Toward Automatic Music Audio Summary Generation from Signal Analysis

Nicolas Ayroles

Michel Blancard

MVA 2014/2015

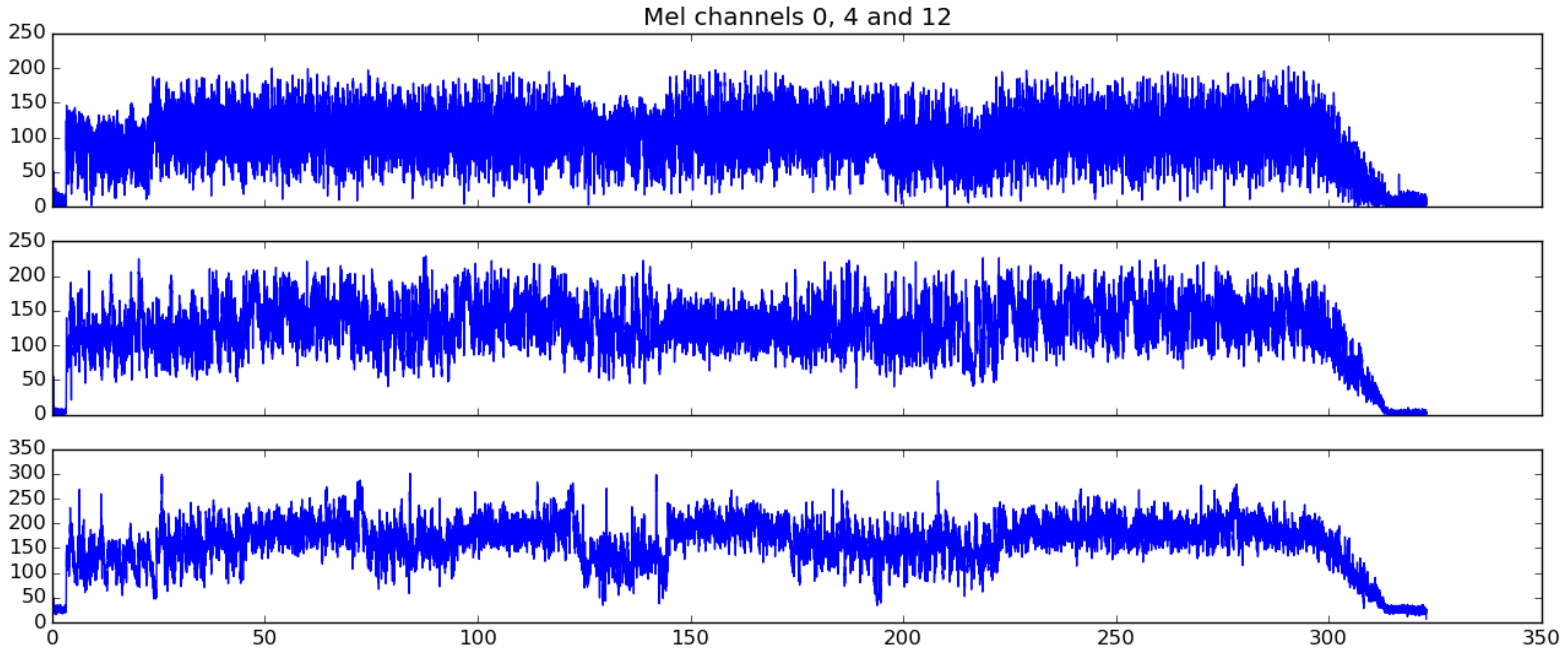
- Music summarization
  - Link similar parts (“states”)
  - Separate dissimilar parts

- Applications
  - Online distribution plateform
  - Music search engines

- Summary of the pipeline
  - First pass : find initial states
  - Second pass : K-means and HMM

Mel filter bank

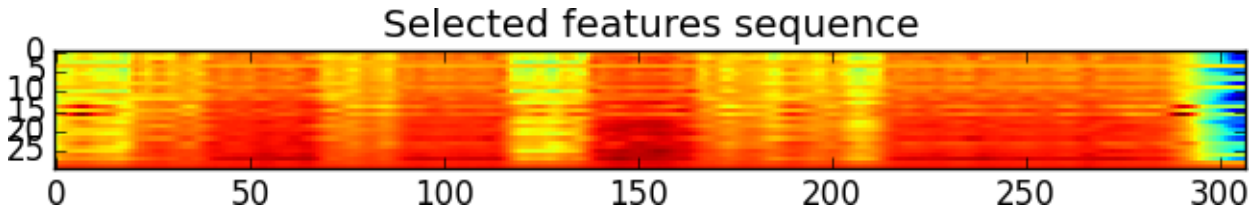
- 44100Hz → 100Hz
- 26 channels



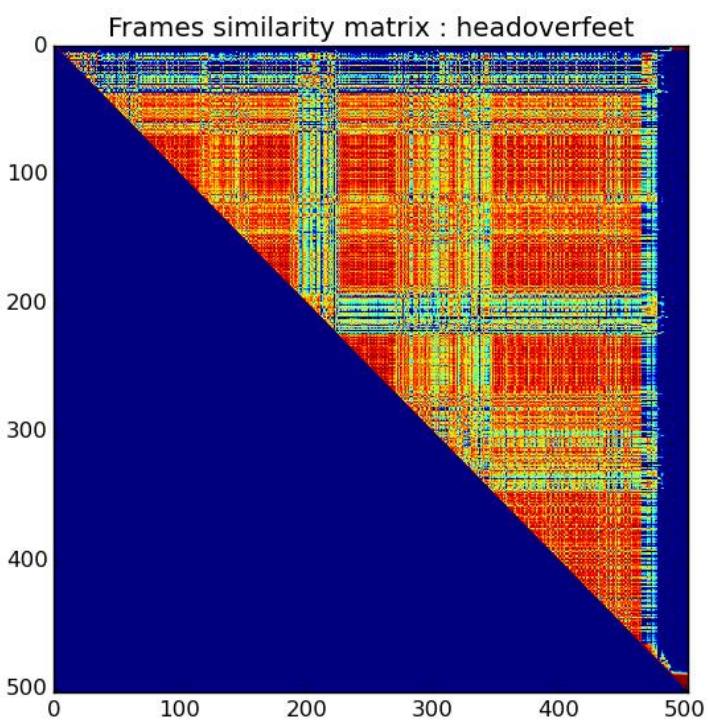
Dynamic features

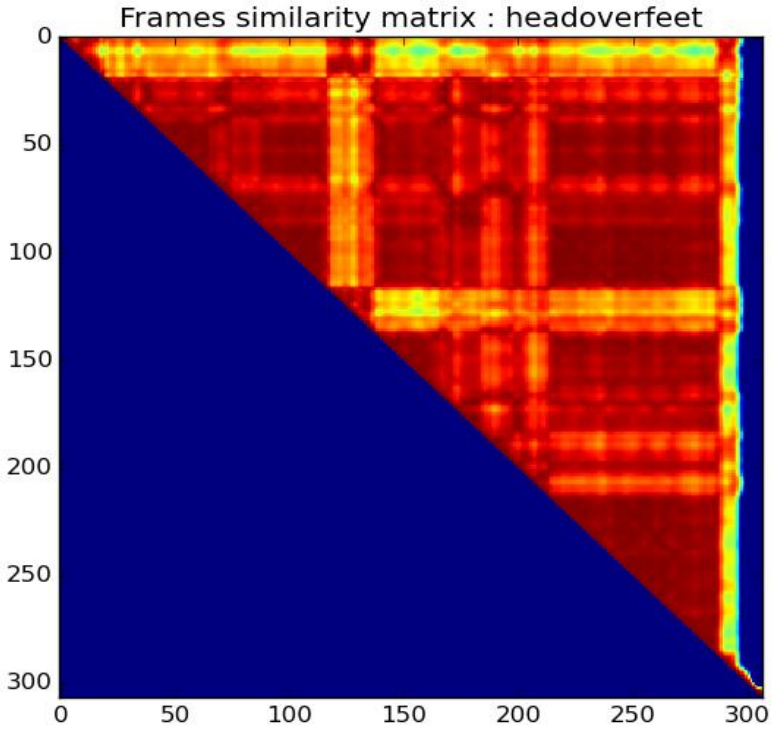
- STFT on the Mel features
- Choice of the window size
- 26x513 features
- Most of them are useless !

- Mutual information
- 20-50 features

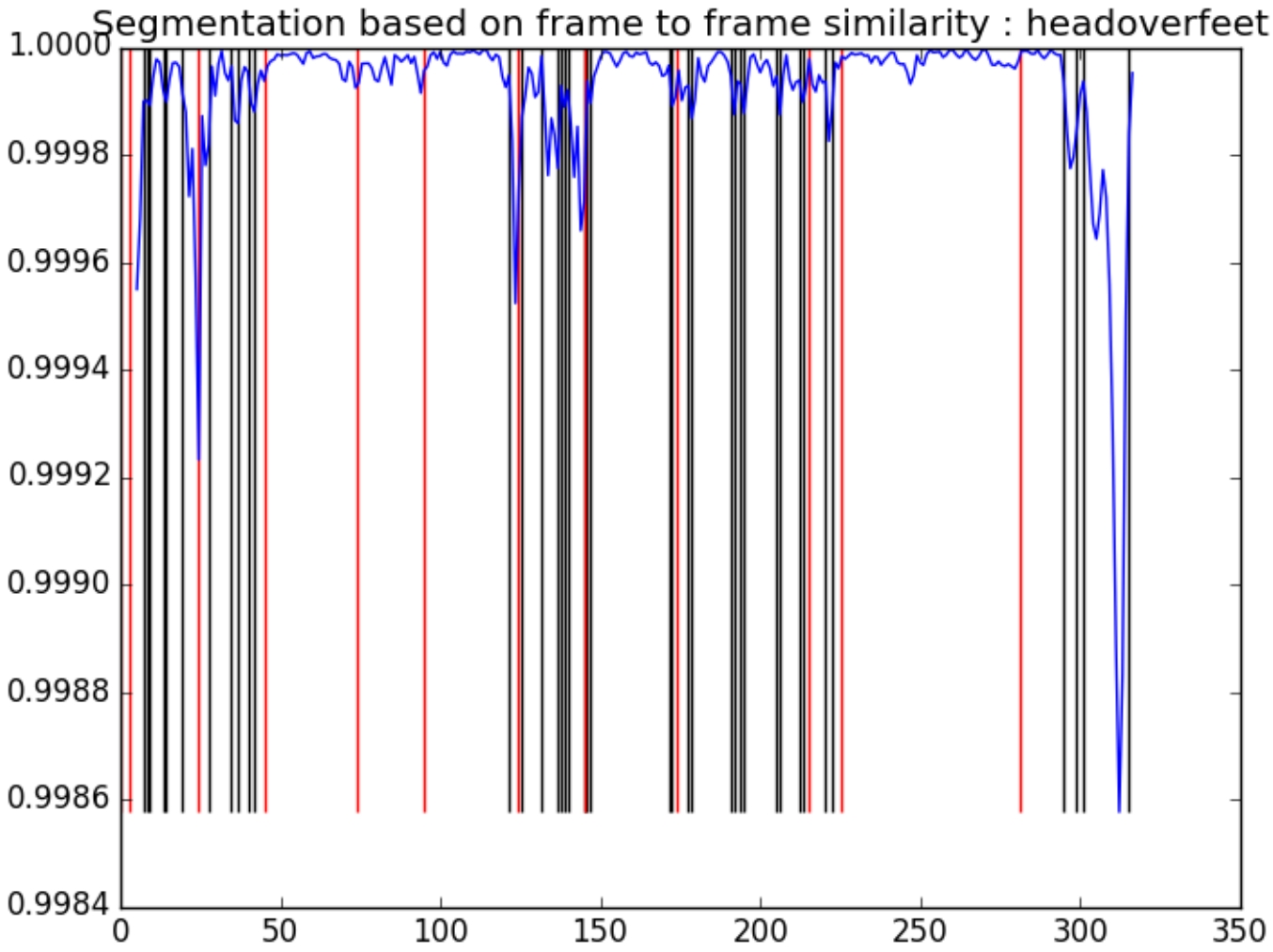


Similarity matrix

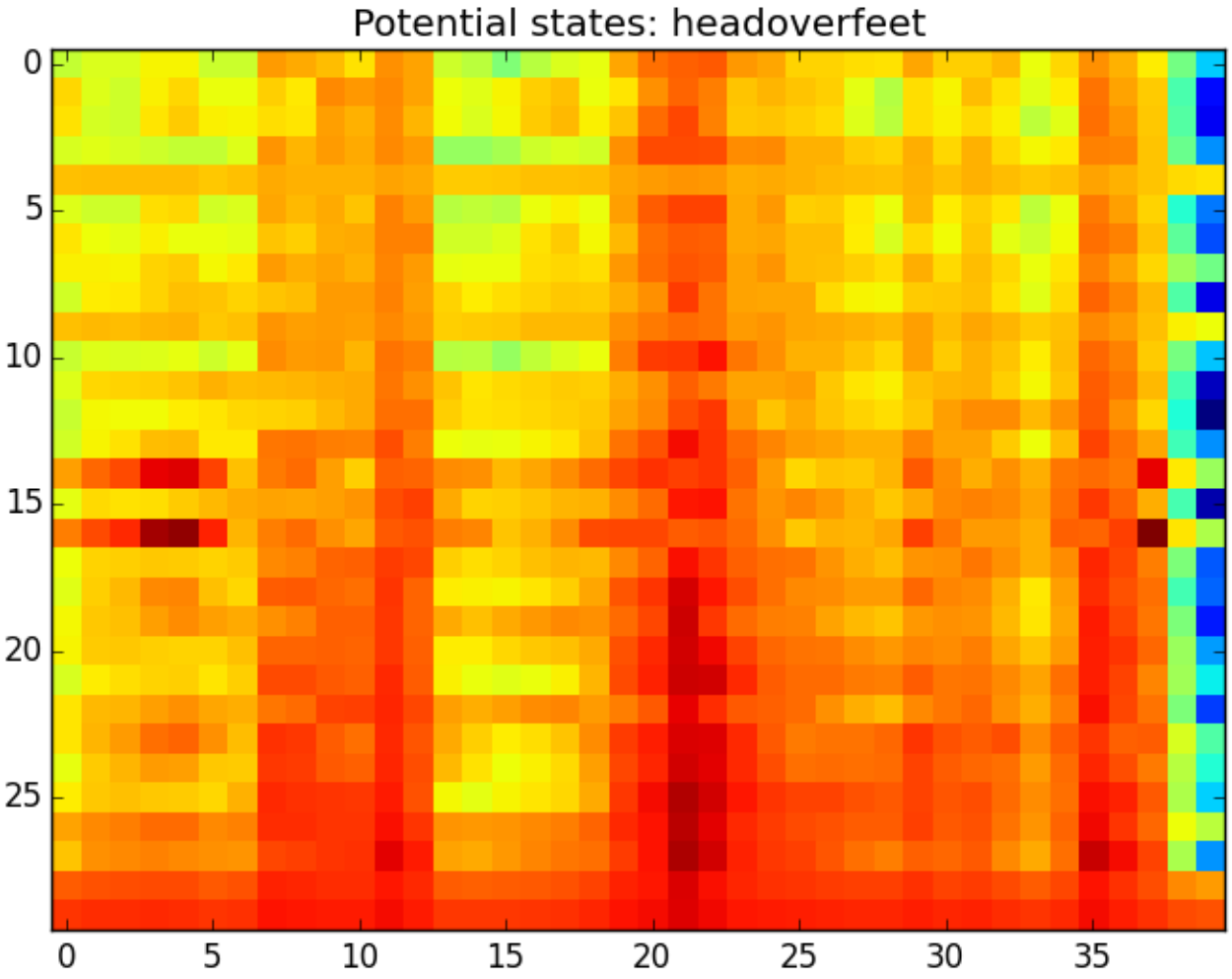




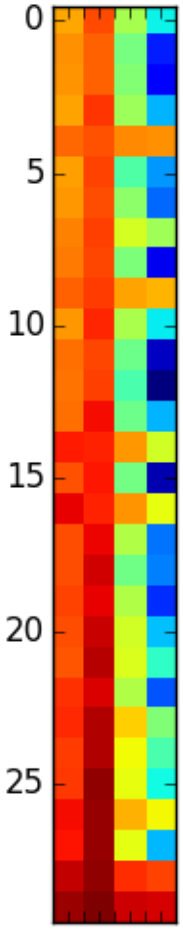
Segmentation



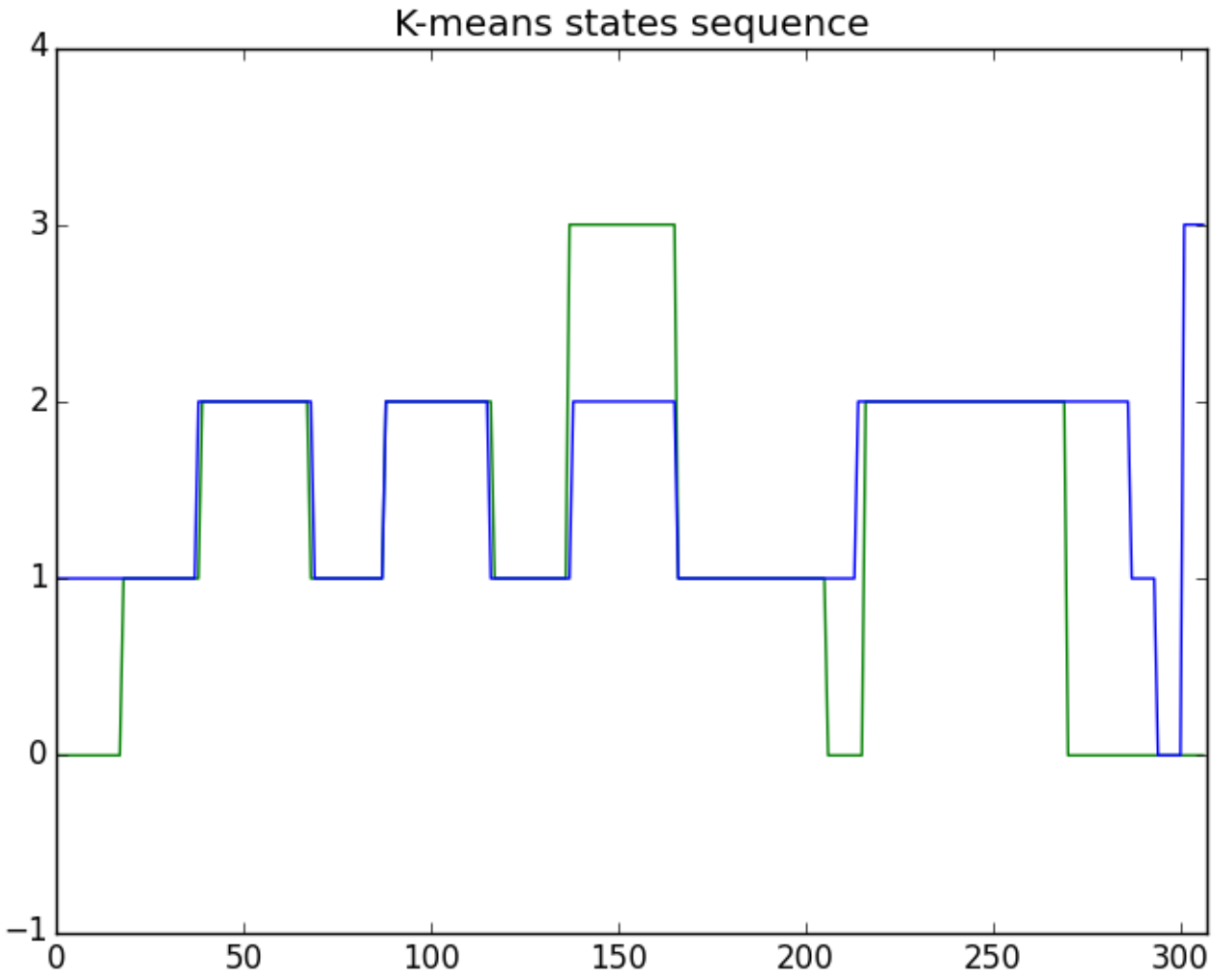
- Average over the segments



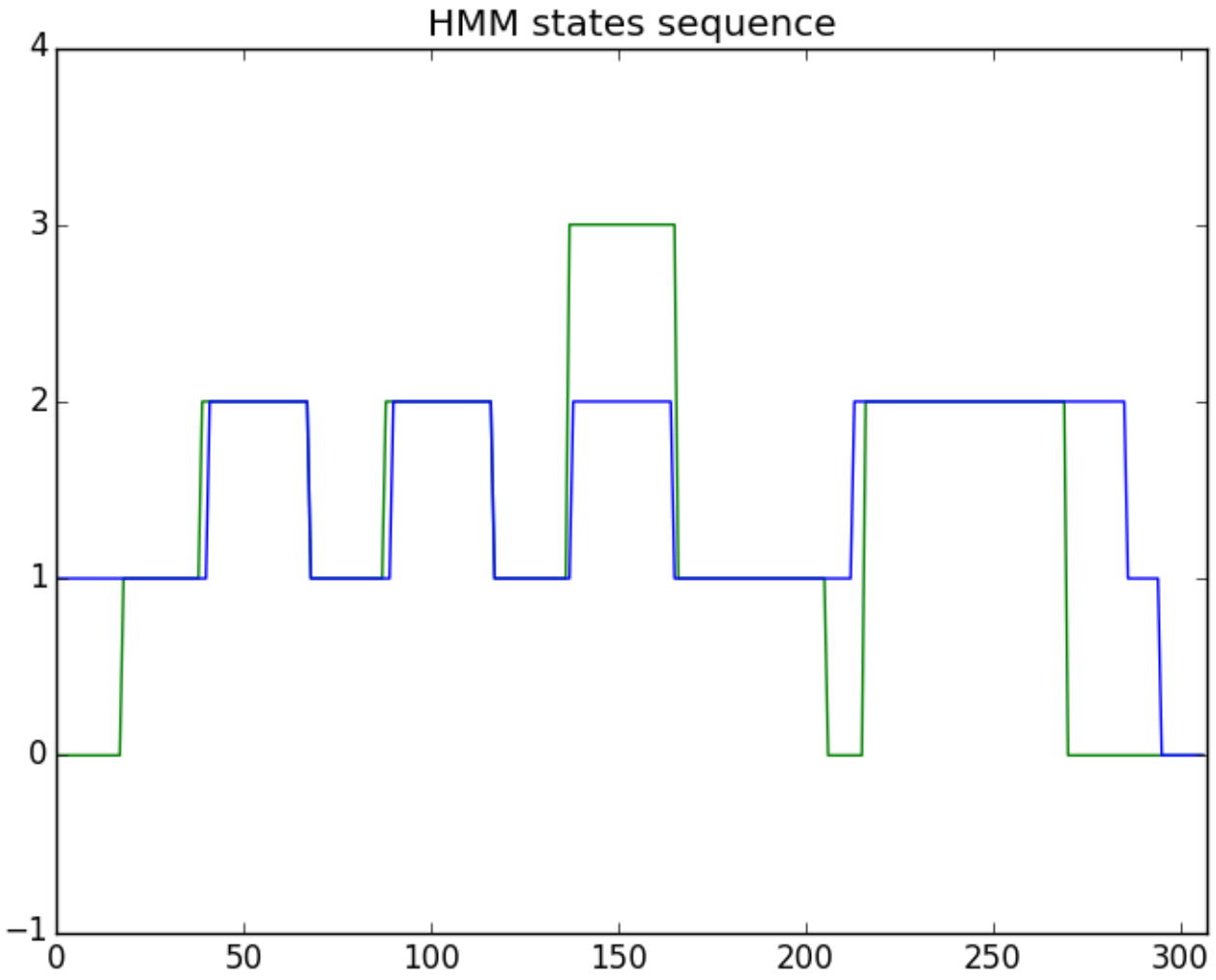
- Regroup the potential states by similarity



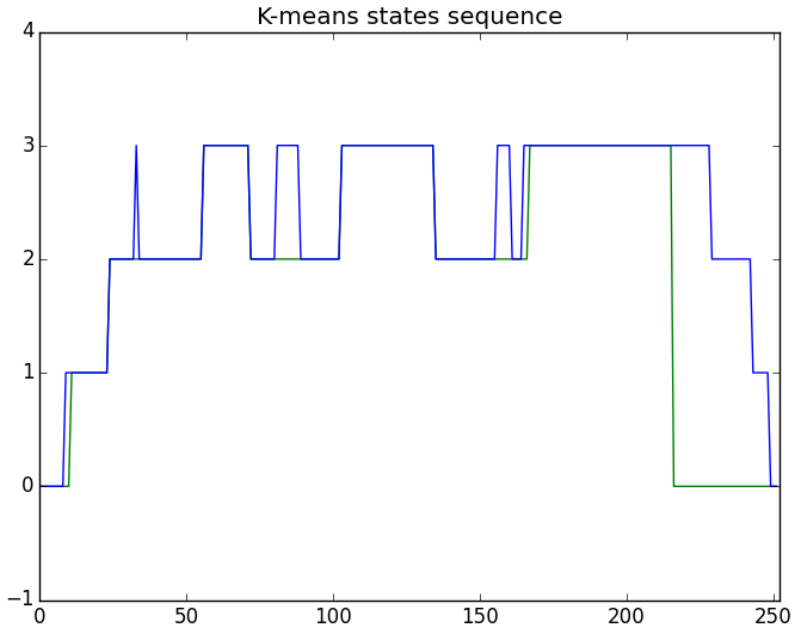
- Assign a state a each feature vector by K-means

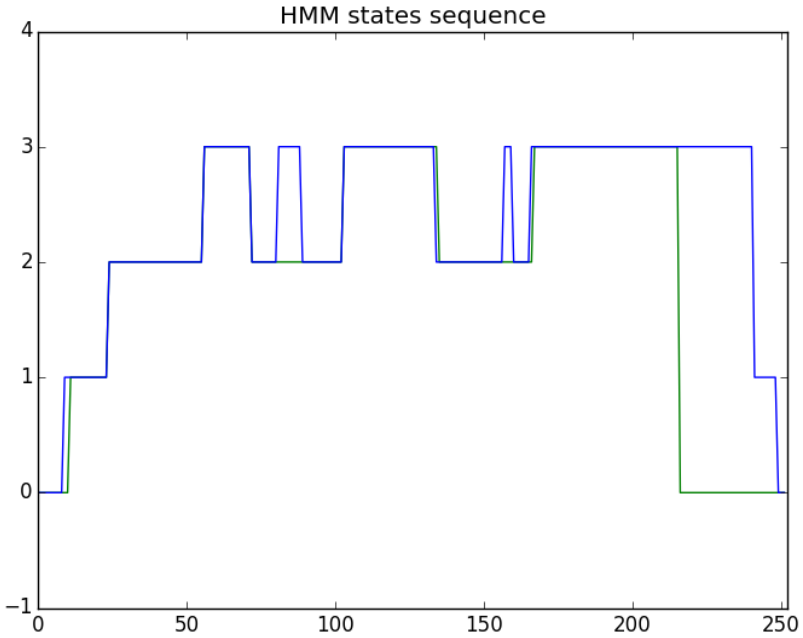


- Take into account the temporal ordering with HMM



Hidden Markov Model





Conclusion

- Results vary over songs
  - Size of the training set
  - Choice of the similarity threshold

- Similarity is sensitive especially for complex structures, many instruments

- Possible improvements and exploration
  - similarity measure
  - audio features