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:
  - Store all tokens and run until all tokens have arrived to end state
  - Find the structural description E maximising
    \[ P(E) = \sum_{m=1}^{M} \sum_{n=1}^{M} A(s_m, s_n) L(s_m, s_n) \]
  - 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.
- Operation parametrised by number of propagated tokens and maximum number of stored tokens.
  - Rapid increase of search space size as a function of number of segmentation point candidates.

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.