MIREX 2013 TEMPO ESTIMATION SUBMISSION: M4 RHYTHMIC FEATURES

Chris Baume
BBC Research and Development
chris.baume@bbc.co.uk

ABSTRACT

This extended abstract contains a concise description of a tempo estimator submitted for evaluation. The algorithm works by generating an onset detection curve, thresholding using a moving average, and calculating the autocorrelation. The peaks are found by using a 95th percentile threshold, and the top two tempi are selected as the two largest common divisors of those peaks.

1. INTRODUCTION

This tempo estimation algorithm is based on a modified version of a rhythmic feature extractor designed and built during the 'Making Musical Mood Metadata' (M4) project [1], which was a collaboration between the BBC, Queen Mary University of London's Centre for Digital Music and music provider I Like Music.

The simple and lightweight algorithm uses a number of techniques described in [2] and [3], combined with a moving average filter. Due to the very fast development time, the algorithm was developed empirically as the implementation was written.

2. ALGORITHM

2.1 Onset Curve

The audio signal is first subjected to a fast Fourier transform using a window of block size b and step size s. The FFT bins for each window are summed to produce the 'intensity' which is then convolved with a half-hanning window (see Equation 1), where L is set as 12.

$$H(w) = 0.5 + 0.5\cos\left(2\pi \cdot \frac{w}{2L-1}\right) \qquad w \in [0, L-1] \tag{1}$$

Subsequently, each of the signals are convolved with a peak-enhancing Canny window (see Equation 2), where L is set as 12 and σ is set as 4.

$$C(w) = \frac{w}{\sigma^2} e^{-\frac{w^2}{2\sigma^2}} \qquad w \in [-L, L] \tag{2}$$

An onset curve is produced by normalising the signal to $\mu = 0$ and $\sigma = 1$ and applying half-wave rectification.

The moving average A of the onset curve O is produced from the mean value of a rectangular window of length $2L+1$ (see Equation 3). L is set using the *moving average window length* parameter.

$$A(x) = \sum_{y=-L}^{L} \frac{O(x+y)}{2L+1} \tag{3}$$

The final onset curve is created by subtracting the moving average and applying half-wave rectification.

2.2 Tempo estimation

The autocorrelation of the difference signal is calculated between delays of $\frac{60}{T_{max}} \cdot \frac{F_s}{s}$ frames and $\frac{60}{T_{min}} \cdot \frac{F_s}{s}$ frames, where T_{min} and T_{max} are the min/max tempo in BPM, F_s is the sample rate and s is the step size.

The peaks of the autocorrelation - P_i - are defined as those which are above a certain threshold, defined as the 95th percentile, and whose value is the maximum within a 7-sample window.

The tempi are defined as the two largest common divisors of the detected peaks, which are determined by finding the two peaks that minimise the distance function in Equation 4.

$$D(P_k) = \sum_{i=1}^{N} \left| \frac{P_i}{P_k} - \text{round}\left(\frac{P_i}{P_k}\right) \right| \tag{4}$$

The chosen peaks are converted to BPM using Equation 5.

$$TT_k = \frac{60}{P_k} \cdot \frac{s}{F_s} \tag{5}$$

The weighting between the two tempi is defined as the ratio of the distance functions, as shown in Equation 6.

$$ST1 = \frac{D(P_1)}{D(P_2)} \tag{6}$$

If only one peak is detected and it represents a tempo above 120 BPM, the second tempo is defined as half. If it is below 120 BPM, the second tempo is double. In both cases, the weighting is 0.5.

3. PARAMETER SELECTION

The parameters of the algorithm were optimised by testing them against the MIREX tempo extraction training set using a grid search. The following parameters were found to perform best:

- Block size: 2048 samples
- Step size: 256 samples
- Moving average window length: 150 frames
- T_{max}: 240 BPM
- T_{min}: 35 BPM

4. IMPLEMENTATION

This algorithm is implemented as a Vamp audio analysis plugin (`vamp-plugins.org`). The original C++ code is available at `github.com/bbcrd/bbc-vamp-plugins` and the modified code used for this evaluation is available on the 'mirex2013' branch. The code is licensed under Apache licence 2.0.

5. REFERENCES

[1] C. Baume, "Evaluation of acoustic features for music emotion recognition", *134th Audio Engineering Society Convention, Rome*, 2013.

[2] L. Lu, D. Liu, H.-J. Zhang "Automatic Mood Detection and Tracking of Music Audio Signals" *IEEE Transactions on Audio, Speech and Language Processing*, Section 3C, Vol. 14, pp. 5-18, 2006.

[3] S. Dixon, "Onset Detection Revisited" *International Conference on Digital Audio Effects (DAFx)*, pp. 133-137, 2006.