# A METHOD OF THE COMPONENT SELECTION FOR THE TEMPOGRAM SELECTOR IN TEMPO ESTIMATION

**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

In the study, we propose a method to select components from datasets to form two kinds of categorizes {reqProcessing, noProcessing} for tempogram selector, which is built from a classifier. The research utilizes the low pass filtering (LPF) as the processing, so there are two tempogram with LPF or without LPF. The trained classifier is used to select the best out of the tempograms, which represents the results - 2 tempi and strength of tempo estimation. The submission is 1st place of at-least-one-correct index and 2nd place of P-score index in MIREX 2013 audio tempo estimation contest.

Index Terms – Tempo Estimation, Tempogram Selector, Component Selection

## 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.

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 and their relative strength are derived from the tempogram. In the third phase, a tempo selector chooses the best solution from the results of two paths: with processing and without processing. In this study, first two phases of the framework is similar to beat tracking work [10] and the third phase is derived from the work [11]. Section 2 describes the flow char and the process of the component selection for tempogram selector.

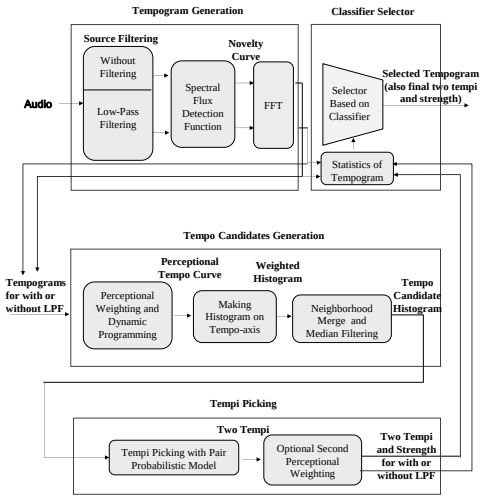


**Figure 1 Flowchart of the tempo estimation system**

## 2. COMPONENT SELECTION FOR THE TEMPOGRAM SELECTOR

Figure 1 illustrates the block diagram of the proposed method. The first 'Tempogram Generation' block has two audio processing paths: with or without the low-pass filtering (LPF). Following the source filtering, the spectral-flux detection function generates the onset detection function [13] of the music over time as in Figure 2 (a). The STFT then transforms the function into a frequency domain to obtain the so-called tempogram as in Figure 2 (b) which is the specified spectrogram of the onset detection function and represents the periodicity of onsets.

In the second block 'Tempo Candidates Generation', a perceptional weighting window is applied to the tempogram. Then, the most prominent tempi are derived by dynamic programming (DP) on the perceptional tempogram along time to obtain the so-called tempo curve (as in Figure 4 (a). The histogram of the tempo curve is made on the tempo-axis; the tempi within the specified threshold are merged and median filtered as new tempo bin.

The third 'Tempi Picking' block uses the pair probabilistic model to calculate the likelihood of all existing pairs and picks the maximum likelihood pair as the last two tempi. At last step, the perceptional weighting is performed for calculating the strength. Finally, the 'Classifier Selector' block uses the tempogram features to select the final tempi and the strength from those of "with LPF" and "without LPF".

### 2.1 COMPONENT SELECTION

The training data preparation and the classification proceed as follows:

1. Training data preparation: This stage verifies every excerpt of the training set to determine whether the LPF could improve the accuracy of the tempo estimation. This selects some of the original excerpts into two classes {"requires LPF", "no LPF"}. Tracks classed as "requires LPF" have their accuracy improved by LPF, while LPF decreases the accuracy of tracks classed as "no LPF". Excerpts, showing no difference of accuracy are excluded from the training process. The process of labeling goes through the whole flow of the tempo estimation by using post processing of the proposed algorithm following the "with and without LPF source filtering block" to extract the low-level features ( $\mu_T$ , $\sigma_T$ , $cv_T$ , $\gamma_T$ and $\kappa_T$ ) for training, and then compares the accuracy to annotate groundtruth classes.
2. Classification: The *k*-Nearest Neighborhood Classifier (*k*-NNC) is used for feature selection, training, and validation. The combinations of the low-level features are formed as the input features to a classifier.

## 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, 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.

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