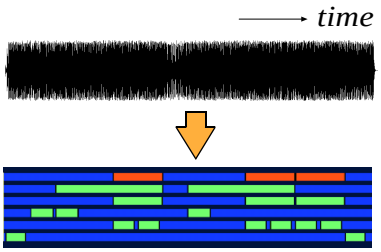
A Chorus-Section Detecting Method for Musical Audio Signals

Introduction



Chorus Section Detection

❑ Chorus (refrain) sections

- Most representative thematic sections in the music structure of popular songs
  Most repeated and memorable portions

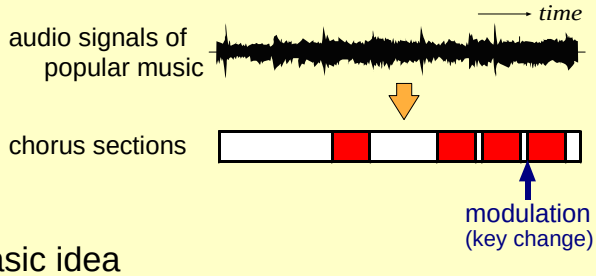
➡ Automatic detection is essential for building a computational model of music understanding

- Useful in various practical applications
  Audio thumbnail, music retrieval, etc.

RefraiD (Refrain Detecting Method)

❑ Exhaustively detect all the chorus sections appearing in a song

- Obtain a list of the start and end points of every chorus section in CD recordings
- Detect modulated chorus sections (with key change)



❑ Basic idea

- Detect without using any prior information about spectral characteristics of chorus sections
- Chorus sections are usually the most repeated sections in popular music

1. Find various groups of repeated sections

2. Output the chorus-like group appearing frequently

❑ Real-time output

- List of chorus sections
- Intermediate-result list of repeated sections
  Usually reflect the music structure of the song
  ex) repetition of "verse A ➡ verse B ➡ chorus"

Previous Work

❑ Previous chorus-detection methods
[Bartsch and Wakefield, 2001][Cooper and Foote, 2002][Logan and Chu, 2000]

- Only extract a single segment from several chorus sections
  Detect a repeated section of a given length as the most representative of a song

➡ Did not address the problem of detecting all the chorus sections in a song identifying both ends of the chorus sections

➡ Not able to deal with modulated repetition

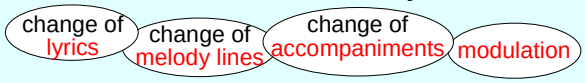
❑ Previous music-summarization methods
[Peeters, Burthe, and Rodet, 2002][Dannenberg and Hu, 2002]

➡ Did not identify both ends of chorus sections

➡ Not able to deal with modulated repetition

Problems

❑ It is rare for repeated sections to be exactly the same

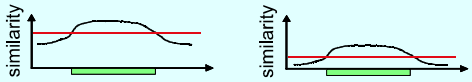


❑ Problem 1: acoustical features and similarity

- Typical power spectrum and MFCC features are liable to change considerably

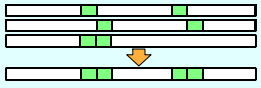
❑ Problem 2: repetition-judgment criterion

- Appropriate criterion of the similarity for judging the repetition depends on the song



❑ Problem 3: integrating repeated sections

- Identify both ends by examining mutual relationships among various repeated sections



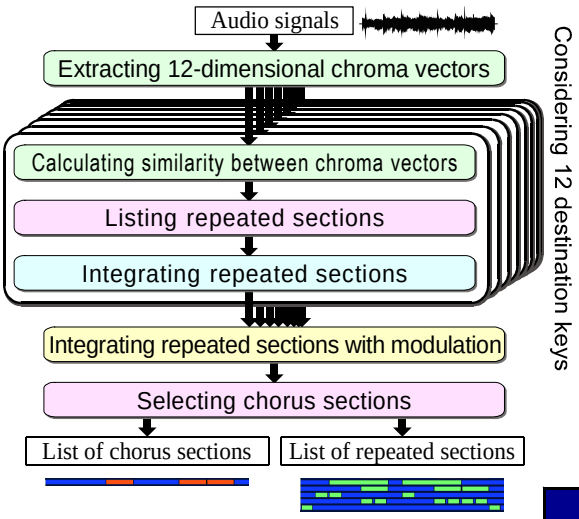
❑ Problem 4: detecting modulated repetition

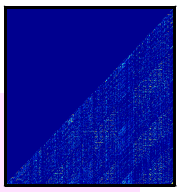
- Acoustic features generally undergo a significant change after modulation (key change)

Masataka Goto [1,2]

[1] "Information and Human Activity," PRESTO, JST
[2] National Institute of Advanced Industrial Sci. and Tech. (AIST)

Overview of RefraiD



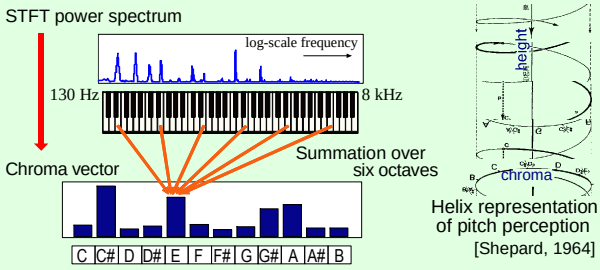


Solution to Problem 1

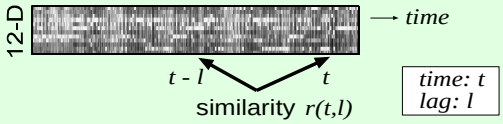
❑ Extracting 12-dimensional chroma vectors

- Sum of power at frequencies of each pitch class

  Capture the overall harmony (pitch-class dist.)



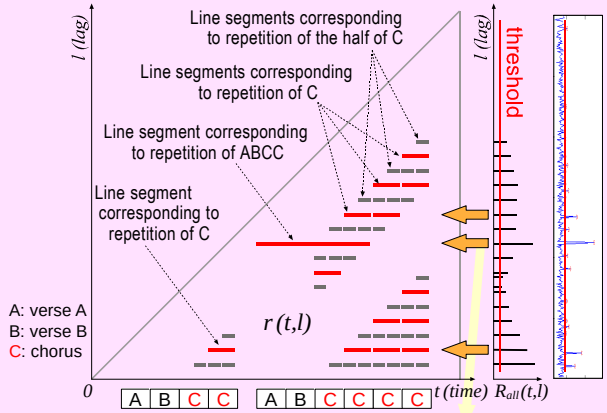
❑ Calculating the similarity between vectors

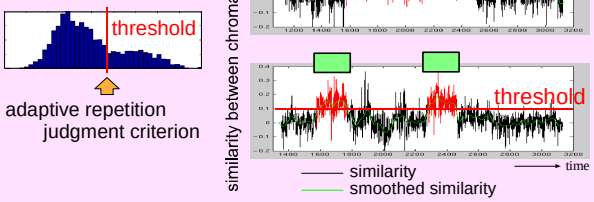


Solution to Problem 2

❑ Listing repeated sections

- Find line segments in the similarity $r(t,l)$





- Automatic threshold selection method

  based on a discriminant criterion [Otsu, 1979]

  Dichotomize peak heights into two classes

  Discriminant criterion measures:

  maximizing class separability (between-class variance)

$$\sum_i \omega_i (\mu_i - \mu_T)$$

ω_i : probability of class occurrence

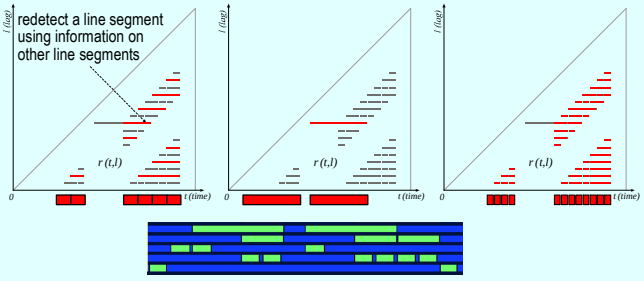
μ_i : class mean of heights

μ_T : total mean of heights

Solution to Problem 3

❑ Integrating repeated sections

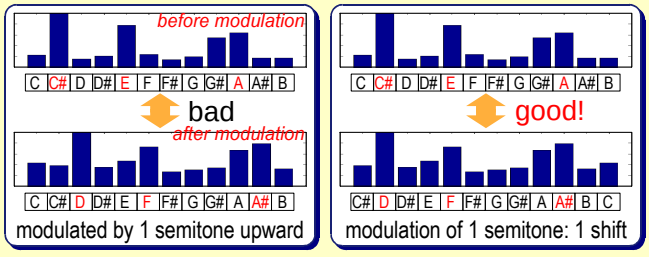
- Group line segments having common sections
  - Revise both ends of sections in each group
  - Redetect missing line segments



Solution to Problem 4

❑ Integrating repeated sections w/ modulation

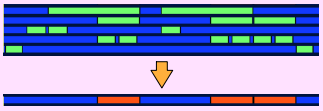
- Shift 12 elements of 12-D chroma vectors



URL of the Masataka Goto's Home Page:
http://staff.aist.go.jp/m.goto/

Selecting Chorus Sections

❑ Evaluate chorus possibility for each group



- # of sections is many
- reliability of section is high
- length of section is long

$$\sum (\text{reliability of section}) \log \frac{(\text{length of section})}{(\text{constant})}$$

• Three assumptions

1. Chorus has an appropriate length (7.7- 40 sec)
2. When there is a long repeated section, the end of it is likely to be the chorus section
3. When a section has half-length repeated sub-sections, it is likely to be the chorus section

Experimental Results

❑ Conditions

• Tested on 100 songs
from *"RWC Music Database: Popular Music"*
RWC-MDB-P-2001 No. 1 - 100 [Goto et al., 2002]

• Correct chorus sections were labeled manually
Develop a music-structure labeling editor

• F-measure: harmonic mean of recall & precision rates
Judged to be correct if F-measure > 0.75

$$\text{F-measure} = \frac{2RP}{R+P} \qquad \text{Recall (R)} = \frac{\text{total length of correctly detected chorus sections}}{\text{total length of correct chorus sections}}$$

$$\text{Precision (P)} = \frac{\text{total length of correctly detected chorus sections}}{\text{total length of detected chorus sections}}$$

❑ Results

• 80 songs out of 100 were correct (F-measure mean: 0.938)

Modulation detection	○	×	○	×
Use of assumptions 2 & 3	○	○	×	×
# of songs (out of 100)	80	74	72	68

Summary

❑ Propose the RefraiD method

• Regard the most repeated sections
as the chorus sections in popular music

• Detect all the chorus sections
with their start and end points

• Detect modulated chorus sections,
which previous methods could not detect

• Robust enough to deal with
real-world audio signals in real time