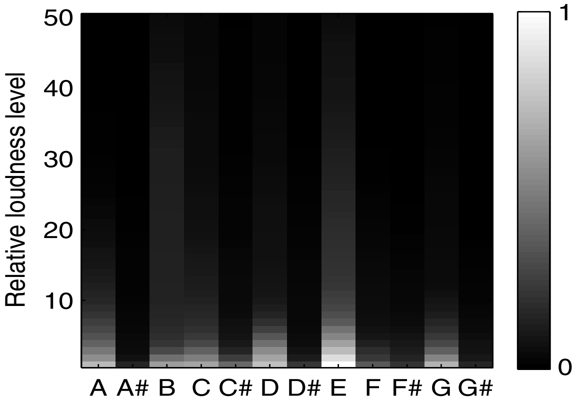

Fig. 2. Chroma histogram derived from the chroma features showed in Fig. 1

1. Elements of CM are normalized in [0, 1], and quantized with a loudness resolution of 50. For example we assign the value 20 to $CM(i,j)$ if 0.38< $CM(i,j)$ <0.42. We denote this quantized chroma feature matrix CM^q.

2. Partition CM^q into 2^N sub-sequences of chroma features with overlap.

3. For each sub-sequence, we use a 2-dimension chroma histogram to summarize the melody and chord pattern. A chroma histogram has 12 columns for 12 chroma bands and 50 rows for loudness resolution. The histogram counts how many times a specific loudness in a specific chroma band was reached or exceeded. The sum of histogram is normalized to 1. Fig. 2 is the chroma histogram derived from the chroma matrix Fig. 1. In the end, a piece of music is represented by 2^N chroma histograms.

It is necessary to bring Step2 for discussion. Compared with computing only one chroma histogram for one song, using a series of chroma histograms is more reasonable. First of all, a series of chroma histograms can preserve both the structure information and the sequence information of CM^q. Second, it is often the case that a song is performed in more than one key. Step2 guarantees that every sub-sequence is in only one key and facilitates the transposition-invariant matching in Section 3. Choosing the parameter N is crucial. If N

is too small, the aforementioned advantages are not apparent. On the other hand if N is too large, the feature will fail to describe the common characteristics shared by a group of songs. Experiments show that, $N = 3$ will lead to a good trade off.

3. MATCHING METHOD

As mentioned above, we obtain a series of chroma histogram for every song, denoted as $S = \{CH^1, CH^2, \ldots, CH^8\}$. We then compare the chroma histograms representation of two different songs. Since similar songs usually have similar structure (e.g. ABAB), to take advantage of structure information, we sequentially compute the distance between 2 corresponding chroma histograms. The formula is as following:

$$Dist(S_1, S_2) = \sum_{i=1}^{8} D(CH_1^i, CH_2^i),$$

Where $S_1 = \{CH_1^i\}$, $S_2 = \{CH_2^i\}$. And $D(.)$ is the transposition-invariant distance, which will be introduced below.

3.1. Tranposition-Invariant Distance

In a general sense, human's perception of a piece of music is irrelevant to its key. For example, people usually think there is little difference between a song in C and its counterpart in G. If we cyclically shift the C version rightwards by 7 semitones, the chroma histograms of these two versions will be exactly the same. Thus, we can derive a transposition-invariant distance measure by using this property of chroma histogram.

For a chroma histogram $CH = \{v(1), v(2), \ldots, v(12)\}$, where $v(i)$ corresponds to the ith column of CH, we define a transposition function as

$$f^1(v(1), v(2), \ldots, v(12)) = \{v(12), v(1), \ldots, v(11)\}.$$

Accordingly, the i-transposed version of CH is $f^i(CH)$, and $f^{12}(CH) = CH$. In the end, we define the transposition-invariant distance of two chroma histograms as:

$$D(CH_1, CH_2) = min_{i \in [0:11]} d(CH_1, f^i(CH_2)),$$

where $d(x, y)$ is the Euclidean distance between x and y.

4. EXPERIMENTS

Two experiments are designed to evaluate the ability of our method to capture the musical pattern. The first one is to identify "cover songs", the alternate interpretations of the same underlying piece of music. Cover songs typically keep the essence of the melody and the chord, but may vary greatly in other aspects such as timbre, tempo and key. In addition, similarity of musical pattern exists not only among songs with the same underlying melody. Human can measure the similarity between songs with different melodies. Thus, a more general experiment is designed to test whether our method can effectively evaluate the similarity of musical pattern among songs with different melodies.

In the experiments, we manually cut off the intro and outro of every song, because these two parts contain little melody information and vary a lot from song to another. Besides, automatically cutting off these large scale non-vocal parts is feasible by using vocal part detection methods. Other similarity measures [1, 2, 3, 4, 5] are also tested for comparison. We use the implementation of these similarity measures provided by Elias Pampalk ($www.ofai.at/\ elias.pampalk/ma/$).

Two metrics are used for evaluation. The first one is the recall on top 5 returns. Each song is compared with all other songs in test sets and the 5 most similar songs are returned. The recall is computed as:

$$recall = \frac{\#hits}{total\#similar\ songs},$$

where $\#hits$ is the number of songs in the same group as the query song on the return list and $total\#similar\ songs$ is the total number of songs in the same group as the query song. A rate of one means all similar songs of the query song are on the return list.

Another metric used for evaluation is intra-inter distance ratio [8]:

$$iir = \frac{intra - distance}{inter - distance},$$

where intra-distance is the average distance within a group and inter-distance is the average distance between all pieces. A ratio of one means that the distance between arbitrary pieces is the same as that between members within a group.

4.1. Cover song identification

We choose 14 pop songs for the first experiment, 6 English songs and 8 Chinese. And particularly, the cover song of Robbie Williams' "Better man" is a Chinese version sung by Sandy Lam. The reason why we choose songs in different languages is that we believe musical pattern should be irrelevant to language.

In this experiment, songs with the same title are categorized in the same group. The recall of every song and the intra-inter distance ratio of each group are computed. By using our method, all songs except two have a recall of one, including two versions of "Better man". The failed 2 songs are of poor sound quality, which makes it difficult to extract good chroma features. Statistics of results of all similarity measures are shown in Fig. 3. It shows that the chroma histogram (CH) outperforms all other measures. One explanation is that in this test set, songs with the same title are quite similar in musical pattern but vary greatly in timbre and rhythm. The result supports the idea that the chroma histogram is able to

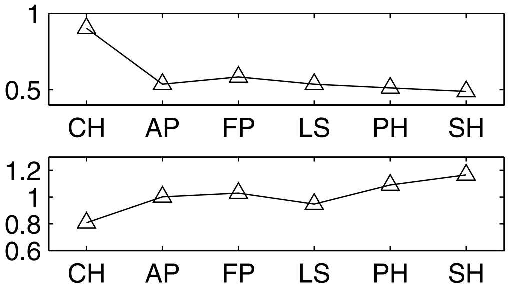

Fig. 3. Results of cover song identification.The top pane shows the average recalls of 6 similarity measures. The bottom pane shows the intra-inter distance ratios of 6 similarity measures.

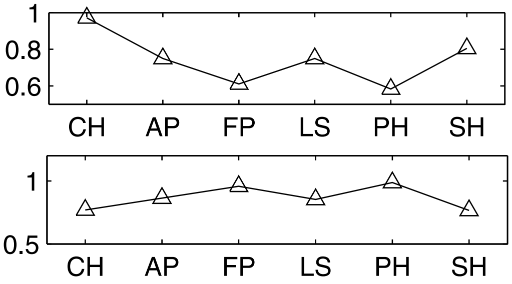

Fig. 4. Results of general evaluation of similarity measure.The top pane shows the average recalls of 6 similarity measures. The bottom pane shows the intra-inter distance ratios of 6 similarity measures.

describe the common character of a song and its cover versions.

4.2. General evaluation of similarity measure

We choose 12 songs from genres of American Folk song, Blues, and Jazz since there are sharp differences of the musical patterns among these 3 genres. Songs from the same genre are regarded in the same group. The results are illustrated in Fig. 4. Rhythm related (PH and FP) measure perform the worst because they cannot characterize the common property of the selected genres. And it is not surprising that timbre related measures (AP, LS and SH) have relatively good performance on this test set, since huge timbre gaps exist among Blues, Jazz and Folk. The performance of our method (CH) is even better than timbre related measures. One possible reason is that, the difference of musical pattern is shaper than the difference of timbre among the selected 3 genres. The result proves that our method is able to measure the similarity of musical pattern of songs with different melodies. This result also indicate that genre classification is a possible application of musical pattern related measures.

5. CONCLUSION AND FUTURE WORK

In this paper, we introduced a novel feature, chroma histogram to capture the musical pattern of a piece of music. A transposition-invariant matching method is proposed to compute the distance between chroma histograms. Compared with other similarity measures which mainly focused on timbre and rhythm, our method efficiently measures the similarity of musical pattern. Different from objective features like timbre and rhythm, musical pattern contains rich information related to human feelings. We believe that many higher level properties of music could be revealed by exploring musical pattern. Therefore our future goal is to design new method to extract additional information from music pattern.

6. REFERENCES

[1] B. Logan and A.Salomon, “A music similarity function based on signal analysis,” *Proc. IEEE Intl Conference on Multimedia and Expo*, 2001.

[2] J.-J. Aucouturier and F. Pachet, “Music similarity measures: What’s the use?,” *Proc Intl Conference on Music Information Retrieval*, 2002.

[3] Elias Pampalk, Simon Dixon, and Gerhard Widmer, “Exploring music collection by browsing different views,” *Proc Intl Conference on Music Information Retrieval*, 2003.

[4] E.D. Scheirer, “Tempo and beat analysis of acoustic musical signals,” *Journal of Acoustical Society of America*, vol. 103, no. 1, pp. 588–601, 1998.

[5] E. Pampalk, A. Rauber, and D. Merkl, “Content-based organization and visualization of music archives,” *Proc ACM Multimedia*, 2002.

[6] Mark A. Bartsch and Gregory H. Wakefield, “Audio thumbnailing of popular music using chroma-based representations,” *IEEE Transactions on multimedia*, vol. 7, no. 1, pp. 96–104, 2005.

[7] Daniel P.W. Ellis and Graham E. Poliner, “Identifying ’cover songs’ with chroma features and dynamic programing beat tracking,” *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2007.

[8] Elias Pamoak, Simon Dixon, and Gerhard Widmer, “On the evaluation of perceptual similarity measures for music,” *Proc. of the 6th conference on Digital Audio Effects*, September 2003.