

MIREX 2013 ONSET DETECTION SUBMISSION: M4 RHYTHMIC FEATURES

Chris Baume

BBC Research and Development

chris.baume@bbc.co.uk

ABSTRACT

This extended abstract contains a concise description of an onset detector submitted for evaluation. The algorithm works by generating an onset detection curve, filtering it with half-Hanning and Canny windows, and thresholding using a moving average. Onsets are defined as the maximum value within a given window size.

1. INTRODUCTION

This onset detection algorithm comes from 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

The audio signal is first subjected to a fast Fourier transform using a window with a block size of 2048 samples and step size of 256. 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.

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

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.

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

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 (see Equation 3) is produced from the mean value of a rectangular window of length $(2L + 1)$ plus a threshold t . L and t are set using the *moving average window length* and *threshold* parameters respectively.

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

The final onset curve is created by subtracting the moving average and applying half-wave rectification. An onset is detected when a sample is the maximum within a given window of length $(2L + 1)$, where L is set by the *onset peak window length* parameter.

3. PARAMETER SELECTION

The undefined parameters have been set up as a grid search, which forms part of the evaluation, using the following ten configurations.

	Onset peak (samples)	Moving average (frames)	Threshold
1	4	200	0.5
2	4	200	1
3	4	150	0.5
4	4	150	1
5	4	100	1
6	6	200	0.5
7	6	200	1
8	6	150	0.5
9	6	150	1
10	6	100	1

Table 1. Parameter configurations for grid search

4. IMPLEMENTATION

This algorithm is implemented as a Vamp audio analysis plugin (vamp-plugins.org). The C++ code is available at github.com/bbcrcd/bbc-vamp-plugins and 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.