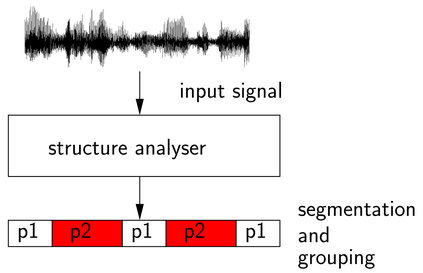
Acoustic Features for Music Piece Structure Analysis

Jouni Paulus, Anssi Klapuri
Department of Signal Processing, Tampere University of Technology, Tampere, Finland

Introduction

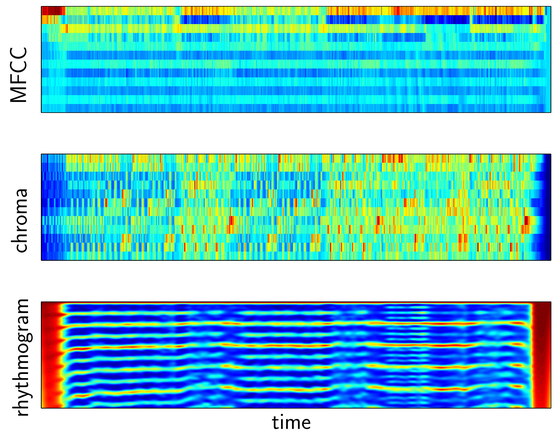
- Evaluate three acoustic features and two distance measures for music structure analysis.
  - Features focused on several time scales.
  - Distance measures defined between structural parts.
- Structure analysis
  - Recover sectional form of the piece.
  - Audio input.
  - Divide into segments (occurrences of parts, such as chorus, verse, etc.).
  - Group segments with similar content (occurrences of same part).



Acoustic features

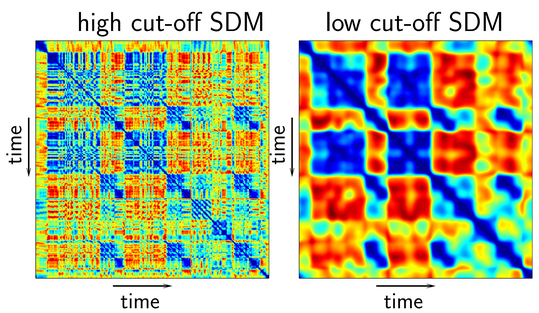
- Important cues in music structure perception:
  - repetitions (especially melodic),
  - change in rhythm, and
  - change in timbre.
- Timbre modelled with mel-frequency cepstral coefficients (MFCCs) → rough shape of spectrum.
- Tonal content modelled with chroma (pitch-class profile).
- Rhythmic content modelled with rhythmogram.
  - Onset accent signal → react to sound events.
  - Autocorrelation in windows of several seconds.

Example features from "Moottoritie on kuuma" by Pelle Miljoona Oy.



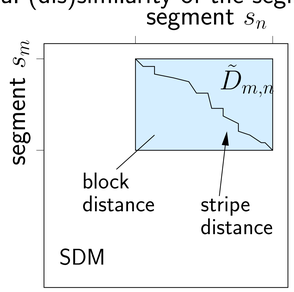
Feature processing

- Features resampled to beat-synchronised frames.
- Temporal filtering with varying cut-off frequencies to focus on different time scales.
- Self-distance matrices (SDMs) for all features.
  - Distance between all frames with cos-distance.
  - Depending on filter cut-off, dark stripes and blocks are formed (examples from MFCCs above).

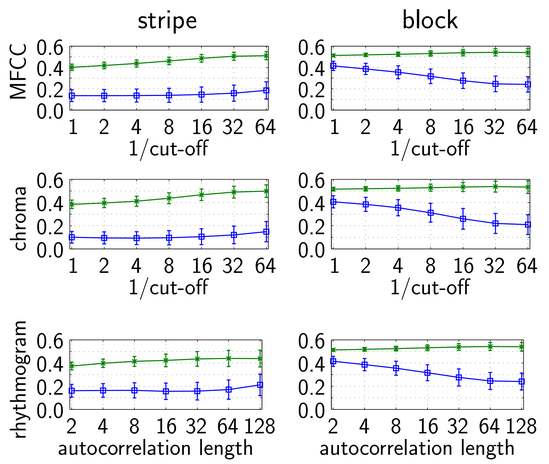


Segment distances

- Two segments s_m and s_n of piece define a submatrix $\tilde{D}_{m,n}$ in SDMs.
- Block distance for segments: average distance value in the submatrix.
  - General feature value (in)consistency during segments.
- Stripe distance: lowest cumulative distance across the submatrix.
  - Sequential (dis)similarity of the segments.

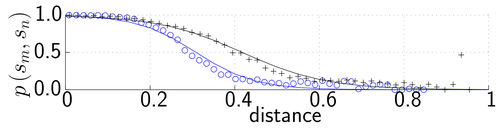


- Analyse distances between segments from same group and from different group, varying the time-scale parameter.
  - Manually annotated set of 557 popular music pieces, TUTstructure07.



Use for structure analysis

- Map distance to probability that the segments belong to same group, $p(s_m, s_n)$ (blocks, stripes).



- Find explanation of structure maximising

$$P(E) = \sum_{m=1}^{M} \sum_{n=1}^{M} A(s_m, s_n) L(s_m, s_n),$$

$$L(s_m, s_n) = \begin{cases} \log(p(s_m, s_n)) & \text{if } g_m = g_n \\ \log(1 - p(s_m, s_n)) & \text{if } g_m \neq g_n \end{cases}.$$

  - $A(s_m, s_n)$: area of $\tilde{D}_{m,n}$
  - g_m: group of segment s_m

Results

- Segmentation given, only group segments:
  → Chroma and MFCC stripes perform very well alone.
- System attempts to determine segmentation:
  → Stripe distance measure performance decreases. Feature/distance measure combinations improve result.

Conclusions

- If segmentation points are accurate, one-feature stripe distance is enough.
- If segmentation points are inaccurate, adding features and utilising different distance measure improve result.
- Different features provide complementary information.