Measuring the Similarity of Rhythmic Patterns

Jouni Paulus, Anssi Klapuri

Tampere University of Technology
Institute of Signal Processing
P.O. Box 553, FIN-33101 Tampere, Finland
{paulus,klap}@cs.tut.fi

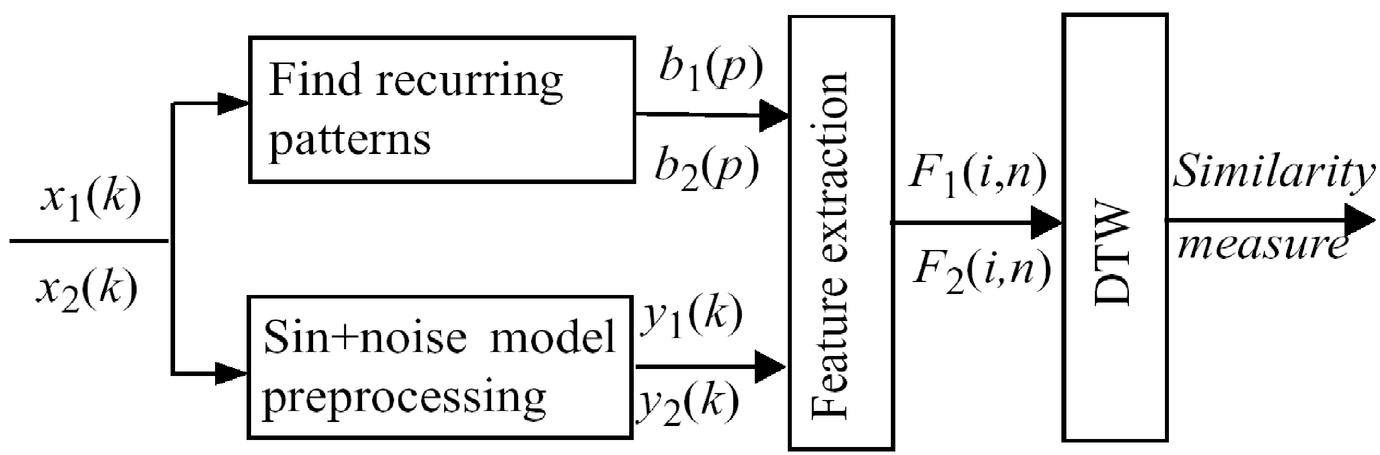
15th October 2002

Introduction

- Pattern induction and matching is important in music information retrieval
  - earlier done for melody
  - in rhythm, mainly beat detection
- The goal is to be able to measure the similarity of two arbitrary rhythmic patterns presented as acoustic signals.
- Similarity of rhythmic patterns?
  - psychological findings: factors like ”meter”, ”rapidity” and ”complexity”
  - used sound set should not effect the result
  - nor should different tempo
- No *a priori* information about the rhythms.

Overview

- Input: two acoustic signals.
- Output: similarity measure in range [0,1].
- Two novel parts that constitute the system core are
  - the musical meter estimation part and
  - the actual similarity measurement part.
- Pre-prosessing
  - sinusoid+noise modelling to separate drums from the source signal
  - drum sounds in noise residual



Pattern segmenting

- Essential step before similarity measurements
  - from continuous time domain signal to chunks that represent patterns
  - brute-force matching of all possible patterns is not practically computable

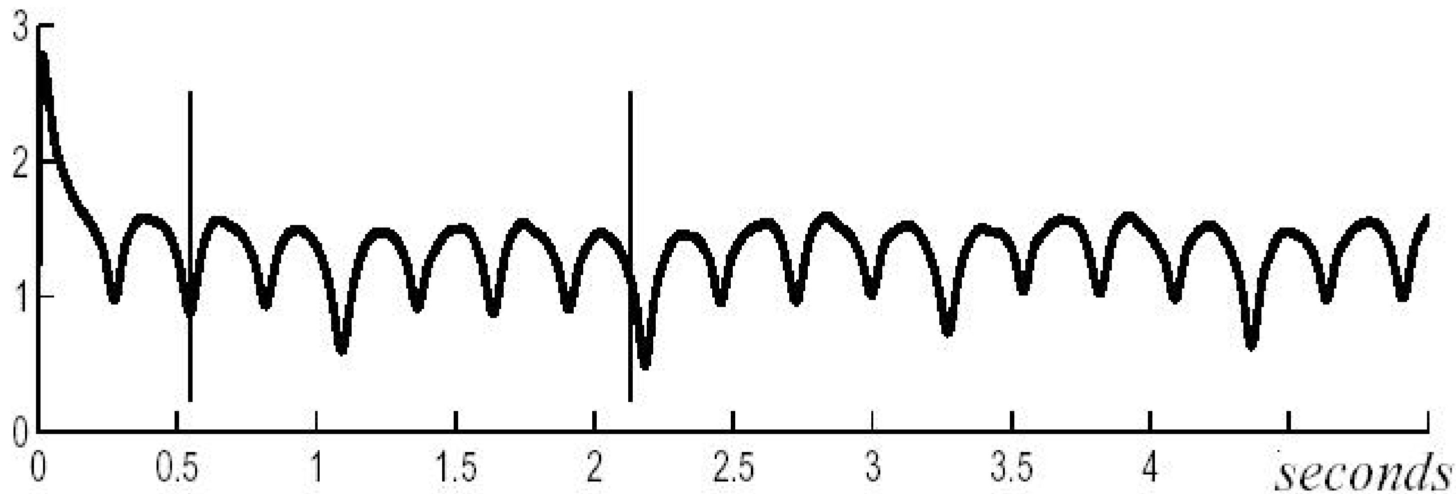
- Musical meter estimation process
  - input: acoustic music signal without pre-processing
  - output:
    - *tatum* period
    - *tactus* period (”beat”)
    - musical *measure* period (pattern length) and phase
      - candidates for pattern boundaries

Pattern segmenting

Periodicity detection

- Signal modelled with amplitude envelopes at 8 subbands
  - model retains the rhythmic percept of most musical signals and significantly reduces the amount of data [Scheirer]
- Periodicity analysis of the amplitude envelopes at subbands, $v_c(k)$
  - use principles of ”YIN”-algorithm, originally proposed by Cheveigné & Kawahara for fundamental frequency estimation
- Amplitude envelopes are summarized across channels
  - resulting to $s(\tau)$
- Function $s(\tau)$ serves as the source of information for musical meter estimation.
- Tatum, beat and pattern lengths are selected from among maxima of $1/s(\tau)$ by restricting the search range (τ) with a log-normal Gaussian prior distribution.
- Pattern phase determined according to the power maxima of the lowest subband.

Pattern segmenting
Periodicity detection



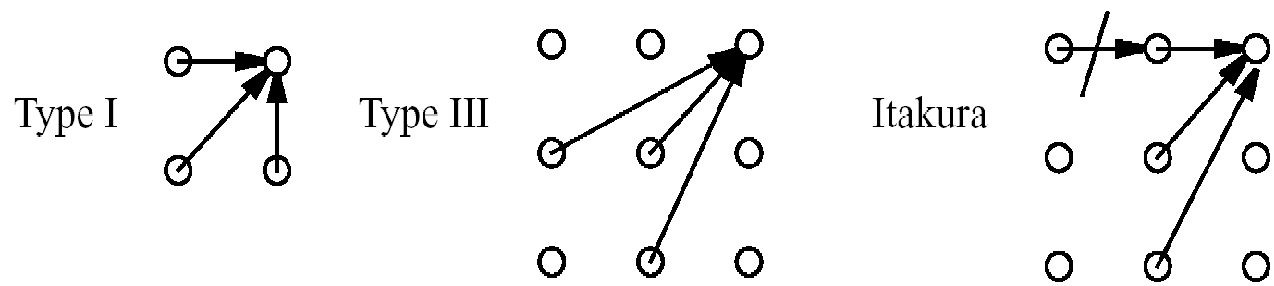
- Figure: typical instance of $s(\tau)$ for a rock piece
  - actual beat and pattern periods indicated with vertical lines

Feature extraction and normalization

- Fundamental features in addition to timing are perceived loudness and brightness.
- Loudness is modelled with log mean-square energy.
- Brightness modelled with spectral centroid.
  - Balancing point of spectral power distribution.
- Also utilization of timbral information was tried with 15 Mel-frequency cepstral coefficients (MFCCs).
- So far, extracted features represent ”tone colour” and loudness in absolute scale,
  - but rhythmic events are perceived in relation to each other.
- Loudness is normalized to zero minima and unity variance.
- Spectral centroid (and MFCCs) are weighted with normalized loudness.
- Resulting features have zero mean and unity variance over time.
- Tone color modeled by only deviations up/down from the average value.
  - attempt to discard sound set information.

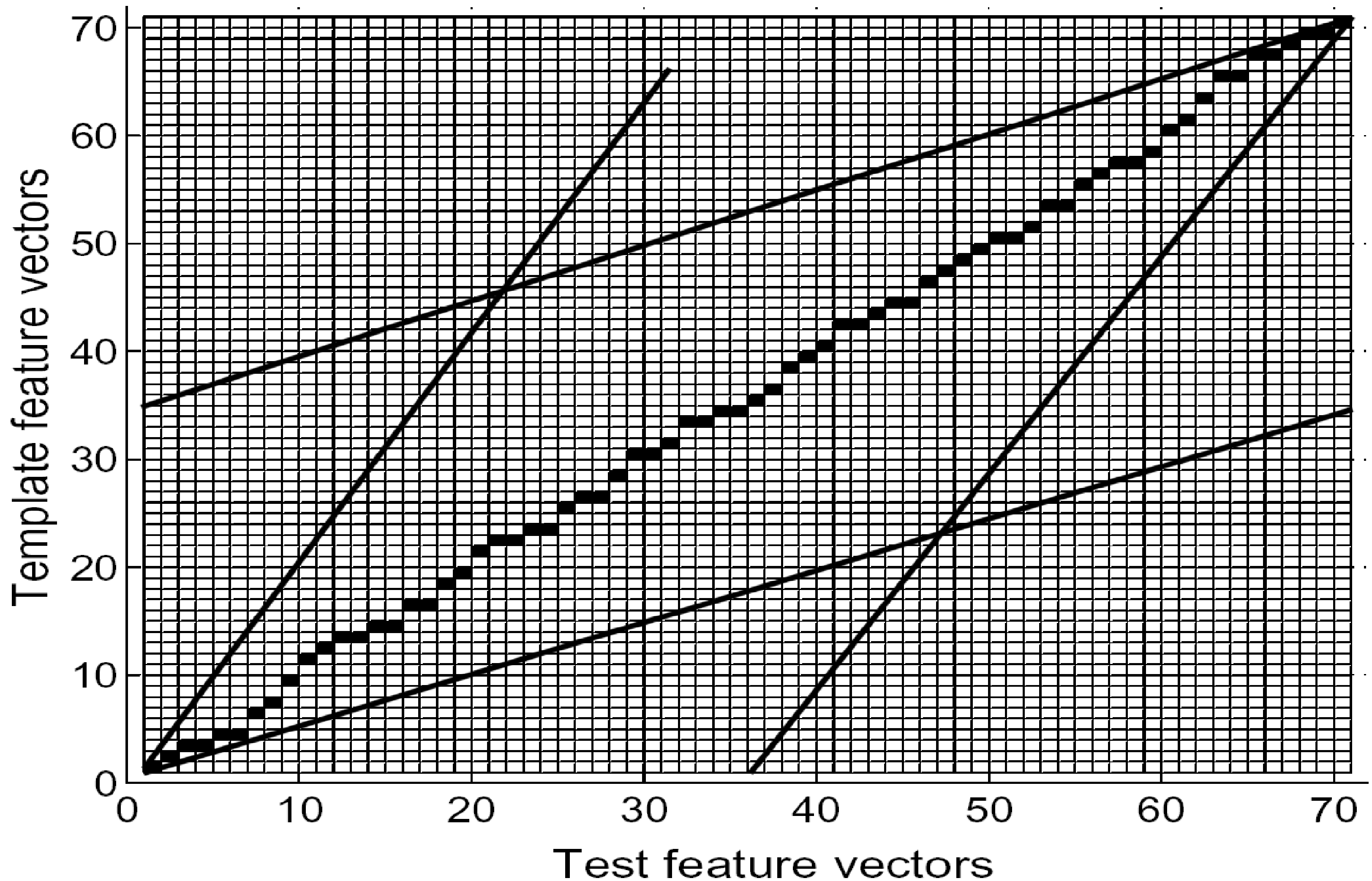
Dynamic time warping

- Tries to find optimal fit for the template and an test vector.
  - The shortest path through a matrix consisting of feature vector fitting points.
- Allows the template and the test vector to be different in length.
- Fitting done by allowing local timing adjustments with local path constraints.



- Global constraints either implied from local constraint or explicitely
  - maximum slope 2 and minimum slope 0.5
- Similarity value is defined to be the normalized inverse of total path length of the optimal path.
  - Path length consists of feature vector distances and timing alignment costs.

An example path



Database

- Database to evaluate the musical meter estimation part
  - acoustic signals: single-channel, 44.1 kHz, 16-bit PCM
  - tactus (beat) and musical measure (i.e. pattern) positions were manually annotated for a one-minute excerpt of each piece
  - done by tapping along with the pieces
  - measure boundaries could be reliably marked by listening for a subset of the pieces only

Genre	Tactus annotated (# of songs)	Patterns annotated (# of songs)
Classical	85	–
Electronic/Dance	27	18
Hip Hop/Rap	12	8
Jazz/Blues	62	19
Rock/Pop	111	61
Soul/RnB/Funk	44	27
World/Folk	24	8
Total	**365**	**141**

Pattern segmenting

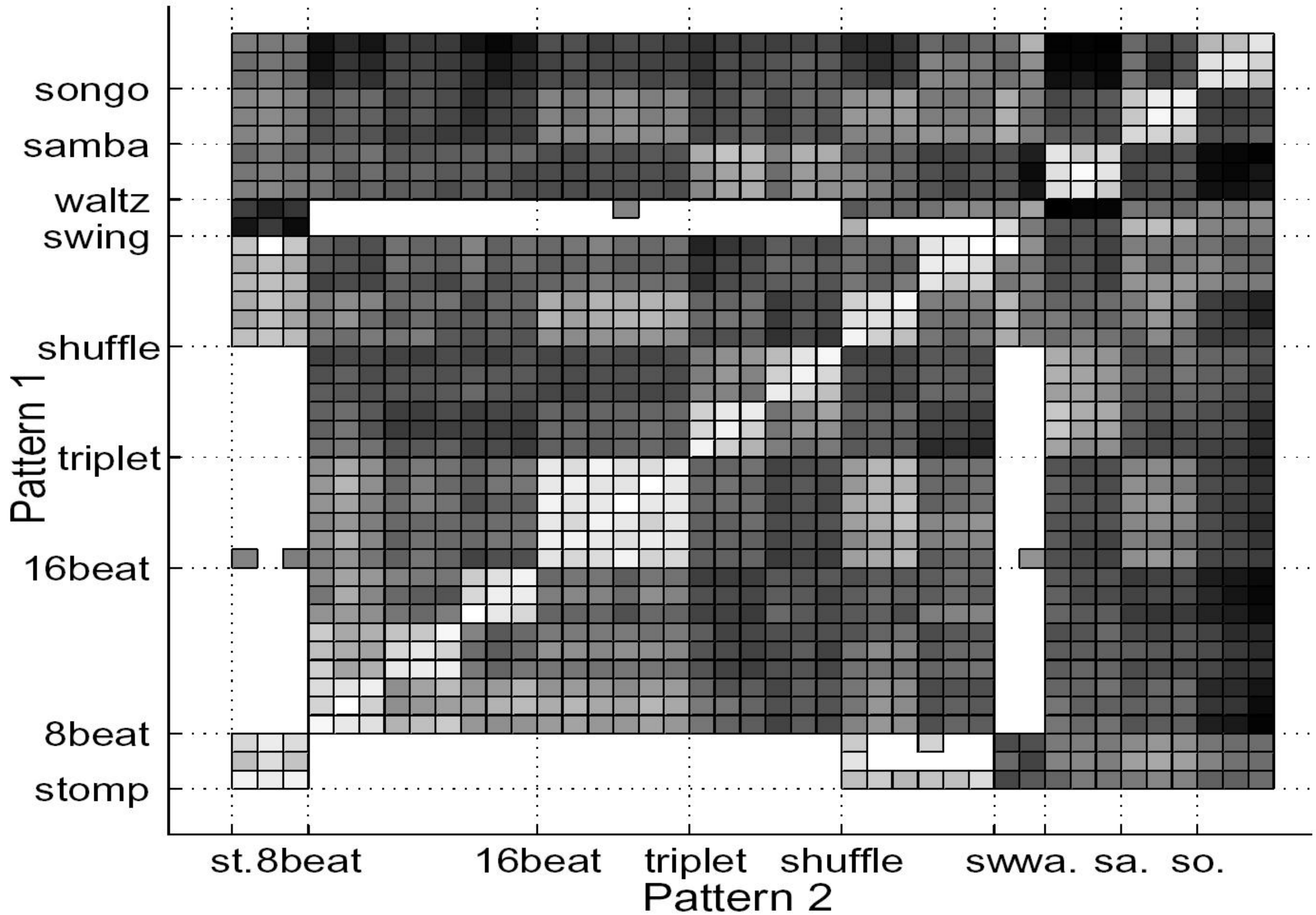
- Input: a 10-second excerpt from each piece
- Correct tactus / measure period was defined to deviate less than 10 % from the reference
- Tactus period was correct for 67 % of the 365 pieces
  - most typical error was tactus period doubling.
- Measure (pattern) period was correct for 77 % of the 141 pieces
  - 17 % : half or double the pattern lengths
  - 6 % : unclassified errors
- Measure information could be annotated only for metrically more clear cases
  - partly explains the performance difference between tactus and measure length estimation
- Pattern phase: correct in approximately half of the cases
  - several pattern boundary candidates have to be considered in comparison

Similarity measurements setup

- Drum pattern database
  - 9 standard rhythm patterns with variations
    - stomp, 8-beat, 16-beat, triplet, shuffle, swing, waltz, samba, songo
    - totalling 14 different patters
  - 3 different drum sets
- Evaluation of the DTW part independently
  - pattern borders are annotated by hand
  - no pre-processing

Drum set	Sounds involved		
1	bass drum	snare	hi-hat
2	bass drum	brush slap snare	ride cymbal
3	bass drum	cross stick	shaker

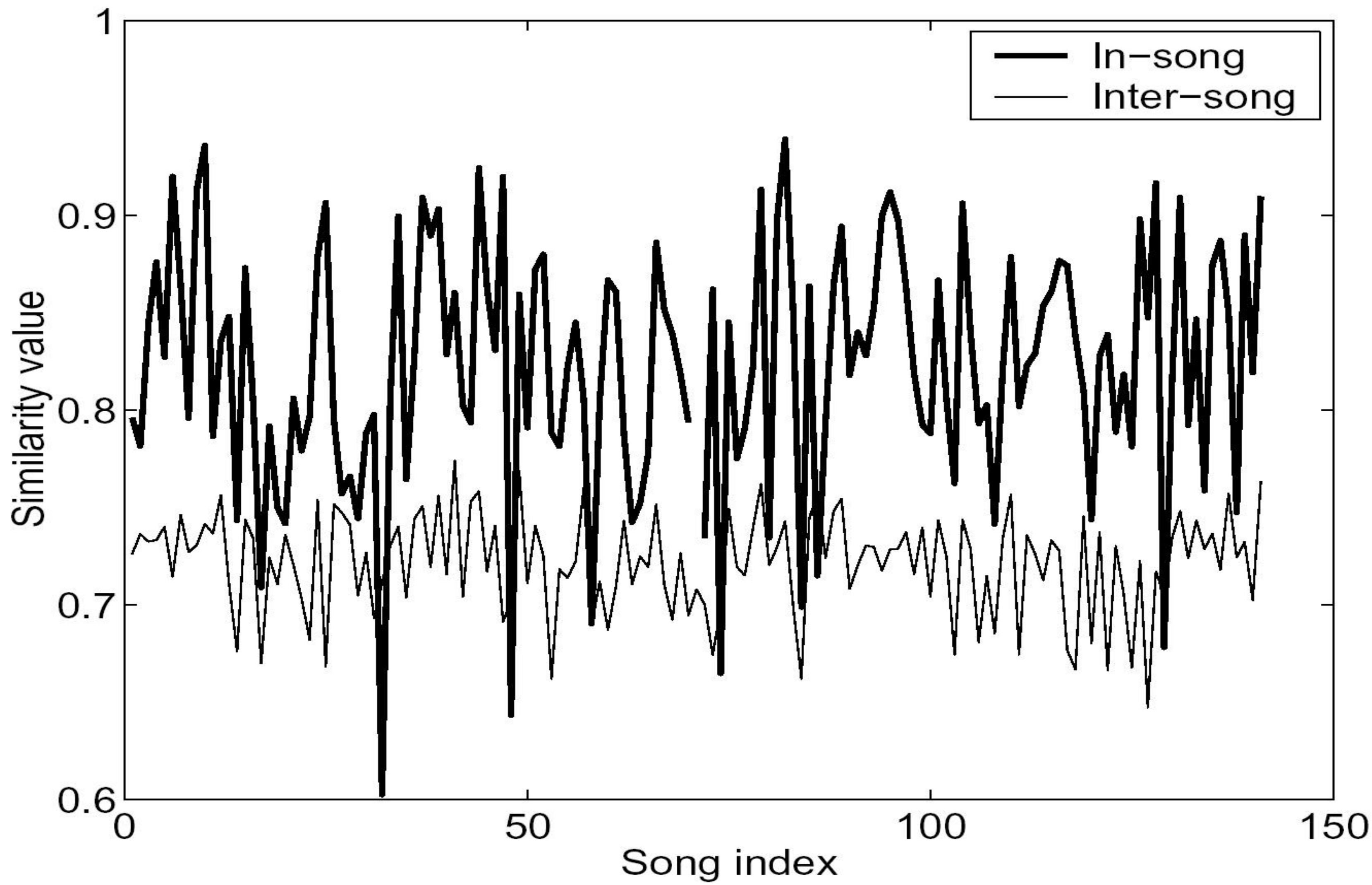
Similarity measurements results



Performance with real music signals

- Same database as in use with meter estimation simulations.
  - 141 songs with pattern borders annotated by hand
- Sinusoid+noise pre-processing used to separate drums from signal.
- Randomly two patterns from each song
  - one to work as template and the other as the test pattern.
- Similarity of all templates against all test patterns was measured.
  - result matrix of size 141*141, where in-song distance is at diagonal
- In-song similarity bigger than the average inter-song similarity, as expected.

Simulations
Performance with real music signals



Conclusions

- Successful pattern extraction from actual music signals.
- Consistent similarity measure assigning for drum patterns
  - even when performed with different sound sets.
- The most successful feature turned out to be the log-energy weighted spectral centroid.
  - Normalized to have zero mean and unity variance.
- Dynamic time warping reconciled tempo differences and small pattern border deviations.
  - High path cost needed to force steady proceeding in time (no tempo fluctuations).