Systematic Exploration of Computational Music Structure Research

Oriol Nieto
Juan Pablo Bello

New York City, NY, USA
August 10th, 2016

PANDORA®

MARL
NYU Music and Audio Research Laboratory

Outline

- MIR task: Structural Segmentation
- MSAF: Music Structure Analysis Framework
- SPAM: Structural Poly-Annotations of Music
- Experiments

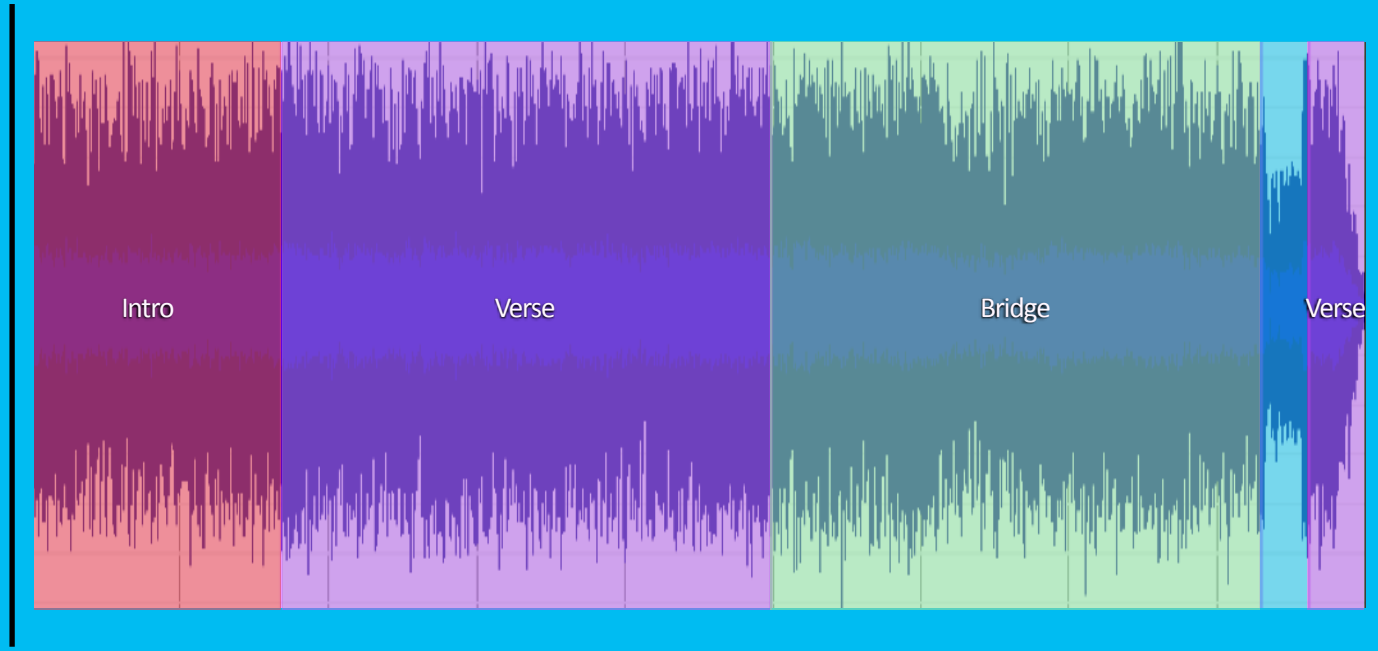
http://creationoutreach.com/sitebuildercontent/sitebuilderpictures/spirals3.jpg

Structural Segmentation

MIR TASK

Goal:

Automatically identify the large-scale non-overlapping music segments from an audio signal



Two Subtasks: Boundary Detection & Structural Grouping (Labeling)

Structural Segmentation

CURRENT CHALLENGES

Ambiguity

Multiple approaches have been proposed, but some not open source

Hard to identify the best solution that best aligns with the final application


MSAF
PANDORA

MSAF

MUSIC STRUCTURE ANALYSIS FRAMEWORK

Open source Python package to facilitate research in the task of Structural Segmentation
- MIT License

Contains all moving parts:
- Feature computation
- Algorithm implementations
- Evaluation metrics
- Human annotated datasets

```
$ pip install msaf
```

Each moving part interchangeable (including boundary and labeling algorithms)

Designed to be easy to use and to extend

MSAF

MUSIC STRUCTURE ANALYSIS FRAMEWORK

Features
CQT
PCP
MFCCs
Tonnetz
(Tempograms)

librosa (McFee 2015)

Results: MIREX 2016


SPAM
PANDORA

SPAM Dataset

STRUCTURAL POLY-ANNOTATIONS OF MUSIC

Data collection:

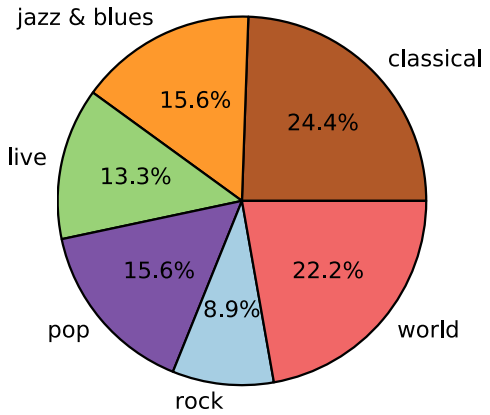
- Select tracks that are highly ambiguous to segment
- Have them annotated by multiple experts

Song selection:

- Run all MSAF algorithms on 2000+ tracks
- Rank them based on their Mean Ground-Truth Precision

Choose 50 tracks:

- 45 most challenging (lowest MGP)
- 5 least challenging (highest MGP)



Annotations:

- **5 different annotations** per track
- Using the **JAMS** format
- Pre-computed features available

Experiments

WITH MSAF AND SPAM

Default parameters:

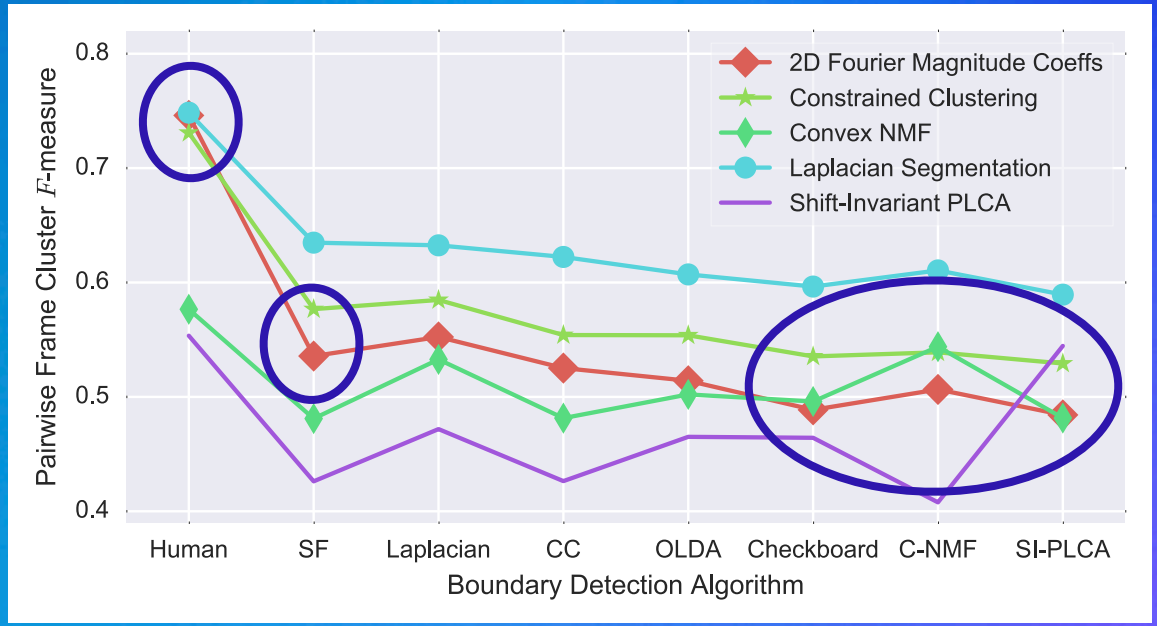
Features	Evaluation
PCP (Beat-synchronous)	Hit Rate @ 3 seconds
	Pairwise Frame Clustering

Experiments - Algorithms

ON THE BEATLES DATASET

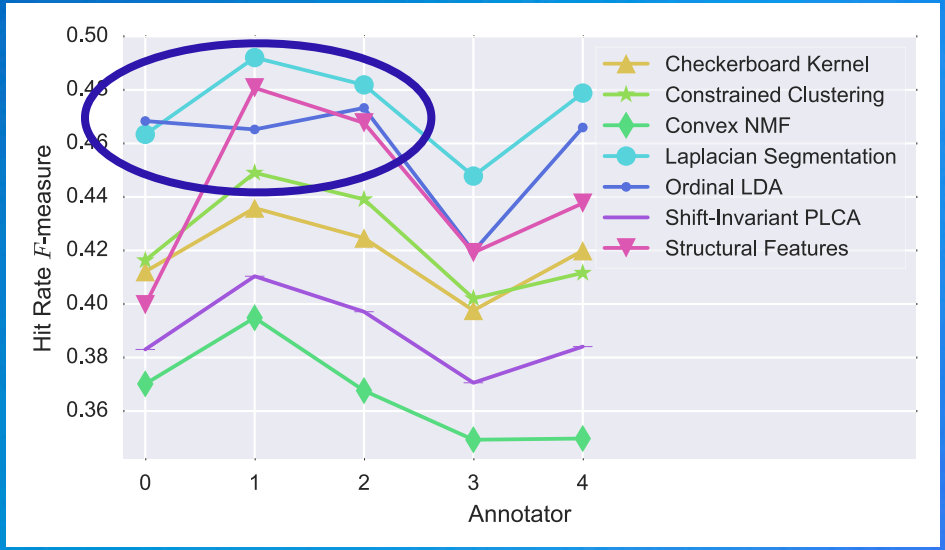
Some **label** algorithms are more robust to the quality of the **boundaries** than others (eg 2D-FMC)

Label algorithms are ranked differently based on the **boundaries**



Experiments - Annotations

ON THE SPAM DATASET



Single “ground-truth” for boundaries
can be misleading due to **ambiguity**

More experiments in the Paper

The Future

Include more algorithms in MSAF

Extend MSAF to other music structure tasks:

- Pattern discovery
- Hierarchical structure (already in development)

Have more open source systems à la MSAF for other MIR tasks (see McFee 2016)

Conclusions

Presented **MSAF**: open source framework to facilitate research in music structural segmentation

`pip install msaf`

(https://github.com/urinieto/msaf)

Presented **SPAM**: poly-annotated dataset to better approach the ambiguity problem of music structure

https://github.com/urinieto/msaf-data

Experiments suggest:

- Label algorithms depend on quality of the boundaries
- Relying in single human reference may be misleading

https://github.com/urinieto/msaf-experiments

THANK YOU!

References

Foote, J. (2000). Automatic Audio Segmentation Using a Measure Of Audio Novelty. In Proc. of the IEEE International Conference of Multimedia and Expo (pp. 452–455). New York City, NY, USA.

Humphrey, E. J., Salamon, J., Nieto, O., Forsyth, J., Bittner, R. M., & Bello, J. P. (2014). JAMS: A JSON Annotated Music Specification for Reproducible MIR Research. In Proc. of the 15th International Society for Music Information Retrieval Conference (pp. 591–596). Taipei, Taiwan.

Levy, M., & Sandler, M. (2008). Structural Segmentation of Musical Audio by Constrained Clustering. IEEE Transactions on Audio, Speech, and Language Processing, 16(2), 318–326. http://doi.org/10.1109/TASL.2007.910781

McFee, B., Raffel, C., Liang, D., Ellis, D. P. W., McVicar, M., Battenberg, E., & Nieto, O. (2015). librosa: Audio and Music Signal Analysis in Python. In Proc. of the 14th Python in Science Conference (pp. 1–7). Austin, TX, USA.

McFee, B., & Ellis, D. P. W. (2014a). Learnign to Segment Songs With Ordinal Linear Discriminant Analysis. In Proc. of the 39th IEEE International Conference on Acoustics Speech and Signal Processing (pp. 5197–5201). Florence, Italy.

McFee, B., & Ellis, D. P. W. (2014b). Analyzing Song Structure with Spectral Clustering. In Proc. of the 15th International Society for Music Information Retrieval Conference (pp. 405–410). Taipei, Taiwan.

Nieto, O., & Jehan, T. (2013). Convex Non-Negative Matrix Factorization For Automatic Music Structure Identification. In Proc. of the 38th IEEE International Conference on Acoustics Speech and Signal Processing (pp. 236–240). Vancouver, Canada.

Nieto, O., & Bello, J. P. (2014). Music Segment Similarity Using 2D-Fourier Magnitude Coefficients. In Proc. of the 39th IEEE International Conference on Acoustics Speech and Signal Processing (pp. 664–668). Florence, Italy. http://doi.org/10.1109/ICASSP.2014.6853679

Weiss, R., & Bello, J. P. (2011). Unsupervised Discovery of Temporal Structure in Music. IEEE Journal of Selected Topics in Signal Processing, 5(6), 1240–1251.