MIREX 2013: ESSENTIA MULTI FEATURE BEAT TRACKER

José R. Zapata[1], Matthew E.P. Davies[2] , Dimitry Bogdanov[1] and Emilia Gómez[1]
[1]Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
[2]Sound and Music Computing Group, INESC TEC, Porto, Portugal
joser.zapata@upf.edu, mdavies@inescporto.pt, dmitry.bogdanov@upf.edu, emilia.gomez@upf.edu

ABSTRACT

The Multi-feature Beat tracker Essentia implementation uses five different onset detection functions to estimate the beats of a musical audio signal using only one beat tracker algorithm, where the beat tracker output is selected using a committee technique. This is a C++ implementation of the algorithm ZDG1 (five onset detection function), submitted to MIREX 2012 audio beat tracking task.

1. INTRODUCTION

Based on the beat tracking selection from a committee of state of the art algorithms presented in [12] [9], we use five different onset detection functions as input signal to the Degara Beat Tracker [5]. Each output is considered as a committee member and in each case the beat tracker output chosen to represent the committee is selected automatically as the one which most agrees with the remainder of the committee (Maximum Mutual Agreement, MaxMA).

2. ESSENTIA

Essentia 2.0, an extensive open-source library for audio analysis and audio-based music information retrieval released under the Affero GPL license and well suited for both research and industrial applications. In its core, Essentia is comprised of a reusable collection of algorithms to extract features from audio. The available algorithms include audio file input/output functionality, standard digital signal processing (DSP) building blocks, filters, generic algorithms for statistical characterization, and spectral, temporal, tonal and high-level music descriptors. The library is written in C++, which provides considerable performance benefits. Importantly, it also includes Python bindings to facilitate the usage of the library for the users who are familiar with the matlab/python environment.

The design of Essentia is focused on robustness, time and memory performance, and ease of extensibility. The algorithms are implemented keeping in mind the use-case of large-scale computations on large music collections. One may develop his own executable utility with a desired processing flow using Essentia as a C++ library. Alternatively a number of executable extractors are included with Essentia, covering a number of common use-cases for researchers, for example, computing all available music descriptors for an audio track, extracting only spectral, rhythmic, or tonal descriptors, computing predominant melody and beat positions, and returning the results in yaml/json data formats.

The multi feature beat tracker is in the Essentia [1] rhythm descriptors by the function *BeatTrackerMultiFeature()*

3. BEAT TRACKING SYSTEM

The Beat tracking estimation is computed five times, each time with a different onset detection function. the output is selected using the Maximum Mutual Agreement.

3.1 Onset Detection Function

The onset detections used are:

- Complex Spectral Difference [6]
- Energy Flux [10]
- Harmonic function [8]
- Sub Band weight [4]
- Mel Auditory Features [7]

3.2 Period Estimation

The beat tracking system estimates the beat period and phases independently. To extract the sequence of periods from the beat period salience observation signal. The beat period is estimated by passing the autocorrelation function of the onset detection function through a shift invariant comb filterbank matrix. Then the information of the beat period is used to recover the beat alignment by passing the onset detection function through a comb filter matrix. The system assumes the beat period like a slowly varying process and the transition probabilities are modeled using a Gaussian distribution of fixed standard deviation. For a complete description of the beat period estimation method see [1, 5].

[1] http://essentia.upf.edu/

3.3 Beat Tracking

The Degara Beat tracking probabilistic model [5], using hidden Markov model, takes as input parameters the phase observation signal and the beat period estimation, returning the set of beat time estimates. This beat estimation process is done with each onset detection function described in section 3.1, getting five beat estimations for the same audio file.

3.4 Output Selection with Maximum Mutual Agreement (MaxMA)

The Mean Mutual Agreement (MMA) follows the Query by Committee concept [11] which selects the most informative set of samples from a database based on the mutual (dis-)agreement between a designated committee of learners. Given a committee of beat trackers, the low MMA between their estimated beat sequences (see Figure 1) on a music database was shown to indicate difficult samples for beat tracking, by being strongly correlated with low performance against the ground truth on this data [9].

As depicted in Figure 1 and proposed in [9], the MMA of a sample is computed by using the beat estimations of N beat trackers on a musical piece, measuring the mutual agreement (MA) among their estimated beat sequences, and retrieving the mean of all $N(N-1)/2$ mutual agreements:

$$MMA = \frac{1}{N(N-1)/2} * \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} MA_{i,j}. \quad (1)$$

The Maximum Mutual Agreement (MaxMA) refer to the MA of the beat sequence that highly agrees with the others:

$$MaxMA = max_i \left(\sum_{j=1, j\neq i}^{N} MA_{i,j} \right), i = 1, \ldots, N. \quad (2)$$

In order to measure the MA between each pair of estimated beat sequences a beat tracking, the Information Gain measure [3] evaluation criteria was selected. The output selection is done using the Maximum mutual agreement presented in [12] [9].

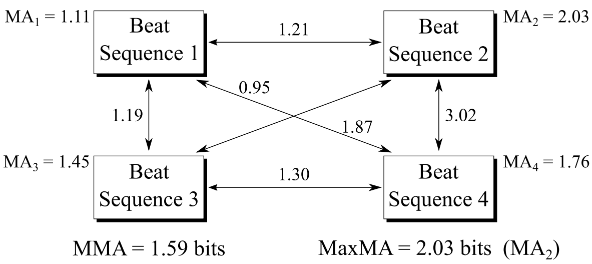

Figure 1. Example calculation of the MMA and MaxMA for a musical piece with the beat sequences estimated from a committee of four beat trackers.

4. RESULTS

5. ACKNOWLEDGEMENTS

We would like to thank all the authors (N. Degara) of the available matlab code for their support. Universidad Pontificia Bolivariana (Colombia) and Colciencias.

6. REFERENCES

[1] M. E. P. Davies and M. D. Plumbley, "Context-dependent beat tracking of musical audio" *IEEE Transactions on Audio,Speech, and Language Processing*, vol. 15, no. 3, pp. 1009–1020, 2007.

[2] M. E. P. Davies, N. Degara, and M. D. Plumbley, "Evaluation methods for musical audio beat tracking algorithms," Queen Mary University of London, Centre for Digital Music, Tech. Rep. C4DM-TR-09-06, 2009.

[3] M. E. P. Davies, N. Degara and M. D. Plumbley, "Measuring the performance of beat tracking algorithms using a beat error histogram," *IEEE Signal Processing Letters*, vol. 18, no. 3, pp. 157–160, 2011.

[4] M. E. P. Davies, M. D. Plumbley and D. Eck, "Towards a Musical Beat Emphasis Function," In *Proceedings of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2009)*, New Paltz, NY, October 18-21, pp. 61-64, 2009

[5] N. Degara, E. Argones, A. Pena, S. Torres-Guijarro, M. E. P. Davies and M. D. Plumbley, "Reliability-Informed Beat Tracking of Musical Signals," *IEEE Transactions on Audio, Speech and Language Processing,* Vol. 20, pp. 290–301, 2012.

[6] S. Dixon, "Onset detection revisited," in *6th International Conference on Digital Audio Effects (DAFx-06)*, Montreal, Canada, Sept. 1820, pp. 133137,2006.

[7] D. P. W. Ellis, "Beat tracking by dynamic programming," *Journal of New Music Research*, vol. 36, no. 1, pp. 51–60, 2007.

[8] S. Hainsworth and M. Macleod. "Onset detection in musical audio signals". In *Proceedings of International Computer Music Conference (ICMC)*, pp. 136–166, Singapore. 2003.

[9] A. Holzapfel, M. E. P. Davies, J.R. Zapata, J.L. Oliveira and F. Gouyon, "Selective sampling for beat tracking evaluation," *IEEE Transactions on Audio, Speech and Language Processing,* Vol. 20, No. 9, P. 2539–2548, Nov 2012.

[10] J. Laroche, "Efficient tempo and beat tracking in audio recordings," *Journal of the Audio Engineering Society,* vol. 51, no. 4, pp. 226–233, 2003.

[11] H. S. Seung, M. Opper and H. Sompolinsky "Query by committee," in *Proc. of the 5th Annual Workshop on Computational learning theory,* pp. 287–294, 1992.

[12] J.R. Zapata, A. Holzapfel, M. E. P. Davies, J.L. Oliveira and F. Gouyon, "Assigning a Confidence Threshold on Automatic Beat Annotation in Large Datasets," In *Proceedings of 13th International Society for Music Information Retrieval Conference (ISMIR 2012),* Porto. 2012.