MUSIC STRUCTURAL SEGMENTATION BY COMBINING HARMONIC AND TIMBRAL INFORMATION (MIREX 2011)

Ruofeng Chen
Georgia Tech Center for Music Technology
Georgia Institute of Technology
ruofengchen@gatech.edu

Ming Li
Institute of Acoustics
Chinese Academy of Sciences
liming@mail.ioa.ac.cn

ABSTRACT

We propose a novel model for music structural segmentation aiming at combining harmonic and timbral information. We use two-level clustering with splitting initialization and random turbulence to produce segment labels using chroma and MFCC separately as feature. We construct a score matrix to combine segment labels from both aspects. Finally Non-negative Matrix Factorization and Maximum Likelihood are applied to extract the final segment labels. By comparing sparseness, our method is capable of automatically determining the number of segment types in a given song.

1. SEGMENTATION ALGORITHM

This music structure segmentation algorithm is based on the model described in [1]. (1) Chroma and MFCC features are extracted from audio, making use of the algorithms in MIRToolbox [1]. (2) A two-level clustering algorithm is designed to calculate window-based segment labels using either chroma or MFCC as feature. The two-level clustering algorithm involves random turbulence module so it outputs different segmentation results each time. Repeat the two-level clustering algorithm to get T segmentation results using chroma and T segmentation results using MFCC (we call them chroma solution and MFCC solution). (3) A score matrix representation is designed to count how many times two windows have identical segment labels in both chroma solution and MFCC solution. (4) Non-negative Matrix Factorization (NMF) is applied to the score matrix to approximate the score matrix with $W \times H$, rank = 3, 4, 5. (5) Sparseness is calculated over all columns of the three Hs, and the H with the highest average sparseness is picked out. (6) Maximum Likelihood is applied to all columns of the picked H to get final segmentation result (we call it final solution). (7) Post-process is attached to final solution to remove isolated short segments and differentiate "intro" and "outro". The flowgraph is shown on the next page. The two-level clustering algorithm is expanded in detail.

The only difference between this algorithm and the one described in [1] is the attached post-process module at the rear end. Please refer to [1] for detailed description of the algorithm.

[1] https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox

2. REFERENCES

[1] R. Chen, M. Li: "Music Structural Segmentation by Combining Harmonic and Timbral Information," *ISMIR*, Miami, Florida, 2011.

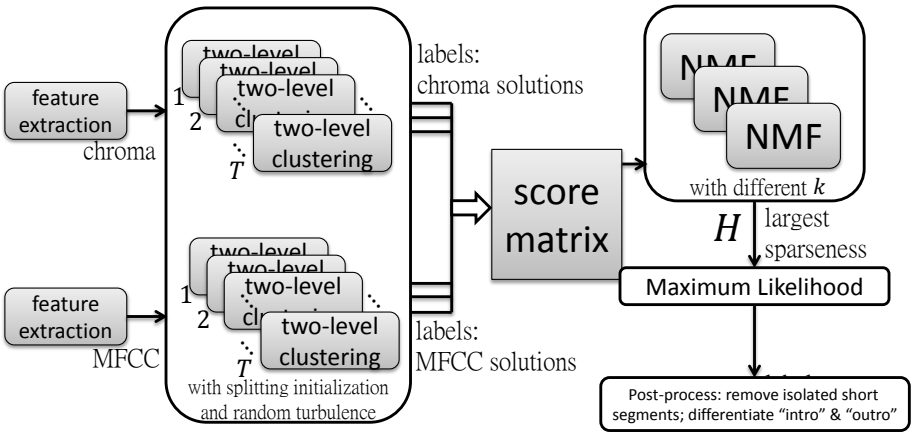

Figure 1. The complete flowgraph of the algorithm.

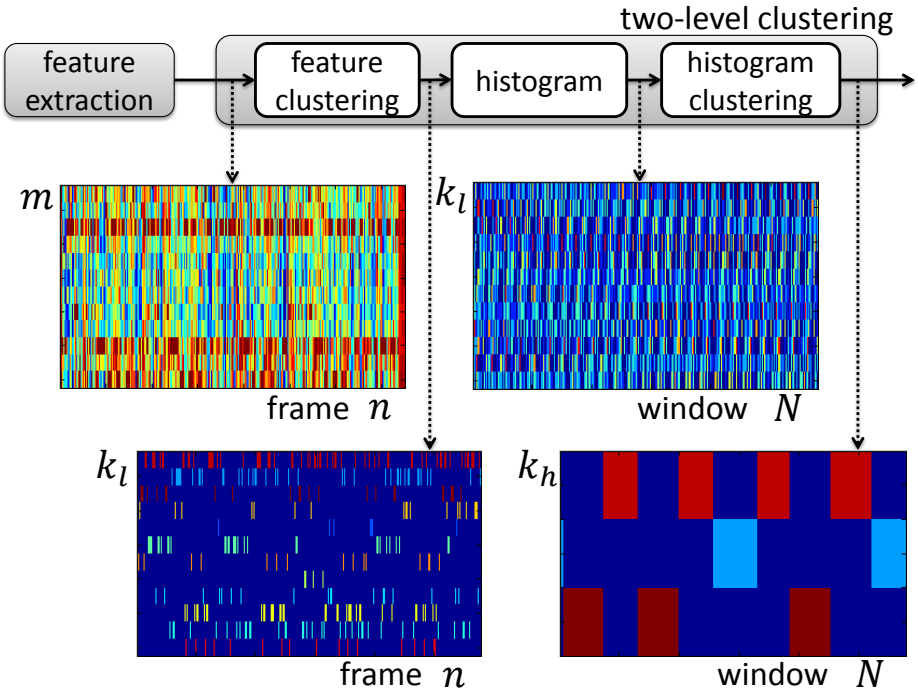

Figure 2. The flowgraph and illustration of intermediate results of two-level clustering. The lower two graphs' colors only illustrate different labels for better looking.