STATE OF THE ART REPORT: AUDIO-BASED MUSIC STRUCTURE ANALYSIS

Jouni Paulus* (Fraunhofer Institute for Integrated Circuits IIS, Germany)
Meinard Müller (Saarland University and Max-Planck-Institut für Informatik, Germany)
Anssi Klapuri (Queen Mary University of London, UK)

13.8.2010

*Work done when author was with the Department of Signal Processing, Tampere University of Technology, Finland

Overview

1. Introduction
2. Feature representation
3. Self-distance matrix
4. Approaches
   - Novelty
   - Homogeneity
   - Repetition
   - Combined
5. Evaluation
6. Conclusions

Introduction
Structure

- "Music is organized sound." – Edgard Varèse

- Organization present on many levels:
  - Notes forming phrases and chords
  - Chord sequences
  - On largest level, musical sections, parts (e.g, chorus, verse)
- Within this paper, *music structure analysis* refers to the process of recovering a description of the sectional form.

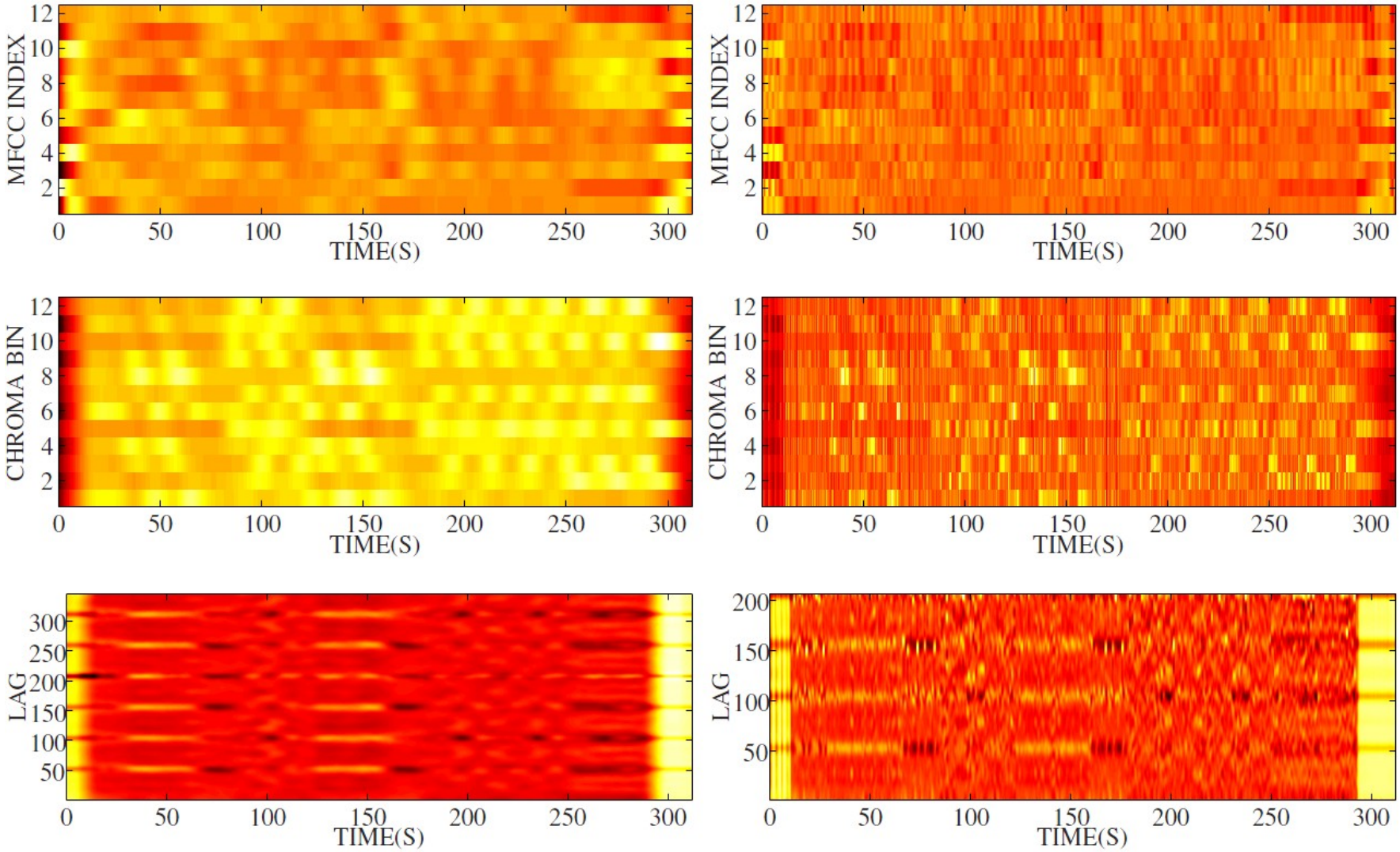
Introduction
Instantiations of structure in music

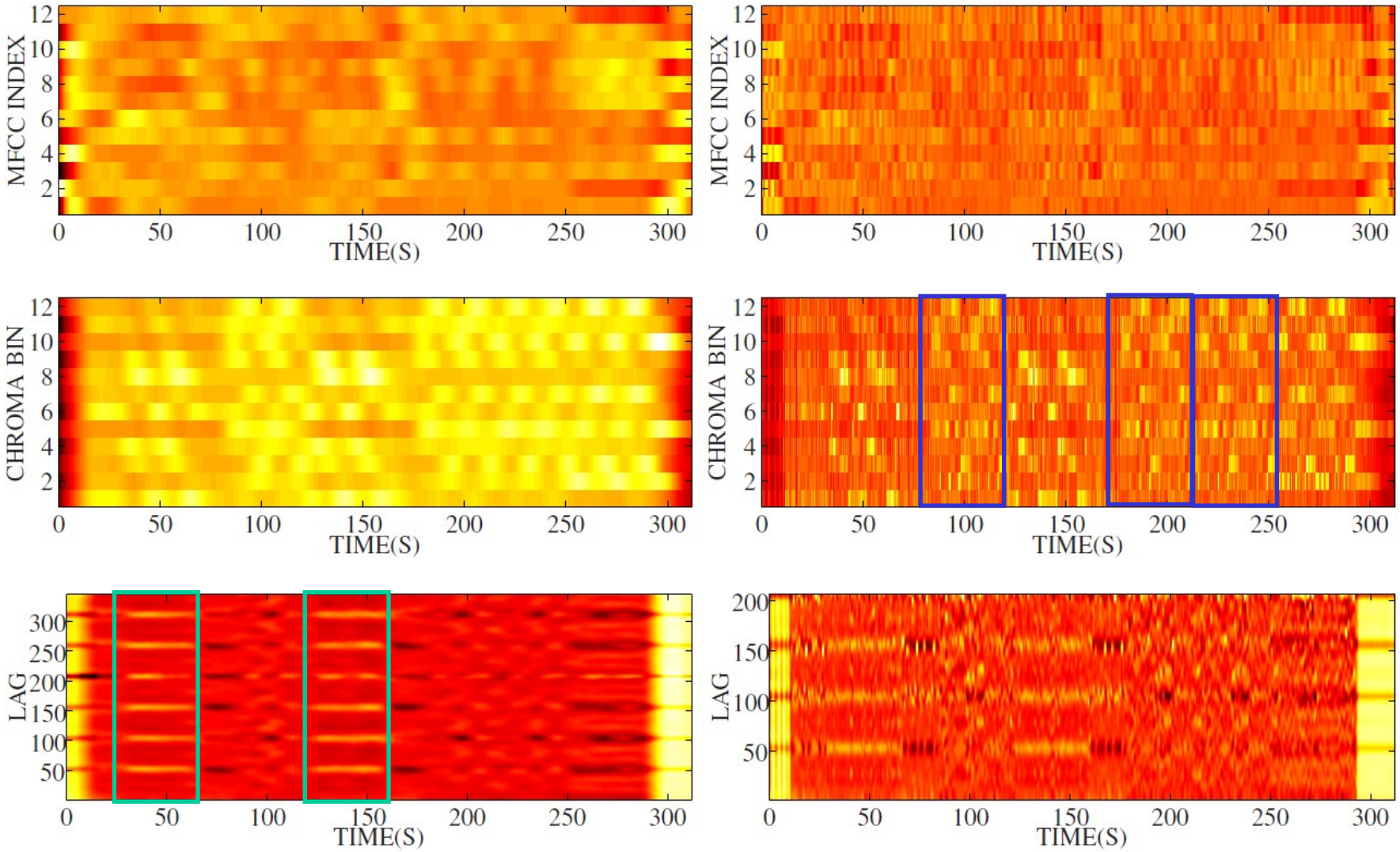
- Relationships between musical elements
  - Temporal sequences (e.g., melodies)
  - Repetitions (e.g., rhythmic and harmonic patterns, also variations)
  - Contrasts (e.g., loud and soft parts)
  - Homogeneity within a musical part (e.g., instrumentation, tempo, or harmonic content)

- Analysis aims at revealing these (and other, hidden) relationships

Feature representation
Common acoustic features

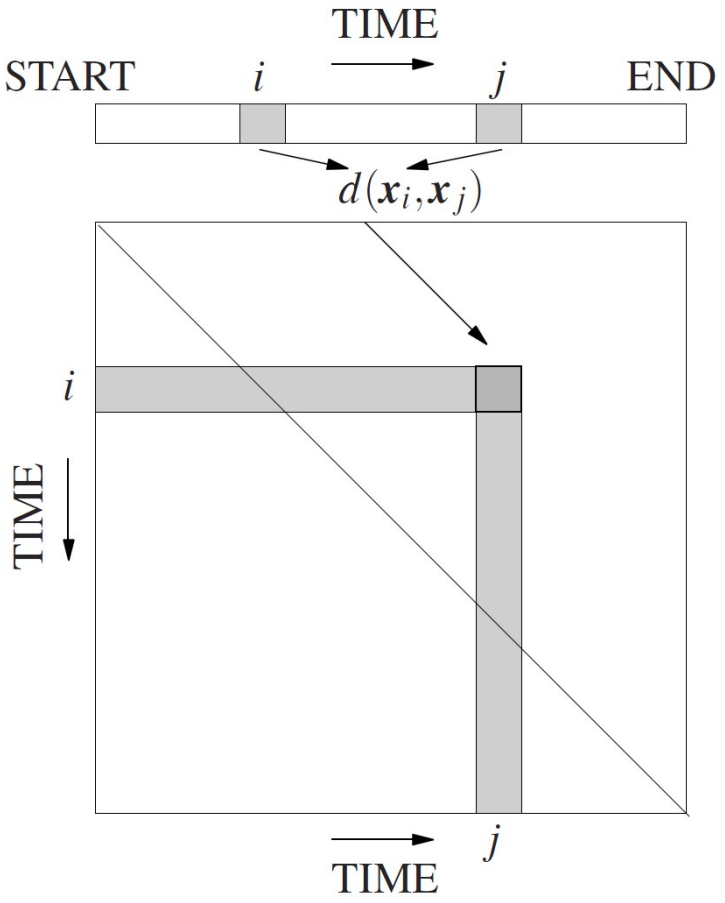
- Humans use different musical aspects simultaneously to deduce the structure
- Musical texture, timbre
  - Related to instrumentation
  - Often using coarse shape of spectrum (e.g., MFCCs)
- Pitched content
  - Melodies, chords, harmonies
  - Often estimating energy at each semitone and folding result to one octave (e.g., chroma)
- Rhythmic content
  - Tempo, drum patterns
  - Often a periodicity estimate of onset accentuation pulse (e.g., tempogram, rhythmogram, beat spectrogram, dynamic features)





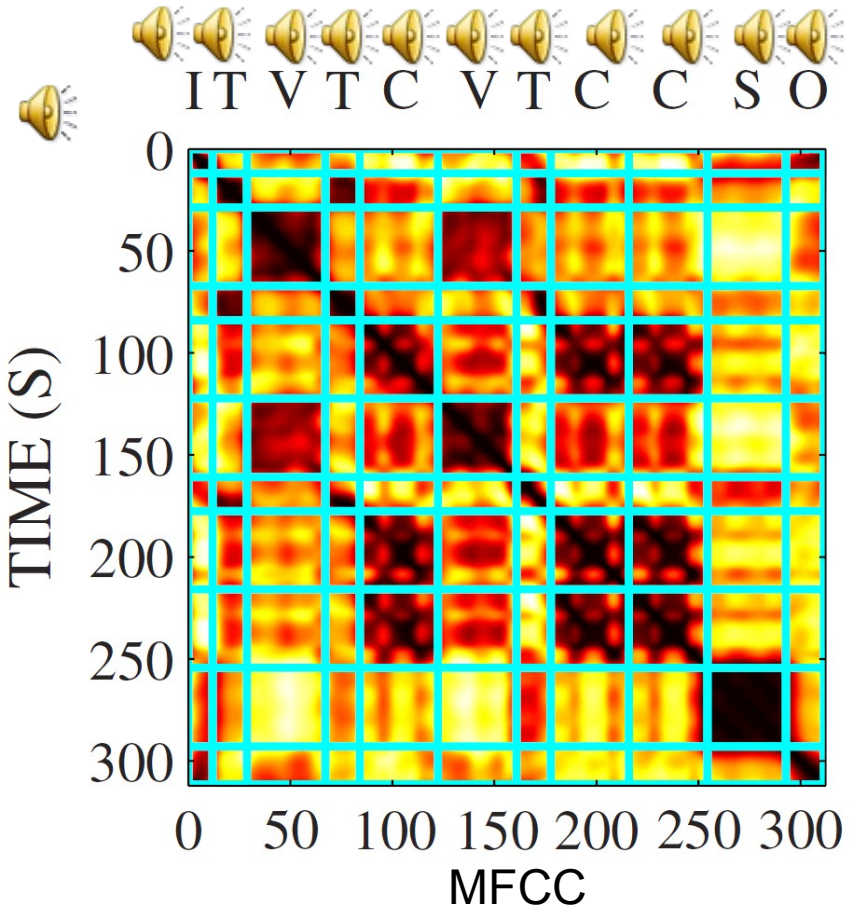
Mid-level representation
Self-distance matrix (self-similarity matrix)

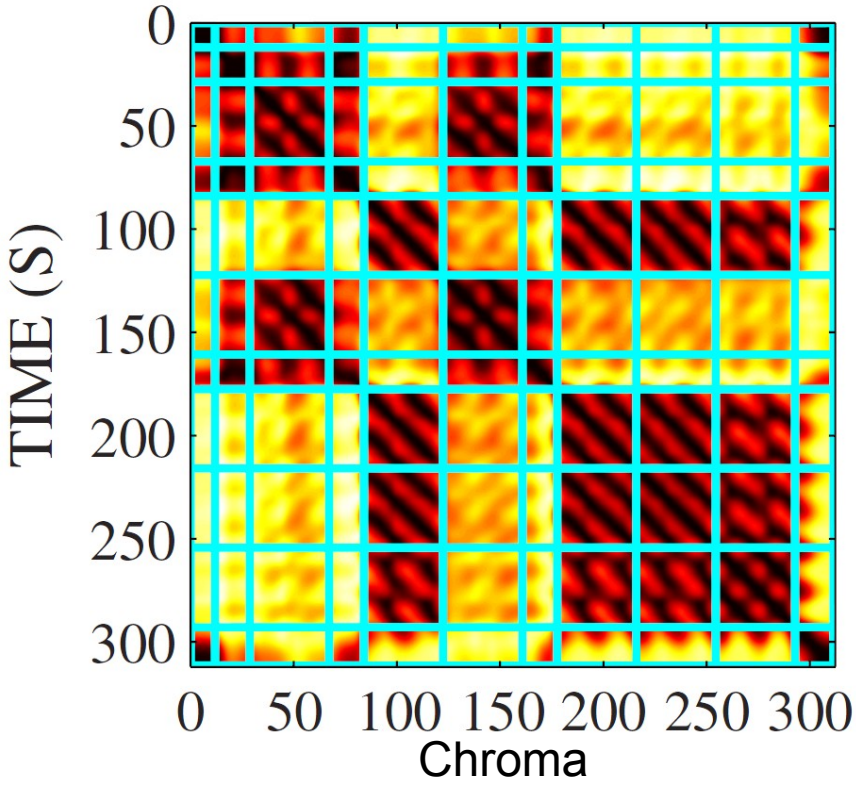
- Common mid-level representation
- Comparing each frame with all other frames
  - Each element describing the dissimilarity of two frames (or a sequence of frames)
- Informative patterns
  - Stripes for repeated sequences
  - Blocks for homogenous segments



Mid-level representation

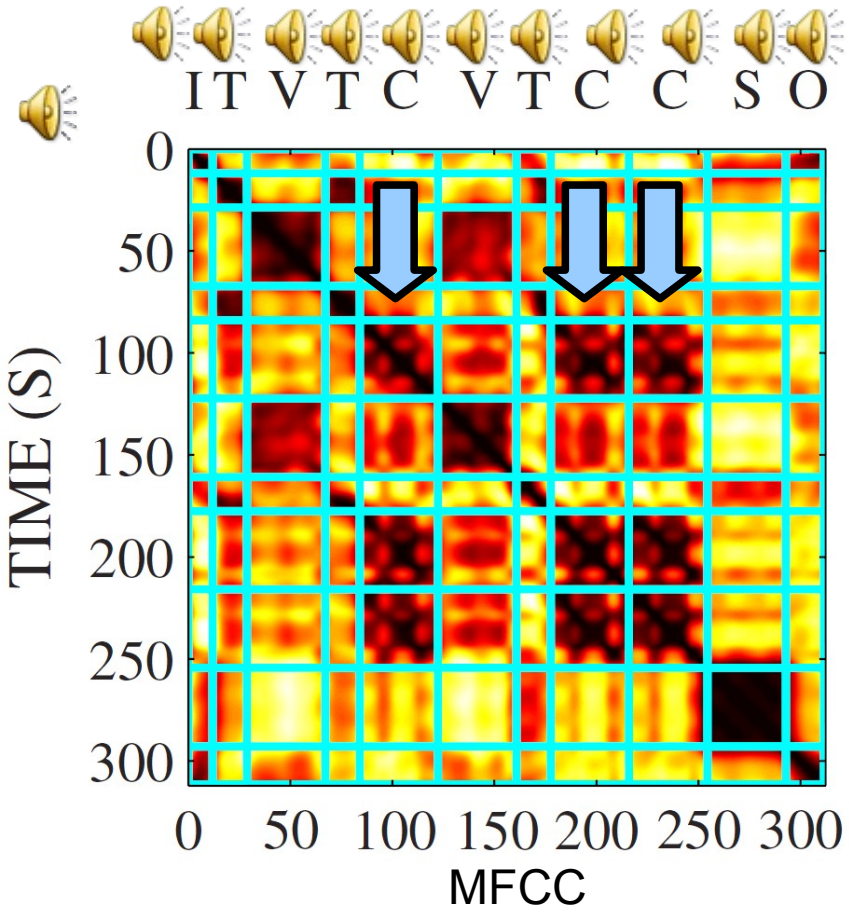
Self-distance matrix examples

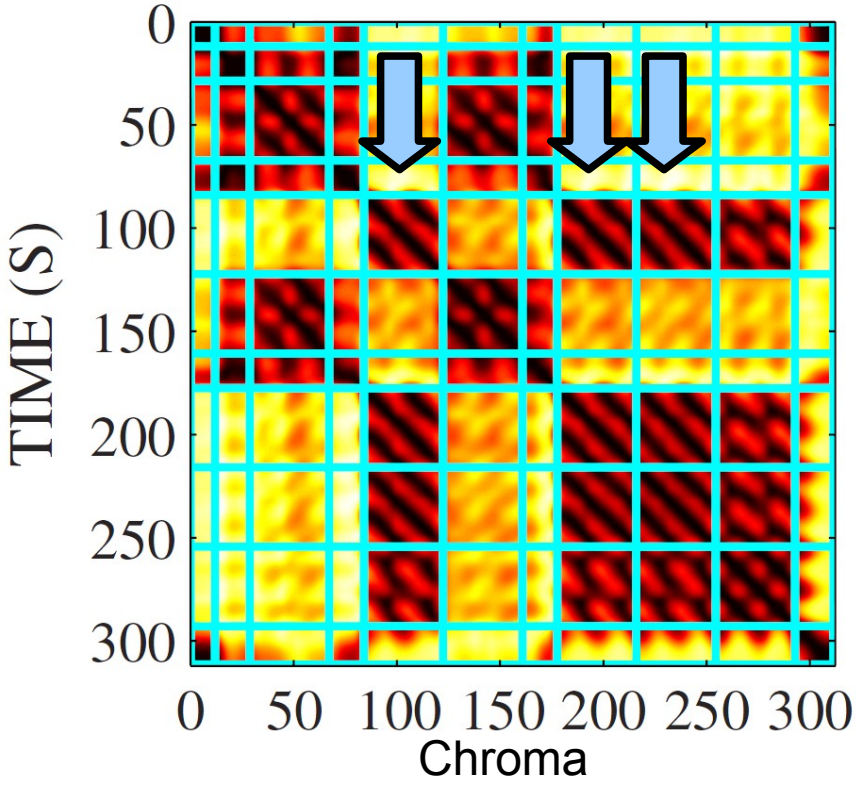




Mid-level representation

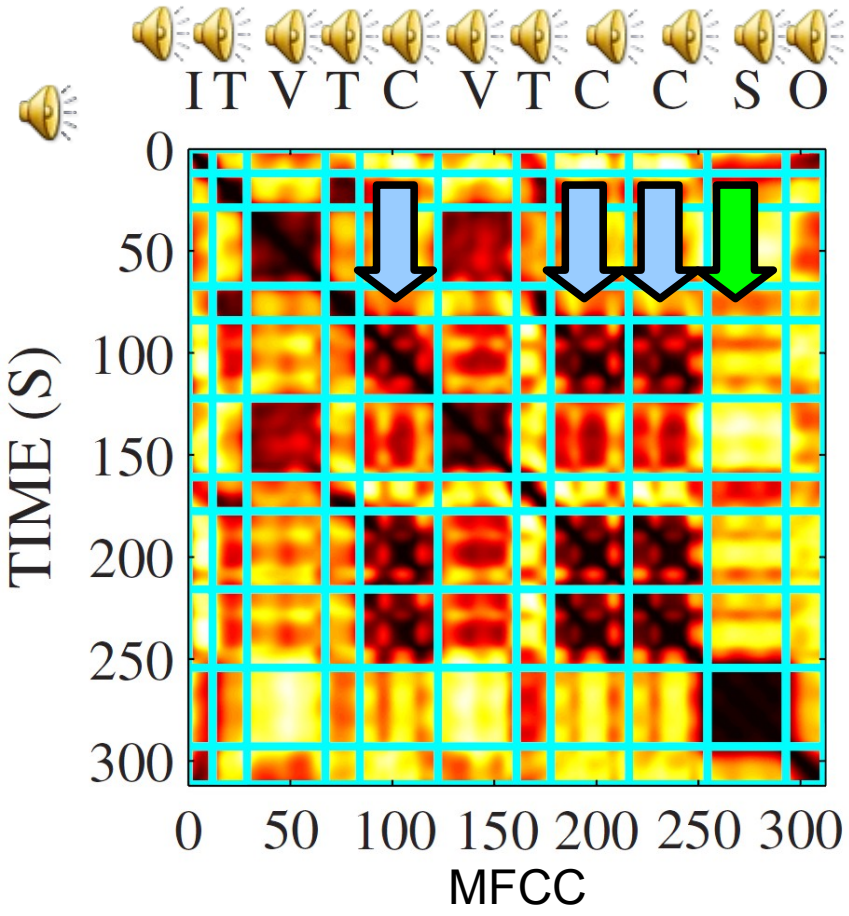
Self-distance matrix examples

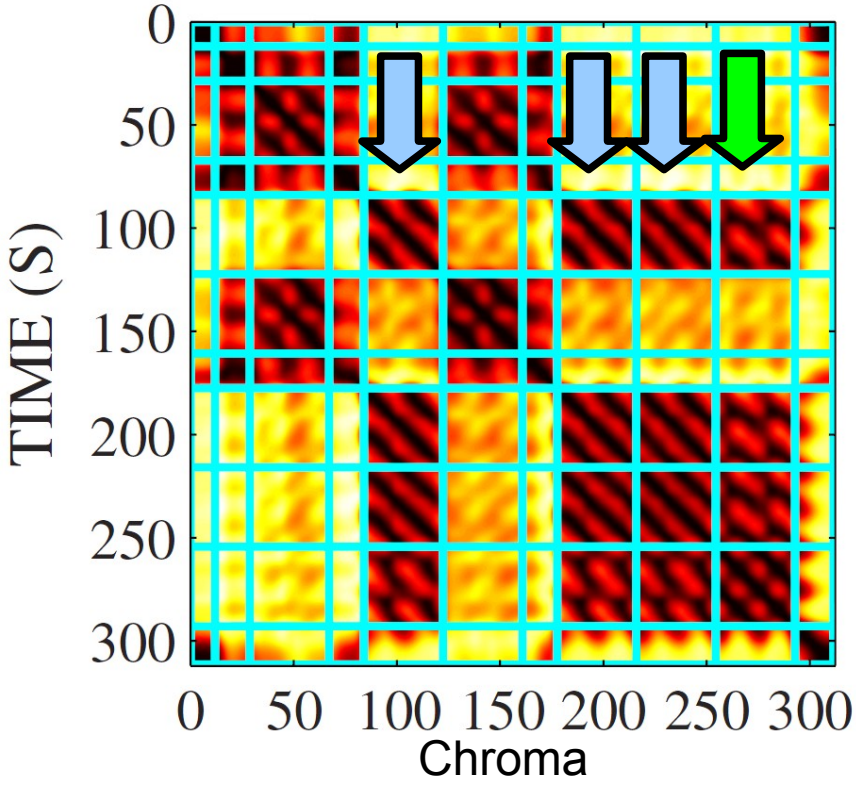




Mid-level representation

Self-distance matrix examples





Approaches
Categorization

- Proposed categorization
  - Novelty-based approaches (points of high contrast)
  - Repetition-based approaches
  - Homogeneity-based approaches

- An earlier division into
  - Sequence approaches: There exists sequences that are repeated during the piece (stripes in SDMs)
  - State approaches: Piece is produced by a state machine, each state produces distinct observations

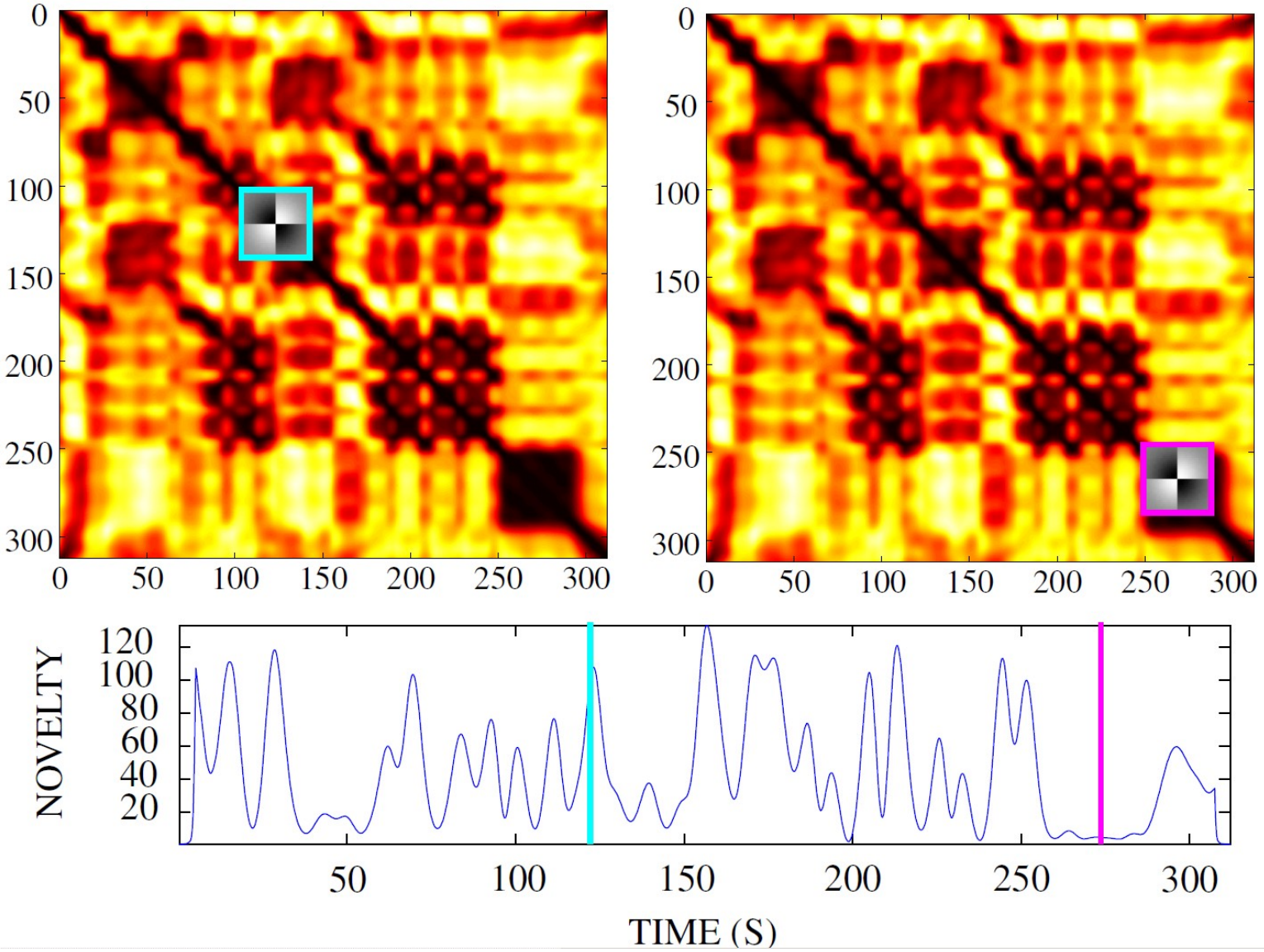
Introduction
Instantiations of structure in music

- Relationships between musical elements
  - Temporal sequences (e.g., melodies)
  - Repetitions (e.g., rhythmic and harmonic patterns, also variations)
  - Contrasts (e.g., loud and soft parts)
  - Homogeneity within a musical part (e.g., instrumentation, tempo, or harmonic content)

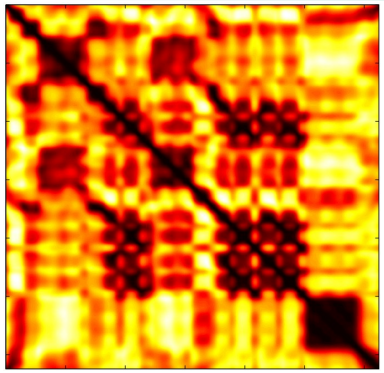
- Analysis aims at revealing these (and other, hidden) relationships

Approaches
Novelty

- Locate points of high contrast (~part boundaries)
- Various methods
  - 2D corner point detection in SDM using a checkerboard kernel matrix
  - Information theoretic approaches (e.g., BIC)
  - Classifiers
- Employed as the first step in many more complex methods


0
50
100
150
200
250
300
NOVELTY
120
100
80
60
40
20
TIME (S)

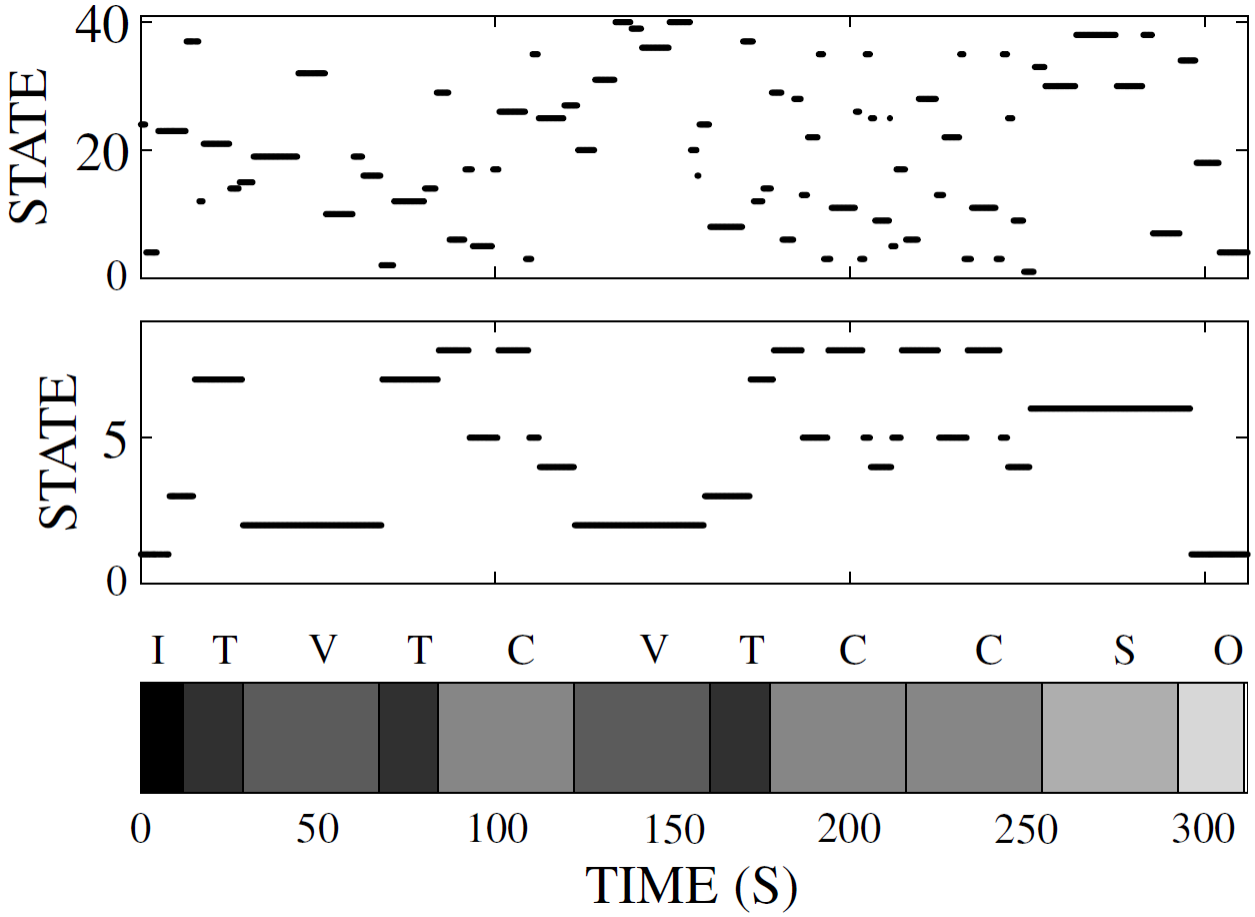
Approaches
Homogeneity



- Locate segments that are homogenous in some aspect
  - E.g., similar timbral characteristics
  - ”Locate blocks in SDM”
- Methods
  - Parametrize segments found by a novelty method, e.g., by feature mean(s) and (co)variance(s), and cluster segments
  - Hidden Markov models
    - Each state produces observations from a distribution
    - Only few states → each state is a musical part
    - Large number of states → vector quantisation
      - Fragmentation problem requiring post-processing

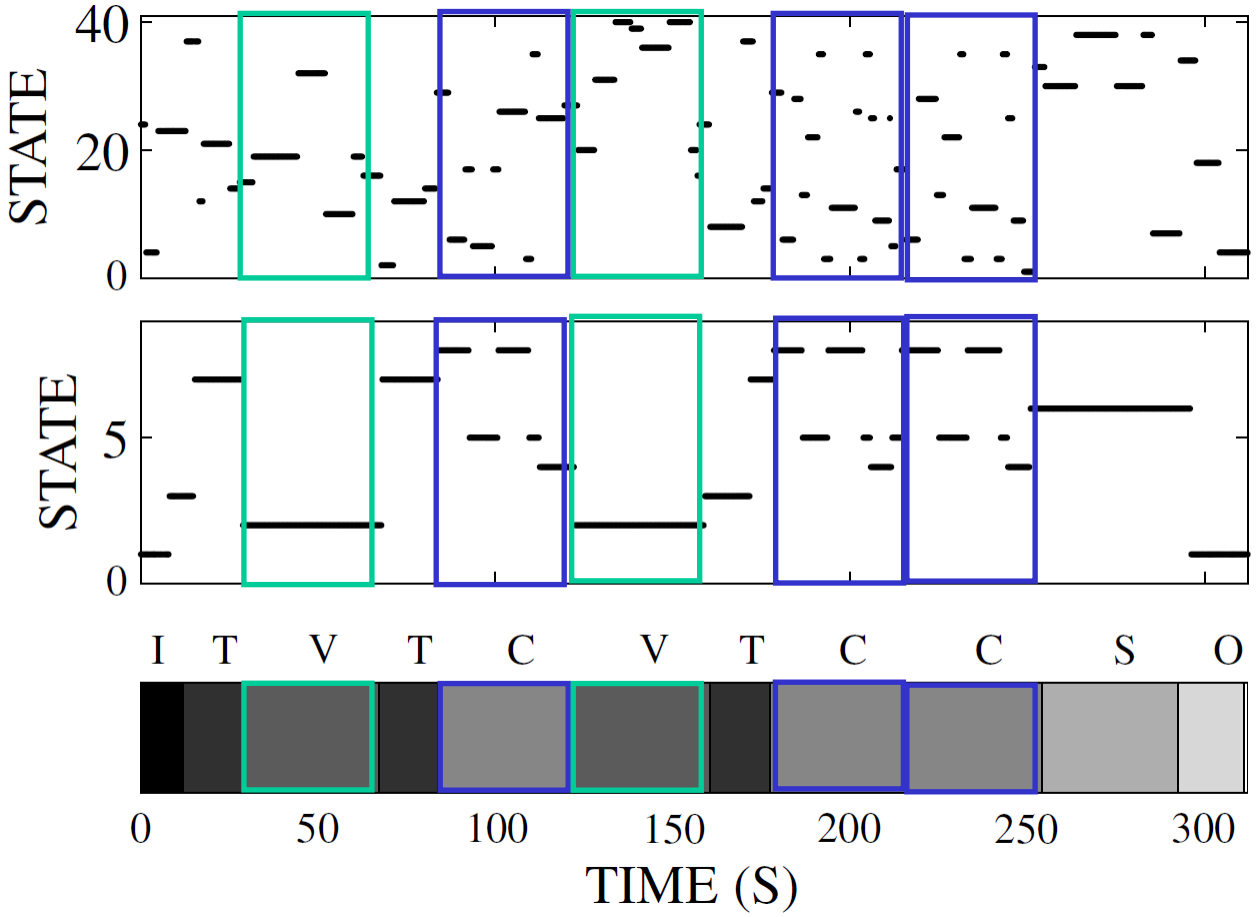
Approaches

HMM example (8 and 40 states)



Approaches

HMM example (8 and 40 states)

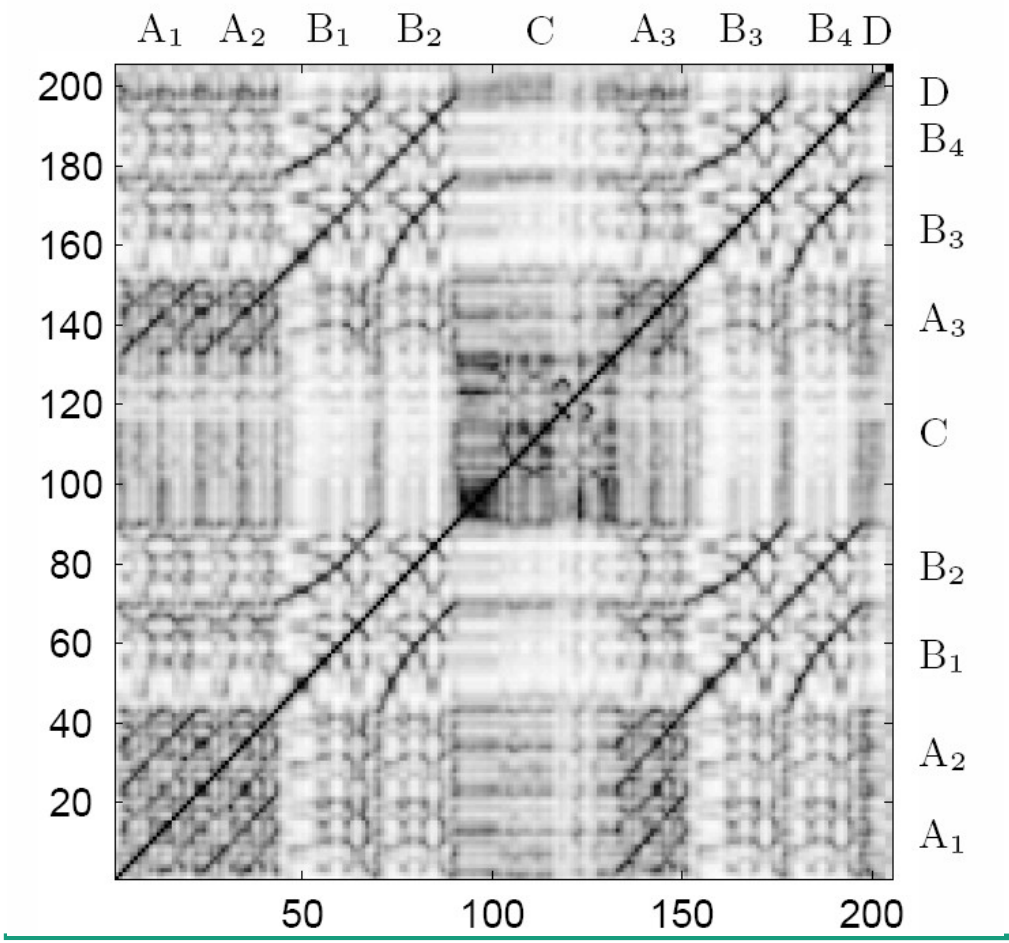


Approaches
Repetition

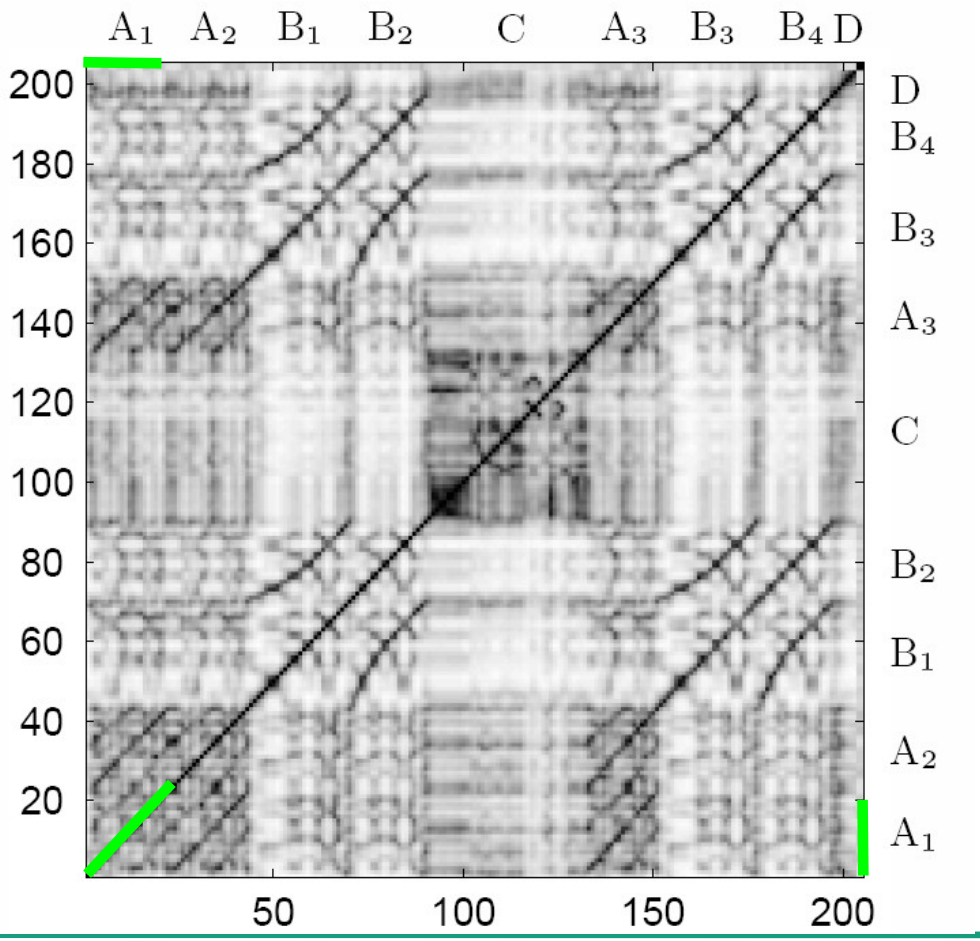
- Locate repeated sequences → order important

- Methods
  - String processing on feature vectors or state sequences
  - Often: locating stripes from SDMs
    - Problems with gaps, curvature, and lack of contrast
      - → Image processing filters for enhancement
    - Binarisation
    - Each stripe defines two segments: original and a repeat
      - → Transitivity logic problems for more comprehensive description

Approaches
Repetition example

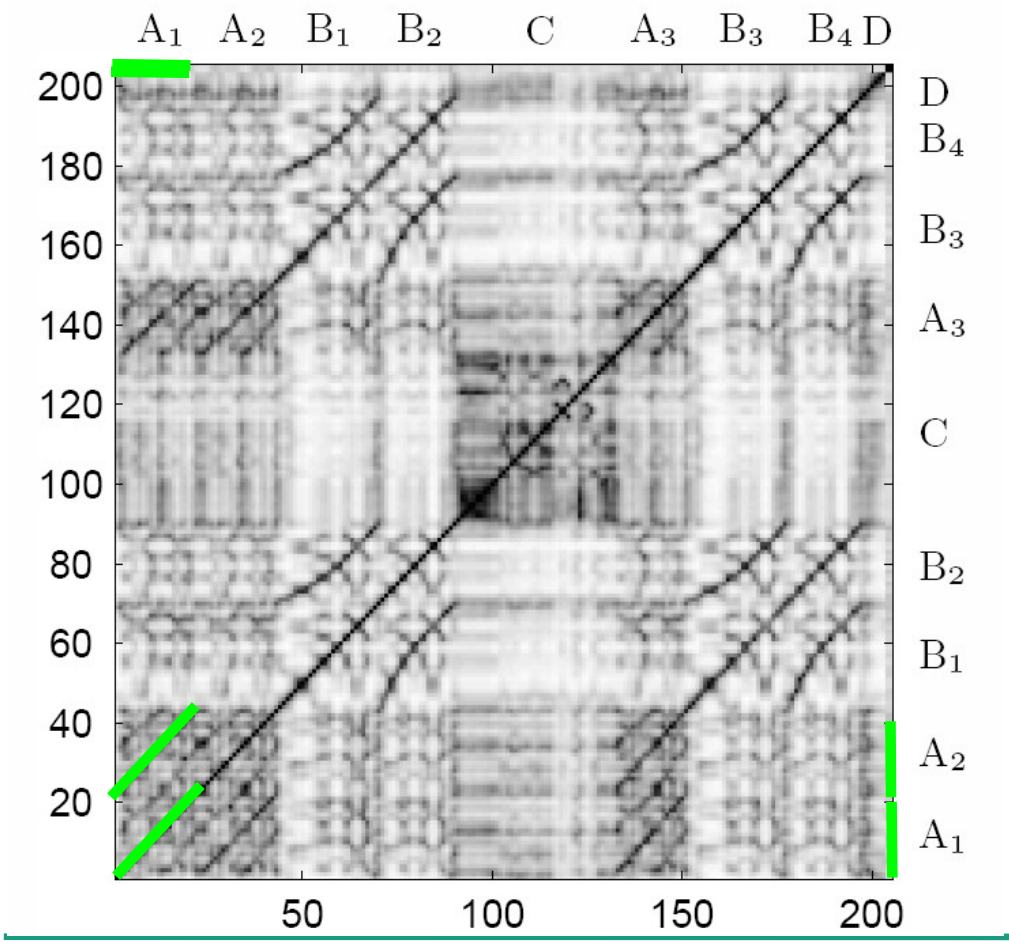


Approaches
Repetition example

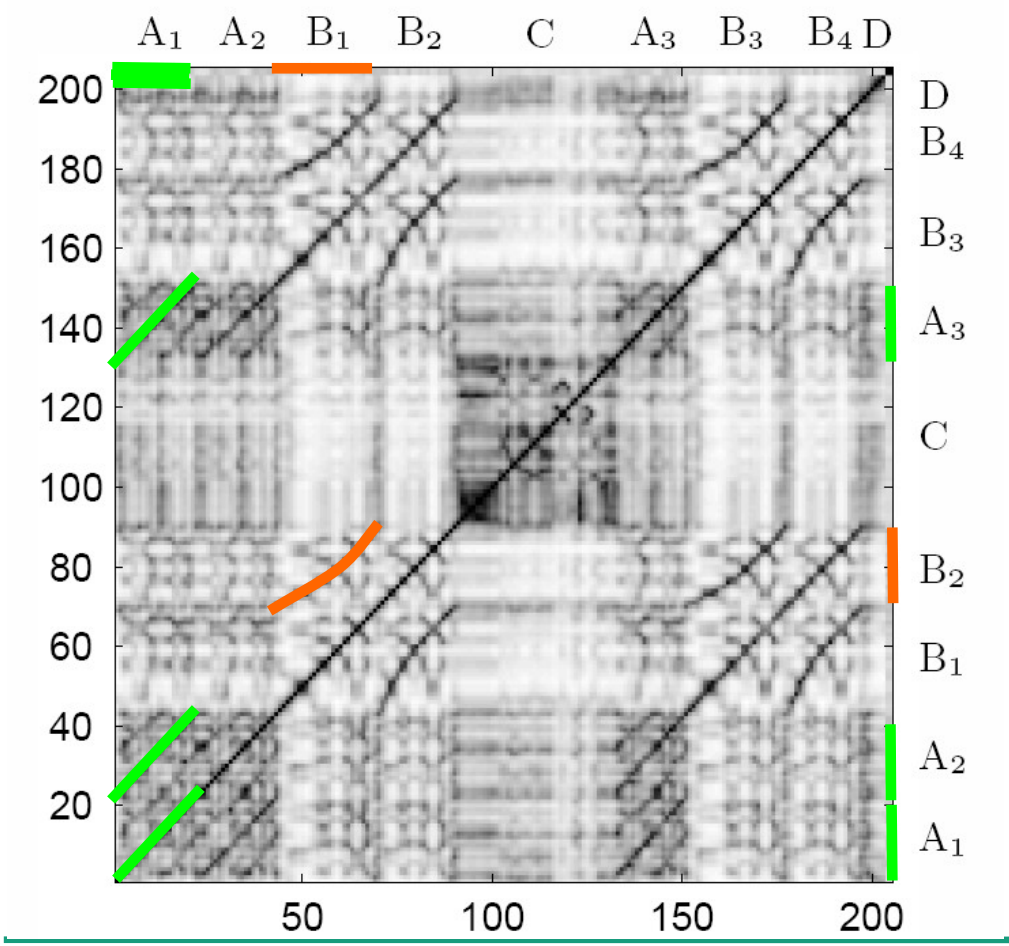


Approaches

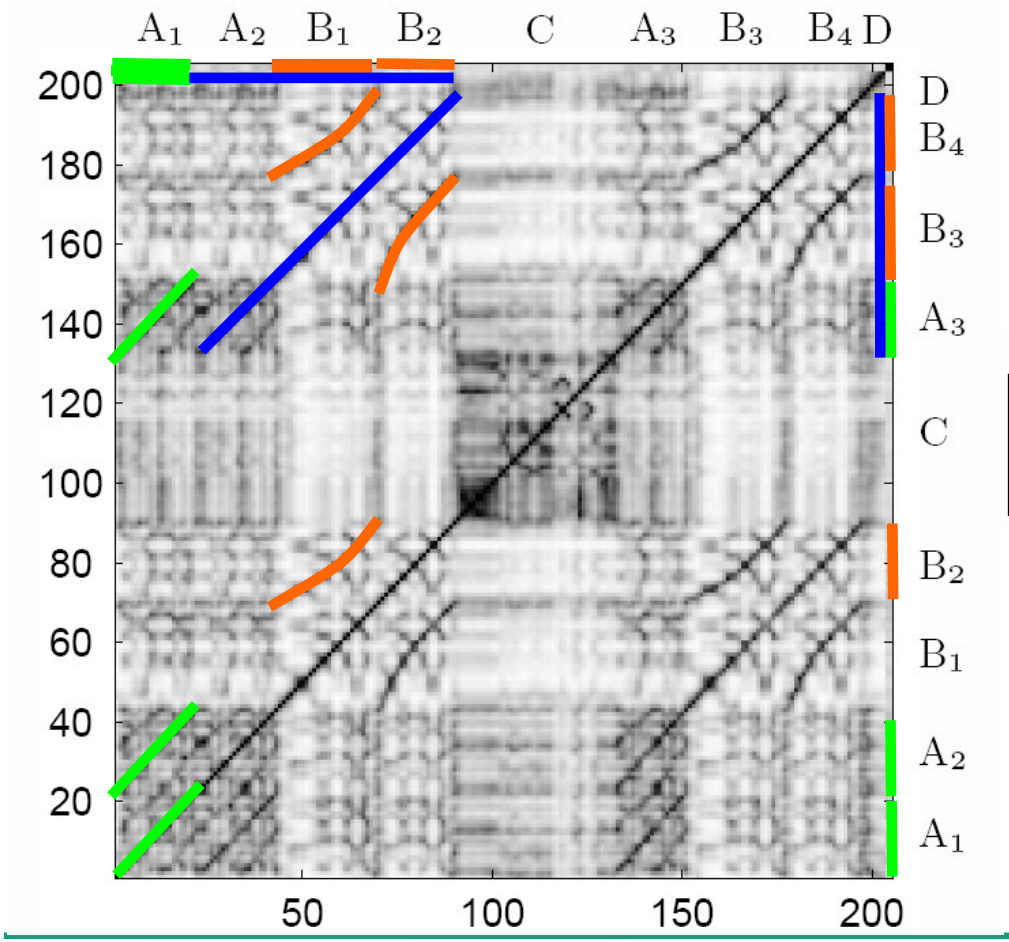
Repetition example

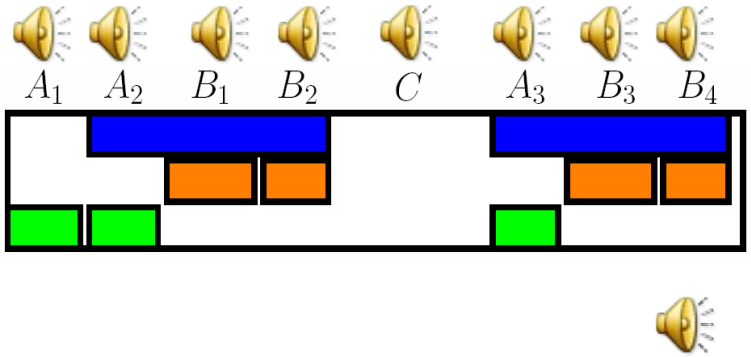


Approaches
Repetition example



Approaches
Repetition example





Approaches
Combined

- For humans, structure is multidimensional

- A combined approach utilises multiple acoustic features simultaneously
  - Timbral, pitched, and rhythmic aspects

- Combining homogeneity (e.g., instrumentation) and sequential repetition (e.g., melodies) in segment clustering
  - Segmentation not fixed, but deduced along clustering
  - Novelty used to limit the search space

Approaches

Methods discussed in paper

Author / publication	Task	Acoustic features	Approach	Method
Aucouturier et al. [4]	full structure	spectral envelope	homogeneity	HMM
Barrington et al. [7]	full structure	MFCC / chroma	homogeneity	dynamic texture model
Bartsch & Wakefield [8]	thumbnailing	chroma	repetition	stripe detection
Chai [13]	full structure	chroma	repetition	stripe detection
Cooper & Foote [15]	summarisation	magnitude spectrum	homogeneity	segment clustering
Dannenberg & Hu [17]	repetitions	chroma	repetition	dynamic programming
Eronen [23]	chorus detection	MFCC+chroma	repetition	stripe detection
Foote [24]	visualization	MFCC		self-similarity matrix
Foote [25]	segmentation	MFCC	novelty	novelty vector
Goto [31]	repetitions	chroma	repetition	stripe detection (*RefraiD*)
Jehan [36]	pattern learning	MFCC+chroma+loudness	homogeneity	hierarchical SDMs
Jensen [38]	segmentation	MFCC+chroma+rhythmogram	novelty	diagonal blocks
Levy & Sandler [41]	full structure	MPEG-7 timbre descriptor	homogeneity	temporal clustering
Logan & Chu [43]	key phrase	MFCC	homogeneity	HMM / clustering
Lu et al. [44]	thumbnailing	constant-Q spectrum	repetition	stripe detection
Maddage [46]	full structure	chroma	homogeneity	rule-based reasoning
Marolt [48]	thumbnailing	chroma	repetition	RefraiD
Mauch et al. [50]	full structure	chroma	repetition	greedy selection
Müller & Kurth [56]	multiple repetitions	chroma statistics	repetition	stripe search & clustering
Ong [57]	full structure	multiple	repetition	RefraiD
Paulus & Klapuri [59]	repeated parts	MFCC+chroma	repetition	cost function
Paulus & Klapuri [62]	full description	MFCC+chroma+rhythmogram	combined	fitness function
Peeters [63]	full structure	dynamic features	homogeneity	HMM, image filtering
Peeters [64]	repeated parts	MFCC+chroma+spec. contrast	repetition	stripe detection
Rhodes & Casey [70]	hierarchical structure	timbral features	repetition	string matching
Shiu et al. [72]	full structure	chroma	repetition	state model stripe detection
Turnbull et al. [75]	segmentation	various	novelty	various
Wellhausen & Höynck [78]	thumbnailing	MPEG-7 timbre descriptor	repetition	stripe detection

Approaches

Author / publication ed in paper

Author / publication		features	Approach	Method
		envelope	homogeneity	HMM
		chroma	homogeneity	dynamic texture model
			repetition	stripe detection
Chai [13]	full structure	chroma	repetition	stripe detection
Cooper & Foote [15]	summarisation	magnitude spectrum	homogeneity	segment clustering
Dannenberg & Hu [17]	repetitions	chroma	repetition	dynamic programming
Eronen [23]	chorus detection	MFCC+chroma	repetition	stripe detection
Foote [24]	visualization	MFCC		self-similarity matrix
Foote [25]	segmentation	MFCC	novelty	novelty vector
Goto [31]	repetitions	chroma	repetition	stripe detection (*RefraiD*)
Jehan [36]	pattern learning	MFCC+chroma+loudness	homogeneity	hierarchical SDMs
Jensen [38]	segmentation	MFCC+chroma+rhythmogram	novelty	diagonal blocks
Levy & Sandler [41]	full structure	MPEG-7 timbre descriptor	homogeneity	temporal clustering
Logan & Chu [43]	key phrase	MFCC	homogeneity	HMM / clustering
Lu et al. [44]	thumbnailing	constant-Q spectrum	repetition	stripe detection
Maddage [46]	full structure	chroma	homogeneity	rule-based reasoning
Marolt [48]	thumbnailing	chroma	repetition	RefraiD
Mauch et al. [50]	full structure	chroma	repetition	greedy selection
Müller & Kurth [56]	multiple repetitions	chroma statistics	repetition	stripe search & clustering
Ong [57]	full structure	multiple	repetition	RefraiD
Paulus & Klapuri [59]	repeated parts	MFCC+chroma	repetition	cost function
Paulus & Klapuri [62]	full description	MFCC+chroma+rhythmogram	combined	fitness function
Peeters [63]	full structure	dynamic features	homogeneity	HMM, image filtering
Peeters [64]	repeated parts	MFCC+chroma+spec. contrast	repetition	stripe detection
Rhodes & Casey [70]	hierarchical structure	timbral features	repetition	string matching
Shiu et al. [72]	full structure	chroma	repetition	state model stripe detection
Turnbull et al. [75]	segmentation	various	novelty	various
Wellhausen & Höynck [78]	thumbnailing	MPEG-7 timbre descriptor	repetition	stripe detection

Approaches

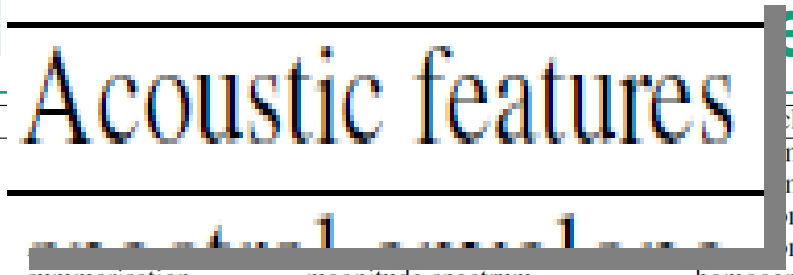
Metho[illegible]ssed in paper

Task

Author / publication	[illegible]	[illegible]coustic features	Approach	Method
Aucouturier et al. [4]		[illegible]ectral envelope	homogeneity	HMM
Barrington et al. [7]		[illegible]FCC / chroma	homogeneity	dynamic texture model
Bartsch & Wakefield [8]		[illegible]roma	repetition	stripe detection
Chai [13]	full structure	chroma	repetition	stripe detection
Cooper & Foote [15]	summarisation	magnitude spectrum	homogeneity	segment clustering
Dannenberg & Hu [17]	repetitions	chroma	repetition	dynamic programming
Eronen [23]	chorus detection	MFCC+chroma	repetition	stripe detection
Foote [24]	visualization	MFCC		self-similarity matrix
Foote [25]	segmentation	MFCC	novelty	novelty vector
Goto [31]	repetitions	chroma	repetition	stripe detection (*RefraiD*)
Jehan [36]	pattern learning	MFCC+chroma+loudness	homogeneity	hierarchical SDMs
Jensen [38]	segmentation	MFCC+chroma+rhythmogram	novelty	diagonal blocks
Levy & Sandler [41]	full structure	MPEG-7 timbre descriptor	homogeneity	temporal clustering
Logan & Chu [43]	key phrase	MFCC	homogeneity	HMM / clustering
Lu et al. [44]	thumbnailing	constant-Q spectrum	repetition	stripe detection
Maddage [46]	full structure	chroma	homogeneity	rule-based reasoning
Marolt [48]	thumbnailing	chroma	repetition	RefraiD
Mauch et al. [50]	full structure	chroma	repetition	greedy selection
Müller & Kurth [56]	multiple repetitions	chroma statistics	repetition	stripe search & clustering
Ong [57]	full structure	multiple	repetition	RefraiD
Paulus & Klapuri [59]	repeated parts	MFCC+chroma	repetition	cost function
Paulus & Klapuri [62]	full description	MFCC+chroma+rhythmogram	combined	fitness function
Peeters [63]	full structure	dynamic features	homogeneity	HMM, image filtering
Peeters [64]	repeated parts	MFCC+chroma+spec. contrast	repetition	stripe detection
Rhodes & Casey [70]	hierarchical structure	timbral features	repetition	string matching
Shiu et al. [72]	full structure	chroma	repetition	state model stripe detection
Turnbull et al. [75]	segmentation	various	novelty	various
Wellhausen & Höynck [78]	thumbnailing	MPEG-7 timbre descriptor	repetition	stripe detection

Approaches

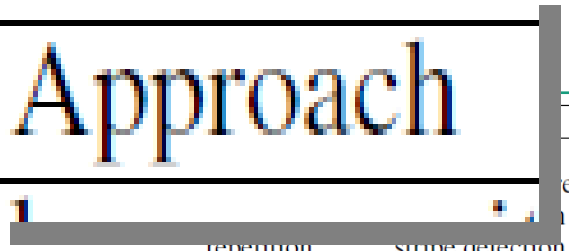
Method



Author / publication			ch	Method
Aucouturier et al. [4]			neity	HMM
Barrington et al. [7]			neity	dynamic texture model
Bartsch & Wakefield [8]			n	stripe detection
Chai [13]			n	stripe detection
Cooper & Foote [15]	summarisation	magnitude spectrum	homogeneity	segment clustering
Dannenberg & Hu [17]	repetitions	chroma	repetition	dynamic programming
Eronen [23]	chorus detection	MFCC+chroma	repetition	stripe detection
Foote [24]	visualization	MFCC		self-similarity matrix
Foote [25]	segmentation	MFCC	novelty	novelty vector
Goto [31]	repetitions	chroma	repetition	stripe detection (*RefraiD*)
Jehan [36]	pattern learning	MFCC+chroma+loudness	homogeneity	hierarchical SDMs
Jensen [38]	segmentation	MFCC+chroma+rhythmogram	novelty	diagonal blocks
Levy & Sandler [41]	full structure	MPEG-7 timbre descriptor	homogeneity	temporal clustering
Logan & Chu [43]	key phrase	MFCC	homogeneity	HMM / clustering
Lu et al. [44]	thumbnailing	constant-Q spectrum	repetition	stripe detection
Maddage [46]	full structure	chroma	homogeneity	rule-based reasoning
Marolt [48]	thumbnailing	chroma	repetition	RefraiD
Mauch et al. [50]	full structure	chroma	repetition	greedy selection
Müller & Kurth [56]	multiple repetitions	chroma statistics	repetition	stripe search & clustering
Ong [57]	full structure	multiple	repetition	RefraiD
Paulus & Klapuri [59]	repeated parts	MFCC+chroma	repetition	cost function
Paulus & Klapuri [62]	full description	MFCC+chroma+rhythmogram	combined	fitness function
Peeters [63]	full structure	dynamic features	homogeneity	HMM, image filtering
Peeters [64]	repeated parts	MFCC+chroma+spec. contrast	repetition	stripe detection
Rhodes & Casey [70]	hierarchical structure	timbral features	repetition	string matching
Shiu et al. [72]	full structure	chroma	repetition	state model stripe detection
Turnbull et al. [75]	segmentation	various	novelty	various
Wellhausen & Höynck [78]	thumbnailing	MPEG-7 timbre descriptor	repetition	stripe detection

Approaches

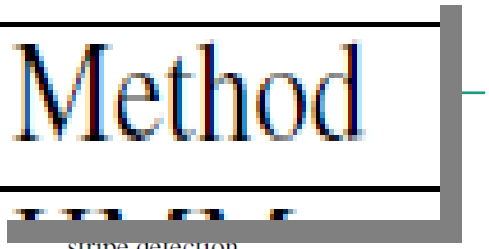
Methods discussed



Author / publication	Task	Acoustic featur		
Aucouturier et al. [4]	full structure	spectral envelo		
Barrington et al. [7]	full structure	MFCC / chrom		e model
Bartsch & Wakefield [8]	thumbnailing	chroma		
Chai [13]	full structure	chroma	repetition	stripe detection
Cooper & Foote [15]	summarisation	magnitude spectrum	homogeneity	segment clustering
Dannenberg & Hu [17]	repetitions	chroma	repetition	dynamic programming
Eronen [23]	chorus detection	MFCC+chroma	repetition	stripe detection
Foote [24]	visualization	MFCC		self-similarity matrix
Foote [25]	segmentation	MFCC	novelty	novelty vector
Goto [31]	repetitions	chroma	repetition	stripe detection (*RefraiD*)
Jehan [36]	pattern learning	MFCC+chroma+loudness	homogeneity	hierarchical SDMs
Jensen [38]	segmentation	MFCC+chroma+rhythmogram	novelty	diagonal blocks
Levy & Sandler [41]	full structure	MPEG-7 timbre descriptor	homogeneity	temporal clustering
Logan & Chu [43]	key phrase	MFCC	homogeneity	HMM / clustering
Lu et al. [44]	thumbnailing	constant-Q spectrum	repetition	stripe detection
Maddage [46]	full structure	chroma	homogeneity	rule-based reasoning
Marolt [48]	thumbnailing	chroma	repetition	RefraiD
Mauch et al. [50]	full structure	chroma	repetition	greedy selection
Müller & Kurth [56]	multiple repetitions	chroma statistics	repetition	stripe search & clustering
Ong [57]	full structure	multiple	repetition	RefraiD
Paulus & Klapuri [59]	repeated parts	MFCC+chroma	repetition	cost function
Paulus & Klapuri [62]	full description	MFCC+chroma+rhythmogram	combined	fitness function
Peeters [63]	full structure	dynamic features	homogeneity	HMM, image filtering
Peeters [64]	repeated parts	MFCC+chroma+spec. contrast	repetition	stripe detection
Rhodes & Casey [70]	hierarchical structure	timbral features	repetition	string matching
Shiu et al. [72]	full structure	chroma	repetition	state model stripe detection
Turnbull et al. [75]	segmentation	various	novelty	various
Wellhausen & Höynck [78]	thumbnailing	MPEG-7 timbre descriptor	repetition	stripe detection

Approaches

Methods discussed in pape



Author / publication	Task	Acoustic features	Approach	
Aucouturier et al. [4]	full structure	spectral envelope	homogen	
Barrington et al. [7]	full structure	MFCC / chroma	homogen	
Bartsch & Wakefield [8]	thumbnailing	chroma	repetition	
Chai [13]	full structure	chroma	repetition	stripe detection
Cooper & Foote [15]	summarisation	magnitude spectrum	homogeneity	segment clustering
Dannenberg & Hu [17]	repetitions	chroma	repetition	dynamic programming
Eronen [23]	chorus detection	MFCC+chroma	repetition	stripe detection
Foote [24]	visualization	MFCC		self-similarity matrix
Foote [25]	segmentation	MFCC	novelty	novelty vector
Goto [31]	repetitions	chroma	repetition	stripe detection (*RefraiD*)
Jehan [36]	pattern learning	MFCC+chroma+loudness	homogeneity	hierarchical SDMs
Jensen [38]	segmentation	MFCC+chroma+rhythmogram	novelty	diagonal blocks
Levy & Sandler [41]	full structure	MPEG-7 timbre descriptor	homogeneity	temporal clustering
Logan & Chu [43]	key phrase	MFCC	homogeneity	HMM / clustering
Lu et al. [44]	thumbnailing	constant-Q spectrum	repetition	stripe detection
Maddage [46]	full structure	chroma	homogeneity	rule-based reasoning
Marolt [48]	thumbnailing	chroma	repetition	RefraiD
Mauch et al. [50]	full structure	chroma	repetition	greedy selection
Müller & Kurth [56]	multiple repetitions	chroma statistics	repetition	stripe search & clustering
Ong [57]	full structure	multiple	repetition	RefraiD
Paulus & Klapuri [59]	repeated parts	MFCC+chroma	repetition	cost function
Paulus & Klapuri [62]	full description	MFCC+chroma+rhythmogram	combined	fitness function
Peeters [63]	full structure	dynamic features	homogeneity	HMM, image filtering
Peeters [64]	repeated parts	MFCC+chroma+spec. contrast	repetition	stripe detection
Rhodes & Casey [70]	hierarchical structure	timbral features	repetition	string matching
Shiu et al. [72]	full structure	chroma	repetition	state model stripe detection
Turnbull et al. [75]	segmentation	various	novelty	various
Wellhausen & Höynck [78]	thumbnailing	MPEG-7 timbre descriptor	repetition	stripe detection

Evaluation
Difficult

- Problem is not accurately defined
  - People have different opinions what the structure of a piece is (e.g., temporal level, similarity between segments)
  - Not all pieces fit to this model
- Evaluation metrics problematic
  - Small differences (similar to between-humans) should be accepted
  - Currently measures adopted from clustering field

Evaluation
MIREX (2009)

- First attempt on the task
- Data donated from OMRAS2 meta data project (kudos!)
- 5 submissions, including multiple clustering approaches, greedy stripe search, and a combination method
- Multiple evaluation measures, frame pair clustering F-measure used as the ”one number”
  - Winner a repetition search approach, but
  - Differences relatively small (F-measure 53-60%)
  - Different evaluation measure produces different ranking
- This year, new methods (NMF, more clustering)
  - But results quite similar to last year (F-measure 49-61%)

Conclusions & Future work

Difficult problem

- Many different approaches proposed
- No single approach currently clearly the best
  - State of the art still quite poor compared to human performance
- Main problem conceptual: task should be defined more accurately
  - More co-operation between engineers and musicologists
- A large, representative, evaluation data set should be collected
  - With multiple partners
- Address also other music than Western pop
  - Some work on classical done

Still plenty of work ahead