Tempo estimation and beat tracking with adaptive input selection

M. E. P. Davies and M. D. Plumbley
Centre for Digital Music
Queen Mary University of London
Mile End Road, London, E1 4NS, United Kingdom
matthew.davies@elec.qmul.ac.uk

Abstract

We present details of our submissions to the Audio Tempo Extraction and Audio Beat Tracking contests within MIREX 2006. The approach we adopt makes use of our existing beat tracking technique with a modified tempo extraction stage, and with the provision of three different onset detection functions to act as input. For each onset detection function we extract potential beat locations and then employ a confidence measure to find the most appropriate input representation for a given audio signal. The beats which yield the highest confidence are extracted as the output of the system.

Keywords: Beat tracking, Tempo extraction, onset detection, MIREX

1. Approach

In this paper we address two aspects of computational rhythmic analysis: i) finding the underlying tempo of a piece of music; and ii) the related topic of extracting beat locations. The output of each algorithm has been compared against the performance of other submitted algorithms with the MIREX 2006 Audio Tempo Extraction [1] and Audio Beat Tracking [2] contests.

A detailed description of the operation of our beat tracking system which incorporates both the extraction of tempo and beat locations may be found in [3]. Therefore within this extended abstract, we merely provide an overview of the operation of each algorithm, in particular addressing those aspects which differ from our previous approach.

1.1. Beat Tracking System

The beat tracking system [3] upon which both MIREX submissions are based can be broken down into four discrete stages:

- The transformation of the input audio signal into an onset detection function (DF).
- The extraction of the beat period by passing the autocorrelation function (ACF) of the DF through a shift-invariant comb filterbank.
- The extraction of the phase of the beats by cross-correlating the DF with an impulse train with impulses at beat period intervals.
- The use of a two-state switching model to track tempo changes and enforce contextual continuity within constant tempo regions.

1.2. Tempo Estimation

The Audio Tempo Extraction contest requires the identification of two perceptual tempi consistent with the performance of multiple human annotators and a salience value describing the strength of the primary metrical level over the secondary level. Within our submission to the MIREX 2005 Audio Tempo Extraction contest (for which the test data and annotations were identical) we extracted the single tempo which most strongly resonated with the periodicities within the input ACF, and inferred the secondary tempo and salience using a simple rule-based approach.

To provide a more robust method, we now apply a peak picking algorithm to the comb filterbank output function. Given a range of possible beat periods with corresponding peak heights, we find the $\log_2$ ratio of all combinations of peak heights, normalising and scaling the results such that a perfect 2:1 ratio will have a minimum absolute score, s_1 of zero,

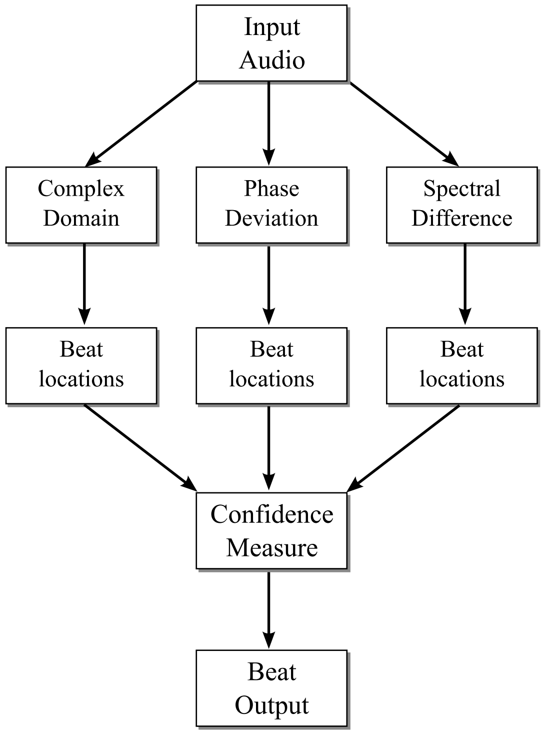

Figure 1. Overview of beat tracking system

$$s_1(i,j) = |(1 - |\log_2(\frac{P(i)}{P(j)})|)| \quad (1)$$

where $P(i)$ and $P(j)$ are the respective periodicities for peaks i and j. To each $s_1(i,j)$ we then add the weighted reciprocal of the sum of the peak heights $A(i)$ and $A(j)$ to give a second score, s_2

$$s_2(i,j) = s_2(i,j) + \frac{\delta}{A(i) + A(j)} \quad (2)$$

where δ is empirically set to 0.01. We then extract the pair of periodicities $P(i)$ and $P(j)$ with the lowest s_2 as those which represent the perceptual tempi of the input, taking the periodicity with the higher peak height as the primary metrical level and the lower as the secondary level. The normalised ratio of peak heights then provides the salience,

$$\text{salience} = \frac{\text{A(i)}}{\text{A(i)} + \text{A(j)}} \quad (3)$$

where in this example $A(i) > A(j)$.

1.3. Beat Tracking with an adaptive input

The task of beat tracking can be considered analogous to the human ability of foot-tapping in time to music. While seemingly intuitive for humans, beat tracking remains a complex task in computational rhythmic understanding. A particular failure of many approaches (for a review see [3]) is the inability to maintain equivalent performance across a wide range of musical genres.

Preliminary experiments in [4] demonstrated that the overall beat tracking performance of our system could be significantly improved by knowing a priori which of several possible onset detection function to use as input, while leaving all other aspects constant. We now extend this concept by attempting to automatically identify the onset detection function most suited to the input signal. The approach we adopt first involves the calculation of three related onset detection functions: i) complex spectral difference; ii) phase deviation; and iii) spectral difference [5].

We then calculate the beat locations for each detection function and then a use *confidence* measure to select which beats are to be taken as the output to the system. The confidence measure is related to the beat tracking evaluation function [2], which we calculate by cross-correlating an impulse train representing the predicted beat locations with each detection function, selecting the beats arising from the strongest correlation as the output. An overview of our approach is shown in fig. 1

2. Results

The results for the Audio Tempo Extraction contest are shown in Table 1 and the results for the Audio Beat Tracking Contest are in Table 2. For each table the overall *P-score* is taken to represent the performance. Further details of the competing algorithms may be found on the respective MIREX web-pages, for Tempo Extraction [1] and for Beat Tracking [2].

Contestant	P-score
Dixon	0.407
Ellis	0.401
Klapuri	0.395
Davies	**0.394**
Brossier	0.391

Table 2. Results table for Audio Beat Tracking

Results indicate that for Tempo Extraction our approach placed 2nd out of 7 entries and was 4th out of 5 entries for Beat Tracking.

Contestant	At least 1 tempo correct	Both tempi correct	P-score
Klapuri	94.29%	61.43%	0.806
Davies	**92.86%**	**45.71%**	**0.776**
Alonso 2	89.29%	43.57%	0.724
Alonso 1	85.71%	45.71%	0.693
Ellis	79.29%	42.86%	0.673
Antonopoulos	84.29%	47.86%	0.669
Brossier	78.57%	50.71%	0.628

Table 1. Results table for Audio Tempo Extraction

Acknowledgements

This research has been partially funded by EPSRC grants GR/S75802/01 and GR/S82213/01. Many thanks also to the MIREX team for overseeing the operation of the contests.

References

[1] "Audio Tempo Extraction - MIREX 2006," [Web site] 2006, [2006 Aug 31], Available: http://www.music-ir.org/mirex2006/index.php/Audio_Tempo_Extraction

[2] "Audio Beat Tracking - MIREX 2006," [Web site] 2006, [2006 Aug 31], Available: http://www.music-ir.org/mirex2006/index.php/Audio_Beat_Tracking

[3] M. E. P. Davies and M. D. Plumbley, "Context-dependent beat tracking of musical audio," *IEEE Transactions on Audio, Speech and Language Processing*,to appear 2007

[4] M. E. P. Davies and M. D. Plumbley, "Comparing mid-level representations for audio based beat tracking," in *DMRN 2005 Digital Music Research Network Summer Conference.*, 2005, pp 36-40.

[5] J. P. Bello, C. Duxbury, M. E. Davies and M. B. Sandler, "On the use of phase and energy for musical onset detection in the complex domain," *IEEE Signal Processing Letters*, vol. 11, no. 6, pp. 553-556, July, 2004