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 1\textsuperscript{st} place of at-least-one-correct index and 2\textsuperscript{nd} 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 duplet or triplet 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 operates 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.

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.
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 classified as “requires LPF” have their accuracy improved by LPF, while LPF decreases the accuracy of tracks classified 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 $(\mu_T, \sigma_T, c_{TV} + y_T$ 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


http://www.musicir.org/mirex/wiki/2012-Audio_Tempo_Estimation


