Music Segment Similarity Using 2D-Fourier Magnitude Coefficients

Oriol Nieto
Juan P. Bello

New York, NY, USA
January, 25th 2013

MARL
NYU Music and Audio Research Laboratory

Overview

- Music Structure Analysis
- 2D-Fourier Magnitude Coefficients
- Experiments
- Conclusions and Discussion

MARL

Overview

- **Music Structure Analysis**
- 2D-Fourier Magnitude Coefficients
- Experiments
- Conclusions and Discussion

MARL

Music Structure Analysis Overview

- Goal:
  - Automatically identify the different segments (or sections) of a musical piece.
- Motivation:
  - Easier intra-piece navigation in music players.
  - Automatic generation of summaries and/or mash-ups.
  - Large-scale musicological research.
- Two subproblems:
  - Estimate the musical boundaries (time points that mark the start/end of a segment).
  - Classify the segments based on their acoustic similarity (e.g. verse, chorus).
- In this work we only focus on the **segment similarity** subproblem in **Western popular music**.

MARL

Segment Similarity

- Music similarity between two segments:
  - Common harmonic or melodic sequences
  - Possible key-transpositions
  - Acceptable phase shifts in patterns
  - Might be played at different tempi
- Beat-synchronous 2D-FMCs are an excellent feature representation candidate:
  - Key-transposition invariance
  - Phase shift invariance
  - Local tempo invariance (beat-sync)
- Previously, in MIR, 2D-FMCs used for Large Scale Cover Song Identification (Bertin-Mahieux 2012, Humphrey 2013)

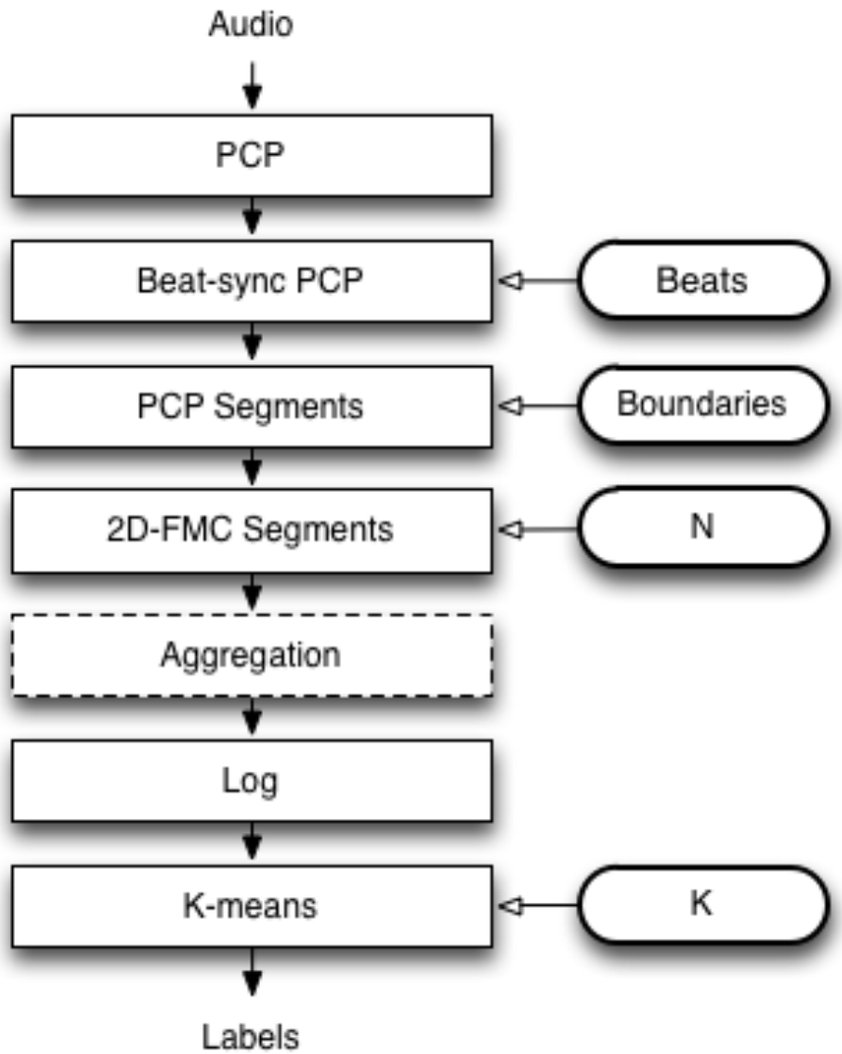
Overview

- Music Structure Analysis
- **2D-Fourier Magnitude Coefficients**
- Experiments
- Conclusions and Discussion

MARL

2D-Fourier Magnitude Coefficients in Music Segment Similarity

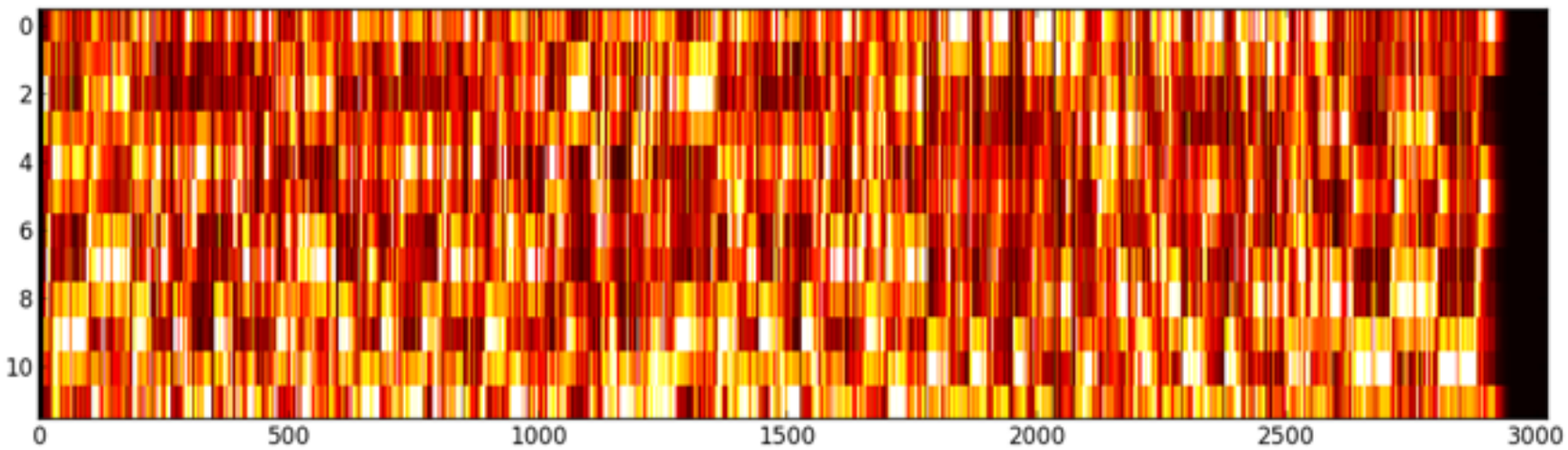
- Pipeline of the system:

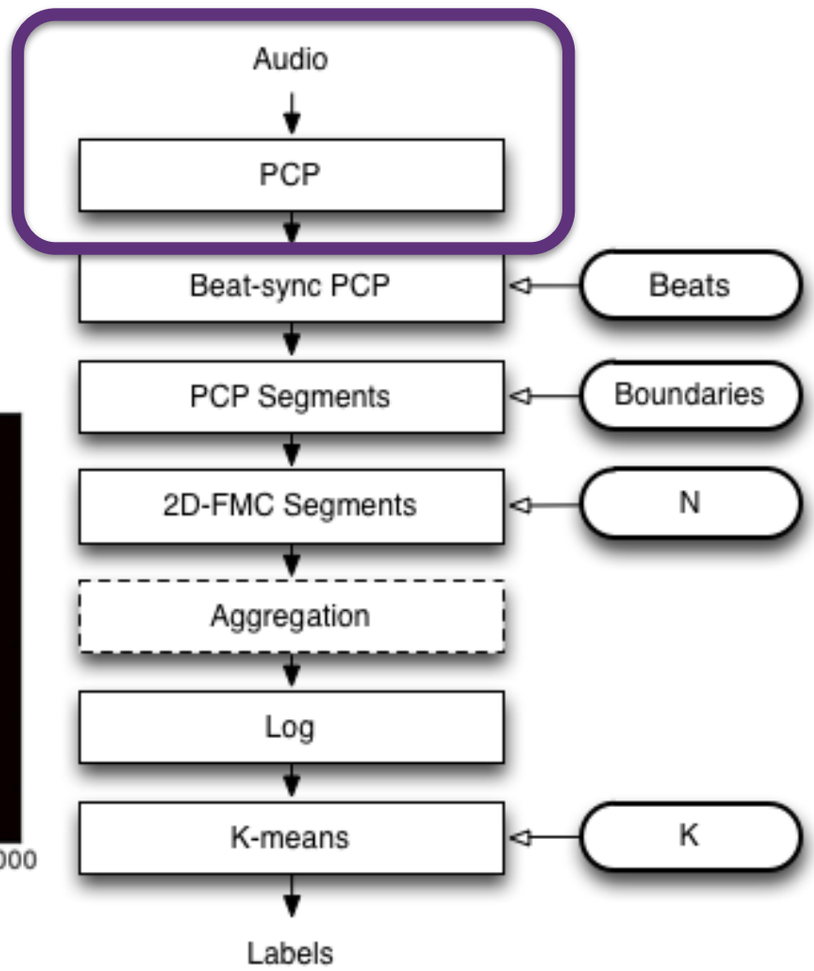


MARL

2D-Fourier Magnitude Coefficients in Music Segment Similarity

- PCP:
  - Pitch Class Profiles (Chromagram)
  - Harmonic representation:

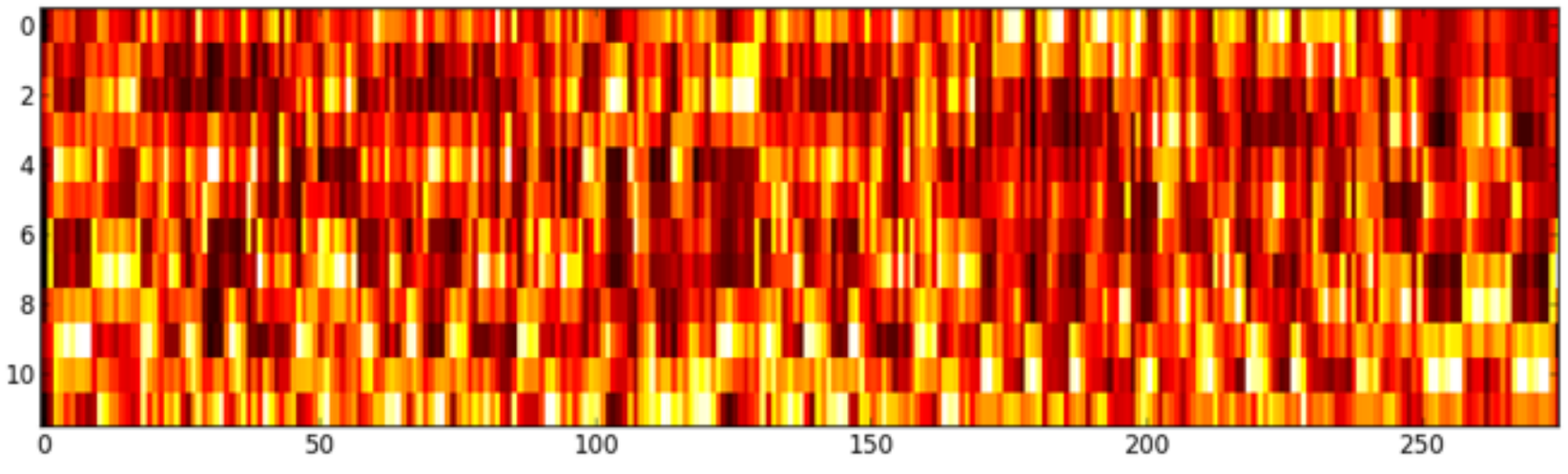


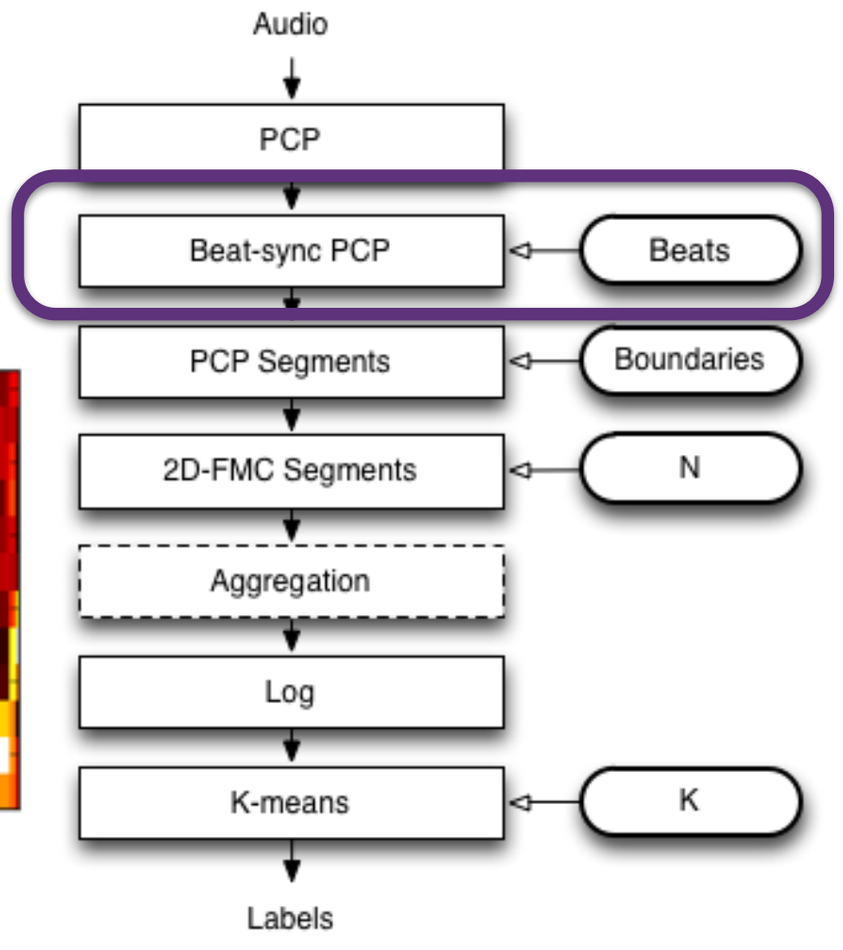


MARL

2D-Fourier Magnitude Coefficients in Music Segment Similarity

- Beat-Synchronous PCP:
  - Obtain the beats using The Echo Nest API
  - Average pitch vectors within beat boundaries (Ellis 2007)

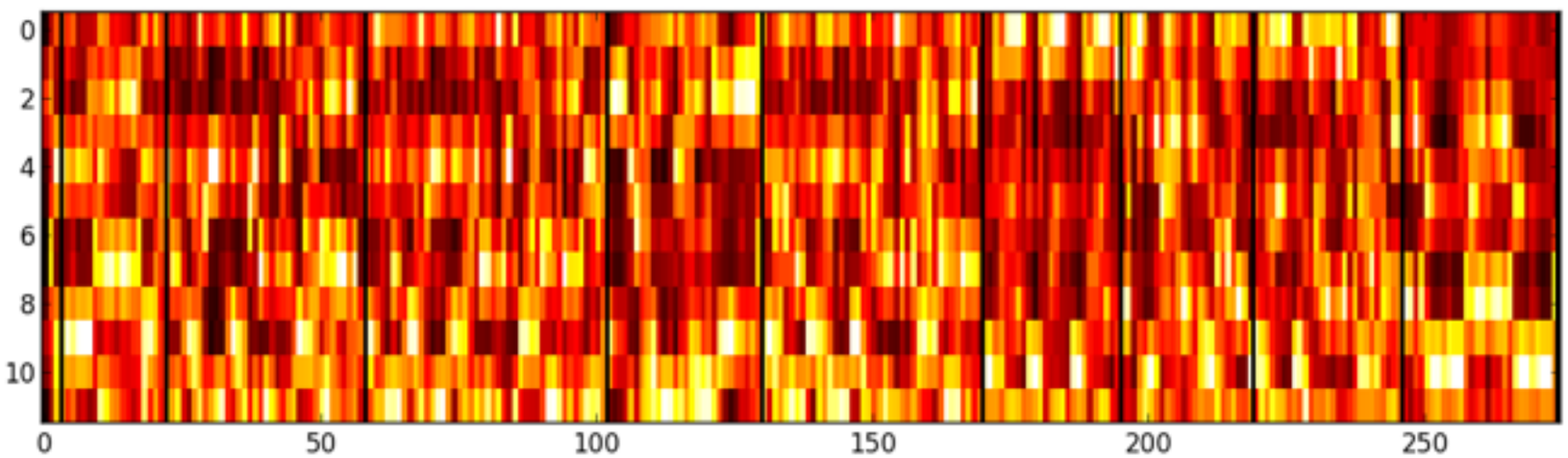


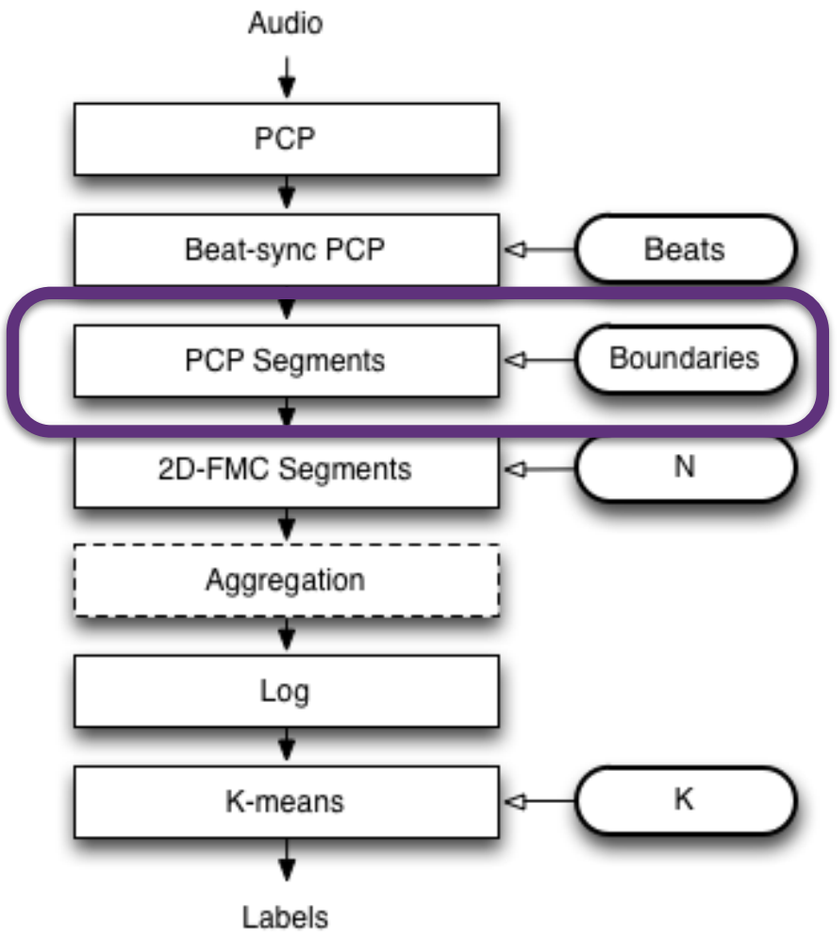


MARL

2D-Fourier Magnitude Coefficients in Music Segment Similarity

- PCP segments:
  - Obtain the segment boundaries using:
    - ground truth
    - Automatic method (Serrà 2012)





MARL

2D-Fourier Magnitude Coefficients in Music Segment Similarity

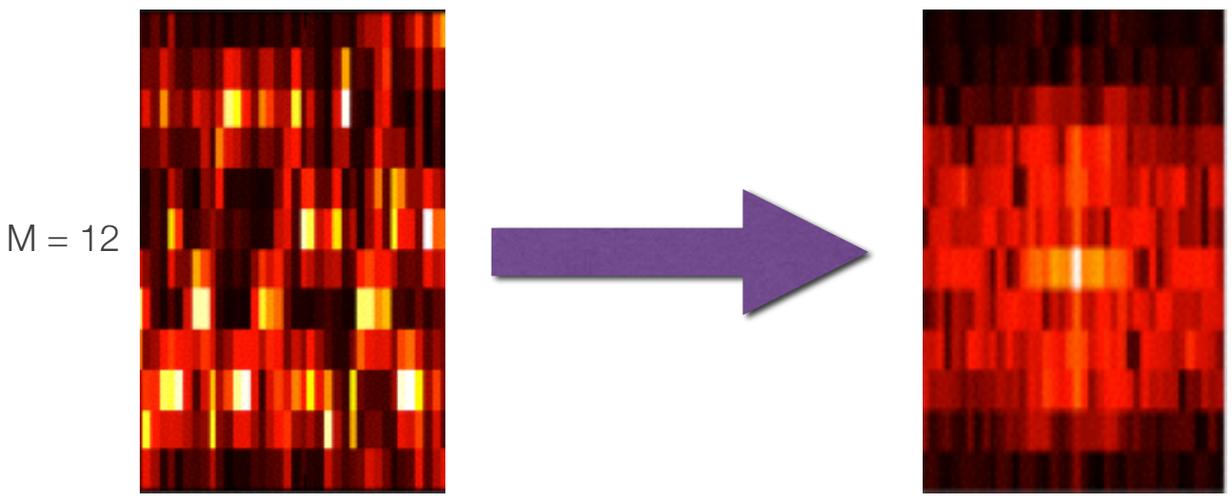
- 2D Fourier Transform:

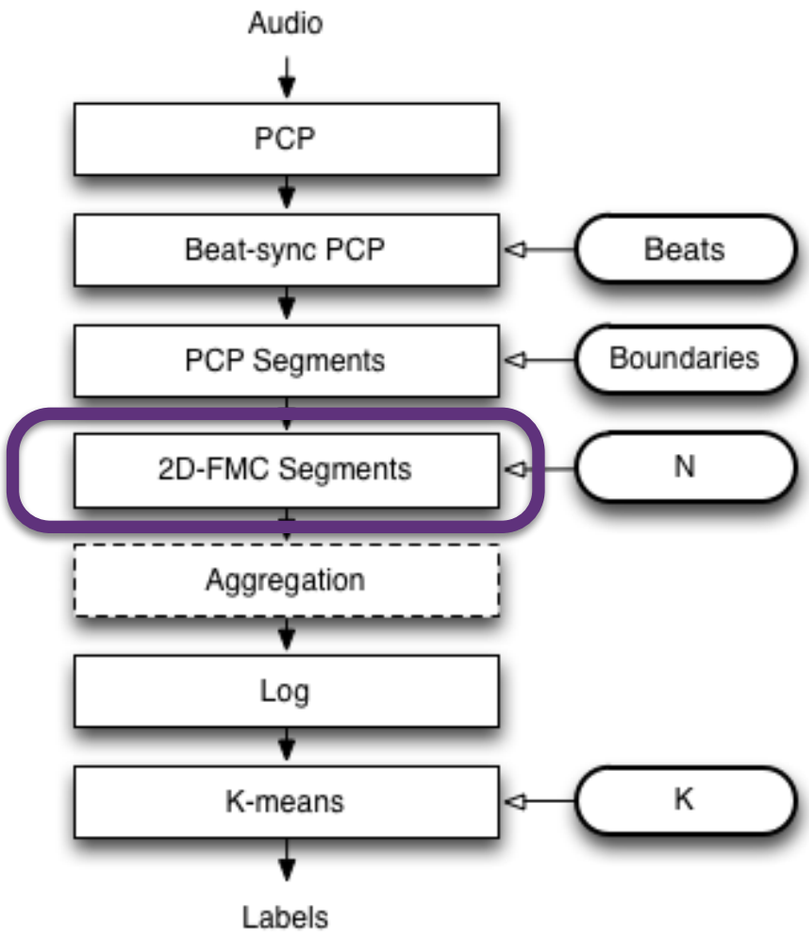
$$X(u,v) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x_i(m,n) e^{-2\pi i \left(\frac{mu}{M} + \frac{nv}{N} \right)}$$

- We **discard the phase**, keep only the Magnitude:
  - Key-transposition invariance
  - Phase-shift invariance

Input: PCP Segment

Output: 2D-FMC patch

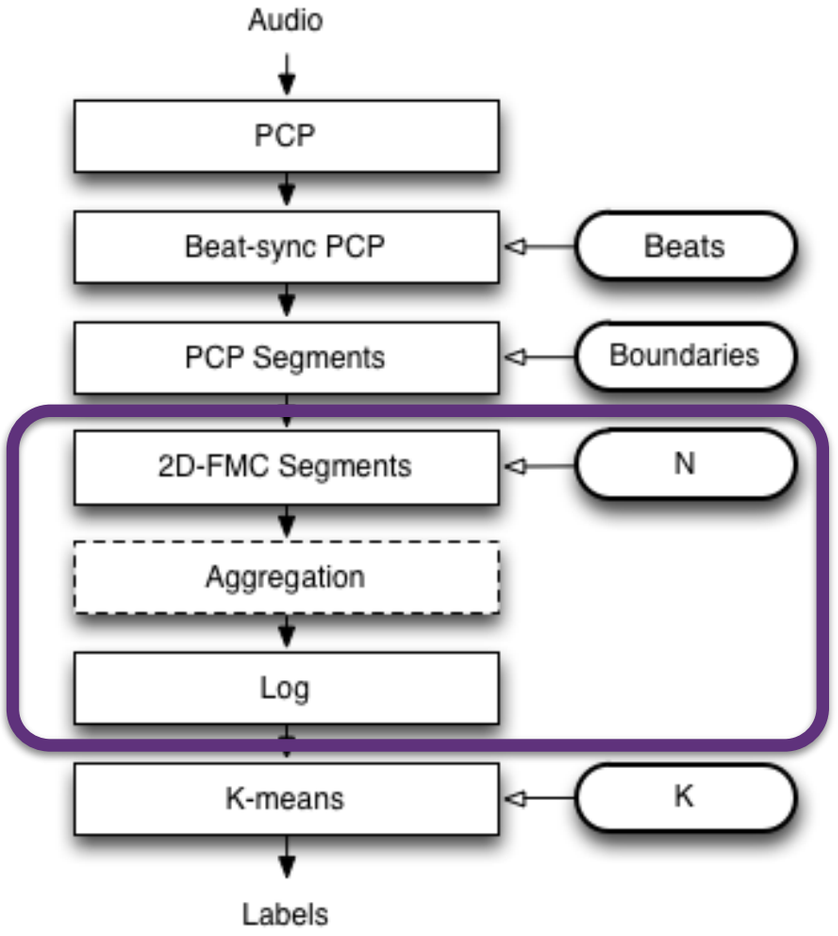




MARL

2D-Fourier Magnitude Coefficients in Music Segment Similarity

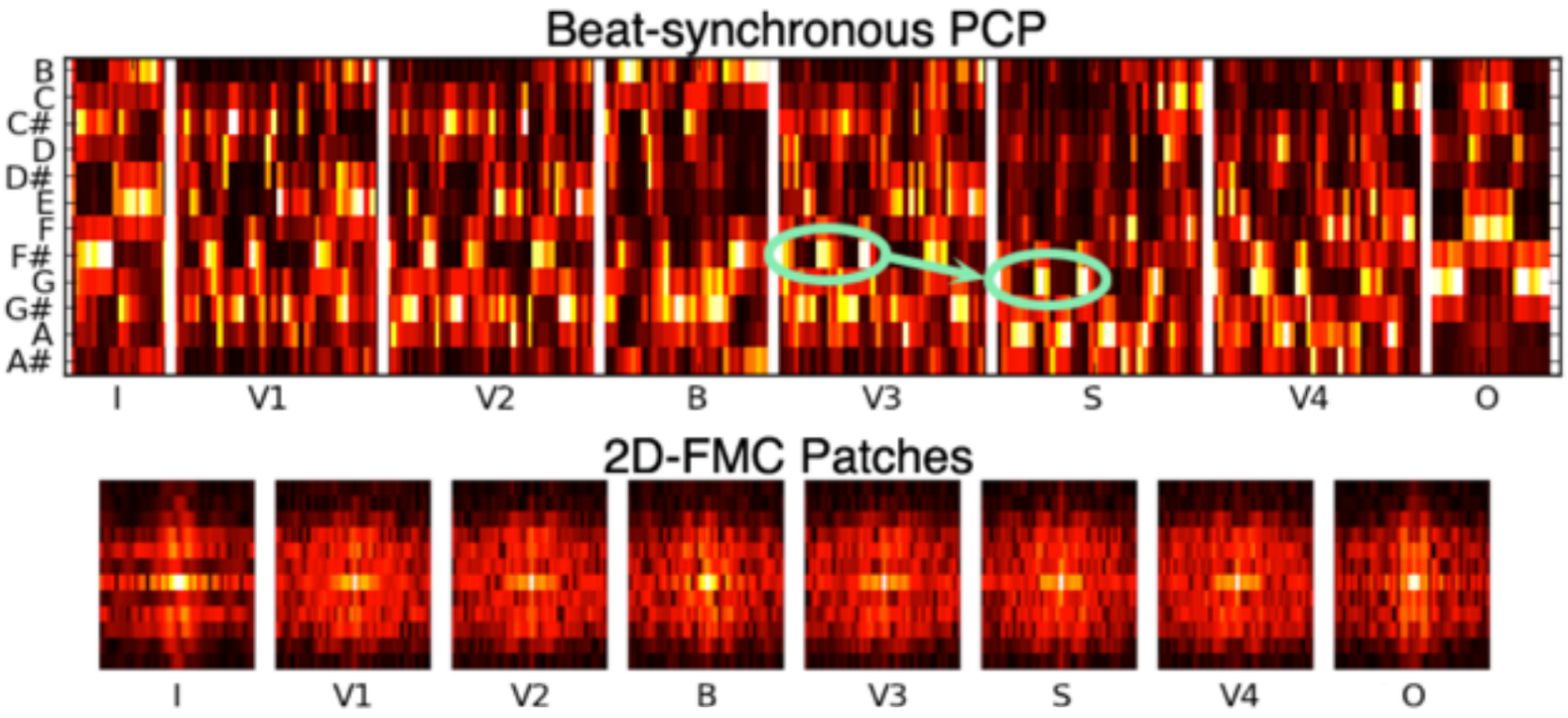
- Segment-synchronous 2D-FMCs:
  - We want to ultimately compare each 2D-FMC segment
  - We need to have 2D-FMC segments of the same size in order to quantify the similarity: **segment-synchronous**
- Segment-synchronization Strategy:
  - Maximum window size
    - N = Maximum segment size
    - Zero pad the rest
    - No aggregation
- Finally, we take the log of the 2D-FMCs

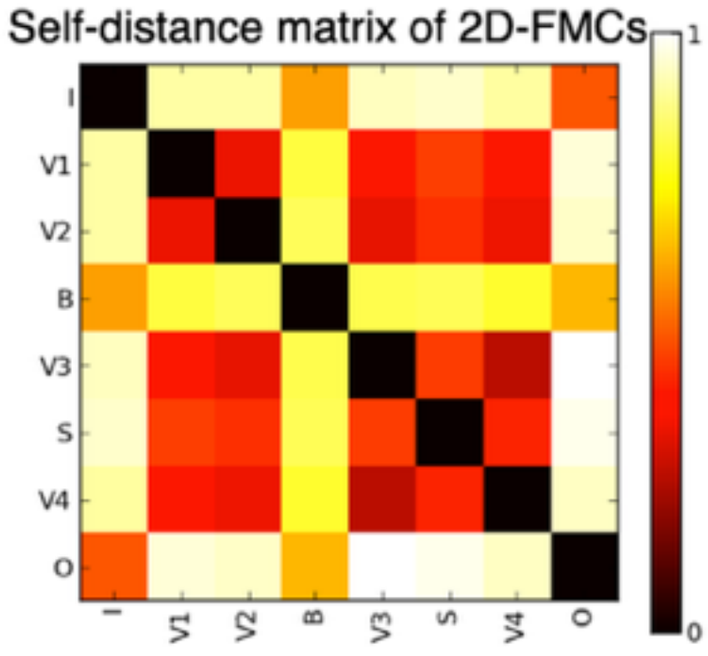


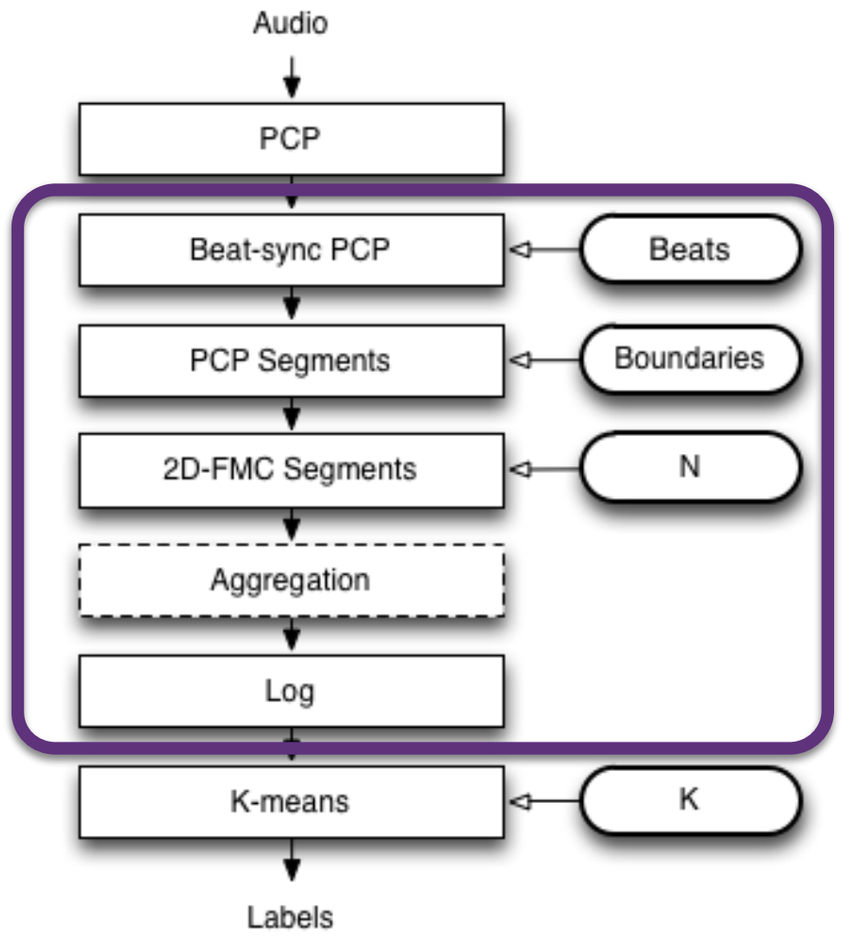
MARL

2D-Fourier Magnitude Coefficients in Music Segment Similarity

- Example (And I Love Her - The Beatles)



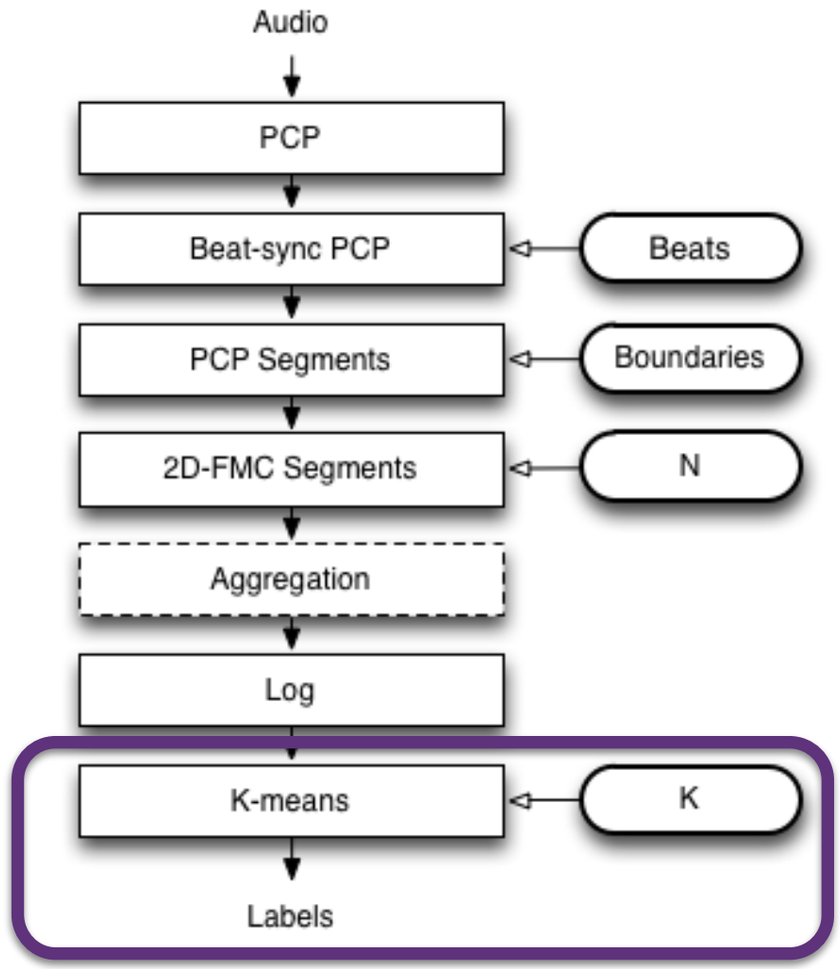




MARL

2D-Fourier Magnitude Coefficients in Music Segment Similarity

- Cluster the segment-synchronous 2D-FMCs:
  - K-means with Euclidean distance
  - We can estimate K:
    - Bayesian Information Criterion (BIC) validates the quality of each partition
    - Run K-means with various K and use the knee point detection method in BIC (Zhao 2008).
  - Output: Labels of the segments



MARL

Overview

- Music Structure Analysis
- 2D-Fourier Magnitude Coefficients
- **Experiments**
- Conclusions and Discussion

MARL

Experiments: Evaluation

- The Beatles Dataset:
  - 180 human annotated tracks

- Evaluation metrics:
- Pairwise clustering (Levy 2008):
  - **Pf** : F-measure
  - **Pp**: Precision
  - **Pr**: Recall

- Entropy scores (Lukashevich 2008):
  - **Sf**: F-measure
  - **So**: Over-segmentation
  - **Su**: Under-segmentation

"Nobody knows what entropy really is, so in any discussion you will always have an advantage" - John von Neumann

MARL

Experiments

- Experiment 1:
  - Ground Truth Boundaries
  - Ground Truth K (number of unique segments per track)

Ntype	Aggr.	P_F	P_P	P_R	S_F	S_o	S_u
Max	–	**81.96**	84.35	81.3	**87.18**	86.27	89.14
Kaiser [10]		80.0	87.0	76.6	–	–	–

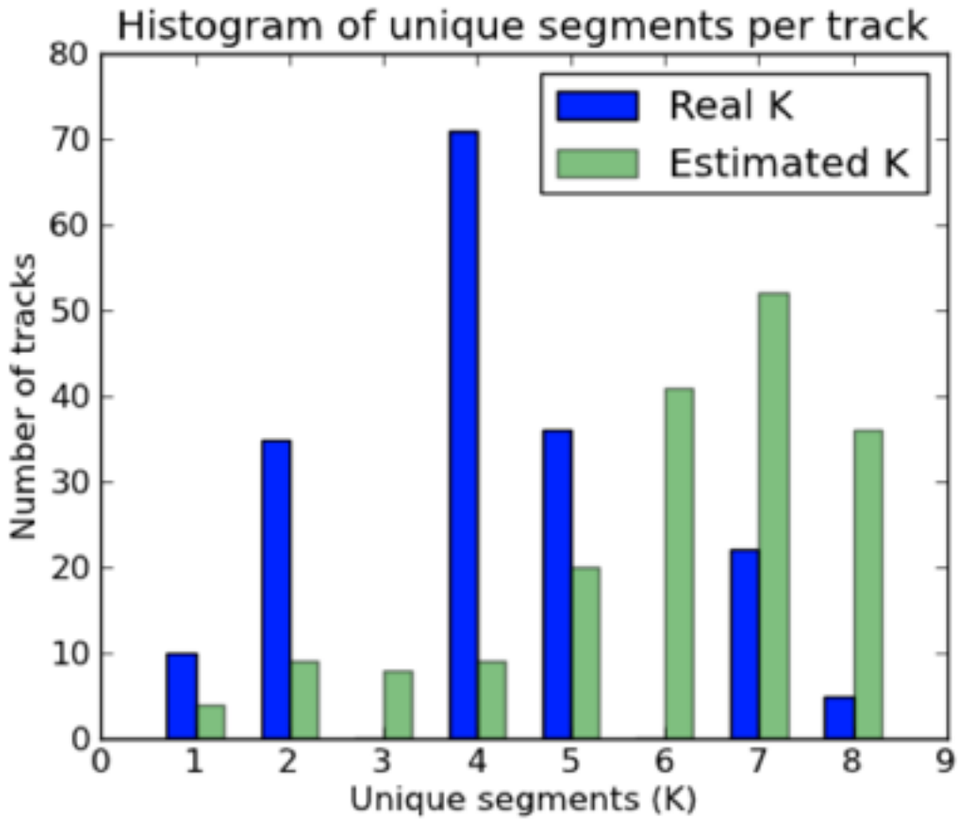
MARL

Experiments

- Experiment 2:
  - Estimate K
  - Ground Truth Boundaries
  - Fixed and automatic K (using the knee method in BIC)

k	P_F	P_P	P_R	S_F	S_o	S_u
3	68.20	55.94	95.03	71.46	94.54	59.66
4	76.12	70.18	88.60	81.20	89.60	76.29
5	**76.83**	80.47	77.93	**83.28**	82.68	85.82
6	72.26	85.14	66.11	81.68	76.14	90.30
auto	71.50	83.93	68.76	80.35	83.39	85.65

- Fixing K=5 yields better F-measures (but this is overfitting to The Beatles)
- Hard task to estimate K given the small amount of segments to cluster per song
- X-means might yield better estimations



MARL

Experiments

- Experiment 3:
  - Estimated Boundaries (Serrà 2012)
  - Fixed and automatic K

k	P_F	P_P	P_R	S_F	S_o	S_u
4	53.93	47.57	67.18	58.76	69.00	53.37
5	54.41	53.83	58.75	63.01	65.82	62.48
6	57.34	64.07	54.49	68.09	65.26	72.95
7	58.31	71.74	51.15	**71.15**	65.01	80.19
auto	57.31	66.68	52.75	68.95	65.99	76.39
Kaiser [10]	60.8	61.5	64.6	–	–	–
Mauch [13]	**66**	61	77	69.48	76	64
Nieto [5]	59.3	48.9	83.2	47.78	49.8	47.8
Paulus [8]	59.9	72.9	54.6	–	–	–
Peiszer [6]*	59.7	61.1	62.3	–	–	–
Weiss [4]	60	57	69	58.84	62	56

  - Worse pairwise clustering measures
  - State of the art entropy scores
  - Significant drop in scores when using estimated boundaries
  - Which metric is better?

MARL

Overview

- Music Structure Analysis
- 2D-Fourier Magnitude Coefficients
- Experiments
- **Conclusions and Discussion**

MARL

Conclusions and Discussion

- Beat-Sync 2D-FMC representation:
  - Key-transposition invariant
  - Phase-shift invariant
  - Local tempo invariant
  - State of the art when using fixed boundaries
  - Hard to estimate K
    - Use X-means?
    - Use more 2D-FMC patches (one per beat)?
  - Future work: Use this method to estimate boundaries

MARL

References

- Bertin-Mahieux, T., & Ellis, D. P. W. (2012). Large-Scale Cover Song Recognition Using The 2D Fourier Transform Magnitude. In Proc. of the 13th International Society for Music Information Retrieval Conference (pp. 241–246). Porto, Portugal.
- Humphrey, E. J., Nieto, O., & Bello, J. P. (2013). Data Driven and Discriminative Projections for Large-scale Cover Song Identification. In Proc. of the 14th International Society for Music Information Retrieval Conference. Curitiba, Brazil.
- 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. doi:10.1109/TASL.2007.910781
- 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.
- Serrà, J., Müller, M., Grosche, P., & Arcos, J. L. (2012). Unsupervised Detection of Music Boundaries by Time Series Structure Features. In Proc. of the 26th AAAI Conference on Artificial Intelligence (pp. 1613–1619). Toronto, Canada.
- Smith, J. B., Burgoyne, J. A., Fujinaga, I., De Roure, D., & Downie, J. S. (2011). Design and Creation of a Large-Scale Database of Structural Annotations. In Proc. of the 12th International Society of Music Information Retrieval (pp. 555–560). Miami, FL, USA.
- 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.

MARL

Questions?

- Beat-Sync 2D-Fourier Magnitude Coefficient representation:
  - Key-transposition invariant
  - Phase-shift invariant
  - Local tempo invariant

MARL