Music Scene Description Project

Toward Audio-based Real-time Music Understanding

Overview

Introduction

❑ Our goal

- Build a real-time system that can understand real-world music signals (CD recordings) in a human-like fashion
  - hum the melody
  - notice a phrase being repeated
  - find chorus sections

- Useful in various applications

  MIR, music production/editing, entertainment, etc.

➡ Brain mechanisms have not been understood
Difficult to implement these abilities

Music Scene Description

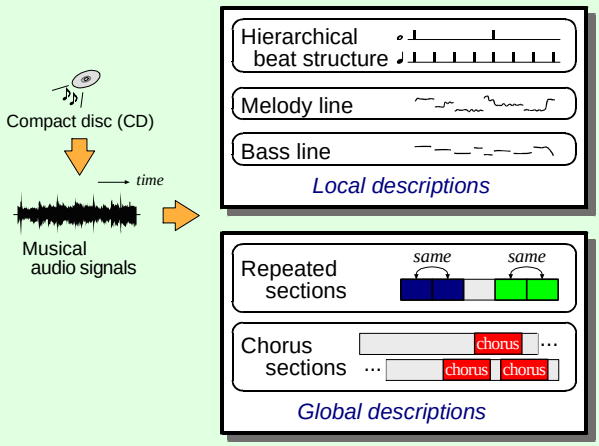
❑ Real-time music-scene-description system

- Obtain descriptions intuitively meaningful to untrained listeners

  from real-world audio signals containing simultaneous sounds of various instruments (w/ or w/o drum-sounds)

- Five descriptions

  Consider what is to be achieved to understand music



❑ Real-time methods

- **Audio-based real-time beat-tracking method**
  [Goto and Muraoka, 1999][Goto, 2001a]
- **Predominant-F0 estimation method for detecting melody and bass lines** *(PreFEst)*
  [Goto, 2001b][Goto, 2003b]
- **Chorus-section detection method** *(RefraiD)*
  [Goto, 2003a][Goto, 2003c]

Previous Work

❑ Two popular approaches

- Sound source separation
- Automatic transcription

➡ Neither separation nor transcription is necessary or sufficient for understanding music

- Human auditory system does not extract each individual audio signal

  Even if a mixture cannot be separated, that the mixture includes certain components can be understood

- Untrained listeners understand music without mentally representing audio signals as scores

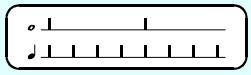
  Even if we could derive separated signals and musical notes, it is still difficult to obtain high-level descriptions like melody and chorus

Beat Structure

❑ Audio-based real-time beat-tracking method

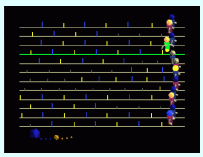
- Recognize a hierarchical beat structure

  Quarter-note and measure levels

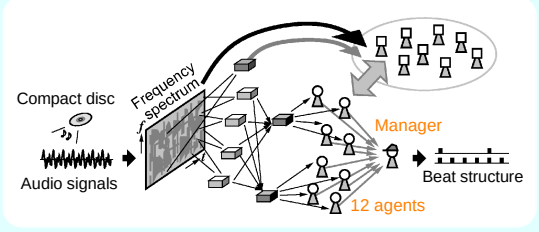


- Advantage

  Track beats above the quarter-note level by using three kinds of musical knowledge
  - onset times
  - chord changes
  - drum patterns



- Multiple agent architecture
- Overview



❑ Beat-tracking application

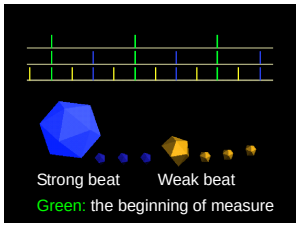
- Beat-driven real-time computer graphics

  Various movements and lighting properties can be changed with musical beats

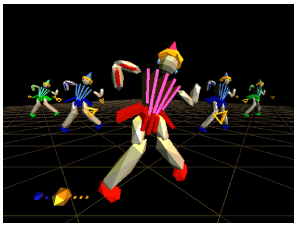
Masataka Goto [1,2]

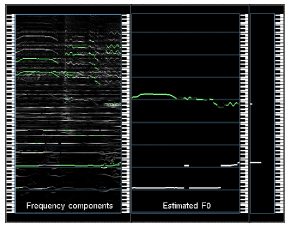
[1] "Information and Human Activity," PRESTO, JST
[2] National Institute of Advanced Industrial Sci. and Tech. (AIST)

Methods

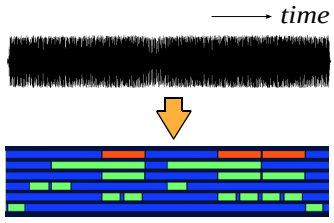


Beat Structure


Beat-driven Dancers



Melody and Bass Lines

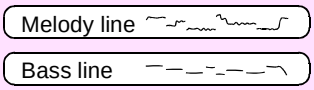


Chorus/Repeated Sections

Melody and Bass Lines

❑ Predominant-F0 estimation method for detecting melody and bass lines *(PreFEst)*

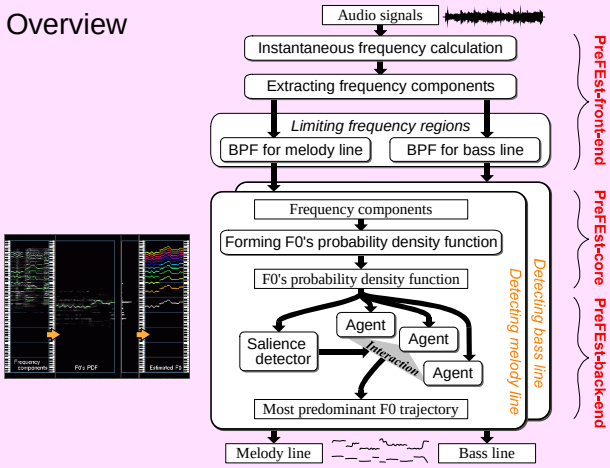
- Estimate the fundamental frequency (F0) of melody and bass lines



- Advantage

  Deal with monaural complex music signals
  - Not assume the number of sound sources
  - Not locally trace frequency components
  - Not rely on F0's frequency component

- Overview

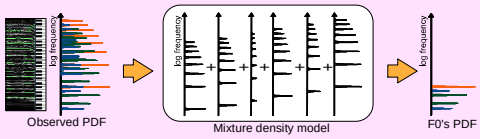


- MAP estimation using the EM algorithm

  (maximum *a posteriori* probability) (expectation-maximization)

  Introduce original mixture density model

  contain every possible harmonic structure with different weights (model parameter)



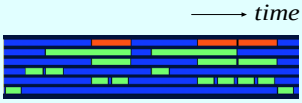
Estimate the F0's PDF (probability density function)

(relative dominance of every possible F0)

Chorus/Repeated Sections

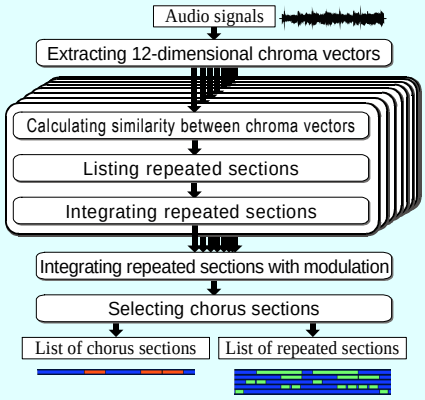
❑ Chorus-section detection method *(RefraiD)*

- Detect all the chorus sections in a song and several repeated sections



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

- Overview

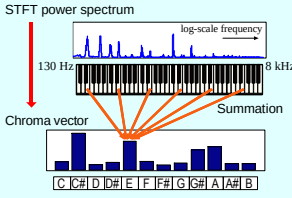


- Regard the most repeated sections as the chorus sections in popular music

  Detect without using prior information about spectral characteristics of chorus sections

- Extract 12-dimensional chroma vectors

  (Sum of power at frequencies of each pitch class)



Detect modulated repetition by shifting chroma vectors

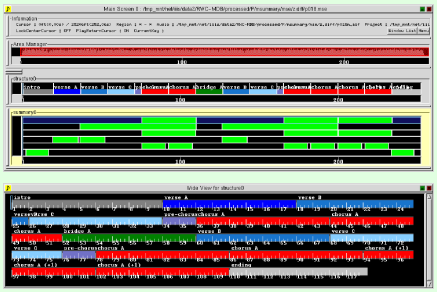
URL of the Masataka Goto's Home Page:
http://staff.aist.go.jp/m.goto/

Metadata

Metadata Editor

❑ Multipurpose music-scene labeling editor

- Hand-label musical pieces with metadata (correct descriptions)
- Evaluate music-scene-description methods



❑ Functions

- Deal with both audio files and SMFs
- Support interactive audio/MIDI playback
- Show subwindows in which any selected descriptions can be displayed and edited
- Support practical editing aids
  - magnifying-glass function
  - region-based cut-and-paste operation
  - cursor movement between context-dependent grid points (e.g., beats)

❑ RWC Music Database: Popular Music

- Hand-labeled chorus sections of all 100 songs (RWC-MDB-P-2001 No.1 - 100)
- Evaluated the RefraiD
  Compare the output with correct sections
  Correctly detected in 80 of the 100 songs

Conclusion

❑ Music Scene Description Project

- Build a music-scene-description system
  Understand real-world audio signals w/o deriving musical scores or separating signals
- Develop a metadata editor
  Enable a user to hand-label audio files/SMFs with descriptions of the music in those files
- Hand-labeled 100 songs of RWC-MDB-P-2001 with their chorus sections