A LOCAL MAXIMUM PEAK PICKING APPROACH TO TEMPO CANDIDATES FOR TEMPO ESTIAMATION

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

ABSTRACT

For the peak picking of tempo candidates, applying k-means clustering on tempo curve is straightforward and leading to good result. But the tempo candidates obtained from tempo curve are limited and lose a lot of information for possible tempi. The study proposes the local maximum peak picking method to increase the number and information of possible tempo candidates. Therefore, the accuracy of tempo estimation increases.

Index Terms – Tempo Estimation, Local Maximum Peak Picking, Tempogram

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 meters [1], which consist of duple or 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 outliner 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 local maximum peak picking along the time. In the last phase, the histogram of those peak picking tempi is analyzed. Some tempo rules [10] are deduced basing on the relative strength of tempogram between tempi, perception assumption of maximum tempo 280 BPM. In this study, first two phases of the framework is similar to beat tracking work [11]. Section 2 describes the details of the mechanism of local maximum peak picking.

2. Tempo Local Maximum Peak picking

In this section, we describe the features to decide the local maximum peak picking tempo candidates and the peak picking process.

2.1 Features for Tempo Candidates Peak picking

In the study, we use strength and count of local maximum tempi in the tempogram as features. The strength and count are the sum of certain range of local maximum tempi which is the center of whole range. The range is defined within eight percent of the local maximum tempo by default. The usual tempogram and local maximum tempi are as Figure 1 (a) and (b).

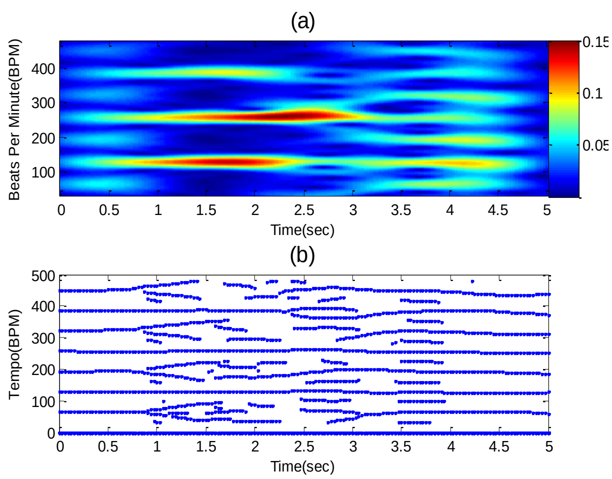

Fig.1 (a) The tempogram obtained from the novelty curve (b) Local maxima of each frame of the tempogram

2.2 Peak picking Process

In each frame, there are many local maximum tempi. First we sort those candidates and picking top N of tempi. In our experiments, the N is within 10 is ok for good results. After the tempi are selected, we use the rules as below to obtain find tempi.

Rule 1: maximum tempo is 280 BPM, minimum tempo is 25 BPM

Rule 2: if $tempo_{ij}$ with the same i are close to each other within some threshold (8%), combine them to be one.

Rule 3: if the number of $tempo_{ij}$ within the same i is ≥ 2, check whether they (called meter pair, acronym as m_p) have double or triple relation with some threshold (8%). Finally, we choose the meter pair with lowest tempi as solution.

3. REFERENCES

[1] F. Gouyon and P. Herrera, "Determination of the meter of musical audio signals: Seeking recurrences in beat segment descriptors," in *Proc. AES 114th Conv.*, Amsterdam, The Netherlands, 2003.

[2] Geoffroy Peeters and Helene Papadopoulos, "Simultaneous Beat and Downbeat-Tracking Using a Probabilistic Framework: Theory and Large-Scale Evaluation" *IEEE Transactions on Speech and Audio Processing,* Vol. 19, No. 6, August 2011.

[3] MIREX 2011 Audio Tempo Estimation. http://www.music-ir.org/mirex/wiki/2012:Audio_Tempo_Estimation

[4] Olivier Lartillot, Tuomas Eerola, Petri Toiviainen, and Jose Fornari, "Multi-feature modeling of pulse clarity: Design, validation, and optimization" in *Proc. ISMIR*, Pennsylvania USA, 2008.

[5] G. Peeters, "Template-based Estimation of Time-Varying Tempo" *EURASIP Journal on Advances in Signal Processing*, Vol. 2007, pages 158–171, 2007.

[6] A.T. Cemgil, B. Kappen, P. Desain, and H. Honing, "On Tempo Tracking: Tempogram Representation and Kalman Filtering" *Journal of New Music Research*, Vol. 28(4), 259-273, 2001.

[7] Parag Chordia and Alex Rae, "Using Source Separation to Improve Tempo Detection" in *Proc. ISMIR*, pages 183–188, Kobe, Japan, 2009.

[8] Antti J. Eronen and Anssi P. Klapuri, "Music Tempo Estimation With k-NN Regression" *IEEE Transactions on Speech and Audio Processing,* Vol. 18, No. 1, January 2010.

[9] Juan Pablo Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, Mark B. Sandler, "A Tutorial on Onset Detection in Music Signals" *IEEE Transactions on Speech and Audio Processing,* Vol. 13, No. 5, September 2005.

[10] Fu-Hai Frank Wu, Jyh-Shing Roger Jang, "A Rule-Based Approach With Meter Estimation To Tempo Estimation For Audio Music" in MIREX 2012.

[11] Fu-Hai Frank Wu, Tsung-Chi Lee, Jyh-Shing Roger Jang, Kaichun K. Chang, Chun Hung Lu, Wen Nan Wang, "A Two-Fold Dynamic Programming Approach to Beat Tracking For Audio Music with Time-Varying Tempo" in *Proc. ISMIR*, Florida, USA, 2011.