TEMPO DETECTION AND BEAT MARKING FOR PERCEPTUAL TEMPO INDUCTION

Geoffroy Peeters

IRCAM – Sound Analysis/Synthesis
1, pl. Igor Stravinsky
75004 Paris – France
peeters@ircam.fr

ABSTRACT

We give a short overview of the tempo detection and beat marking algorithm we have proposed for the MIREX 2005 perceptual tempo induction contest.

Keywords: MIREX 2005, perceptual tempo induction

1 ALGORITHM

A tempo detection and beat marking algorithm (details of it can be found in [1]) has originally been designed for time variable tempo tracking as well as tempo detection for music with and without percussion (classical music). It has been adapted to provide the relevant parameters of the MIREX 2005 perceptual tempo induction.
The system relies on three stages: onset detection, tempo estimation, beat marking.

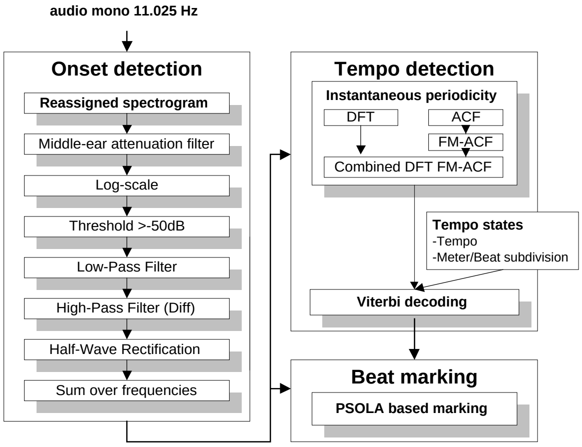

Figure 1 Flowchart of the tempo detection system

1.1 Onset-energy function

An *onset-energy function* is first computed using a proposed reassigned spectral energy flux function. While [2] proposes to measure the variations of the spectrogram over time, we propose to use the reassigned spectrogram [3] which allows to significantly improve temporal and frequency resolution, therefore avoids attack blurring, and allows to better differentiate very close pitches. The reassigned spectrogram consists in reallocating the energy of the ``bins" (frequencies and times) of the spectrogram to the frequency and time corresponding to their center of gravity. The reassigned spectral energy flux is computed in the following steps: a) Compute the reassigned spectrogram; b) apply a filter which simulate the attenuation of the human middle ear; c) convert to log-scale; d) apply a threshold of -50 dB; e) low-pass filter then differentiate each frequency band energy; f) Half-Wave Rectification of each frequency band g) summation over frequencies.

1.2. Tempo estimation

Tempo estimation is performed in two stages: - estimate the dominant periodicities around a specific time – estimate the tempo and meter/beat subdivision paths that best explain the observed periodicities over time.

1.2.1. Periodicities estimation

The dominant periodicities around a specific time are estimated from the local onset-energy function using a combination of Discrete Fourier Transform (DFT) and Auto Correlation Function (ACF). Since the DFT and ACF have inverse octave uncertainties we propose to combine them in a single function. For each frame of the onset-energy function, the DFT and the ACF are computed. The ACF is then mapped to the frequency domain (by considering that the periodicity measure at a given lag is a periodicity measure at the frequency corresponding to the inverse lag). After resampling and interpolation, DFT and ACF are then combined into a product function.

1.2.2. Tempo tracking

Given that the observed periodicity does not only depend on tempo but also on meter and beat subdivision, we estimate the three simultaneously. We consider two meter subdivisions (duple/triple) and two beat subdivisions (simple/compounds). From the observed periodicity we estimate the likelihood of each tempo for each tempo/beat subdivision possibilities. A specific combination of tempo and meter/beat subdivision is called a “tempo state”. We estimate the path of “tempo states” which best explain the observed periodicities over time. This is achieved using a Viterbi decoding algorithm.

1.2. Beat Marking

The beat markers are positioned using a method we previously developed for PSOLA analysis [4]. Two constraints are taken into account: 1) the distance between two markers must be equal to the local tempo period 2) the location of the markers must be close to the local maxima of the onset-energy function. The best solution is found using a least-square algorithm.

2 ADAPTATION TO MIREX 2005 PERCEPTUAL TEMPO INDUCTION

Considering that the audio items of the MIREX contest are of short duration and of constant tempo, the Viterbi decoding algorithm was replaced by a more straightforward search for the maximum of the “tempo states” likelihood in the whole audio item. The most likely “tempo state” is noted A, the second most likely “tempo state” is noted B. The sorting of A and B provides the MIREX tempi ordering T1 and T2. The ratio between the likelihood of T1 and the sum of the likelihoods of T1 and T2 provides the MIREX relative salience/strength. The corresponding MIREX phases are estimated by applying our beat marking algorithm (section 1.2.) with T1 and T2 as input, and retaining the position (in s.) of the respective first beats.

Remark that our algorithm was not designed to output two tempi; B is just the second most likely “tempo state” without any dependence on the previous choice of A; therefore B can have a different meter/beat subdivision.

3 EVALUATION

The evaluation conducted by the MIREX team on the test data (140) files yielded the following result:

- Score (std. deviation): 0.656 (0.223)
- At Least One Tempo Correct: 95.71%
- Both Tempi Correct: 47.86%

ACKNOWLEDGEMENTS

Part of this work was conducted in the context of the European I.S.T. Project Semantic HIFI [5].

REFERENCES

[1] Peeters, G. *Time Variable Tempo Detection and Beat Marking*. in *ICMC*. 2005. Barcelone, Spain.

[2] Laroche, J., *Efficient Tempo and Beat Tracking in Audio Recordings.* J. Audio Eng. Soc., 2003. **51**(4): p. 226-233.

[3] Flandrin, P., *Time-Frequency/Time-Scale Analysis*. 1999, San Diego, California: Academic Press.

[4] Peeters, G., *Modeles et modelisation du signal sonore adaptes a ses caracteristiques locales*, in *PHD Thesis, Universite Paris VI / Ircam*. 2001, Universite Paris VI: Paris, France.

[5] Vinet, H. *The Semantic Hifi Project*. in *submitted to ICMC*. 2005. Barcelona, Spain.