SOLUTION FOR DISCRIMINATION FUNCTION OF TEMPO-PAIR ESTIMATOR

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

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

ABSTRACT

The tempo pair vector (tpv) feature and discrimination function were designed for tempo-pair estimation. The solution for discrimination function and the formation of tempo pair vector were conducted by parameter grid search in preliminary design and experiment stage. Many options are possible for the solution to the task such as multi-class classification, regression and learning-to-rank. The straightforward try of multivariate regression is conducted in this study. The initial result show that the accuracy 0.9357 for at-least-one-correct is less than 0.9517 of parameter sweep method.

1. INTRODUCTION

We have found the long term periodicity (LTP) function could be utilized to generate statistical vector to estimate tempo pair. Although the whole processes include general steps such as filtering preprocessing, onset detection, perceptional weighting, and tempogram generation. Eventually, we could obtain tempo pair vector (tpv) which is histogram between tempo candidates derived from LPT function. In previous work, we deduce a discrimination function and design a Gaussian function with parameters decided by maximizing the ACC2 accuracy for dataset.

In this study, we attempt to solve the problem in close form with regression method. Fortunately, the result in Table 1 shows the difference between parameter grid search (MIREX 2013) and regression (MIREX 2014) have reasonable difference, which could come from un-noticed parameter of other processes, for example, onset detection. In the future work, we will find the real cause and explore the accuracy improvement further. The following section describes the tempo pair estimator.

2. TEMPO PAIR ESTIMATOR

The interesting parts of previous work are tpv feature and the connection to the predominant tempo class. The foundation of this derivation majorly depends on tempogram, which shows the potential periodicity of music. The components of tpv represents the likelihood of the two predominant tempi of the music with the specific ratios between the two tempi in the tempo classes: 'duple', 'triple', '3⁄2', and 'other', which are abbreviated as $\omega_d, \omega_t, \omega_q, \omega_o$, respectively. Not going into detail as the previous work [1], we shortly recapitulate the derivation as follows:

The 'Tempo-pair Estimator' are composed of two blocks as Figure **1** shown. The first block 'Tempogram Generation' has quite standard processing steps, such as onset-detection, short-time Fourier transform (STFT) to transform the onset-detection function (ODF) into a frequency domain to obtain the so-called tempogram to explore periodicity. Then, a perceptual weighting window is applied to the tempogram to reflect human perceptional preference. In the second block 'Tempo-pair Generation', the most salient tempi are derived by summing over the strength of tempogram along time axis to obtain long-term periodicity function and to locate the tempi with local maximum of the LTP function. Then the tempi within the specified threshold are merged as the tempo candidates and the LTP function is normalized to be probability mass function to represent the likelihood of the tempo candidates. Finally, the tempo-pair statistical model to calculate the likelihood of all existing pairs between the tempo candidates and picks the maximum likelihood pair.

2.1 Discrimination Function for Tempo-pair Identifying

We propose a discrimination function to formulate the tempo-pair statistical model in previous work and denoted as:

Table 1 MIREX evaluation result of audio tempo estimation for years of 2013 and 2014

MIREX 2013				MIREX 2014			
Submission Code	P-Score	At-Least-One Correct	Both Correct	Submission Code	P-Score	At-least-One Correct	Both Correct
FW3	0.8263	0.9517	0.55	FW1	0.8185	0.9375	0.55

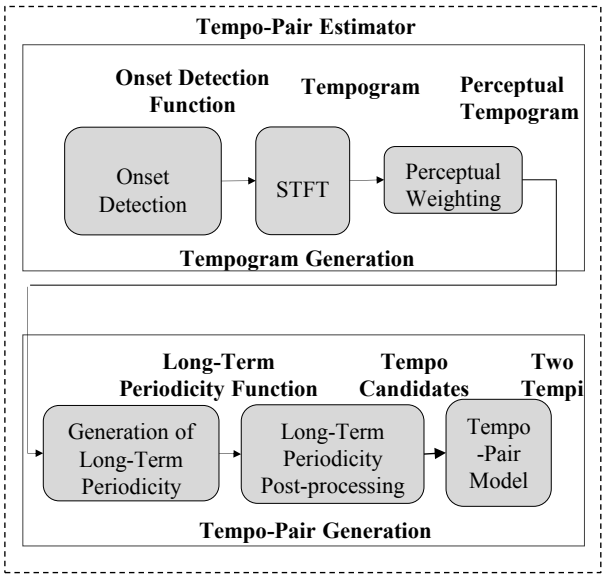

Figure 1. Flowchart of proposed method

- Defining the tempogram vector $tpv = \left[p_d, p_t, p_q, p_o\right]^T$
- Formulating the M-class, which M is equal to four, with the linear discriminant functions

$$\boldsymbol{g_i(tpv) = W_i^T * tpv}$$

, where are W_i weighting vector$\in R^4$; $i \in \{d, t, q, o\}$

- Maximize the specified accuracy ACC of dataset,

$$ACC = \frac{1}{N}\sum_k f_{measure}\left(g_i(tpv_k)\right)$$

, where $f_{measure}$ includes a selection mechanism, and the evaluation metric; k is the index of audio excerpts in training set. The selection mechanism used in the study is to select the class with maximum g_i.

2.2 Solution Moved from Parameter Grid Search to Multivariate Linear Regression

The submission FW1 in MIREX 2013 ate is conducted by a Gaussian function based model to simply presume the weighting matrix as diagonal matrix with component related to tempo-pair count with class C_i as follows and design a grid search to maximize ACC2 for training dataset. The component of tpv is the maximum sum of LTP within class.

$$\boldsymbol{\mu = mean(C_i)}$$

$$\boldsymbol{\rho = var(C_i)}$$

$$\boldsymbol{\widehat{C}_i = \mu - k_i * \rho}$$

$$\boldsymbol{w_i = \widehat{C}_i / \sum_k \widehat{C}_k}$$

, where $i \in \{d, t, q, o\}$.

Because the tempi within different classes has some common tempo, the classes are correlated with each other. The initial step to presume the diagonal weighting matrix is too simplified to an optimized solution. So, among of possible solutions, for example multi-class classification, regression and learning-to-rank, we adopt multivariate linear regression in order to match the form of discrimination function straightforwardly. In order to decouple with the predominant tempo estimation, the measure metrics are calculated by two tempo individually and summed together as output. Therefore the dimension of responses Y is N by 4, and the predicator X is (1+N) by 4, where N is dimension of dataset and the number of tempo-pair class. The response of each component of is evaluated by summing up ACC2 of each tempo in tempo-pair. While the predicator is the component of tpv. To summarize, the model is solved the equation as follows by general linear regression with ordinary multivariate normal estimate.

$$\beta = mvregress(X, Y)$$

In test phase, response $y = [1\ tsv^T] * \beta$ and the estimated tempo-pair is within the class with maximum components.

In addition to regression model for groundtruth with one tempo, there are estimated tempo-pair groundtruth of peer researcher. We try to adapt our model to take advantage of the groundtruth, but no improvement in the submission FW2.

3. REFERENCES

[1] Fu-Hai Frank Wu, Jyh-Shing Roger Jang, "A Supervised Learning Method into Tempo Estimation of Musical Audio", Control and Automation (MED), 2014 Mediterranean conference on, IEEE published.