# A TEMPO-PAIR ESTIMATOR WITH MULTIVARIATE REGRESSION

*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 two submitted algorithms have two different purposes. One includes an extended model to learn tempo pairs by two tempi of the hypothetic ground truth from the downbeat and beat annotation of training set and adding the speed class training. The results are not successful because the hypothesis of ground truth seems to be questionable. On the other hand, The two is an improved version with the multivariate regression and general single-tempo ground truth prove the potential again by the record performance, P-Score 0.8347, compared with those of our previous submissions.

## 1. INTRODUCTION

We have invented the long term periodicity (LTP) function and tempo pair model to estimate tempo pair. The whole processes include general steps such as filtering preprocessing, onset detection, optional perceptional weighting, and tempogram generation. Eventually, we could obtain tempo pair vector (tpv) which is the histogram of two-tempo candidates derived from LPT function. In previous work FW1 of MIREX 2014, we tried to use the multivariate regression to model the discrimination function by maximizing the ACC2 accuracy for dataset with the worse performance than that of grid search with the Gaussian modeling.

In this study, we have two purposes. One (submission FW5) is to improve the performance of the multivariate regression and maximized ACC2 based solution. The other (FW1) is to and a new speed class to improve Both-Correct index, which imply the accuracy of meter and speed of tempo. The results show that FW5 is successful as Table 1 owing to bugs fixed and bigger training sets which the work [1] used; FW4 is not successful because of the inappropriate hypothesis of two tempo groundtruth majorly derived from the downbeat and beat annotation of ISMIR 2004 “Ballroom” and “Songs” datasets. In the following section, we introduce 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 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 [2][3], 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 optionally. 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.

**Table 1 MIREX evaluation result of audio tempo estimation for years of 2014 and 2015**

| MIREX 2014 | | | | MIREX 2015 | | | |
|---|---|---|---|---|---|---|---|
| Submission Code | P-Score | At-Least-One Correct | Both Correct | Submission Code | P-Score | At-least-One Correct | Both Correct |
| FW1 | 0.8185 | 0.9375 | 0.55 | FW5 | 0.8346 | 0.95 | 0.5714 |

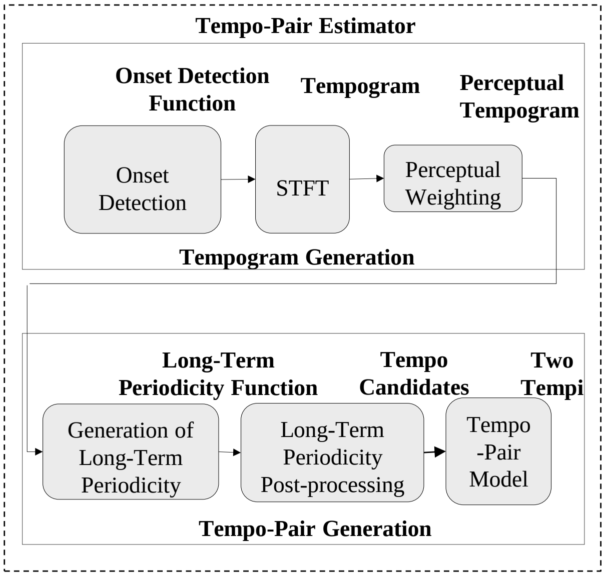

Figure 1. Flowchart of proposed method

### 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:

- 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 Multivariate Linear Regression

Because the tempi within different classes has some common tempo, the classes are correlated with each other. 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.

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

, where $mvregress$ is a matlab function.
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.

### 2.3 Hypothesis of Two-Tempo Ground Truth

Because there is no public dataset available with two groundtruth as the MIREX tempo estimation dataset. We try to generate from the dataset with downbeat and beat annotation such as “Ballroom” dataset. The hypothesis is one of the tempo could be inter beat interval (IBI) and the other could be one-third or half depending on the meter and tempo of excerpts. This hypothesis seems to be simple to count for meter 2/4 and 3/4, 4/4 only and restrict the two tempi in duple or triple relationship.

## 3. REFERENCES

[1] G. Percival and G. Tzanetakis, “Streamlined tempo estimation based on autocorrelation and cross-correlation with pulses”, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22(12):1765–1776, 2014.

[2] 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 Xplore published.

[3] Fu-Hai Frank Wu, “Musical Tempo Octave Error Reducing Based on The Statistics of Tempogram”, Control and Automation (MED), 2015 Mediterranean conference on, IEEE Xplore published.