

Automatic Audio Segmentation: Segment Boundary and Structure Detection in Popular Music

Ewald Peiszer Thomas Lidy Andreas Rauber



Institute of Software Technology & Interactive Systems

Workshop on Learning Semantics of Audio Signals, 2008

- 1 Introduction
- 2 Algorithm
- 3 Evaluation
 - Evaluation Setup
 - Results
- 4 Discussion

Automatic Audio Segmentation



Tasks

- Segment boundaries
- Musical form / structure (ABCDBCDBDA)
- Chorus detection (CD=chorus)
- Audio thumbnailing / summarization (ABCD)
- Semantic labelling
(Intro - verse - prechorus - chorus - verse - prechorus - chorus - verse - chorus/bridge - outro)

Motivation

- Browsing of music collections
- New features for playback devices
- Aid subsequent processing steps

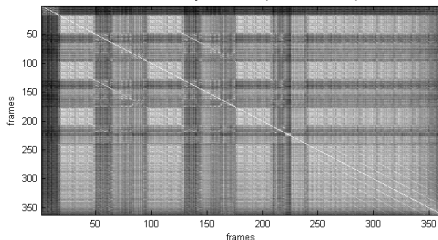
Contributions

- Algorithm for boundary and structure detection
- Evaluation using 109 song corpus
- Flexible XML ground truth file format

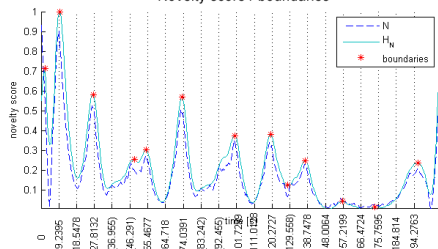
Boundary Detection

- 22,050 Hz audio, beat detection, beat synchronized frames
- Feature extraction
- Self similarity matrix
- Novelty score [Foote]
- Low pass filter
- Local maxima \rightarrow segment boundaries

Similarity matrix S (ds: Euclidean)

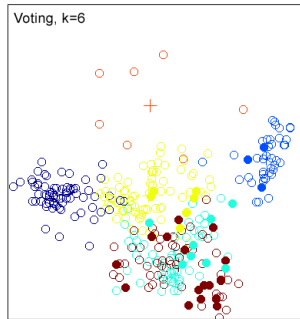
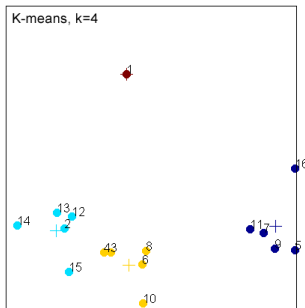


Novelty score / boundaries



Structure Detection

- K-means
- Agglomerative hierarchical clustering
- “Voting”
- Dynamic Time Warping
- Cluster validity index (Dunn, Davies-Bouldin)
- Minimal user input: number of desired segment types



Ground Truth

Main problem

Ambiguity!

- XML ground truth file **SegmXML**
- Alternative names
- Subsegments (two level hierarchical segmentation)
- Semantics → ground truth variants

Britney Spears_-_Hit_Me_Baby_One_More_Time

with subsegments, variant 0



without subsegments



Corpus

- $94 + 15 = 109$ songs
- Genres: rock, pop, dance, R&B, rap
- 60 from [LS07]^a, 47 from [PK06]^b, 14 as *qmul14*, 10 from RWC-Pop
- Realistic but music not free to get and use ☹

^aM. Levy and M. Sandler. Structural segmentation of musical audio by constrained clustering. *IEEE Transactions on Audio, Speech and Language Processing*, 16(1)318–326, 2007.

^bJ. Paulus and A. Klapuri. Music structure analysis by finding repeated parts. In *Proc AMCMM*, pages 59–68, Santa Barbara, California, USA, 2006. ACM Press New York.

A-HA, ABBA, ABBA, Alanis Morissette, Artful Dodger feat. Craig David, Beastie Boys, Beatles, Björk, Black Eyed Peas, Britney Spears, Chicago, Chumbawamba, Coolio, Cranberries, Creedence Clearwater Revival -, Depeche Mode, Desmond Dekkert, Deus, Dire Straits, Eminem ft. Dido, Faith No More, Gloria Gayner, KC and the Sunshine Band t, KoRn, Lucy Pearl, Madonna, Marilyn Manson, Michael Jackson Nick Drake, Nirvana, Nora Jones, Oasis, Pet Shop Boys, Portishead, Prince, Queen Yahna, R.E.M., R Kelly, Radiohead, Red Hot Chili Peppers, Salt N Pepa, Saxon, Scooter, Seal, Shania Twain, Simply Red, Sinhead O Connor, Spice Girls, Suede, ...

Performance Measures

Boundary Detection

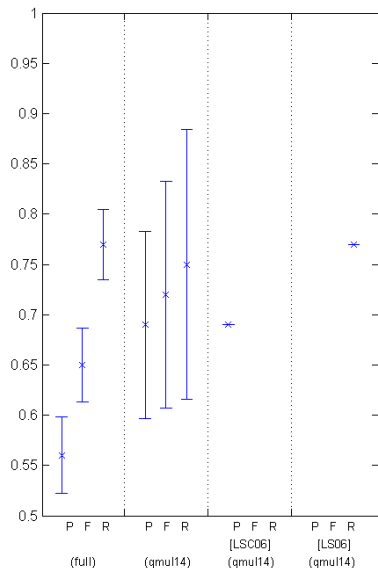
$$P = \frac{|\mathcal{B}_{algo} \cap_w \mathcal{B}_{gt}|}{|\mathcal{B}_{algo}|} \quad (1)$$

$$R = \frac{|\mathcal{B}_{algo} \cap_w \mathcal{B}_{gt}|}{|\mathcal{B}_{gt}|} \quad (2)$$

$$F = \frac{2PR}{P + R} \quad (3)$$

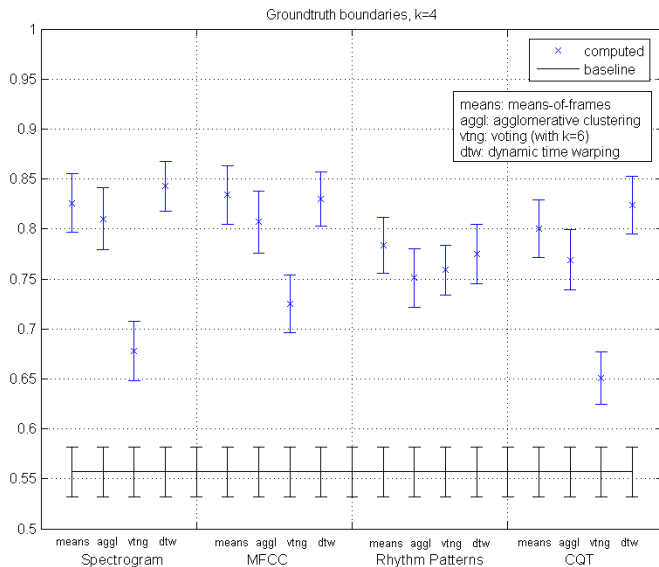
Structure Detection

$$r_f = 1 - ed'_s/t_s \quad (4)$$

Boundary Detection: $F = 0.66 \pm 0.034$ 

[LSC06] M. Levy, M. Sandler, and M. Casey. Extraction of high-level musical structure from audio data and its application to thumbnail generation. In Proc. ICASSP, Toulouse, France, 2006.

[LS06] M. Levy and M. Sandler. New methods in structural segmentation of musical audio. In Proc. EUSIPCO, Florence, Italy, 2006.

Structure Detection: $r_f = 0.707 \pm 0.025$ 

Discussion

- No restricting domain knowledge
- $F = r_f = 1$? Unrealistic!
E.g., Michael Jackson: Black or White. $r_f^{gt} = 0.76$
- Robust against improvement attempts

Michael_Jackson_-_Bad

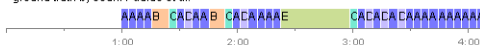
ground truth by Jouni Paulus et al.



ground truth by Mark Levy et al.

**Michael_Jackson_-_Black_Or_White**

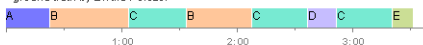
ground truth by Jouni Paulus et al.



ground truth by Mark Levy et al.

**Shania_Twain_-_Youre_Still_The_One**

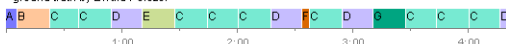
ground truth by Ewald Peiszer



ground truth by Jouni Paulus et al.

**Gloria_Gayner-I_Will_Survive**

ground truth by Ewald Peiszer



ground truth by Mark Levy et al.



Future Work

- Higher level features
- Select parameter values song-by-song
- User input

- Common corpus, groundtruth
- MIREX task?

Summary

- Algorithm for boundary and structure detection
- Large corpus, SegmXML annotations
- Source code

Thank you

Annotation files, source code available from
<http://www.ifs.tuwien.ac.at/mir/audiosegmentation/>

Q&A

Erratum: article, page 10

