TEMPO INDUCTION BY INVESTIGATING THE METRICAL STRUCTURE OF MUSIC USING A PERIODICITY SIGNAL THAT RELATES TO THE TATUM PERIOD

Christian Uhle

Fraunhofer Institute for Digital Media Technology
Langewiesener Strasse 22
98693 Ilmenau
uhle@idmt.fraunhofer.de

ABSTRACT

The estimation of the musical tempo of audio signals is of large interest for various applications, e.g. the synchronization of multi-media content to music and music information retrieval systems. This article describes the author's contribution to the 2nd Annual Music Information Retrieval Evaluation Exchange (MIREX 2005) and gives an evaluation and comparison of the recognition performance.
The presented method computes a representation of the metrical structure of a musical piece as a function of the degree of periodicity of a phenomenal accent signal at lags that relate to the period of the fastest pulse.
This representation is compared to pre-defined metric templates, corresponding to various time signatures and metrical structures. A list of tempo hypotheses is obtained from the best fitting template, from which the tempo is chosen with respect to their salience and a-priori probabilities of tempo values.

Keywords: MIREX, tempo estimation.

1 INTRODUCTION

Automated estimation of the tempo of musical audio signals is of great interest for a large number of applications, e.g. the synchronization of multi-media content to music, music information retrieval, automated transcription and play list generation.

While research on automated tempo estimation has been in progress for some time, current algorithms still seem to be a long way from matching human recognition performance, especially with respect to the ability to handle complex rhythmic content, to track tempo variations in music and to determine the primary metrical level in a metrical structure.

Various authors agree that metrical structure of musical signals is constructed from hierarchical layers of pulse series, corresponding to the tatum and beat period and higher levelled periodicities, e.g. with periods that relate to a musical measure. The term tatum refers to the pulse of the lowest metrical level [1]. Although many tempo induction algorithms focus on the estimation of the most prominent periodicity in a phenomenal accent signal, correlating to the perceived accentuation in a musical auditory stream, this strategy is likely to fail when analysing syncopated rhythms and is subject to tempo octave errors.

According to the theory of hierarchical metrical structure [2], the beat period is related to the period lengths of other metrical levels. It is approximately an integer multiple of the tatum period an divides the length of one musical measure nearly whole-numbered. Therefore, a desirable representation of the periodic behaviours of an accent signal evaluates the salience of periodicity at lags that relate to the tatum period.

2 METHOD

The audio signal is segmented into characteristic and similar regions using Foote's self-similarity method [3]. This pre-processing step is motivated by results of former research [4] and due to the assumption that the tempo may change between different segments, corresponding to a verse or chorus for example. The tempo estimation and beat location process comprises the calculation of the accent signal, the periodicity signal, its evaluation and the beat phase estimation.

2.1 Accent signal calculation

Amplitude envelopes $E_i(t)$, with sub-band index i, of 24 logarithmically spaced sub-bands of the audio signal are calculated by means of Discrete Fourier Transformation and low-pass filtering. A slope signal $D_i(t)$ is computed by means of the relative difference function. The slope signals are weighted with the maximum of an excerpt starting at lag=150 ms of the normalized autocorrelation function (ACF) of the slope signal $D_i(t)$. As a consequence, the influence of sub-bands with less rhythmic behaviour is diminished. The accent signal is calculated by summation of the weighted slope signals of each sub-band. Additionally, note onset times and saliencies are extracted.

2.2 Periodicity signal calculation

The tatum period is estimated using two methods examining note onsets, namely the two-way mismatch error function (TWME) [5] and the greatest common divisor method [6] and two methods processing the accent signal, the TWME and an ACF-based method for overlapping segments of a few seconds length each. The calculated tatum values are clustered using the Sequential Leader Algorithm [7] and a final estimate is obtained from the most representative cluster according to a criterion rating the number of points in each cluster and the rhythmic intensity of the audio segment from which each estimate has been calculated.

A second ACF of the accent signal is calculated on a larger time scale to detect periodicities in the range of a musical measure. By assigning the local maxima of the ACF to lags that relate to the tatum period, a periodicity representation is obtained that reflects the metrical structure of the audio signal. Figure 1 and 2 illustrate an example of an ACF of an accent signal and the corresponding periodicity representation respectively.

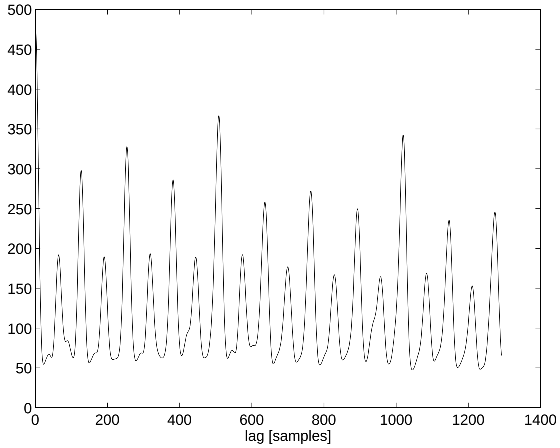

Figure 1: ACF of accent signal

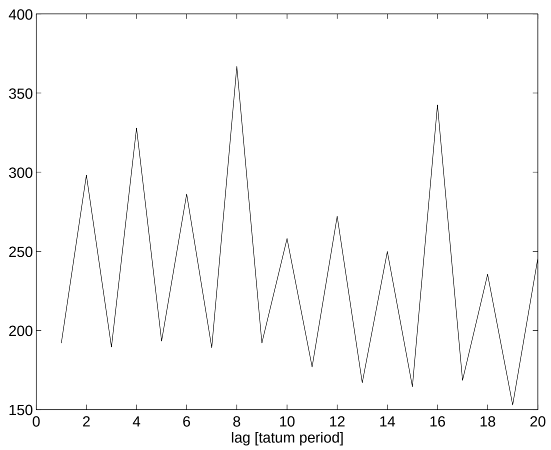

Figure 2: Periodicity representation at lags that relate to the tatum period

2.3 Evaluation of the periodicity signal

The derived periodicity signal is compared to various pre-defined metrical templates by means of their correlation coefficient. A metrical template is defined as a vector describing the periodicity at lags that relate to the tatum period for distinct metrical structures. These templates are heuristically defined after observing a training set and employing musical knowledge. For each template, possible time signatures and relations between beat and tatum are defined.

Among the beat hypotheses the beat length is chosen according to their salience as expressed in the periodicity representation weighted by a-priori probabilities for beat lengths [8].

2.4 Beat phase estimation

The beat phase is estimated by means of TWME by calculating the error function between a equidistant pulse grid and the accent signal.

2.5 Computing a final result from all segments

The final result is obtained from all segment's results by computing a weighted voting mechanism implemented as a weighted histogram taking the energy of the accent signal into account. Finally, the results are cross-checked by examining a beat histogram [9].

2.6 Algorithm 2: Extending the analysis by investigating instrument information

An optional extension of the algorithm is the use of a method for the automated detection of percussive unpitched instruments, described in [10]. Detected and classified occurrences of drum instruments seem to be a valuable accent signal for rhythmic analysis of various musical signals. From the event list, drum patterns are obtained and evaluated with respect to musical meter. The processing method is described in detail in [11].

3 RESULTS

The performance of the algorithms were evaluated and compared using a data set with 140 audio files. The evaluation metric and conditions have been worked out among the participants of the contest. The participants were asked to provide two tempo period and phase estimates and the relative salience of the two estimates. A P-score has been defined evaluating the congruency between reference and estimated values.

Both submitted algorithms yielded a P-score of 0.675% and rank 2 and 3 respectively. There are only minor differences in the evaluation metric among the best ranking algorithms.

3.1 Comparison of both algorithms

The evaluation of instrument information did not improve the recognition performance and increased the processing load. This particular result differs from other experiments using a larger data base. Implementation errors can not be excluded as a possible source of error, but a more detailed analysis of the results is needed.

3.2 Particular strength and weakness of the two algorithms

The submitted algorithm performs better in the categories "Both tempos correct" and "Mean absolute differences of scored saliencies" then in category "At least one tempo correct". The explanation of this behaviour is found in the nature of the algorithm. The calculation of the described periodicity signal is prone to errors, and every error made at this stage propagates in the final result.

Once a periodicity representation is calculated correctly, the rhythmic features describing the metrical structure are estimated robustly.

3.3 Differences to last years contest

At first glance, the results of this years contest compared to last years contest show an improvement of tempo estimation algorithm in general. But this conclusion is more than doubtful, since the data sets used for evaluation differ significantly at least in size. Without having a closer view on the items and the results of the algorithms per item, no clearer statements can be made. After listening to a large amount of the last years data base (all items in the collection “songs”), it turned out that these items cover a wide range of musical styles and genres and represent a challenging collection for tempo estimation algorithms. The limitation of the amount of test data used in this year’s contest is caused by the fact that the reference tempos were measured by evaluation of the tapped tempos of a large number of listeners, which is a time consuming process.

The MIREX 2005 is clearly a great success for all participants. The growing number of contributions indicates a large interest of the music information retrieval community in the comparison of the performance of the developed tools and methods.

REFERENCES

[1] J. Bilmes. "Timing is of essence: perceptual and computational techniques for reüpresenting, learning and reproducing expressive timing in percussive rhythms", Massachusetts Institute of Technology, MSc thesis, 1993.

[2] G. Cooper and L. Meyer. "The rhythmic structure of music", University of Chicago Press, 1963.

[3] J. Foote. "Automatic audio segmentation using a measure of audio novelty", IEEE Int. Conf. on Multimedia and Expo, 2000.

[4] C. Uhle. "Generation of musical scores of percussive un-pitched instruments from automatically detected events", 116th AES Convention, 2004.

[5] F. Gouyon, P. Herrera and P. Cano. "Pulse-dependent analysis of percussive music", 22nd AES Conference, 2002.

[6] J. Seppänen. "Tatum grid analysis of musical signals", IEEE Workshop on Applications of Digital Signal Processing to Audio and Acoustics, 2001.

[7] J. Hartigan. "Clustering Algorithms", Wiles and Sons, 1975.

[8] R. Parncutt. "A perceptual model of pulse salience and metrical accent in musical rhythms", Music Perception, vol. 11, no. 4, 1994.

[9] G. Tzanetakis and P. Cook. "Musical Genre classification of audio signals", IEEE Transactions on Speech and Audio Processing, 2002.

[10] C. Dittmar and C. Uhle. "Further steps towards drum transcription of polyphonic music", 116th AES Convention, 2004.

[11] C. Uhle and C. Dittmar. "Drum pattern based genre classification of popular music", 116th AES Convention, 2004.