A RULE-BASED APPROACH WITH METER ESTIMATION TO TEMPO ESTIMATION FOR AUDIO MUSIC

Fu-Hai Frank Wu  
Department of Computer Science  
National Tsing Hua University  
Hsinchu, Taiwan  
frankwu@mirlab.org

Jyh-Shing Roger Jang  
Department of Computer Science  
National Taiwan University  
Taipei, Taiwan  
jang@mirlab.org

Jui-Yu Hung  
Multimedia Applications Lab. TL  
Chunghwa Telecom Co., Ltd.  
Taoyuan, Taiwan  
jyhung@cht.com.tw

ABSTRACT

For audio music with time-varying meters, tempo estimation is a challenging task. This paper proposes several new low-level features of tempogram to determine pulse the classification of duple/triple meters, which can guide the determination of the overall tempo. After the judgment of meters, then we apply k-means clustering on tempo curve to perform tempo estimation.

Index Terms – Tempo Estimation, Low-level Features, Tempogram, Tempo Curve, Duple/Triple Meter

1. INTRODUCTION

Tempo is an essential rhythmic element in music. However, automatic tempo estimation is still a challenging task, especially when the music has time-varying tempi and different duple/triple beats between regularly recurring accents (or downbeats [2]). There are different rhythmic levels such as measure, beat, and tatum which influence human perception about tempi. Music tempi are subjective to listeners. Listeners identify beat positions, which then form the sensation of tempi. The tempi defined by a listener are usually presented in the listener’s tapping the feet or clapping the hands. Such tempi are called perceived tempi, which is sometimes different from the tempi of music notation.

In the twenty MIREX06 training excerpts [3], these include a mix of genres and tempo ranges, and annotation of two tempi representing the highest peaks of distribution of perceived tempi annotated by a group of listeners. There are non-duple meters in the excerpts, so the two tempi could have duple or triple relation. There are audio excerpts with quite low pulse clarity, while novice listeners have difficulty to tapping the beats regularly and are hard to obtain clear tempo. Gouyon et al. [1] propose a method to discriminate duple and triple meters of audio signals. They extract two types of low-level features which are named as frame descriptor and beat segment descriptor. Then they use feature selection techniques to reduce the number of descriptors. The beat segment descriptors are used to compute periodicity by ACF (Auto Correlation Function) with beat lag indexes. Lartillot et al. [4] use dozens of descriptors computed by detection function of the state-of-the-art researches and set up a composite model to explain the judgments of pulse clarity from those descriptors.

There are important previous studies that attempted to deal with tempo estimation. Peeters [5] proposed a reassigned spectral flux to detect onset events. The rhythmic meter, beat, and tatum are estimated by meter/beat templates and a Viterbi algorithm. Cemgil et al. [6] model the tempo estimator as stochastic dynamic system. Tempi are treated as hidden state variable and estimated by Kalman filter which operated on tempogram. The tempogram representation interpreted as the response of comb filter bank and is analogue to the wavelet transform. Chordia and Rae [7] use probabilistic latent component analysis (PLCA) to do source separation. Each source is treated as a component which is analyzed to obtain the tempo candidates by autocorrelation-based methods. All of the tempo candidates with information of pulse clarity from different components are clustered to do final tempo estimation. Eronen and Klapuri [8] use K-NN regression with a resampling step for periodicity vectors of training data. They also proposed a method to remove outlier in training process.

Under these mechanisms, the approach of our tempo estimation is accomplished by three phases. In the first phase, the onset strength [9] of music along time, called novelty curve, is generated to indicate the possible rhythmic pattern. In the second phase, the quasi-periodic patterns in novelty curve are analyzed to discover the possible tempi. The novelty curve is transformed into frequency domain to obtained tempi information, so call tempogram. The most prominent tempi are derived by DP to obtain so-called tempo curve. In the last phase, the values of tempo curve are clustering by K-means clusters for different cluster number from 2 to 4. The centers of those clusters are analyzed and manipulated by some learning rule from MIREX06 excerpts. Those rules are based on relative strength and ratio between tempi, assumption of maximum tempo 280 BPM and the strength statistics of tempogram. In this study, first two phases of the framework is similar to beat tracking work [10]. Section 2 describes the details of the low-level features and the tempo rule for final tempi peaking.
2. LOW-LEVEL FEATURES AND TEMPO RULES

In this section, we present the low level features to obtain estimated meters and the proposed tempo rules derived from using the training dataset of the MIREX06 Audio Tempo Estimation contest and operated on tempo candidates [11].

2.1 Low-level Features of Tempogram

In our study, we use statistics of the tempogram as low-level features, including tempogram mean (μT), tempogram standard deviation (σT), and tempogram sparseness (rT), which is defined as the ratio of mean over standard deviation (μT/σT).

For classification of meter, we use classifier with leave-one-out (LOO) to train the excerpts. Table I show the meter accuracy results of different low-level feature with 7 different combination {rf, μT, σT, (rf, μT), (rf, σT), (μT, σT), (rf, μT, σT)}. Figure 5 show the scatter plot with decision boundary of best result (rf, μT).

Table I.

<table>
<thead>
<tr>
<th>Features</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>rf, μT, σT</td>
<td>0.75</td>
</tr>
<tr>
<td>rf, μT</td>
<td>0.85</td>
</tr>
<tr>
<td>rf, σT</td>
<td>0.85</td>
</tr>
<tr>
<td>μT, σT</td>
<td>0.9</td>
</tr>
<tr>
<td>rf, μT, σT</td>
<td>0.85</td>
</tr>
<tr>
<td>rf, μT</td>
<td>0.15</td>
</tr>
</tbody>
</table>

2.2 Tempo Rules

The major tempo screening rules are obtained by the learning process from ground truth, which the rules combine the information from maximum tempo limit, tempoij, count(tempoij), rf, and μT to do the tempo estimation.

The rules are listed as below.

Rule 1: maximum tempo is 280 BPM, minimum tempo is 25 BPM
Rule 2: if tempoij with the same i are close to each other within some threshold (8%), combine them to be one.
Rule 3: the tempo operator (OP) for tempoij are in the set {*,2, *,3, /3} where * means multiply, l means divide.
Rule 4: if the number of tempoij within the same i is ≥ 2, check whether they (called meter pair, acronym as mj) have double or triple relation with some threshold (8%). Finally, we choose the meter pair with lowest tempi as solution.
Rule 5: Using (rf, μT) as a priori for clarity and meter judgment
Rule 6: if the number of tempoij within the same i is = 1, the tempi are {tempoij , tempoij or tempoij} which pair is decided by estimated double/triple meter prior obtained from low-level features.

3. REFERENCES