MSAF: MUSIC STRUCTURE ANALYSIS FRAMEWORK

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

ABSTRACT

We present a transparent framework to facilitate the analysis, evaluation, and comparison of algorithms for music structure analysis. This Music Structure Analysis Framework (MSAF) includes several algorithm implementations and it is built upon existing packages to extract features, assess estimations, and store annotation data.

1. THE FRAMEWORK

In this section we describe the self-contained, open-source framework written in Python [1] that facilitates and encourages research in area of music structure analysis.

1.1 The Task

Music Structure Analysis (MSA) has been explored in MIR since the late 90s, and it is still one of the most active in MIREX [13]. Its goal is to identify the large-scale, non-overlapping segments of a given audio signal (e.g., verse, chorus). Typically, this task is divided in two subproblems: *boundary detection*, where the exact times of the beginning and end of the segments are extracted; and *structural grouping*, where the different segments defined by the previously detected boundaries are classified based on their acoustic similarity. These two subproblems are evaluated differently, and some algorithms only focus at solving one of these two subproblems (e.g., boundary detection [1], structural grouping [9], or both [5]).

1.2 The Features

MSAF pre-computes a set of features using `librosa` [8], such that the same exact input is used across algorithms, thus assuring that the performance of the implementations can be easily compared independently of the impact of the specific set of parameters of the feature extraction process. The current version of MSAF supports Chromagrams, MFCCs, Tonnetz, and Constant-Q Spectrum, all of them standard in the MSA task. These features depend on the additional analysis parameters such as sampling rate, FFT size, and hop size, which can also be modified in MSAF. Additionally, the beat-tracker in `librosa` can be employed to aggregate all the features at a beat level, thus obtaining beat-synchronous features. This process, which is common in MSA, reduces the number of the features while maintaining good accuracy in terms of structure discovery assuming most segment boundaries fall at beat positions. Nevertheless, MSAF can operate both on beat- or frame-synchronous features, and additional features can be easily introduced based on the requirements of future algorithms.

[1] `https://github.com/urinieto/msaf`

Algorithm	Boundary	Structure
2D-Fourier Magnitude Coeffs [9]	No	Yes
Checkerboard Kernel [1]	Yes	No
Constrained Cluster [3]	Yes	Yes
Convex NMF [10]	Yes	Yes
Laplacian Segmentation [5]	Yes	Yes
Ordinal LDA [6]	Yes	No
Shift Invariance PLCA [15]	Yes	Yes
Structural Features [12]	Yes	No

Table 1: Algorithms included in MSAF classified based on the subproblems that they approach in their current implementations.

1.3 The Algorithms

As aforementioned, MSA algorithms can be classified based on the subtask that they aim to solve. Currently, MSAF contains the eight different implementations listed in Table 1: seven that detect boundaries, and five that group structure.

While the implementations of [3, 5, 6, 15] are forked from their original open-source repositories, the rest of the algorithms are implemented based on their original publications due to the lack of publicly available implementations. This might introduce some differences in the results given the mentioned difficulty of reporting every single implementation detail. Regardless, MSAF is designed to be extensible, such that the inclusion of further algorithms is straightforward, thus encouraging researchers to include their novel methods in order to easily assess and share them.

1.4 The Evaluation

The Python package `mir_eval` [11] is used to evaluate each of the two subproblems of MSA. More specifically, MSAF uses by default the Hit Rate F-measure and the Standard Median Deviations [14] to assess the quality of the boundaries, and outputs the Pairwise Frame Clustering [3] and the Normalized Conditional Entropies [4] for the evaluation of the segment labels. Additionally, the T-measures [7], recently introduced and implemented in `mir_eval`, are used to evaluate multilevel segmentations such as the ones produced by [5] and [6].

1.5 The Data

JAMS [2] is the default format used in MSAF to read annotations and read/write estimations. In order to facilitate reproducibility, the following standard datasets have already been parsed to the JAMS format and are included in the MSAF repository: The Beatles TUT, SALAMI, and Isophonics. Additionally, a reduced dataset of five heavy metal tracks from the band Sargon released under a Creative Commons License is added to the repository. This Sargon dataset includes both annotations and audio files.

2. CONCLUSIONS

We have introduced an open-source framework that facilitates the task of analyzing, assessing, and comparing multiple implementations of MSA algorithms. This framework includes the processes of feature extraction, algorithm implementation, evaluation, and datasets (one of which includes audio), such that it is ready to be run as a standalone eco-system. Moreover, its design allows the easy inclusion of further MSA algorithms, in order to encourage usage from the research community for this task. In the future, we hope to see similar frameworks to facilitate research for other MIR tasks.

3. 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] Eric J. Humphrey, Justin Salamon, Oriol Nieto, Jon Forsyth, Rachel M. Bittner, and Juan P. Bello. JAMS: A JSON Annotated Music Specification for Reproducible MIR Research. In *Proc. of the 15th International Society for Music Information Retrieval Conference*, pages 591–596, Taipei, Taiwan, 2014.

[3] Mark Levy and Mark Sandler. Structural Segmentation of Musical Audio by Constrained Clustering. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2):318–326, February 2008.

[4] Hanna Lukashevich. Towards Quantitative Measures of Evaluating Song Segmentation. In *Proc. of the 10th International Society of Music Information Retrieval*, pages 375–380, Philadelphia, PA, USA, 2008.

[5] 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.

[6] 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.

[7] Brian McFee, Oriol Nieto, and Juan Pablo Bello. Hierarchical Evaluation of Music Segment Boundary Detection. In *Proc. of the 16th International Society of Music Information Retrieval*, pages 406—–412, Taipei, Taiwan, 2015.

[8] Brian Mcfee, Colin Raffel, Dawen Liang, Daniel P. W. Ellis, Matt Mcvicar, Eric Battenberg, and Oriol Nieto. librosa: Audio and Music Signal Analysis in Python. In *Proc. of the 14th Python in Science Conference*, pages 1–7, Austin, TX, USA, 2015.

[9] 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.

[10] 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.

[11] Colin Raffel, Brian Mcfee, Eric J. Humphrey, Justin Salamon, Oriol Nieto, Dawen Liang, and Daniel P. W. Ellis. mir_eval: A Transparent Implementation of Common MIR Metrics. In *Proc. of the 15th International Society for Music Information Retrieval Conference*, pages 367–372, Taipei, Taiwan, 2014.

[12] 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.

[13] Jordan B. L. Smith and Elaine Chew. A Meta-Analysis of the MIREX Structure Segmentation Task. In *Proc. of the 14th International Society for Music Information Retrieval Conference*, Curitiba, Brazil, 2013.

[14] Douglas Turnbull, Gert Lanckriet, Elias Pampalk, and Masataka Goto. A Supervised Approach for Detecting Boundaries in Music Using Difference Features and Boosting. In *Proc. of the 5th International Society of Music Information Retrieval*, pages 42–49, Vienna, Austria, 2007.

[15] Ron Weiss and Juan Pablo Bello. Unsupervised Discovery of Temporal Structure in Music. *IEEE Journal of Selected Topics in Signal Processing*, 5(6):1240–1251, 2011.