Automatic Extraction of Musical Structure using Pitch Class Distribution Features

Bee Suan Ong, Emilia Gómez & Sebastian Streich
(presented by **Òscar Celma**)

Music Technology Group, Universitat Pompeu Fabra
http://mtg.upf.edu

Music Technology Group

Outline

1. Introduction & System Architecture
2. Feature Extraction
3. Structural Analysis
4. Evaluation Strategy
5. Results
6. Conclusions

Music Technology Group

Outline

1. Introduction & System Architecture
2. Feature Extraction
3. Structural Analysis
4. Evaluation Strategy
5. Results
6. Conclusions

Introduction

- From the Invited Talk (S. Baumann):
  - Chat with G. Peeters (2006)

  "(ISMIR2006) is the chroma features' year, forget about MFCC"
- Aim of this talk:
  - Compare 3 different approaches to compute chroma vectors (12-dimensional: C, C#, D, ...) for harmonic information.
  - Use chroma features for structural analysis
  - Get the best result (out of the 3) and compare it with state of the art for structural anaylsis (Chai 2006).

Motivation

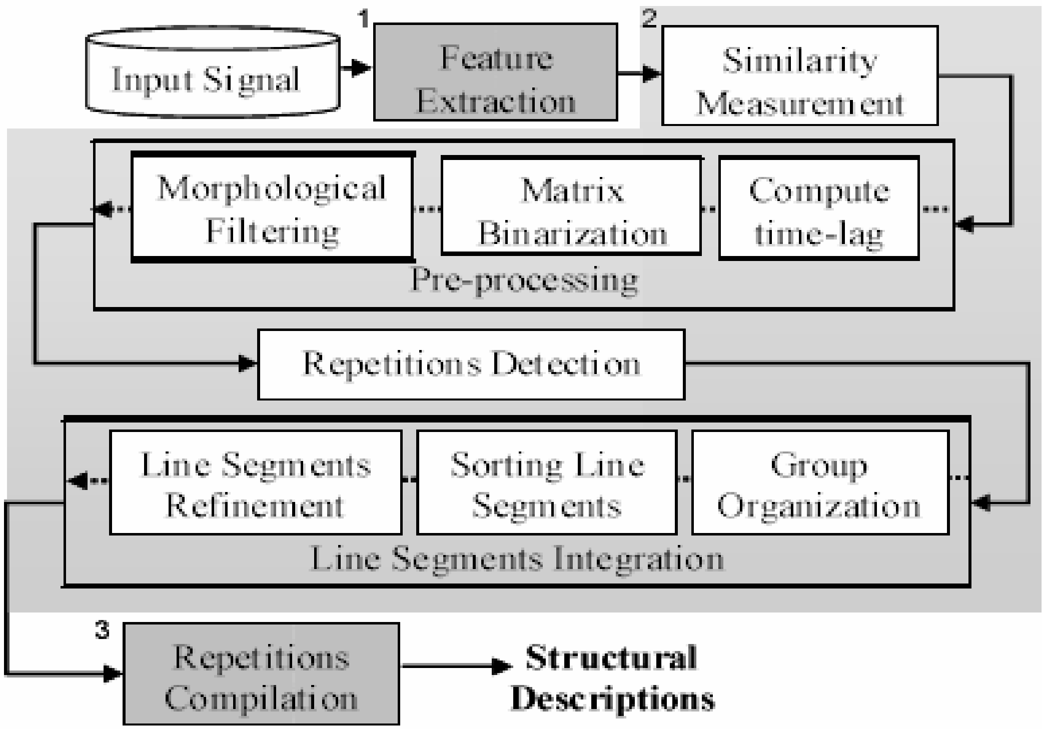
- Structure is a relevant aspect of music
- It is important to consider musical structure to *learn the semantics of audio signals*
- Melody & harmony are related to musical structure & are captured by chroma or pitch class distribution features
- These features have been shown to be preferred to timbre-related features (MFCCs) in this context *(Lu et alt. 04, Bartsch & Wakefield 05, Chai 06)*
- There are different approaches to obtain chroma features, no comparison until now between them

System architecture

Aim: analyze how the use of different approaches to extract low-level tonal descriptors affects the performance of the automatic music structural analysis

Inspired on Goto (03) with three main steps:

1. Feature extraction
2. Structural analysis
3. Repetitive segment compilation



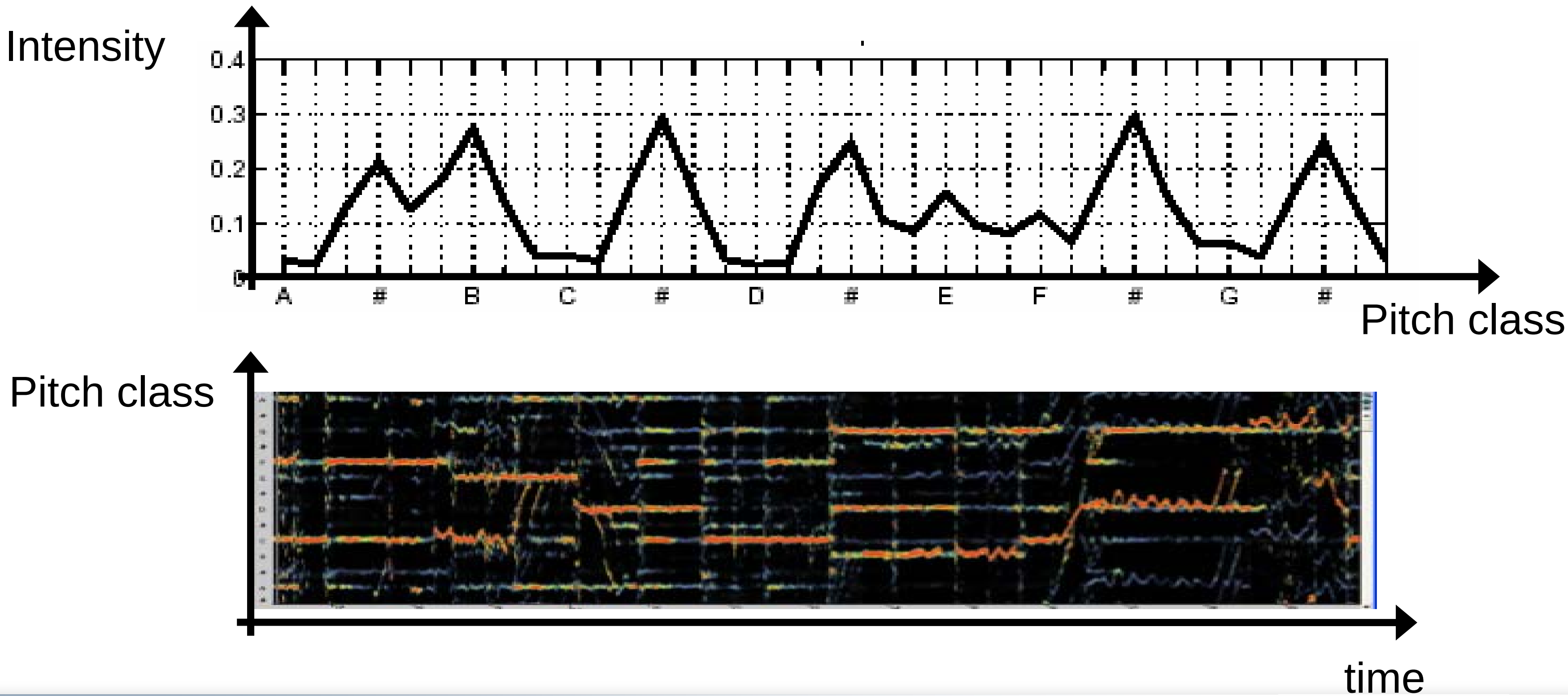
Music Technology Group

Outline

1. Introduction & System Architecture
2. Feature Extraction
3. Structural Analysis
4. Evaluation Strategy
5. Results
6. Conclusions

Pitch class distribution features

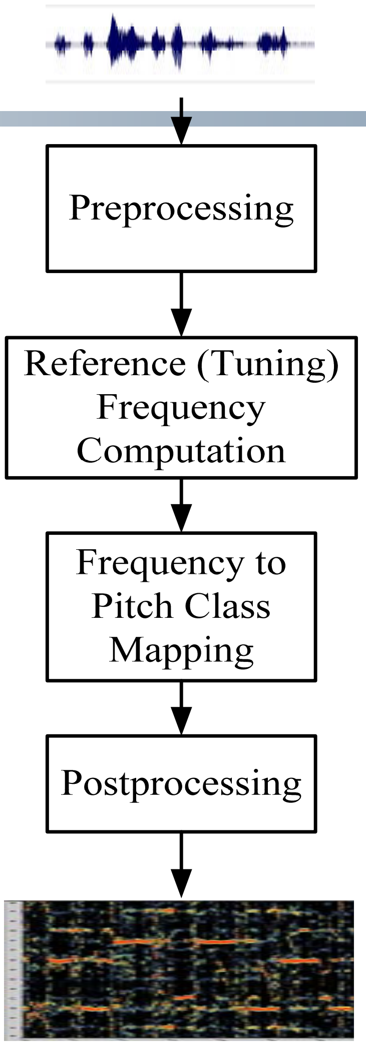
- Measure of the relative intensity of each of the 12 semitones of the equal-tempered scale



Music Technology Group

Tonal Feature Extraction

- Requirements:
  - Sensitivity to tonal similarity &
  - Independence with respect to timbre
- The audio signal is segmented into overlapped frames
  (size 100 ms,
  hopsize 11.6 ms)
- Compared methods:
  - Constant-Q profiles (CQP) (Brown 91)
  - Pitch class profiles (PCP) (Fujishima 1990)
  - Harmonic pitch class profiles (HPCP) (Gómez 2006)



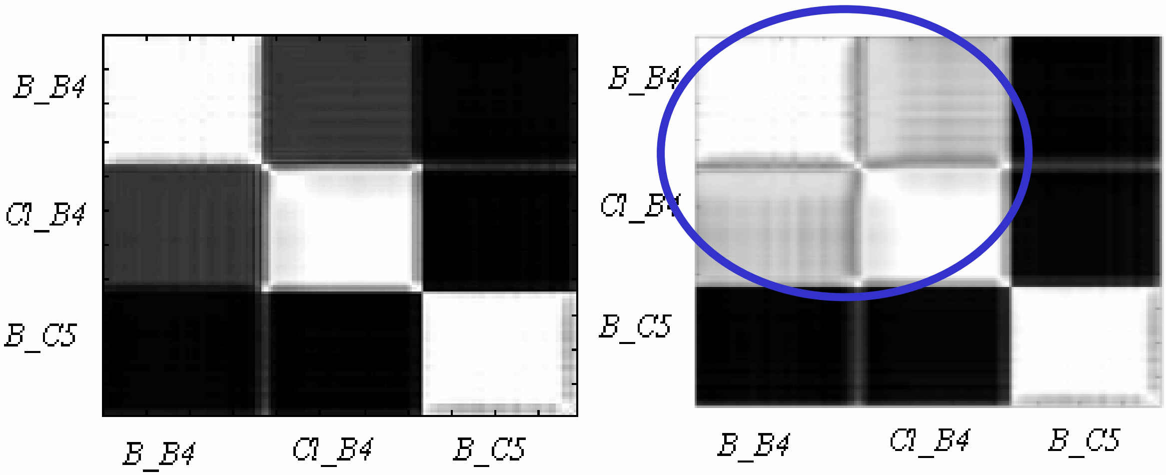
Overall block diagram

Music Technology Group

Approaches for Tonal Feature Extraction

	Preprocessing	Frequency to Pitch-Class Mapping	Postprocessing
CQP (Brown 91)	▪ Constant-Q Transform	▪ Weight: spectral magnitude	▪ Normalization by the maximum value
PCP (Fujishima 00)	▪ FFT	▪ Weight: squared value of spectral magnitude	
HPCP (Gómez 06)	▪ FFT ▪ Peak detection ▪ Tuning frequency estimation	▪ Weight: squared value of spectral magnitude (each frequency contributes to neighbor HPCP bins within a 4/3 semitone window) ▪ Consideration of harmonic frequencies	

Octave vs. Non-octave Mapping of Tonal Features



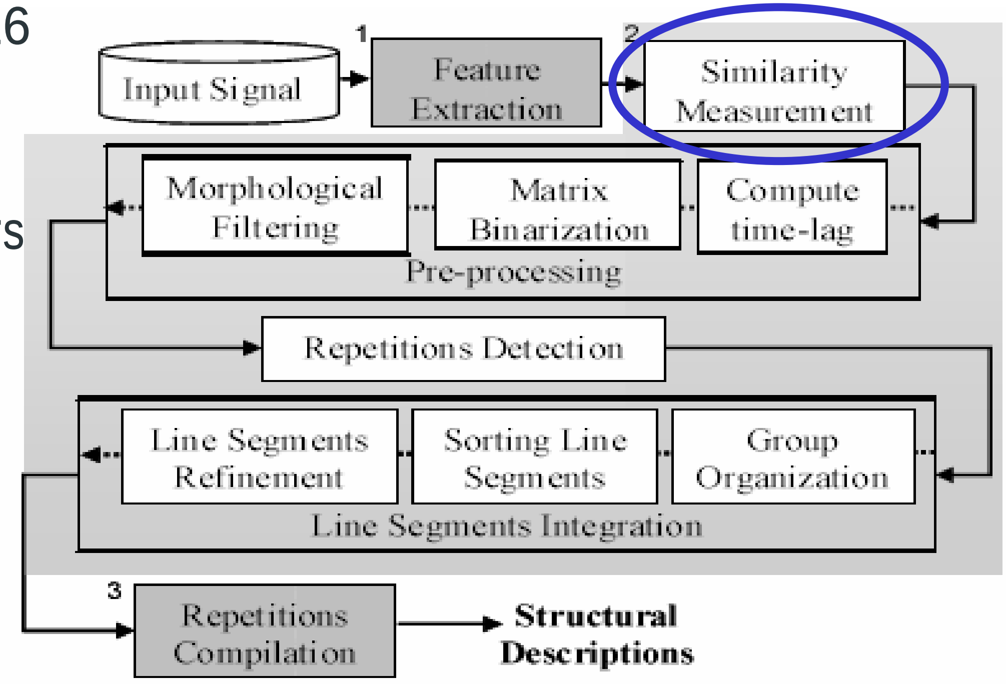
Self-similarity matrices of three notes: B4 played by a bassoon (B_B4), B4 by a clarinet (Cl_B4), and C5 by a bassoon (B_C5).
(left image) Constant-Q extracted directly from 5 octaves of musical notes
(right image) mapped into 1 octave

Observation: octave mapping of tonal features are more sensitive to tonal similarity instead of timbre similarity (different from Lu et alt. 04)

Outline

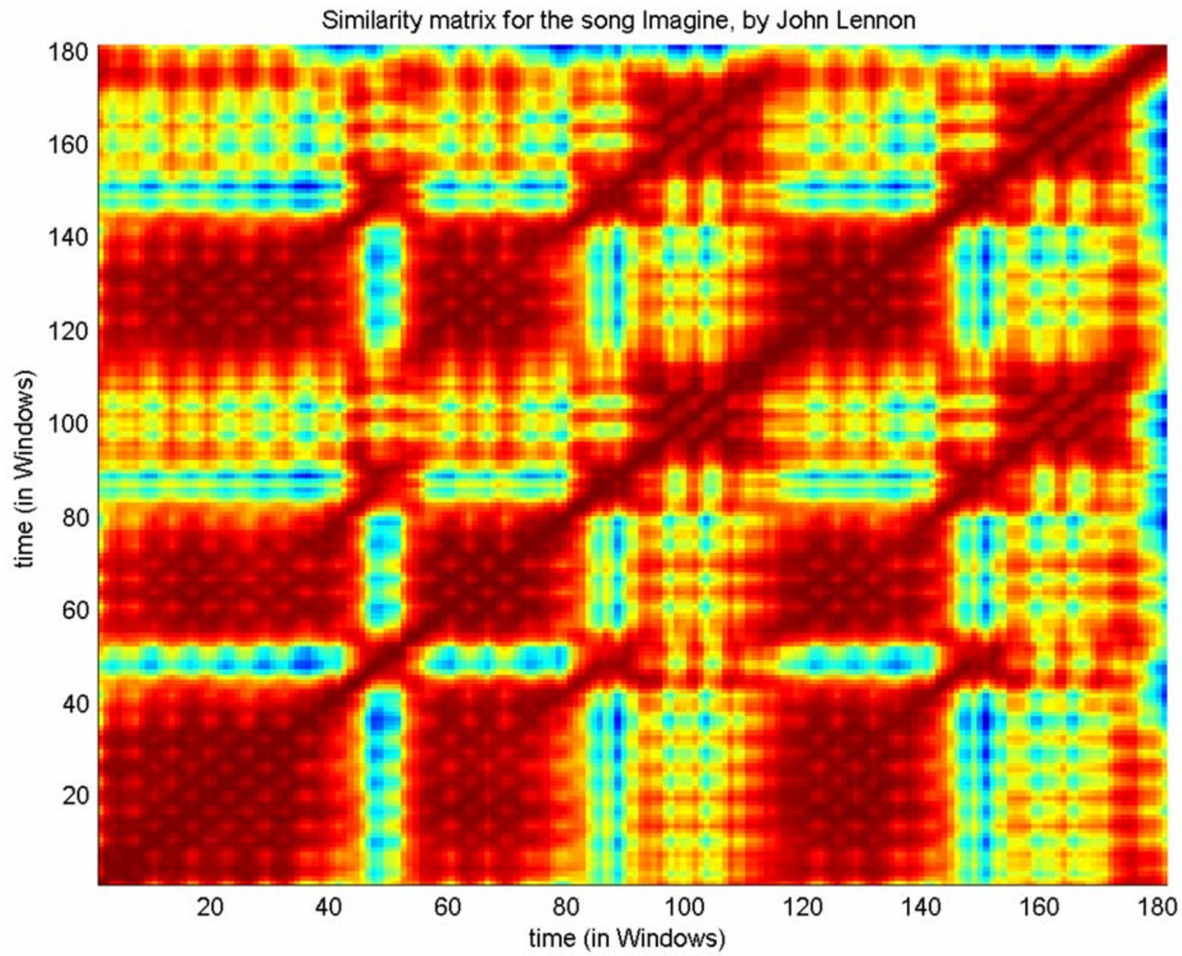
Structural Analysis: similarity

1. Average over 10 frames (116 ms)
2. Computation of similarity matrix using chroma vectors (Foote 00)



Music Technology Group

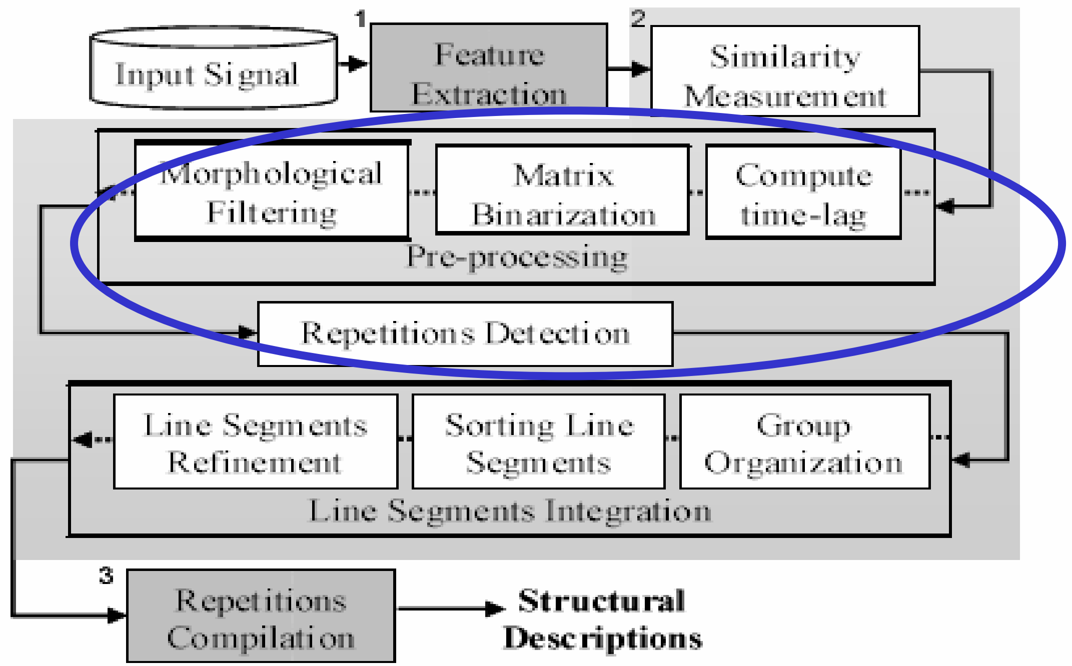
Structural Analysis: similarity



Music Technology Group

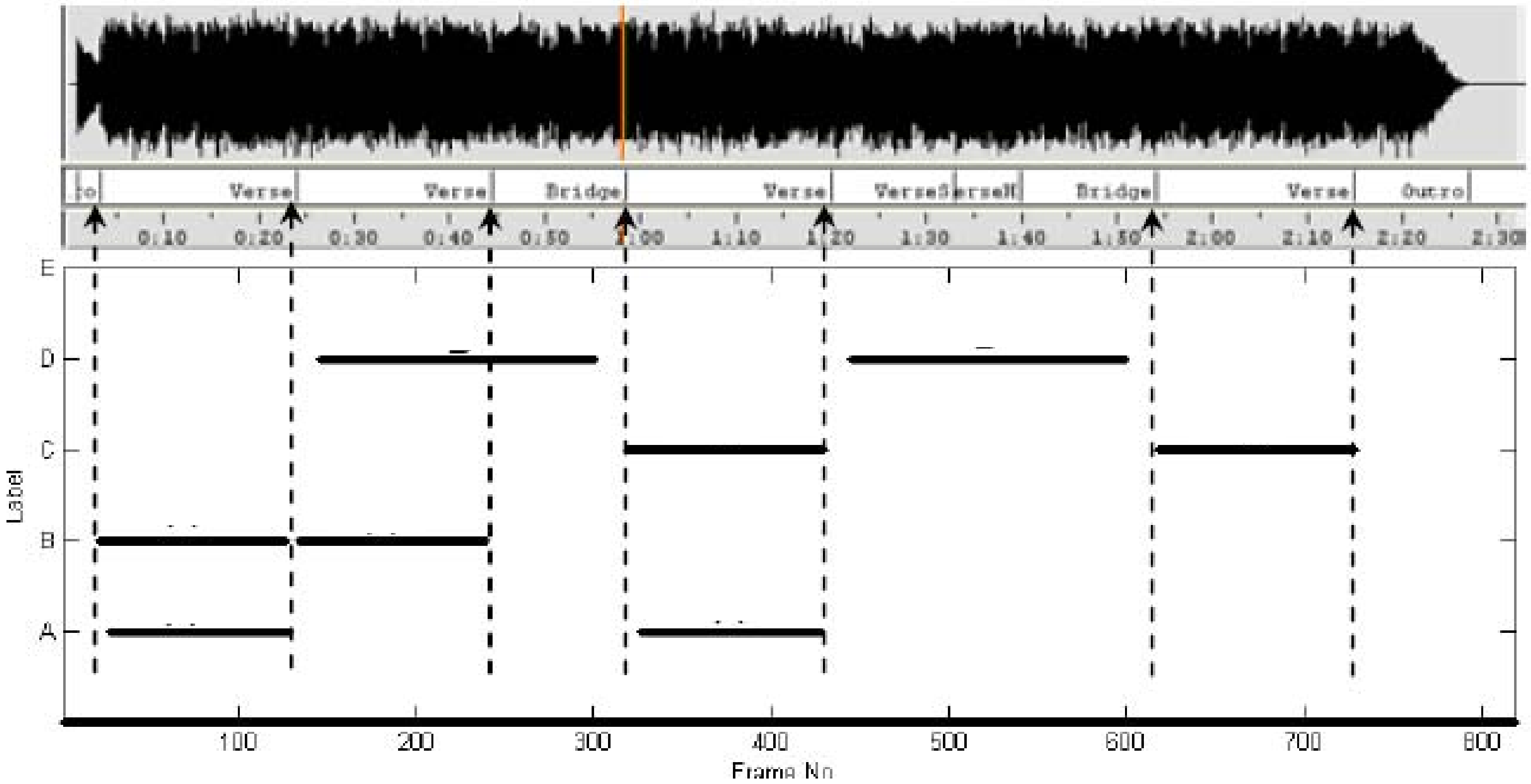
Structural Analysis: repetitions location

Location of repetitions (adaptation of (Goto 03)) only considering segments which are longer than 4 sec.



Music Technology Group

Structural Analysis: repetitions location

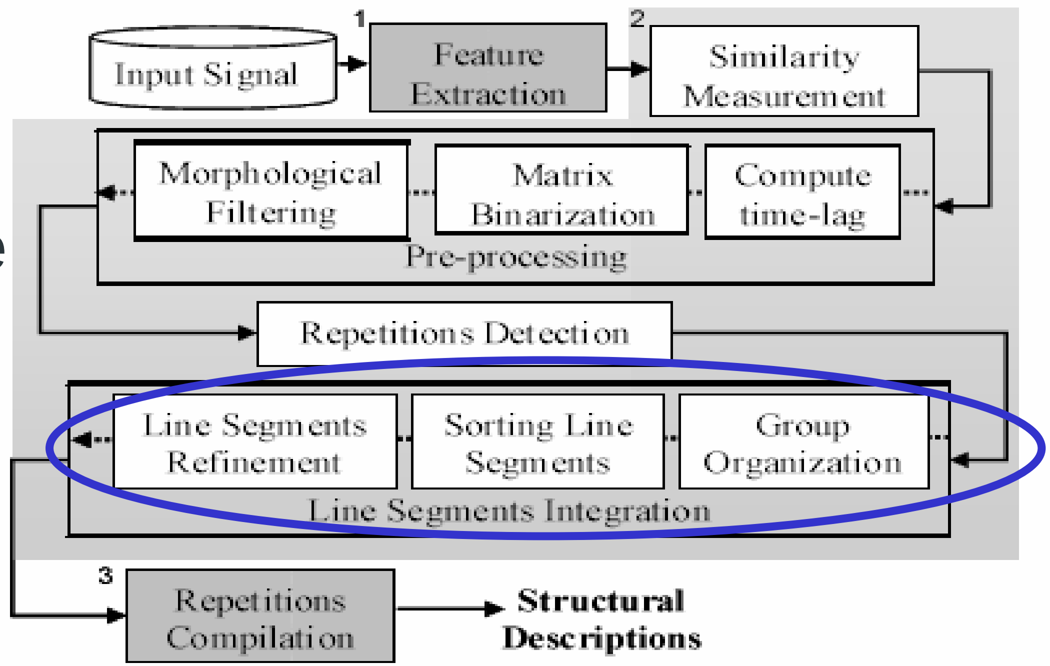


Detected repetitions correspond to the ground truth annotation of *A Hard Day's Night.*

Music Technology Group

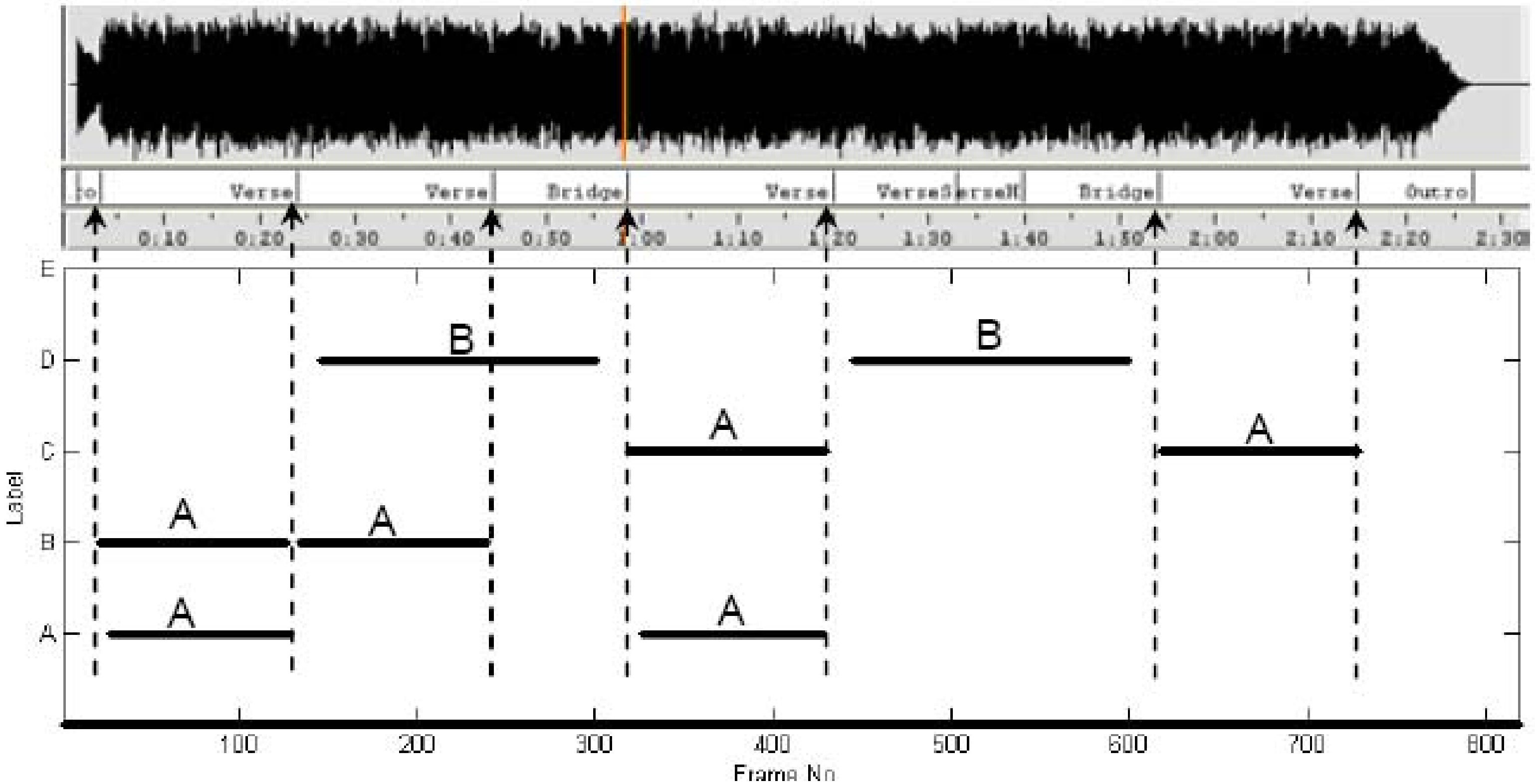
Structural Analysis: segment labeling

- Segment labeling by computing a distance measure between segments based on the extracted chroma features



Music Technology Group

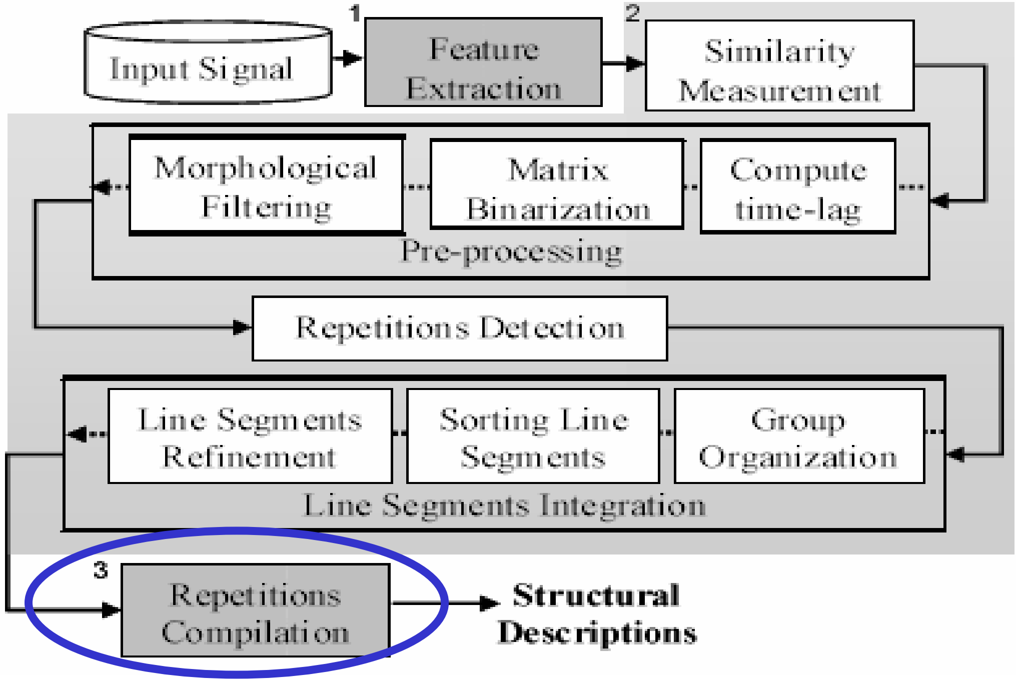
Structural Analysis: segment labeling



Detected repetitions correspond to the ground truth annotation of *A Hard Day's Night.*

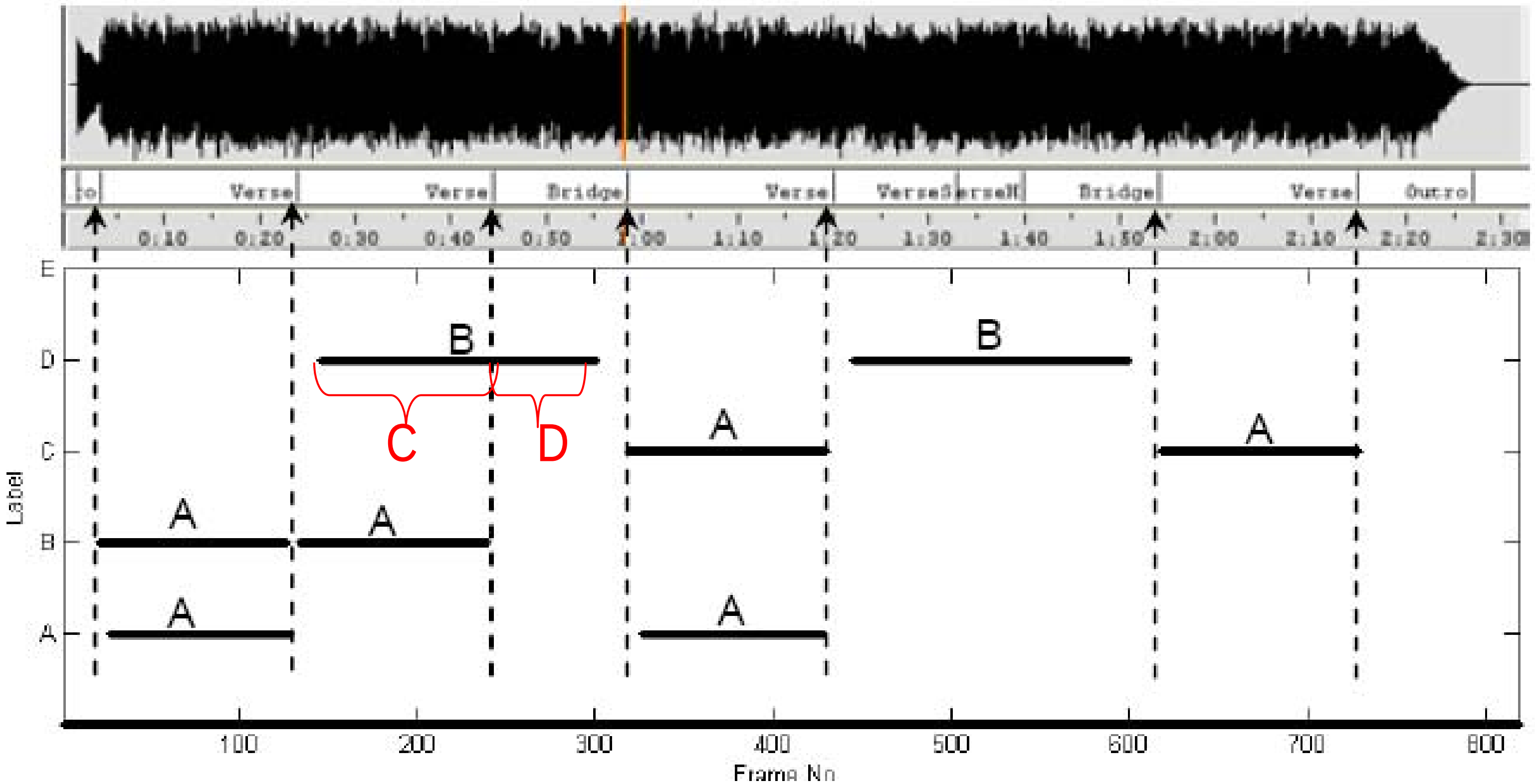
Structural Analysis: repetitions compilation

- If there is an overlap, (e.g. A & B):
  - Overlapped sections are given a new label (e.g. C)
  - Non-overlapped sections are given another new labels (e.g. D)
  - Unlabelled sections are given a new label (e.g. E, F,...)



Music Technology Group

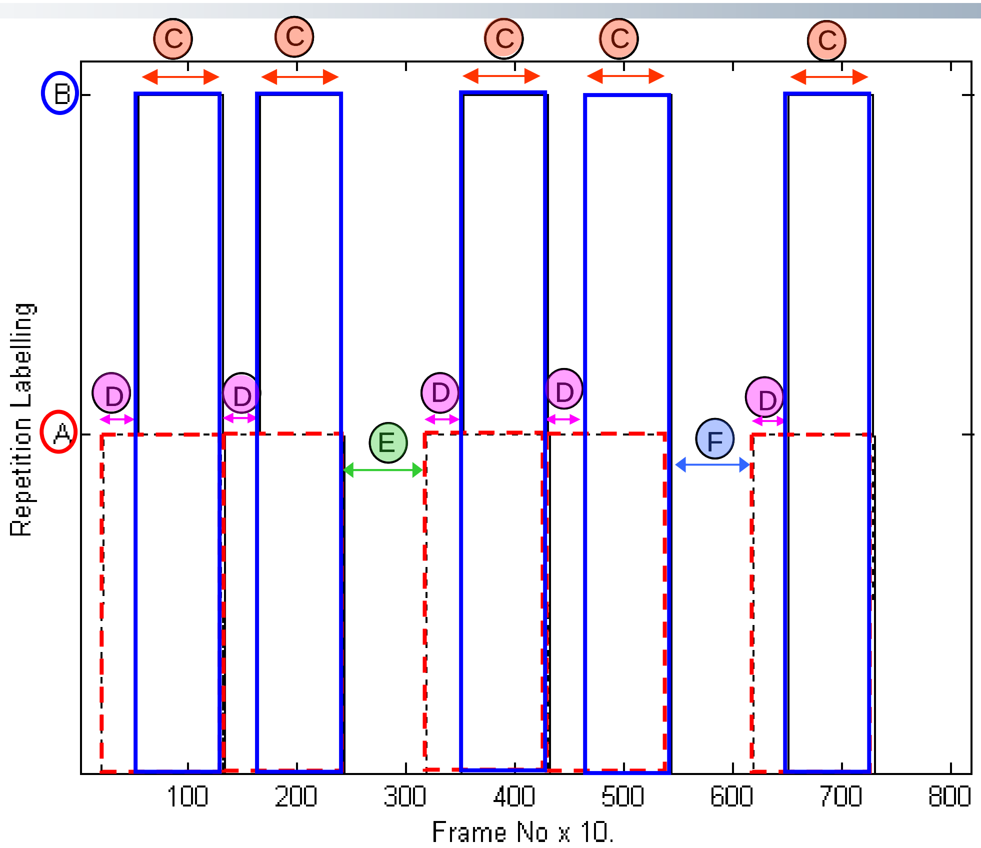
Structural Analysis: repetitions compilation



Detected repetitions correspond to the ground truth annotation of *A Hard Day's Night.*

Music Technology Group

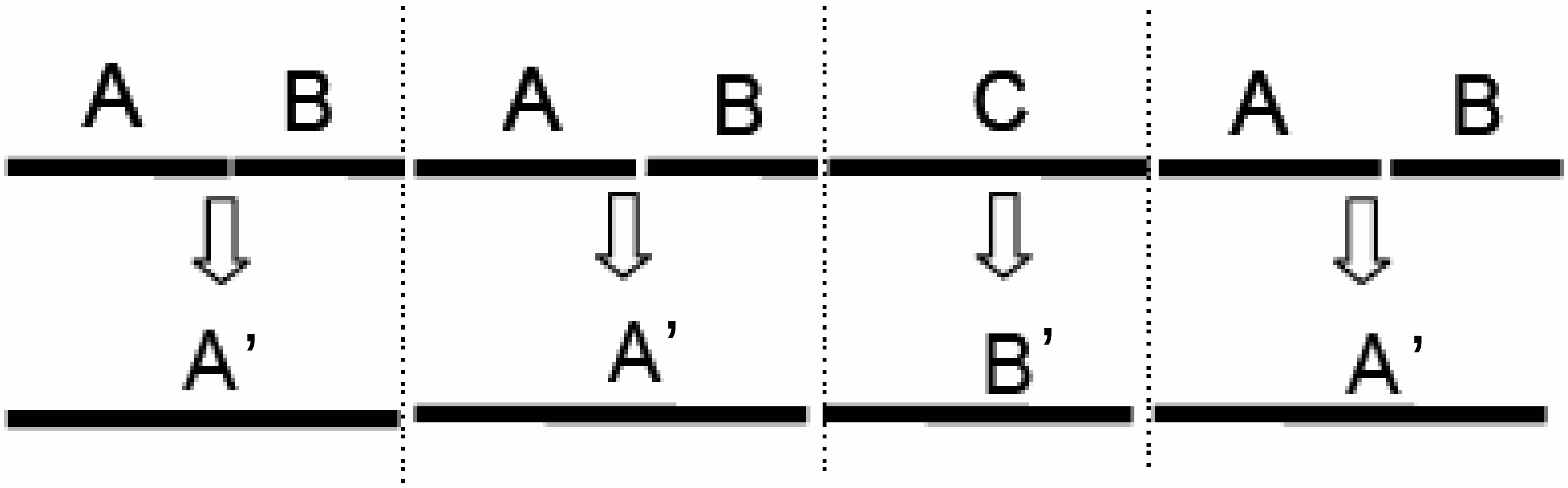
Structural Analysis: repetitions compilation



Repetitive segments compilation process with generated new labels.

Structural Analysis: integration

- Finally, we combine all the repeated series of labels, so that the total length is shorter than 25 s



Music Technology Group

Outline

1. Introduction & System Architecture
2. Feature Extraction
3. Structural Analysis
4. Evaluation Strategy
5. Results
6. Conclusions

Music Technology Group

Ground truth

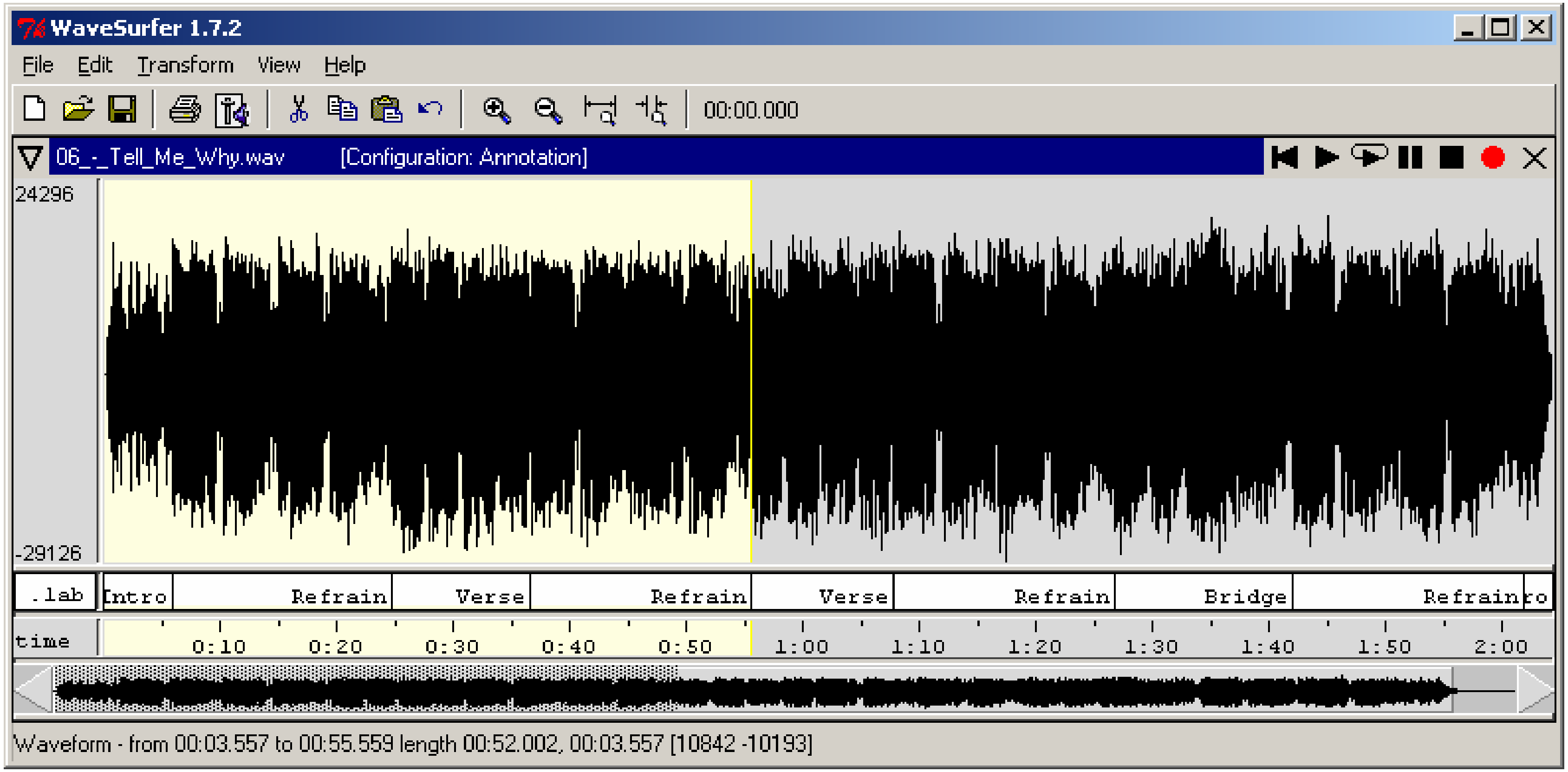
- Data format: 44.1 KHz, 16-bit mono
- Test data: 2 test sets

  (i) 56 songs from The Beatles 70s album → comparison of 3 approaches for computing chroma vectors

  (ii) 26 songs by The Beatles from the year 1962-1966, same collection used by (*Chai 06*) → best approach vs results by *(Chai 06)* using HMMs
- Sections are manually labelled (i.e. intro, versus, chorus, bridge, outro) according to Allan W. Pollack's "Note On" Series (*06*), supervised by a composer

Music Technology Group

Manual labels



Music Technology Group

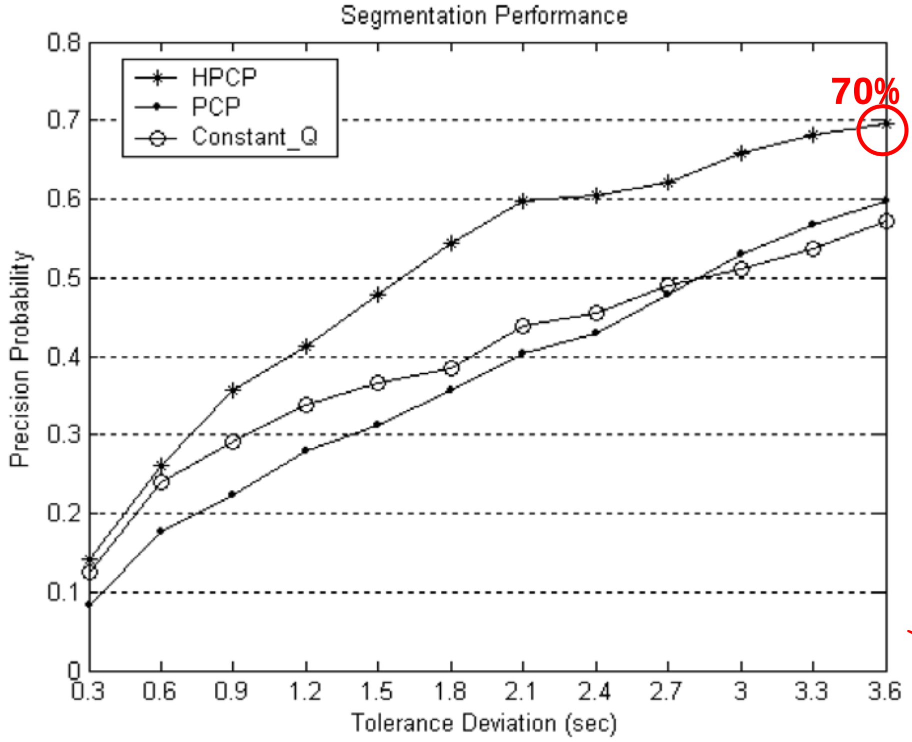
Evaluation measures

- Precision / recall measures
- Different degree of tolerance deviation of segment boundaries: between 0.3 to 3.6 sec.

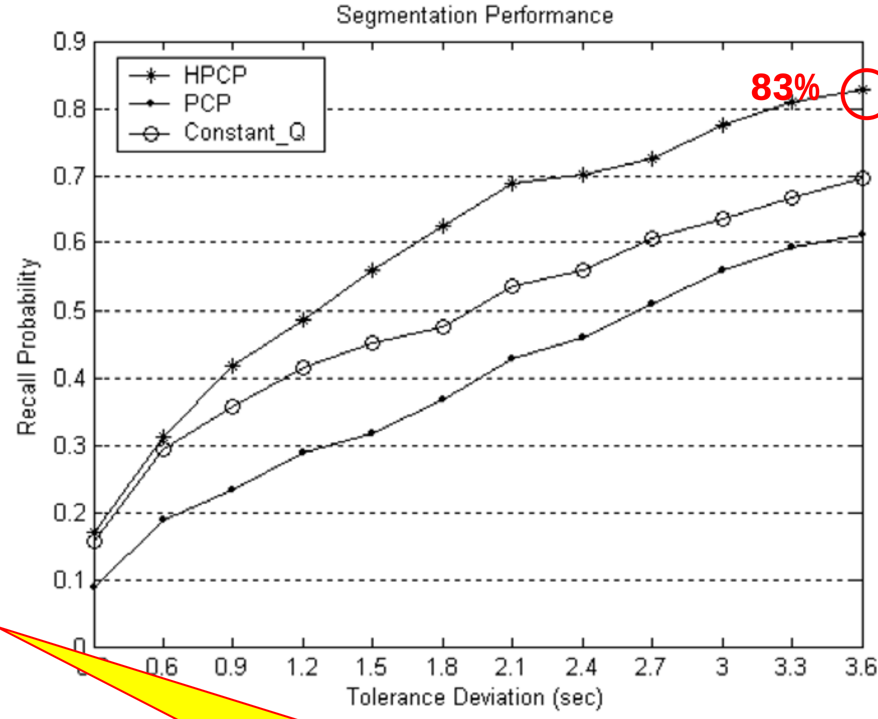
Outline

1. Introduction & System Architecture
2. Feature Extraction
3. Structural Analysis
4. Evaluation Strategy
5. Results
6. Conclusions

Results: Test Set I



Precision vs tolerance deviation (sec) for test set I



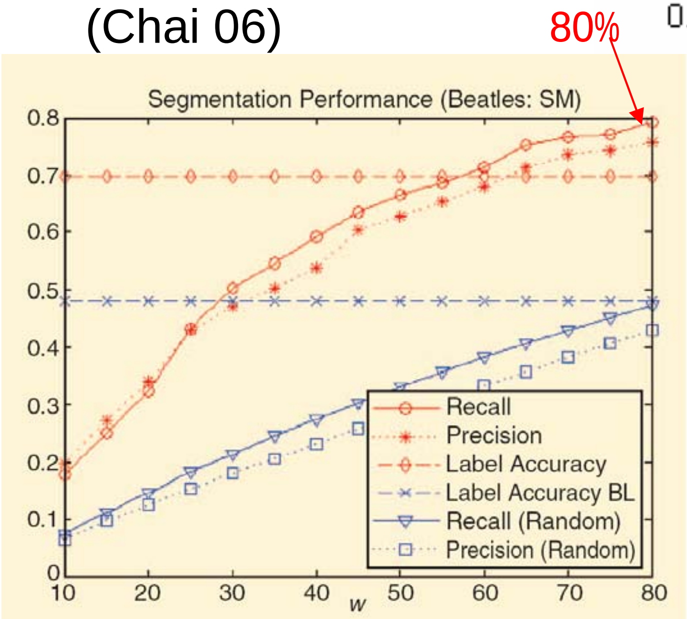
R... on

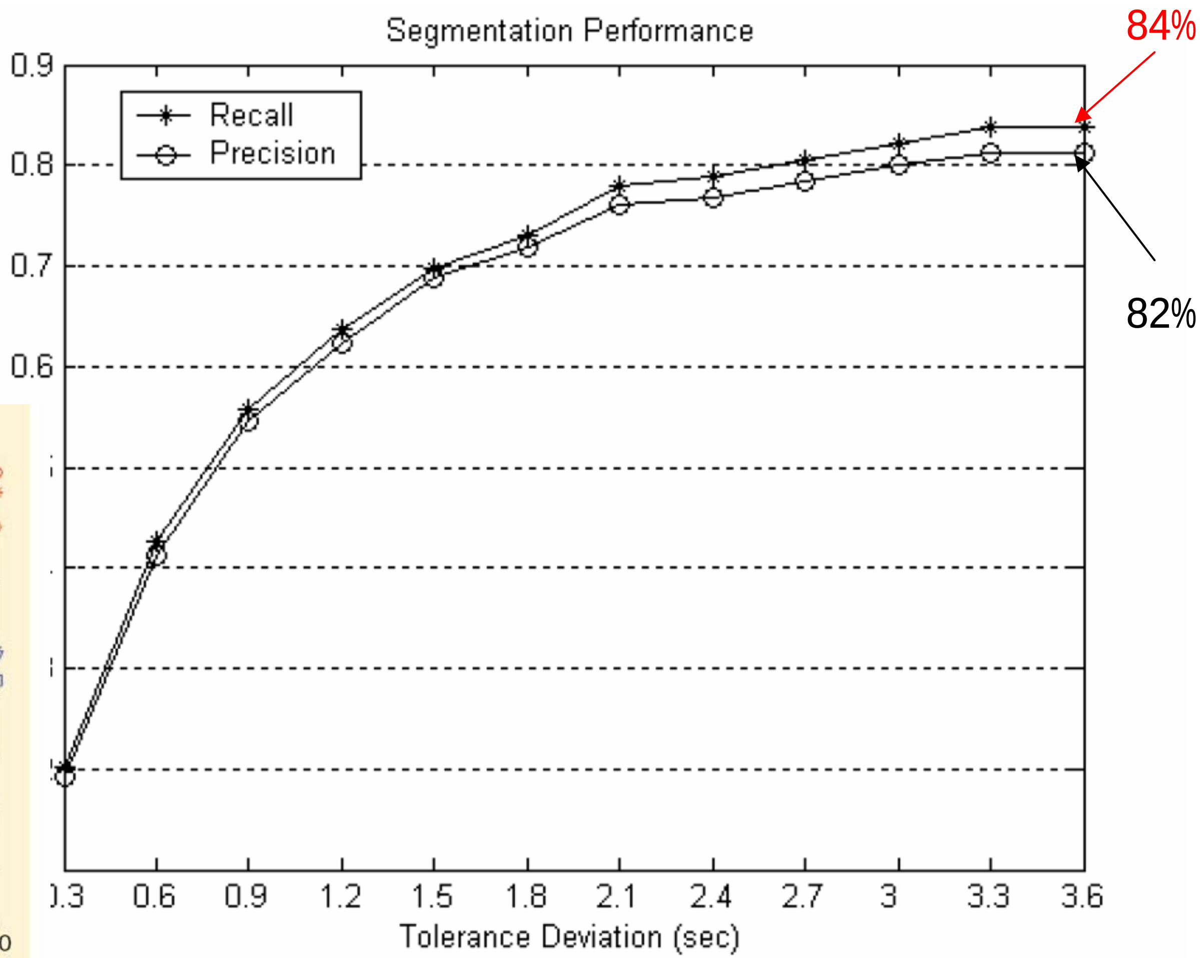
T-test analysis concludes that the differences between HPCP and the other compared tonal descriptors are statistically significant beyond the 99% confidence level with the *p*-values<0.01.

Results: Test Set II

Recall and precision rates of **HPCP** vs tolerance deviation (sec) for test set II

- Slightly better than reported in (Chai 06)





Music Technology Group

Outline

1. Introduction & System Architecture
2. Feature Extraction
3. Structural Analysis
4. Evaluation Strategy
5. Results
6. Conclusions

Music Technology Group

Conclusions

- We have presented an objective evaluation of the performance of structural analysis using different approaches for tonal feature extraction
- We conclude that the employed approach has an influence on the performance
- With our approach (HPCP), we obtain better results given the consideration of harmonic frequencies & spectral modeling (peak estimation).
- We have shown the validity of the approach by comparing to current state of the art *(Chai 06)*

Future Work

- Make use of higher-level analysis techniques (e.g. beat detection or phrase detection) for better segment truncation
- Improve generality, evaluate the performance using varied musical genres

Questions?

- egomez@iua.upf.edu
- beesuan@iua.upf.edu
- sstreich@iua.upf.edu