

# INFLUENCE OF INPUT FEATURES IN PERCEPTUAL TEMPO INDUCTION

**Fabien Gouyon**

Music Technology Group,  
Universitat Pompeu Fabra, Barcelona  
fgouyon@ia.upf.es

**Simon Dixon**

Austrian Research Institute for  
Artificial Intelligence, Vienna  
simon@oefai.at

We report on four entries to the perceptual tempo induction contest, part of the Music Information Retrieval Evaluation eXchange (MIREX) of the International Symposium on Music Information Retrieval (ISMIR) 2005. In a first section we give details of the algorithms and differences between them. A second section provides an evaluation of the algorithms.

**Keywords:** MIREX 2005, perceptual tempo induction

## 1 ALGORITHMS

A total of 4 algorithms have been submitted. They consist in a set of Matlab functions and a binary file compiled under the Windows operating system. In accordance with the functional framework for rhythm description systems proposed in [2], the algorithms consist in three main processing blocks: *feature list creation* from audio, *periodicity function computation* and *parsing*.

Low-level features are computed on a frame-by-frame basis. Algorithms account for 4 different sets of features. *Algorithm0* uses 13 features: the magnitude-normalized derivative of the energy in the 8 frequency bands proposed by Dixon et al. [1] (note that Dixon et al. [1] use the derivative, normalizing by the magnitude yields significant improvements) and the magnitude-normalized derivative of 5 spectral features (the mean of the spectral peaks and the spectrum geometric mean, kurtosis, low-frequency energy ratio, mean and skewness). *Algorithm1* uses 8 features: *Algorithm0* energy features. *Algorithm2* uses 9 spectral features: the magnitude-normalized derivative of the the spectral peaks mean, harmonic centroid and harmonic deviation and the spectrum flatness, geometric mean, kurtosis, low-frequency energy ratio, mean and skewness. *Algorithm3* uses 13 features: the derivative of the MFCCs. For more details on the relevance of different feature sets, see [3].

All algorithms implement the autocorrelation as periodicity function.

Parsing the periodicity function and inferring the most salient tempo (T1, and the meter as a byproduct) is done similarly as in [1]: prominent peaks are collected from each periodicity function. The algorithm then considers all pairs of peaks as possible beat/measure combinations, and computes the fit of all periodicity peaks to each hy-

pothesis, using a weighted sum, where the weights represent the likelihood of each metrical unit appearing as a strong periodicity, given the meter [1]. The second most salient tempo (T2) is chosen as the periodicity (differing from T1) with the highest weight. The normalised relative salience/strength of T1 (ST1) is computed from T1 and T2 weights. The phases of T1 and T2 (P1 and P2) are computed by correlation of pulse trains with feature lists.

## 2 EVALUATION

### 2.1 On the training data

We implemented the *P-score* proposed on the contest webpage ([http://www.music-ir.org/mirexwiki/index.php/Audio\\_Tempo\\_Extraction](http://www.music-ir.org/mirexwiki/index.php/Audio_Tempo_Extraction)) and tested our algorithms on the available training data (20 files). On June 27th, the best score obtained with this evaluation metrics is 0.695 (*Algorithm0*).

### 2.2 On the test data

The evaluation conducted by the MIREX team on the test data (140 files) yielded the following results. *Algorithm0*: 0.670, *Algorithm1*: 0.649, *Algorithm3*: 0.645 and *Algorithm2*: 0.607.

A comparison with other algorithms submitted to the contest can be found at <http://www.music-ir.org/evaluation/mirex-results/>. At publication time, more detailed analyses of the results were not available but they will probably be made in a near future as a joint effort of the contest participants.

## References

- [1] S. Dixon, E. Pampalk, and G. Widmer. Classification of dance music by periodicity patterns. In *Proc. ISMIR*, 2003.
- [2] F. Gouyon and S. Dixon. A review of automatic rhythm description systems. *Computer Music Journal*, 29(1), 2005.
- [3] F. Gouyon, G. Widmer, and X. Serra. Acoustic cues to beat induction: A machine learning perspective, *in press*. 2005.