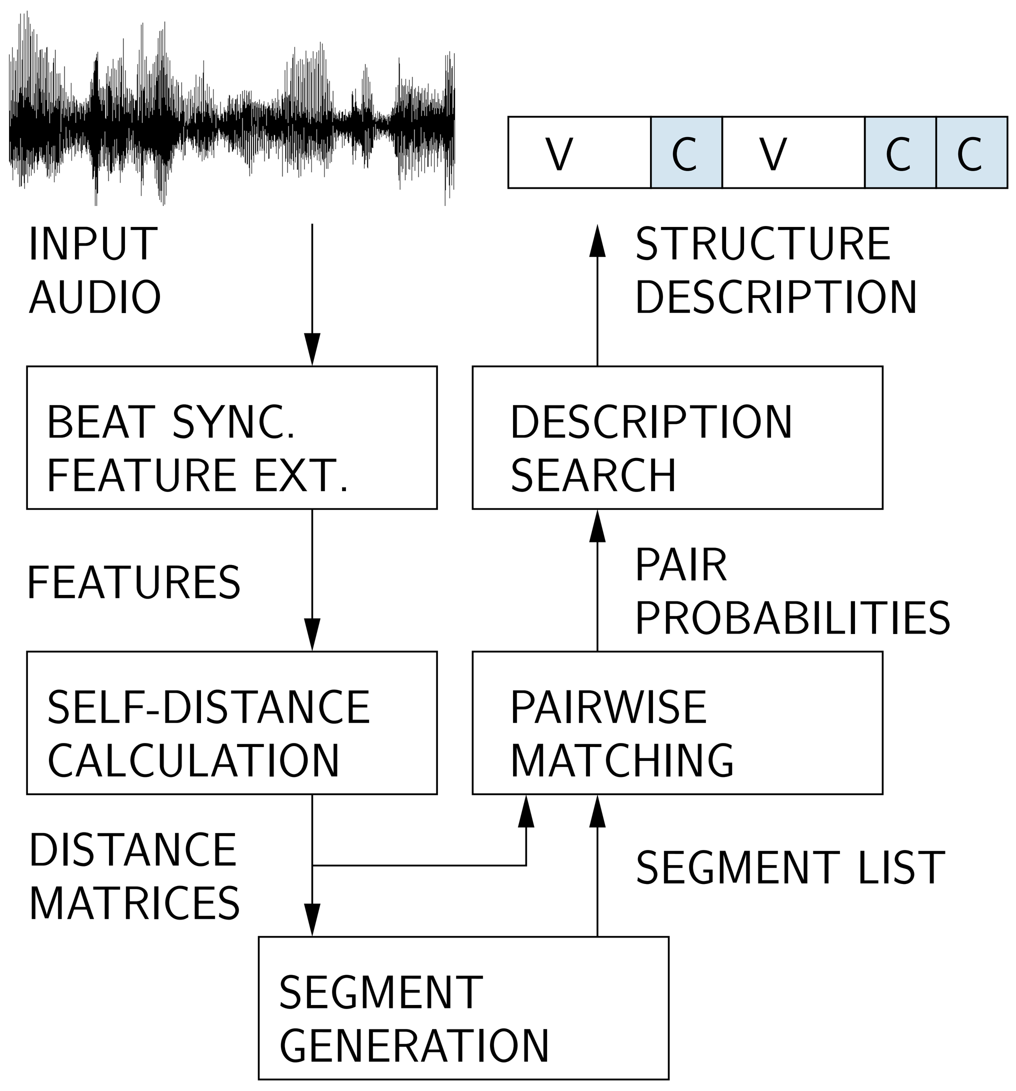
Music Structure Analysis with a Probabilistic Fitness Function in MIREX2009

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

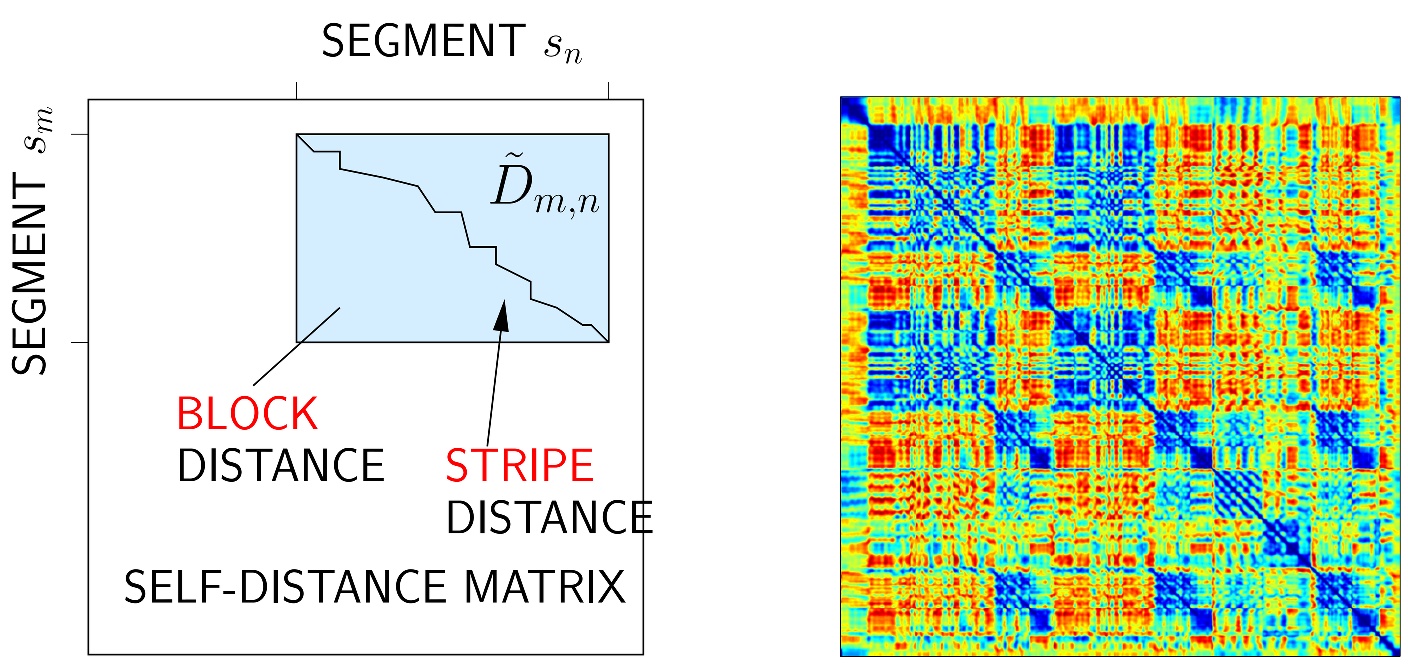
Introduction

- Structure analysis: from audio input
  - find segmentation to musical parts (e.g., chorus and verse), and
  - group segments with similar content.

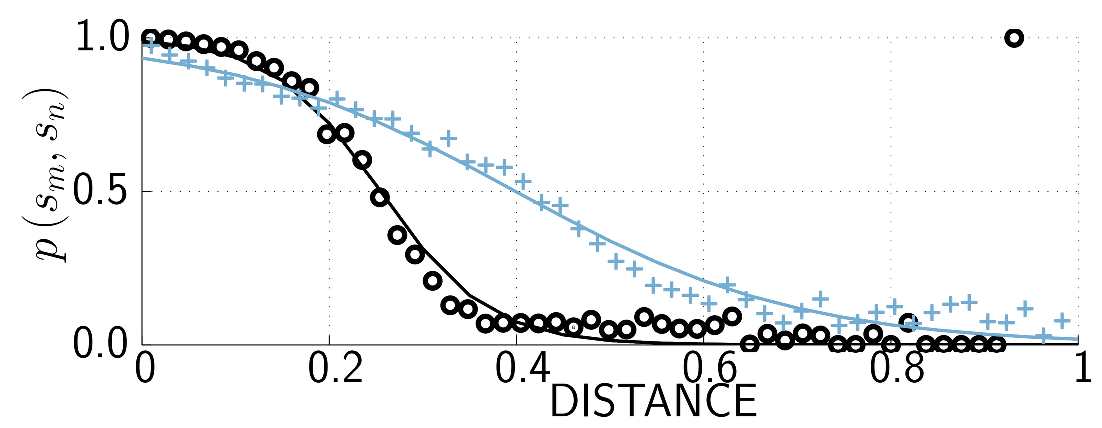


Audio front-end

- Three acoustic features for different aspects:
  - general timbre → MFCCs,
  - tonal / harmonic content → chroma (MF0 salience based)
  - rhythmic content → rhythmogram (onset accent autocorrelation).
  - Each feature focused on two temporal scales.
- Self-distance matrices from cos-distance between all beat frame pairs.
- Distance measures for segment pairs:



- Map distances to probability that the segments belong to same group. (E.g., stripes, blocks, + and ○ empirical values, line sigmoidal fit.)



Optimisation task

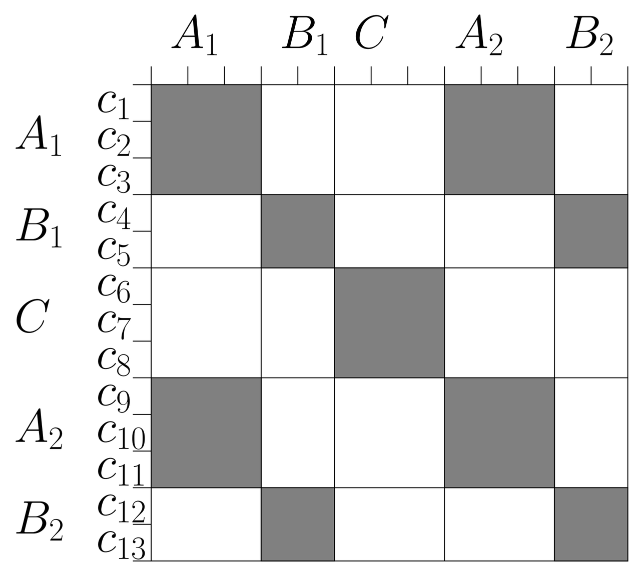
- Find the structural description E maximising

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

where

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

$A(s_m, s_n)$: submatrix area, and g_n: group of segment s_n.



- Formulate task as searching the optimal path through a directed acyclic graph.
  - Each candidate segment & group combination is a state.
  - Transition allowed only between consecutive segments.
- Problem: Rapid increase of search space size as a function of number of segmentation point candidates.

Search algorithm

- States contain an ordered buffer of tokens. At each iteration
  - the N best tokens are propagated and removed from the buffer,
  - arriving tokens are inserted to the buffer, and
  - only the M best tokens are stored for next iteration.
- Tokens store travelled state sequence.
- Operation parametrised by number of propagated tokens and maximum number of stored tokens.
  - Controllably greedy.
  - Finds a solution quickly, iterations increase search scope and may produce better solutions.
  - Store all tokens and run until all tokens have arrived to end state → exhaustive search.

Results

- Over- and under-segmentation scores 59.3% and 79.0% indicate tendency for over-segmentation.
- Frame pair clustering precision (74.1%), recall (46.2%), and F-measure (54.0%) support this assumption.
- Segment boundary detection precision (24.3%), recall (32.3%), and F-measure (27.1%) with 0.5 s allowed deviation indicate relatively accurate segmentation.
  → Most likely the method under-estimates probabilities of segment pairs to be of the same part.

- Method details in J. Paulus and A. Klapuri. Music structure analysis using a probabilistic fitness measure and a greedy search algorithm. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(6):1159–1170, Aug. 2009.