MIREX 2014 ENTRY: 2D FOURIER MAGNITUDE COEFFICIENTS

Oriol Nieto and Juan P. Bello
Music and Audio Research Lab
New York University
{oriol, jpbello}@nyu.edu

ABSTRACT

This extended abstract describes an algorithm submitted to the task of structural segmentation in MIREX 2014. The method uses 2D Fourier Magnitude Coefficients (2D-FMC) in order to label musical segments based on their acoustic similarity. The segments are previously identified by three different boundary retrieval algorithms, and consequently three variations of this 2D-FMC algorithm were submitted, one for each boundary identification technique. We use beat-synchronous harmonic and timbral features to extract the boundaries, while only harmonic information is used by the 2D-FMC method to group the segments. This implementation is open source and available for public download [1] .

1. INTRODUCTION

The task of music segmentation can be divided in two main subtasks: the identification of segment boundaries and the grouping (or labeling) of these segments based on their acoustic similarity. It has been shown that the quality of the boundaries strongly impacts the grouping process [9], therefore separating these process might be beneficial for the assessment of the latter subtask. We have submitted an algorithm to MIREX that uses 2D Fourier Transform Magnitude Coefficients (2D-FMCs) to approximate a solution to the labeling subtask of music segmentation. This method is formally presented in [5], and it needs previously computed boundary information in order to produce the output. These boundaries are estimated by using previously existing algorithms: Checkerboard-like kernel [4], Convex NMF [6], and Structural Features [9]. Therefore, three different flavors of our 2D-FMC algorithm were submitted, one for each boundary technique associated with it.

[1] https://github.com/urinieto/SegmenterMIREX2014

2. AUDIO FEATURES

The three different versions of the algorithm use Essentia [1] to compute the audio features from a one-channel audio file sampled at 44.1 kHz. More specifically, we use Harmonic Pitch Class Profile features —a harmonic representation that is intended to be less noisy than regular chromagrams—, and a Hann analysis window of size 2048 samples with a 50% overlap. Additionally, we also use Essentia to compute 13 Mel Frequency Cepstral Coefficients to capture the timbre of the track. We combine the HPCP with the MFCC by simply stacking them and producing a 25 dimensional feature vector. Furthermore, we use the Multi Feature Beat Tracker [10] implementation included in Essentia to estimate the beats, and the aggregation described in [3] to obtain the beat-synchronous representation, which becomes the input to the boundary algorithms.

3. BOUNDARY ALGORITHMS

In this section we briefly discuss how the selected algorithms to estimate the segment boundaries operate. The Checkerboard-like kernel [4] is one of the most established and successful methods to retrieve boundaries. It focuses on the novel-based segments (see [7] for an explanation of the different types of structural segments) by using the self-similarity across the audio features and running a Gaussian kernel shaped like a checkerboard (i.e. positive in the two top-right and bottom-left parts and negative otherwise). This yields a novelty curve from which peak picking can be performed, detecting a boundary for each peak found.

The Convex NMF method [6] uses a convex variant of the matrix factorization method of NMF in order to divide the audio features into meaningful clusters or segments. As opposed to the Checkerboard-like kernel method, this algorithm focuses on homogeneous-based boundaries, thus allowing us the exploration of how our 2D-FMC method differs when inputting different types of segments.

Finally, the recently published Structural Features method [9], yields some of the best boundary results in the literature. It uses a variant of the lag matrix [2] in order to obtain the so-called *structural features*. These features are differentiated to obtain a novelty curve, from which the boundaries can be obtained analogously as in the Checkerboard-like kernel algorithm. This technique retrieves homogeneity and novelty-based segment boundaries, which should result in a better performance of our

2D-FMC method.

4. 2D-FMC FOR MUSIC STRUCTURE ANALYSIS

Ideally, an audio representation should have the following properties in order to capture the similarity between segments: (i) key-transposition invariant (as long as the same harmonic progressions occur, the key should be irrelevant), (ii) phase-shift invariant (a motive might occur at the beginning or at the end of the segment, and yet be clustered under the same label), and (iii) local-tempo invariance (tempo variations should not be a relevant factor when quantizing segment similarity). These three characteristics are inherent in the 2D-Fourier Magnitude Coefficients, thus becoming an excellent type of representation for approaching the labeling subproblem of music structure analysis.

These 2D-FMCs are computed from the HPCP representation for each segment. In order to obtain 2D-FMC patches of the same size, we zero-pad to the maximum segment length. Afterwards, the similarity is obtained by simply running k-means on the 2D-FMC patches using the Euclidean distance. The number of unique segments k can be estimated using the BIC knee-detection method [8]. For more details, we refer the reader to the original publication of this algorithm [5].

5. ALGORITHM VERSIONS

- **NB1**: Checkerboard-like kernel for boundaries and 2D-FMC for labels.
- **NB2**: Structural Features for boundaries and 2D-FMC for labels.
- **NB3**: Convex NMF for boundaries and 2D-FMC for labels.

6. ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under grant IIS-0844654.

7. REFERENCES

[1] Dmitry Bogdanov, Nicolas Wack, Emilia Gómez, Sankalp Gulati, Perfecto Herrera, Oscar Mayor, Gerard Roma, Justin Salamon, José Zapata, and Xavier Serra. Essentia: An Audio Analysis Library for Music Information Retrieval. In *Proc. of the 14th International Society for Music Information Retrieval Conference*, pages 493–498, Curitiba, Brazil, 2013.

[2] Roger B. Dannenberg and Masataka Goto. Music Structure Analysis from Acoustic Signals. In David Havelock, Sonoko Kuwano, and Michael Vorländer, editors, *Handbook of Signal Processing in Acoustics*, pages 305–331. Springer, New York, NY, USA, 2008.

[3] Daniel P. W. Ellis and Graham E. Poliner. Identifying 'Cover Songs' with Chroma Features and Dynamic Programming Beat Tracking. In *Proc. of the 32nd IEEE International Conference on Acoustics Speech and Signal Processing*, pages 1429–1432, Honolulu, HI, USA, 2007.

[4] Jonathan Foote. Automatic Audio Segmentation Using a Measure Of Audio Novelty. In *Proc. of the IEEE International Conference of Multimedia and Expo*, pages 452–455, New York City, NY, USA, 2000.

[5] Oriol Nieto and Juan Pablo Bello. Music Segment Similarity Using 2D-Fourier Magnitude Coefficients. In *Proc. of the 39th IEEE International Conference on Acoustics Speech and Signal Processing*, pages 664–668, Florence, Italy, 2014.

[6] Oriol Nieto and Tristan Jehan. Convex Non-Negative Matrix Factorization For Automatic Music Structure Identification. In *Proc. of the 38th IEEE International Conference on Acoustics Speech and Signal Processing*, pages 236–240, Vancouver, Canada, 2013.

[7] Jouni Paulus, Meinard Müller, and Anssi Klapuri. Audio-Based Music Structure Analysis. In *Proc of the 11th International Society of Music Information Retrieval*, pages 625–636, Utrecht, Netherlands, 2010.

[8] Dan Pelleg and Andrew Moore. X-means: Extending K-means with Efficient Estimation of the Number of Clusters. In *Proc. of the 17th International Conference on Machine Learning*, pages 727–734, Stanford, CA, USA, 2000.

[9] Joan Serrà, Meinard Müller, Peter Grosche, and Josep Lluís Arcos. Unsupervised Music Structure Annotation by Time Series Structure Features and Segment Similarity. *IEEE Transactions on Multimedia, Special Issue on Music Data Mining*, 16(5):1229 – 1240, 2014.

[10] José R. Zapata, Matthew E. P. Davies, and Emilia Gómez. Multi-feature Beat Tracking. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(4):816–825, 2013.