MIREX: MSAF V0.1.0 SUBMISSION

Oriol Nieto
Pandora Media, Inc.
onieto@pandora.com

ABSTRACT

This MIREX submission contains a set of implementations included in the Music Structure Analysis Framework (MSAF) of previously published structural segmentation algorithms.

1. MSAF V0.1.0

MSAF is an open source framework written in Python to facilitate research in the field of structural segmentation [7]. The latest release [1] contains five boundary detection and three labeling algorithms published under an MIT license, described as follows:

- **2D-FMC** [6]: Labeling technique that makes use of 2D Fourier Magnitude Coefficients to quantify the similarity between music segments.
- **Checkerboard** [1]: Boundary detection algorithm that applies a checkerboard kernel across the diagonal of a self-similarity matrix to obtain the most prominent boundaries in terms of novelty.
- **CNMF** [8]: Algorithm that approaches both subproblems of boundary detection and labeling by factorizing the given features using a convex variant of the standard non-negative matrix factorization technique.
- **Laplacian** [2]: This method also approaches both subproblems by combining local cues with long-term representations and analyzing the eigenvectors of the Laplacian graph.
- **OLDA** [3]: Boundary detection technique that employs supervised learning (Ordinal Linear Discriminant Analysis) to learn a latent structural repetition space optimized for music structure.
- **SF** [9]: Boundary detection method that makes use of the Structural Features, a set of hand-crafted descriptors that aim at retrieving novelty, homogeneous, and repetitive boundaries.

For the current MIREX submission we make use of all these algorithms, combined as follows:

- **ON1**: Checkerboard + 2D-FMC
- **ON2**: OLDA + 2D-FMC
- **ON3**: SF + 2D-FMC
- **ON4**: CNMF
- **ON5**: Laplacian

Note that the 2D-FMC method is employed when a technique only approaches the boundary detection problem. This algorithm is one of the fastest and most effective to label the estimated segments, as shown in MIREX 2014 [5].

The features employed for all of the algorithms are the Constant-Q Transform (CQT) spectrograms, which are computed using librosa [4]. The parameters are the following: sampling rate of 22050 kHz, FFT size of 4096 samples, hop size of 1024 samples, and 84 frequency bins for the CQT matrix. Moreover, the features are synchronized to estimated beats computed using librosa's default beat tracker. Finally, all algorithms' parameters are set to their default values in MSAF v0.1.0.

This submission was generated using the `run_mirex.py` script contained in the MSAF repository [2].

2. REFERENCES

[1] 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.

[2] Brian McFee and Daniel P W Ellis. Analyzing Song Structure with Spectral Clustering. In *Proc. of the 15th International Society for Music Information Retrieval Conference*, pages 405–410, Taipei, Taiwan, 2014.

[3] Brian McFee and Daniel P. W. Ellis. Learnign to Segment Songs With Ordinal Linear Discriminant Analysis. In *Proc. of the 39th IEEE International Conference on Acoustics Speech and Signal Processing*, pages 5197–5201, Florence, Italy, 2014.

[4] Brian Mcfee, Colin Raffel, Dawen Liang, Daniel P. W. Ellis, Matt Mcvicar, Eric Battenberg, and Oriol Nieto.

[1] https://github.com/urinieto/msaf/releases/tag/v0.1.0

[2] https://github.com/urinieto/msaf/blob/master/examples/run_mirex.py

librosa: Audio and Music Signal Analysis in Python. In *Proc. of the 14th Python in Science Conference*, pages 1–7, Austin, TX, USA, 2015.

[5] Oriol Nieto and Juan P Bello. MIREX 2014 Entry: 2D Fourier Magnitude Coefficients. In *Music Information Retrieval Evaluation eXchange*, Taipei, Taiwan, 2014.

[6] 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.

[7] Oriol Nieto and Juan Pablo Bello. Systematic Exploration of Computational Music Structure Research. In *Proc. of the 17th International Society for Music Information Retrieval Conference*, pages 547–553, New York City, NY, USA, 2016.

[8] 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.

[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.