A STATISTIC LEARNING APPROACH TO TEMPO ESTIMATION FOR AUDIO MUSIC

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

ABSTRACT

Automatic beat tracking and tempo estimation are challenging tasks, especially for audio music with non-binary tempo or weak percussion. This paper proposes a K-means clustering approach to handle tempo estimation with one-third/triple tempo or weak percussion. In particular, the first stage is to compute the tempo curve from the tempogram by DP(Dynamic Programming). Then use K-mean clustering to extract tempo candidates with 2 to 4 clusters. Finally, make tempo rules discovery to match ground truth of audio dataset by learning approach. The tempo estimation algorithm could almost obtain maximum 1.0 pscore value for mirex2006 tempo training dataset with 20 excerpts in length of 30 seconds.

Index Terms – Tempo Estimation, Tempogram , Tempo Curve, K-means Clustering, Dynamic Programming

1. INTRODUCTION

Tempo is essential element in music. Such information is useful in several applications such as query by tempo (querying a large database based on tempo). However, automatic tempo estimation is still challenging tasks when the music has time-varying tempos.

There are several important previous studies that attempted to deal with tempo estimation. Peeters [1] proposed a reassigned spectral flux to detect onset events. The study also proposed a combination of DFT and frequency-mapped autocorrelation to estimate periodicity. The rhythmic meter, beat, and tatum are estimated by meter/beat templates and a Viterbi algorithm. Cemgil et al. [2] 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. Groshe and Muller [3] used the novelty curve to generate predominant local pulse (PLP) for estimating time-varying tempos. The PLP curve is mid-level representation. The PLP curve solved the problem of noisy novelty curve and obtained more prominent tempogram. Klapuri et al. [4] used the bandwise time-frequency method to obtain accentuation information, then used comb filter resonators and probabilistic models to estimate pulse width and phase of different music meters, including tatum, tactus, and mesurement.

Music tempi are subjective to listener. Listener identifies beat positions, then, those beat positions comprise tempi. The tempi defined by listener usually reflected to tap the feet or clap the hands by listener. Such tempi are called perceived tempi to differentiate the tempi of music notation. Under this mechanism, the approach of tempo estimation is accomplished by three phases. In the first phase, the onset strength [5] of music along time, called *novelty curve*, is generated to indicate the possible positions of beat. The study adds a statistic way to qualify the periodicity of novelty curve. 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 picked up by some learning rule from dataset. Those rules are based on relative strength between tempi, assumption of maximum tempo and the distribution of tempi value. The dataset is mirex2006 tempo training dataset with 20 excerpts in length of 30 seconds which is provided by MIREX wiki. The dataset includes non-binary(triple) meters and weak beat position. The values of tempo curve and strength statistic data of tempogram have information to discriminate from excerpt with binary tempi and strong periodicity tempi. The confidence level of those rules are high for the dataset. Because the whole process is effective for the dataset so far, the pscore value for this dataset could approach almost 1.0 maximum. Due to limitation of the dataset, the rules by learning from those dataset could be not general for broaden genre or variation of tempi range.

In this study, the three-phase framework is similar to beat tracking work [6]. The previous work have good result for time-verying tempi The remainder of this paper is organized as follows. Section 2 describes the details of the proposed framework.

2. SYSTEM DESCRIPTION

The proposed tempo estimation system is shown in Figure 1.

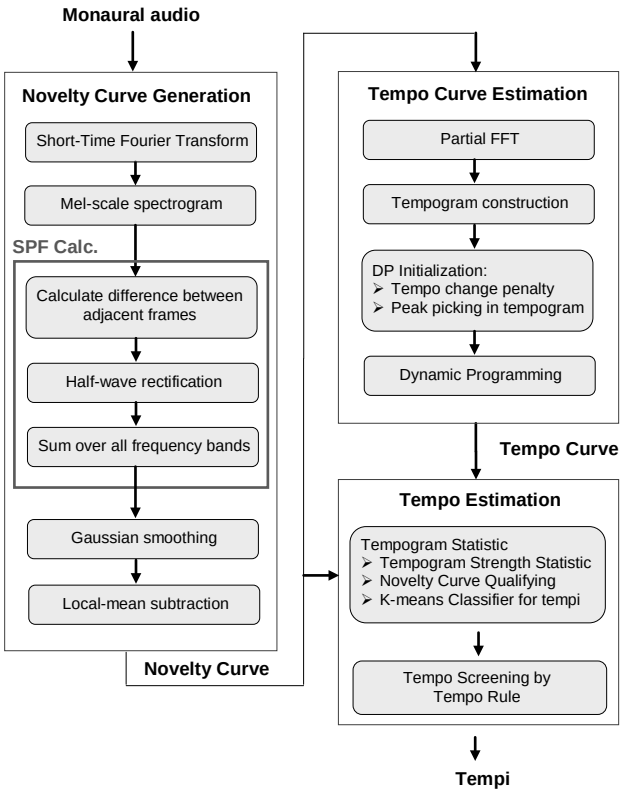

Figure 1. Flowchart of Tempo Estimation System

The first block computes the novelty curve, while the second block generates the tempogram and estimates the tempo curve from the novelty curve [6]. The major difference is the penalty setting is lower than the beat estimation for rule learning and tempo estimation. In the third block, tempi are estimated by using the information from previous two blocks and tempo rules. The concept of tempo estimation block will be explained in the following subsection.

2.1 Tempo Estimation

The stage of this phase begins from strength statistic of tempogram. Mean(μ) and standard deviation(ρ) are obtained by the statistic process. The criterions of different expert are setting by learning from ground truth of the dataset. One of the criterions is for the qualification of novelty curve. If it isn't pass the criterion, the novelty curve will be regenerated from original monaural audio signal with different filtering parameter. Then regenerate the tempogram and tempo curve.

Figure 2 shows typical outputs of tempo curve in tempo curve estimation phase. The tempo curve is obtained by little penalty $\theta = 0.00\,5$ for tempo change. The cluster centers are also shown in the Figure 2. K-means classifier is used to get the candidate of tempi with clusters number from 2 to 4. Those centers of clusters are passed to tempo screening rules to get final two tempi values. The major tempo screening rule obtained by the learning process is that the double of lower center is within the threshold of higher center value. The one-third/triple tempo rule is tempogram strength rati $= \mu/\rho$ larger than a threshold and ρ is less than a threshold.

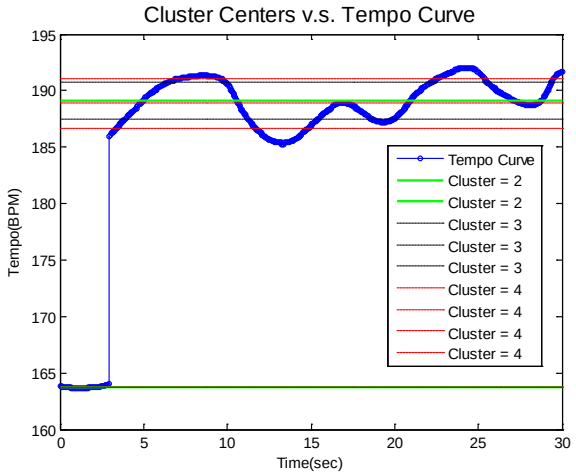

Figure 2. Cluster Center and TempoCurve(with $\theta = 0.005$)

3. REFERENCES

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

[2] 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.

[3] P. Grosche and M. Müller, "A Mid-level Representation for Capturing Dominant Tempo and Pulse Information in Music Recordings" in *Proc. ISMIR*, pages 189–194, Kobe, Japan, 2009.

[4] M.F. McKinney, D. Moelants, M.E.P. Davies and A. Klapuri, "Evaluation of audio beat tracking and music tempo extraction algorithms," *Journal of New Music Research*, Vol. 36, no. 1, pp. 1-16, 2007.

[5] 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

[6] 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, Accepted.