Lecture

Music Processing

Audio Structure Analysis

Meinard Müller
International Audio Laboratories Erlangen
meinard.mueller@audiolabs-erlangen.de

Music Structure Analysis

- Music segmentation
  - pitch content (e.g., melody, harmony)
  - music texture (e.g., timbre, instrumentation, sound)
  - rhythm
- Detection of repeating sections, phrases, motives
  - song structure (e.g., intro, versus, chorus)
  - musical form (e.g., sonata form, rondo form)
- Detection of other hidden relationships

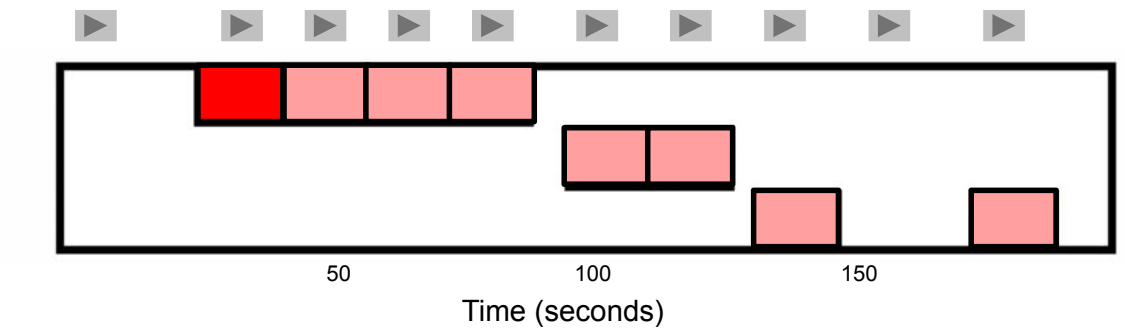
Repetition-Based Audio Structure Analysis

- Extract the repetitive structure of a given audio recording
- Often corresponds to musical form of the underlying piece
- The thumbnail is the most repetitive segment

Repetition-Based Audio Structure Analysis

- Extract the repetitive structure of a given audio recording
- Often corresponds to musical form of the underlying piece
- The thumbnail is the most repetitive segment

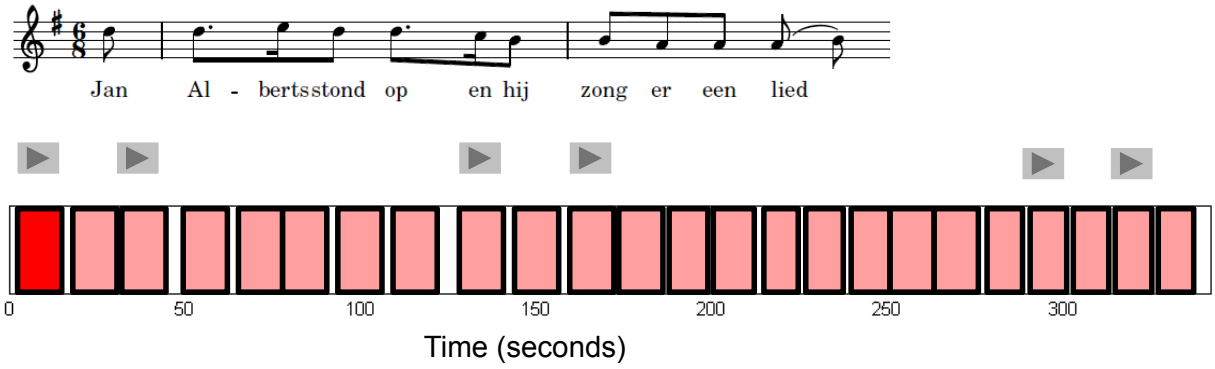
Example: Zager & Evans “In The Year 2525”



Repetition-Based Audio Structure Analysis

- Extract the repetitive structure of a given audio recording
- Often corresponds to musical form of the underlying piece
- The thumbnail is the most repetitive segment

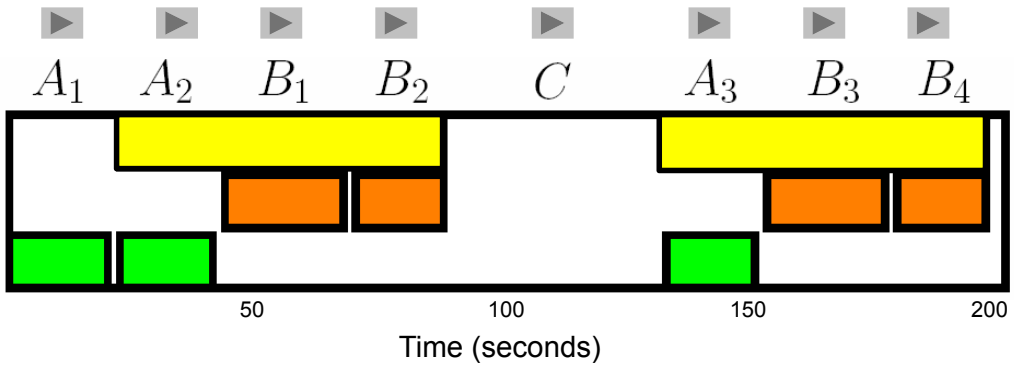
Example: Folk Song Field Recording (Nederlandse Liederenbank)



Repetition-Based Audio Structure Analysis

- Extract the repetitive structure of a given audio recording
- Often corresponds to musical form of the underlying piece
- The thumbnail is the most repetitive segment

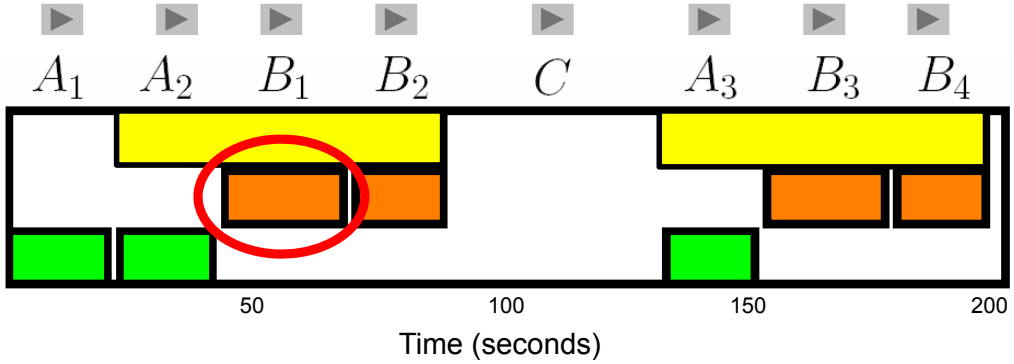
Example: Brahms Hungarian Dance No. 5 (Ormandy)



Repetition-Based Audio Structure Analysis

- Extract the repetitive structure of a given audio recording
- Often corresponds to musical form of the underlying piece
- The thumbnail is the most repetitive segment

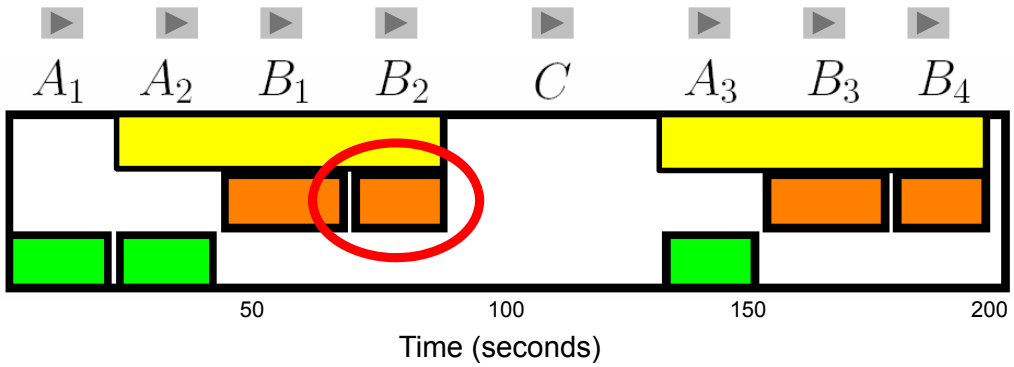
Example: Brahms Hungarian Dance No. 5 (Ormandy)



Repetition-Based Audio Structure Analysis

- Extract the repetitive structure of a given audio recording
- Often corresponds to musical form of the underlying piece
- The thumbnail is the most repetitive segment

Example: Brahms Hungarian Dance No. 5 (Ormandy)



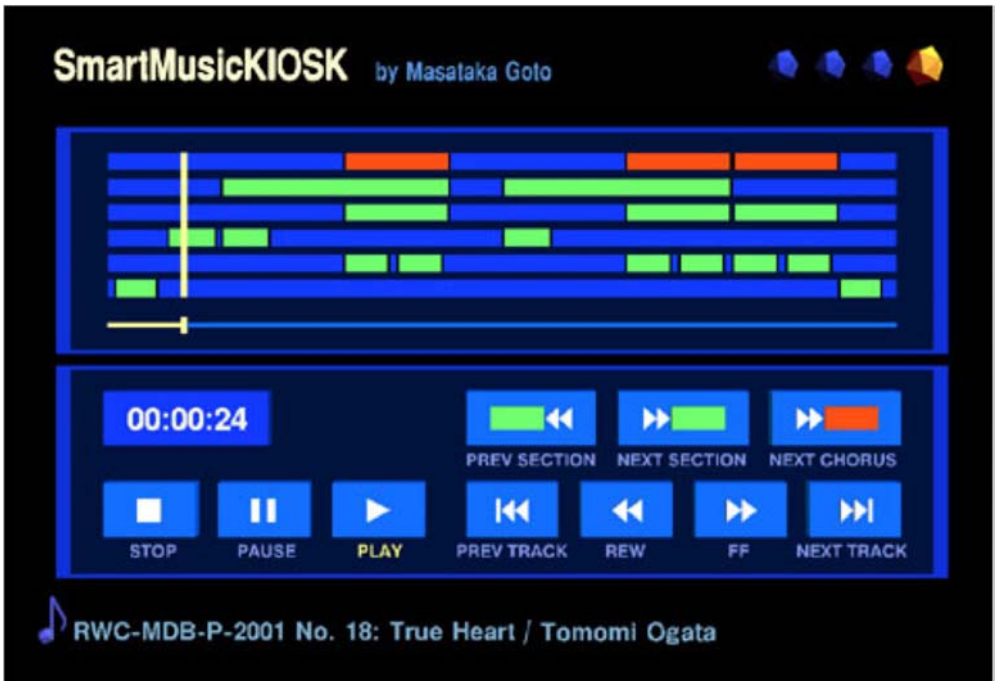
Repetition-Based Audio Structure Analysis

- Extract the repetitive structure of a given audio recording
- Often corresponds to musical form of the underlying piece
- The thumbnail is the most repetitive segment

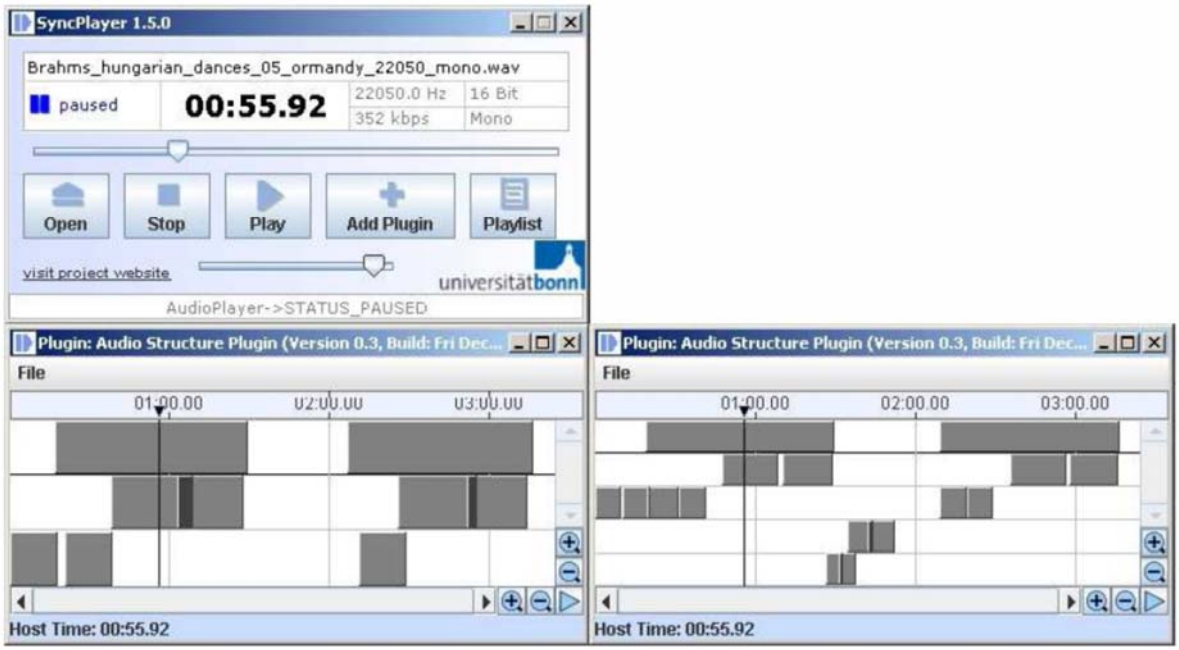
Lots of previous work such as:

- Dannenberg/Hu (ISMIR 2002)
- Peeters/Burthe/Rodet (ISMIR 2002)
- Cooper/Foote (ISMIR 2002)
- Goto (ICASSP 2003)
- Chai/Vercoe (ACM Multimedia 2003)
- Lu/Wang/Zhang (ACM Multimedia 2004)
- Bartsch/Wakefield (IEEE Trans. MM 2005)
- Goto (IEEE Trans. Audio 2006)
- Müller/Kurth (EURASIP 2007)
- Rhodes/Casey (ISMIR 2007)
- Peeters (ISMIR 2007)
- Paulus/Klapuri (IEEE TASLP 2009)
- Paulus/Müller/Klapuri (ISMIR 2010)
- Müller/Grosche/Jiang (ISMIR 2011)
- …

System: SmartMusicKiosk (Goto)



System: SyncPlayer/AudioStructure



Basic Procedure

- Audio features
- Cost measure and cost matrix
  - ⇝ self-similarity matrix
- Path extraction (pairwise similarity of segments)
- Global structure (clustering, grouping)

Basic Procedure

- Audio $\rightsquigarrow$ $V := (v^1, v^2, \ldots, v^N)$
- v^n = 12-dimensional normalized chroma vector
- Local cost measure $c : \mathbb{R}^{12} \times \mathbb{R}^{12} \to \mathbb{R}$

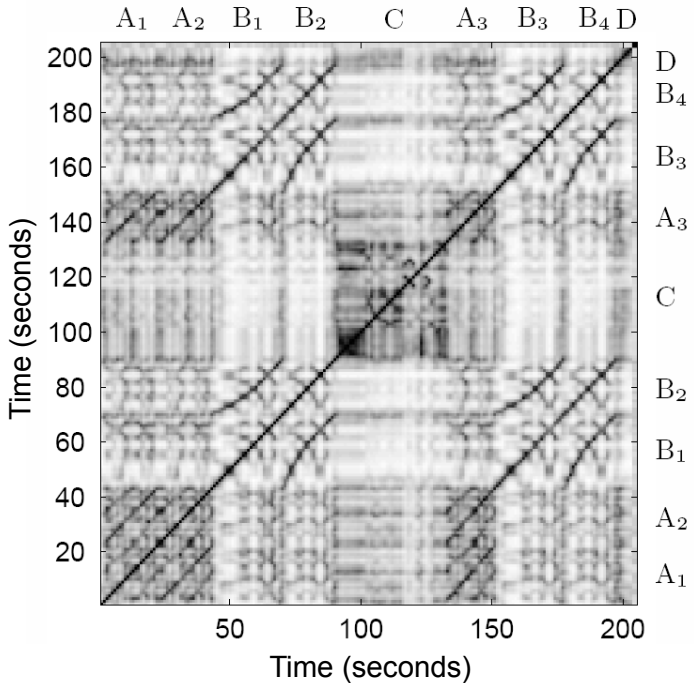
$$c(v^n, w^m) := 1 - \langle v^n, w^m \rangle$$

- $N \times N$ cost matrix $C(n, m) := c(v^n, w^m)$

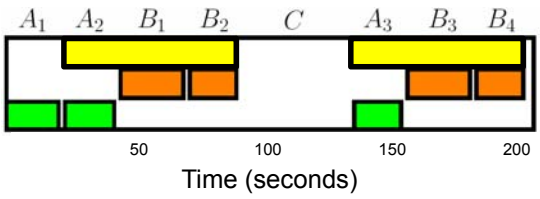
$\rightsquigarrow$ quadratic self-similarity matrix

Basic Procedure

Self-similarity matrix

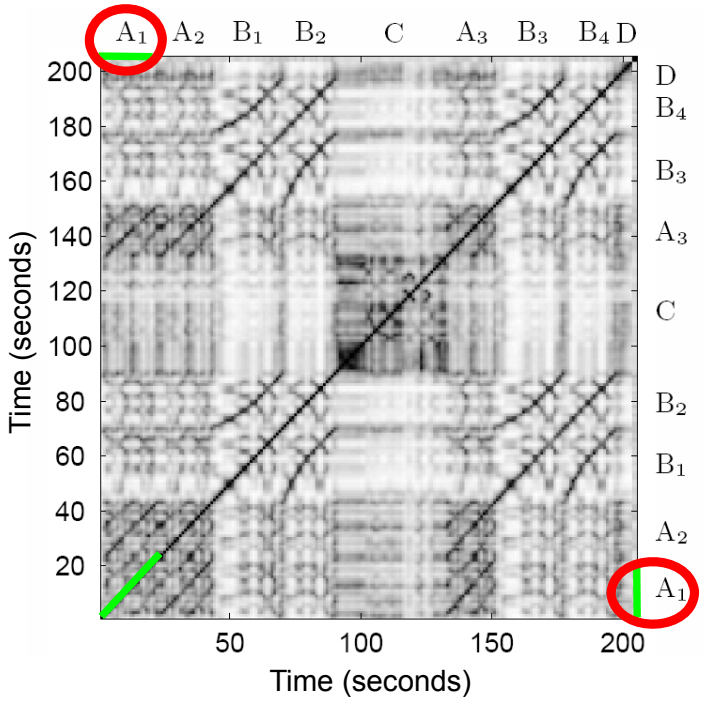


Similarity structure

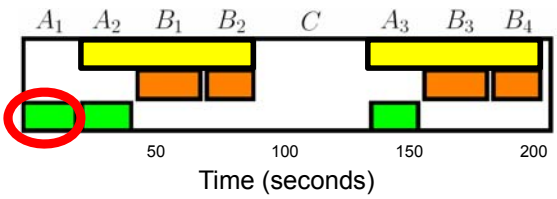


Basic Procedure

Self-similarity matrix

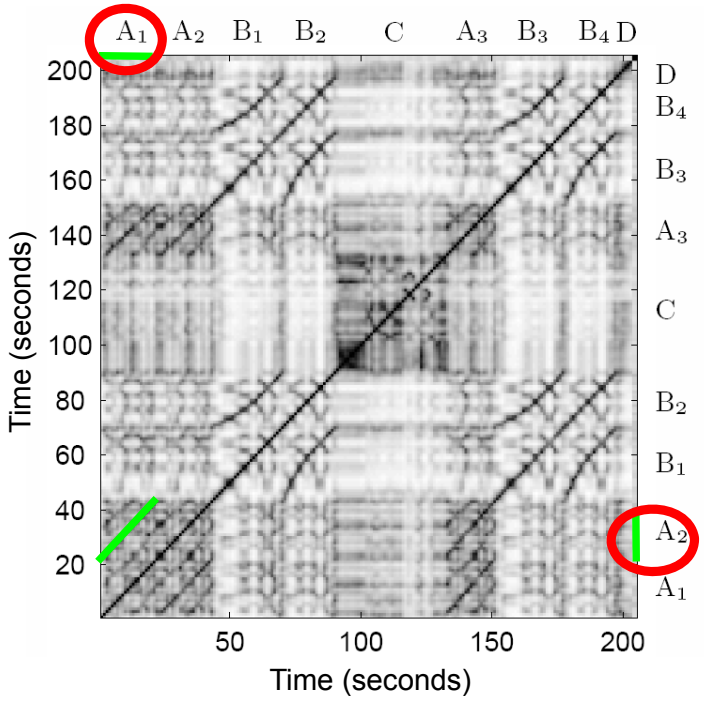


Similarity structure

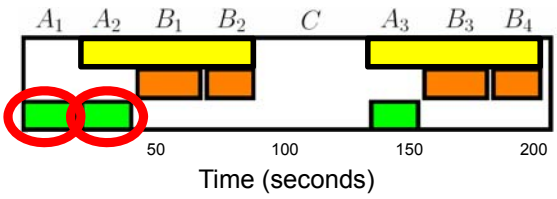


Basic Procedure

Self-similarity matrix

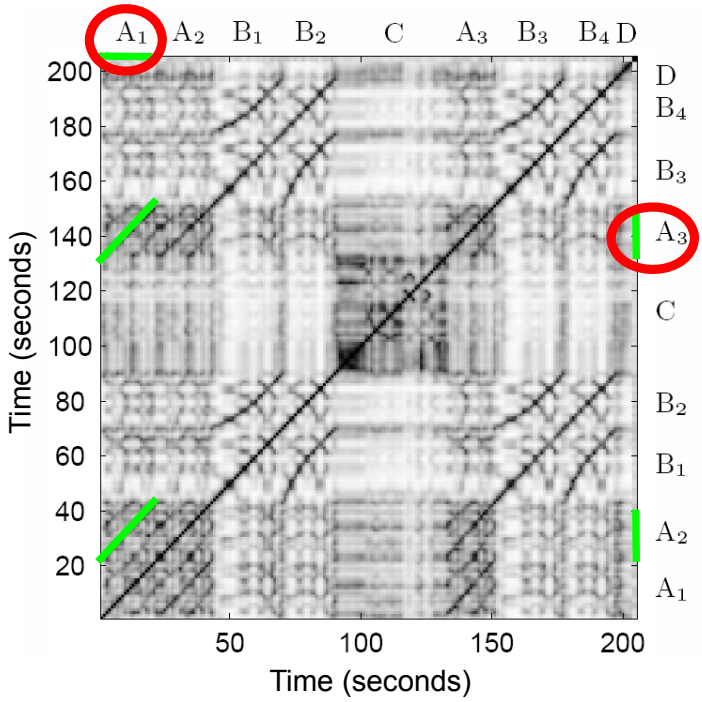


Similarity structure

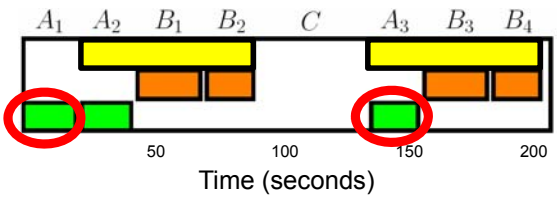


Basic Procedure

Self-similarity matrix

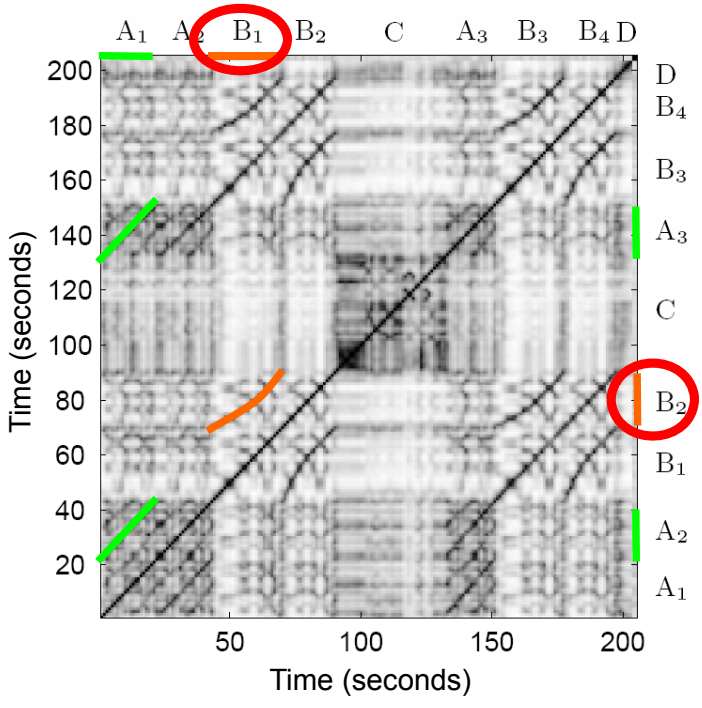


Similarity structure

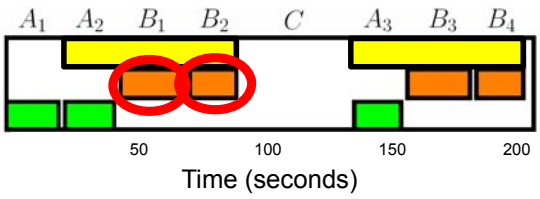


Basic Procedure

Self-similarity matrix

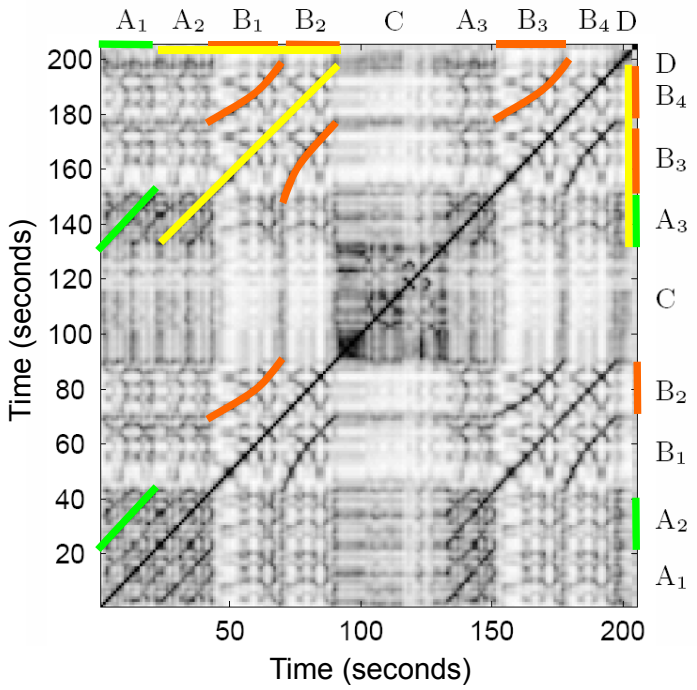


Similarity structure

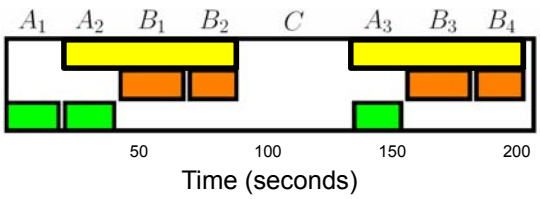


Basic Procedure

Self-similarity matrix

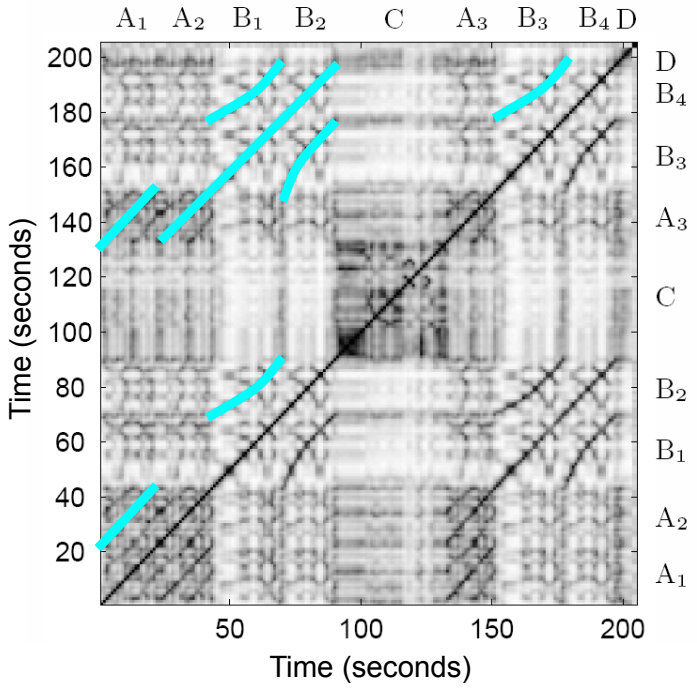


Similarity structure

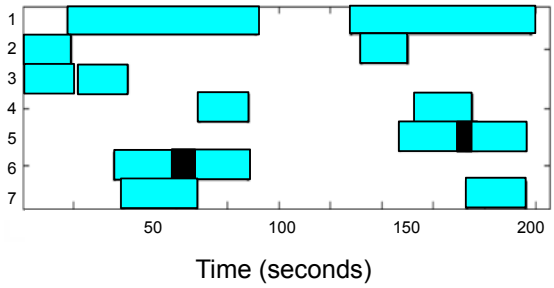


Basic Procedure

Self-similarity matrix

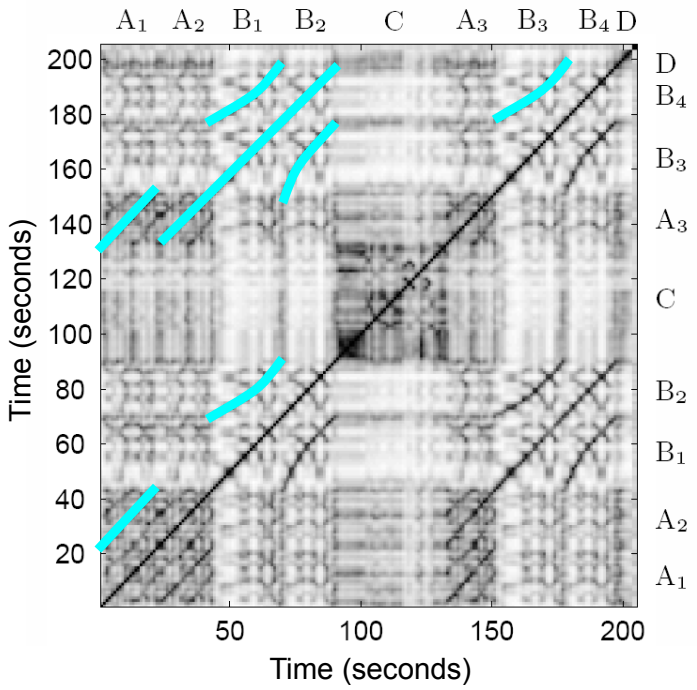


Path relations

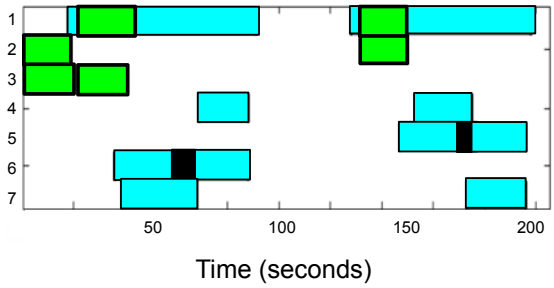


Basic Procedure

Self-similarity matrix



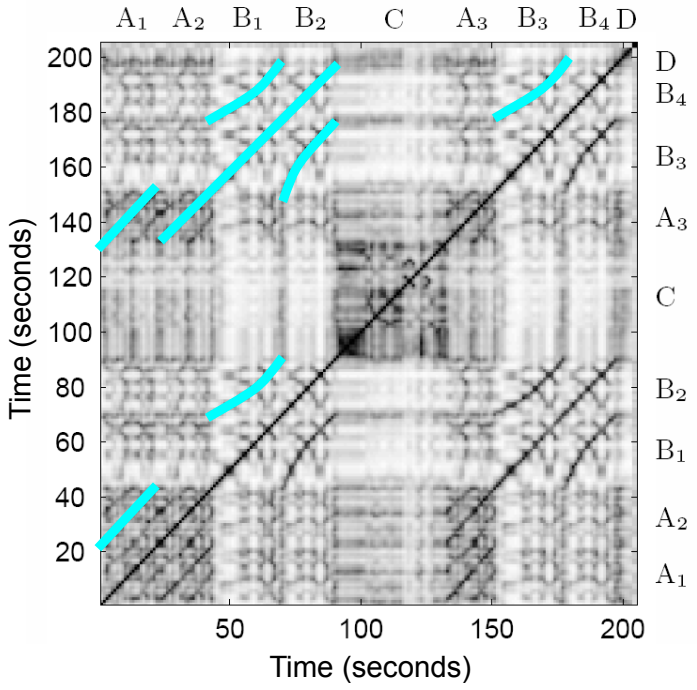
Path relations



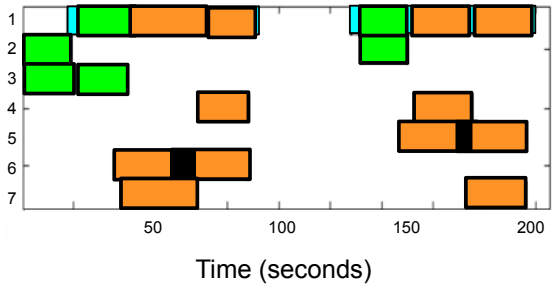
Grouping / Transitivity

Basic Procedure

Self-similarity matrix



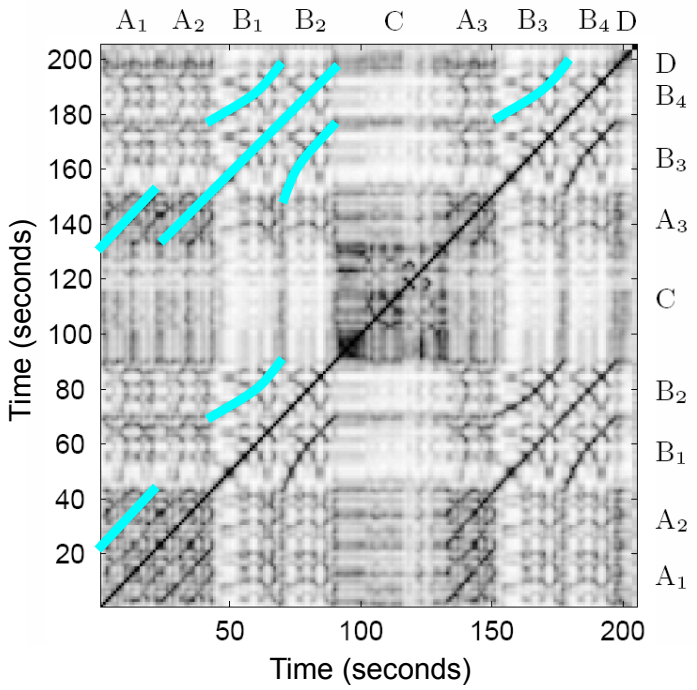
Path relations



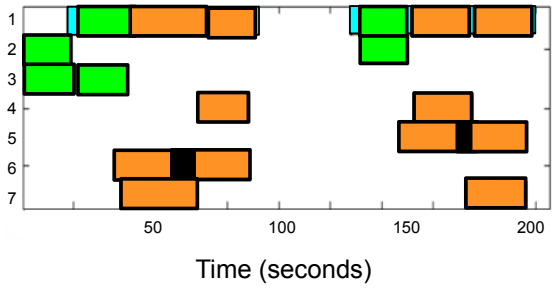
Grouping / Transitivity

Basic Procedure

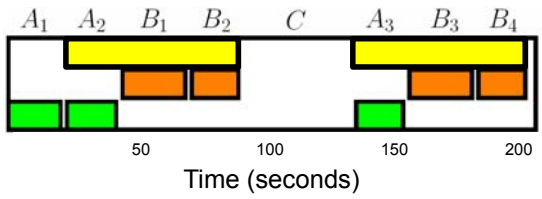
Self-similarity matrix



Path relations



Grouping / Transitivity



Matrix Enhancement

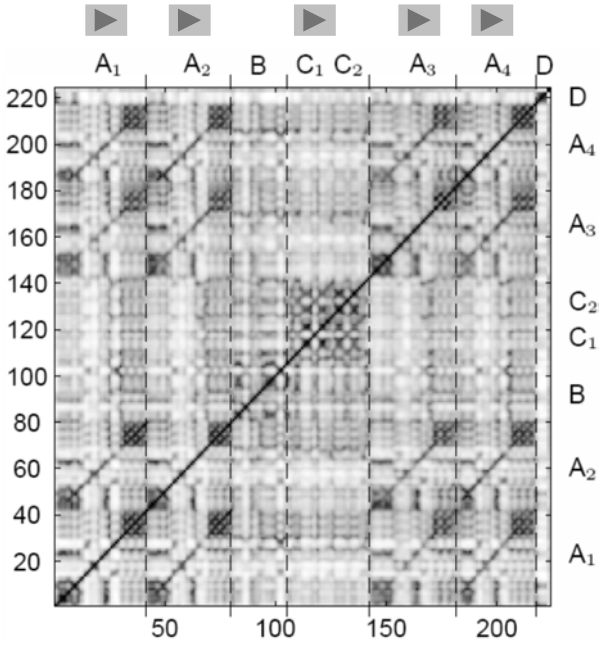
Challenge: Presence of musical variations

- Fragmented paths and gaps
- Paths of poor quality
- Regions of constant (low) cost
- Curved paths

Idea: Enhancement of path structure

Matrix Enhancement

Shostakovich Waltz 2, Jazz Suite No. 2 (Chailly)



Matrix Enhancement

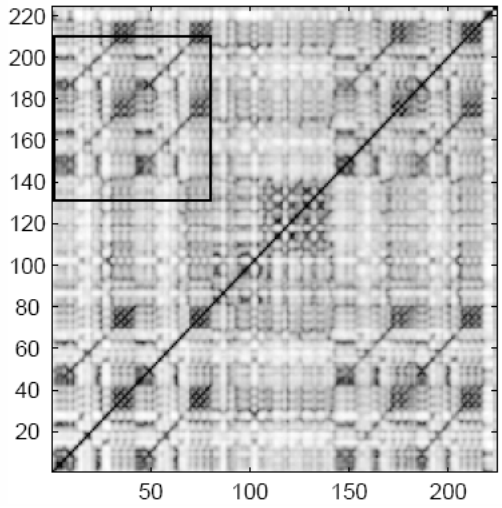
Idea: Usage of contextual information (Foote 1999)

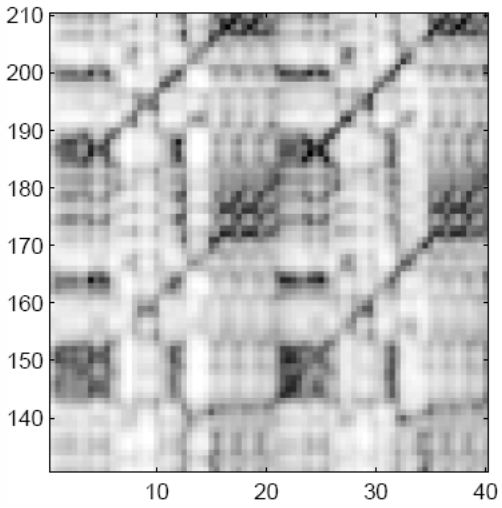
$$C_L(n,m) := \frac{1}{L} \sum_{\ell=0}^{L-1} c(v_{n+\ell}, v_{m+\ell})$$

- Comparison of entire sequences
- $L =$ length of sequences
- $C_L =$ enhanced cost matrix

$\rightsquigarrow$ smoothing effect

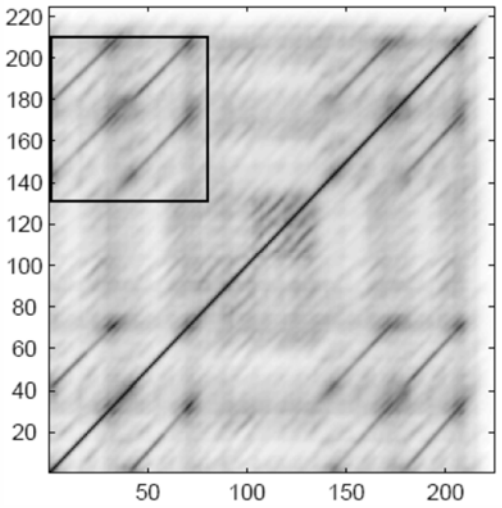
Matrix Enhancement (Shostakovich)

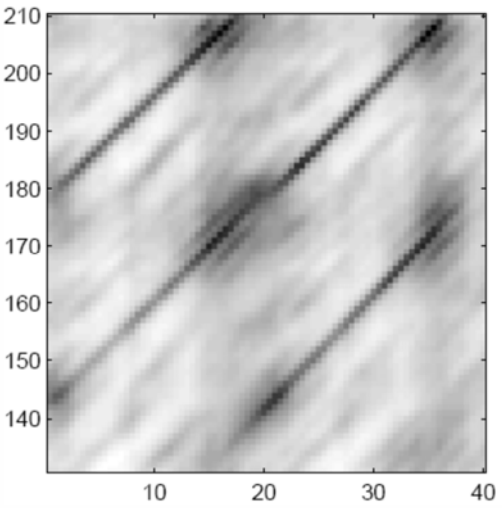




Cost matrix C

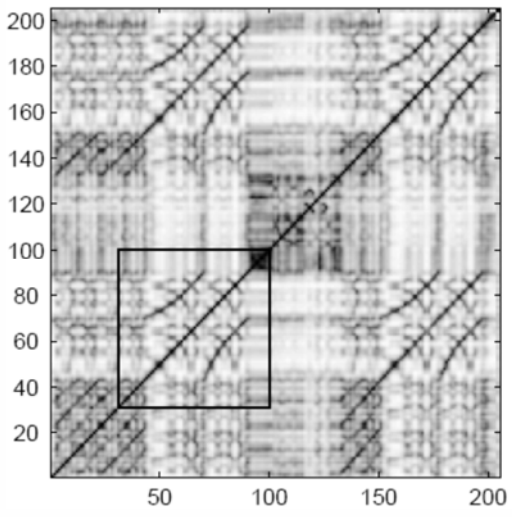
Matrix Enhancement (Shostakovich)

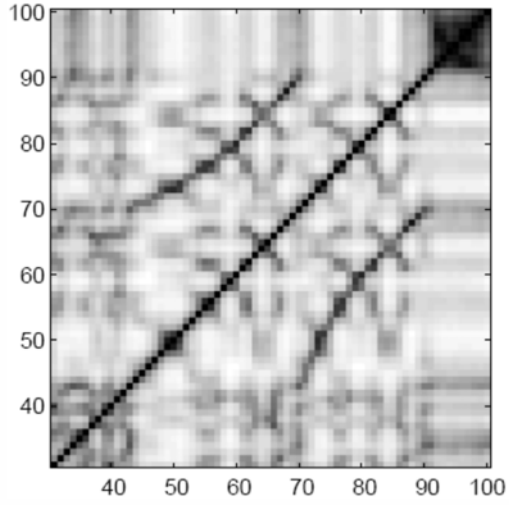




Enhanced cost matrix C_L

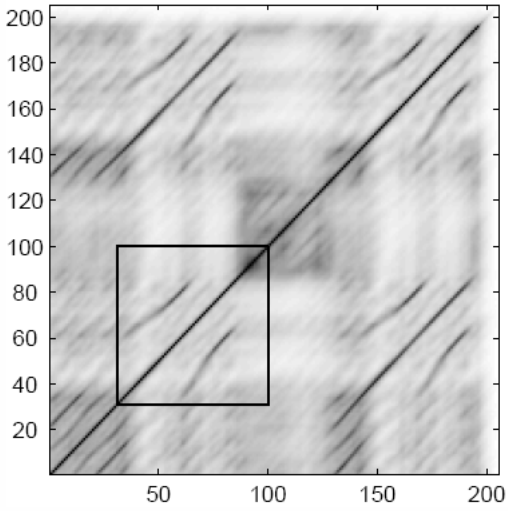
Matrix Enhancement (Brahms)

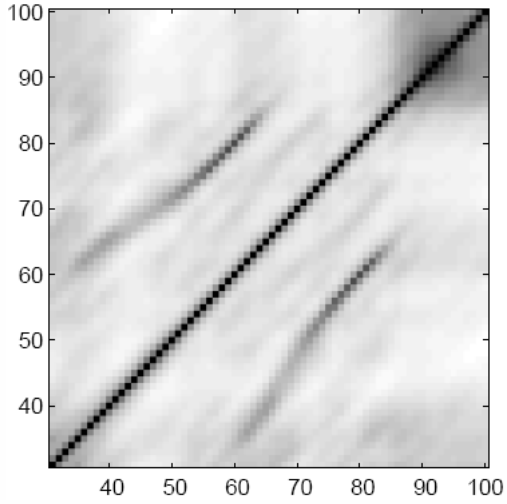




Cost matrix C

Matrix Enhancement (Brahms)





Enhanced cost matrix C_L

Problem: Relative tempo differences are smoothed out

Matrix Enhancement

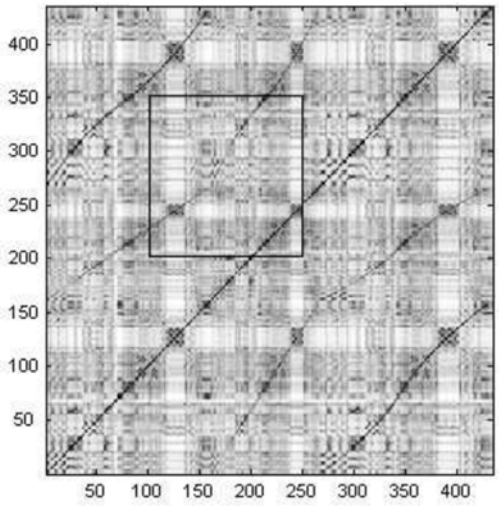
Idea: Smoothing along various directions
and minimizing over all directions

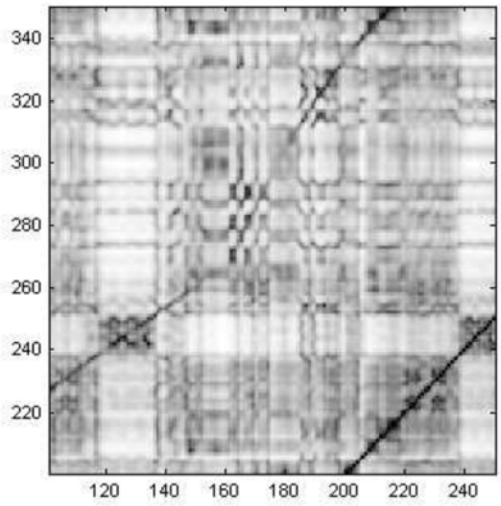
$$C_L^{\min}(n,m) := \min_k C_L^{\text{slope}_k}(n,m)$$

- $\text{slope}_k = k$th direction of smoothing
- $C_L^{\text{slope}_k}$ = enhanced cost matrix w.r.t. slope_k
- Usage of eight slope values

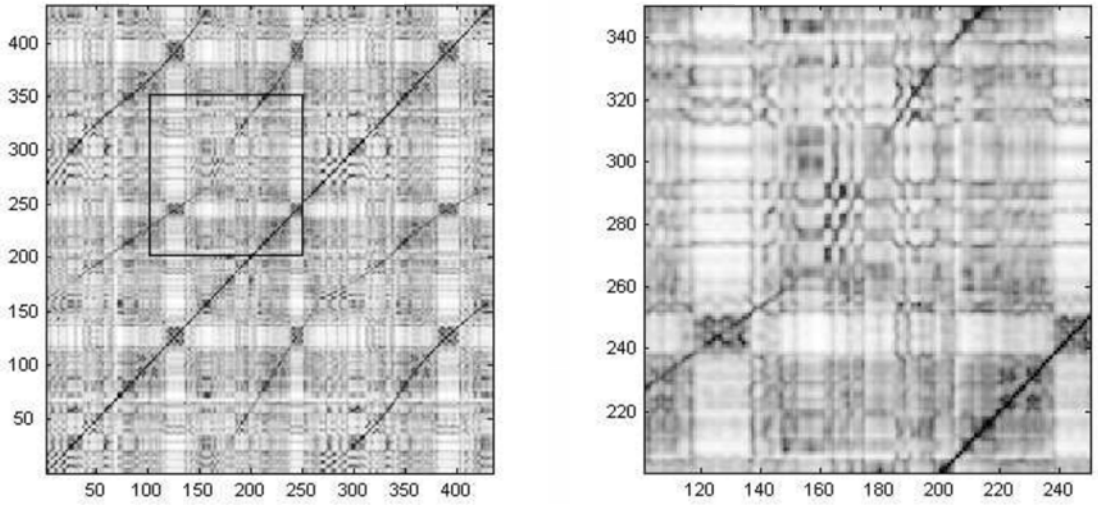
$\rightsquigarrow$ tempo changes of -30 to +40 percent

Matrix Enhancement



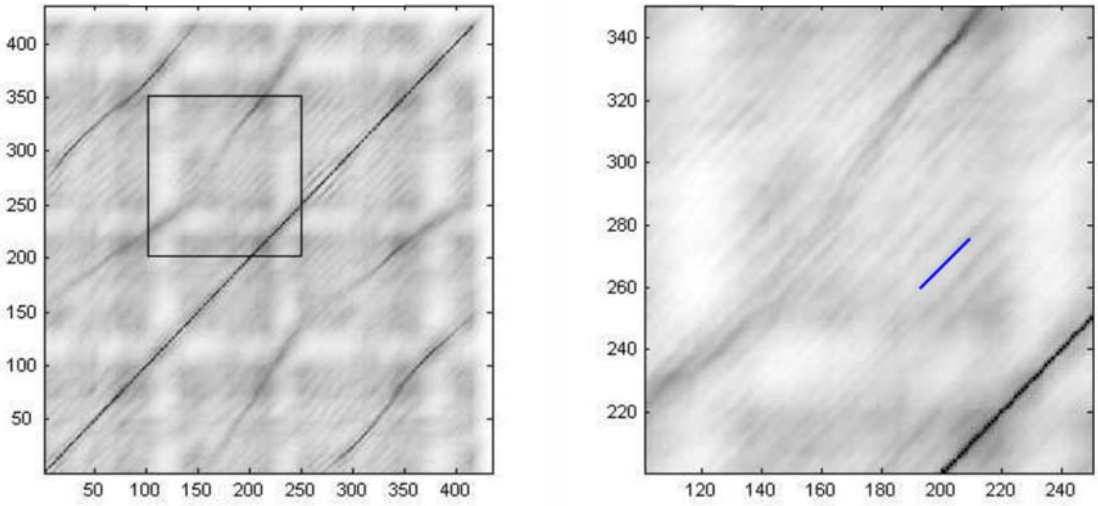


Matrix Enhancement



Cost matrix C

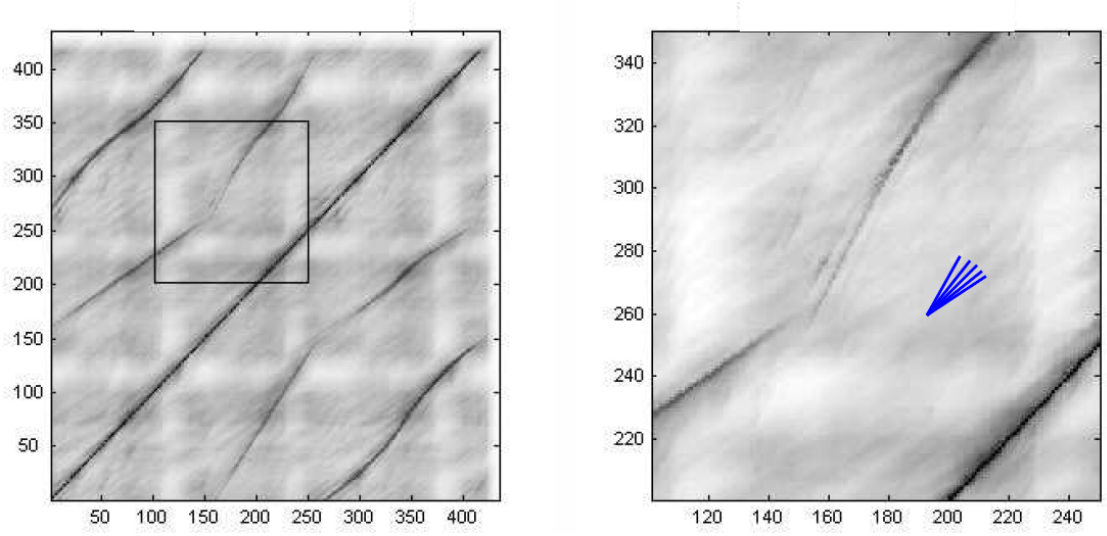
Matrix Enhancement



Cost matrix C_L with $L = 20$

Filtering along main diagonal

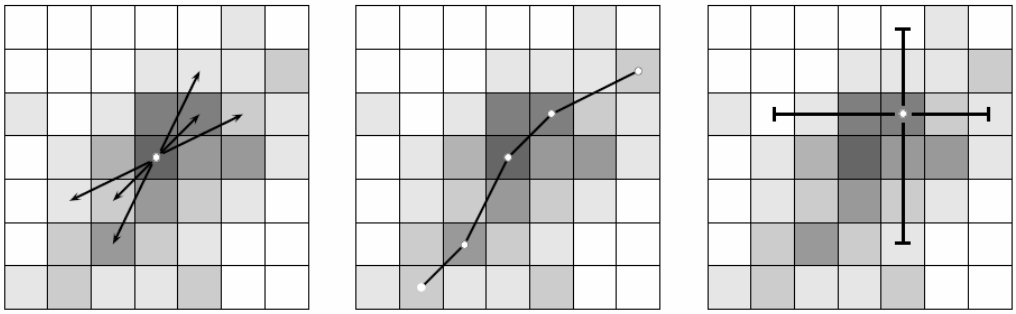
Matrix Enhancement



Cost matrix $C_L^{\min}$ with $L = 20$

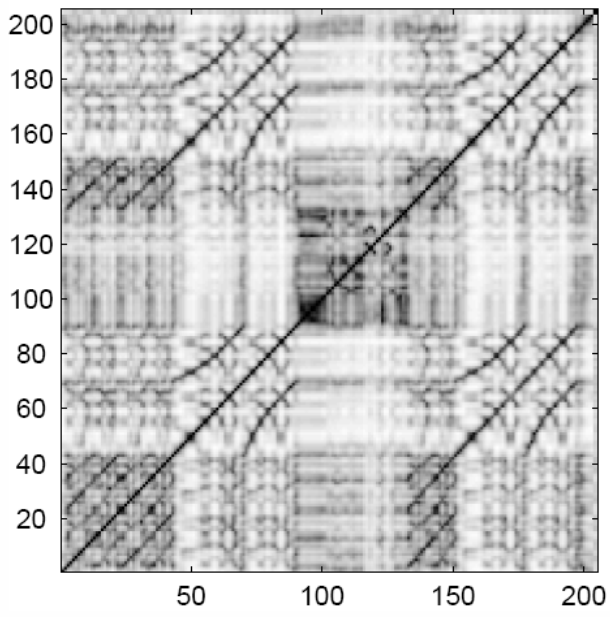
Filtering along 8 different directions and minimizing

Path Extraction



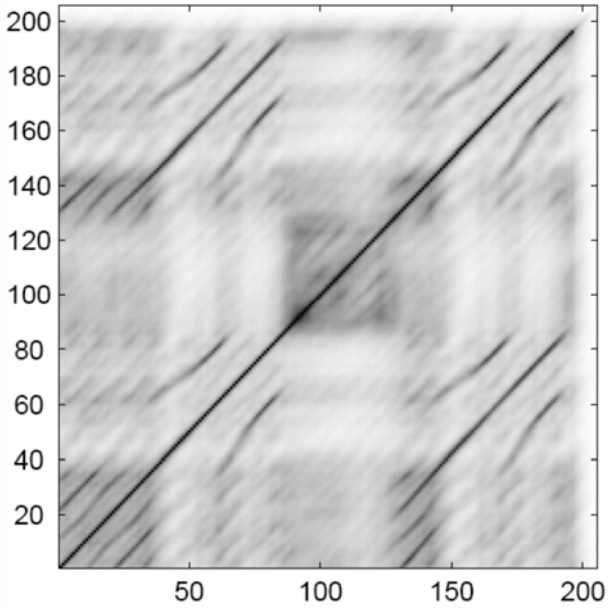
- Start with initial point
- Extend path in greedy fashion
- Remove path neighborhood

Path Extraction



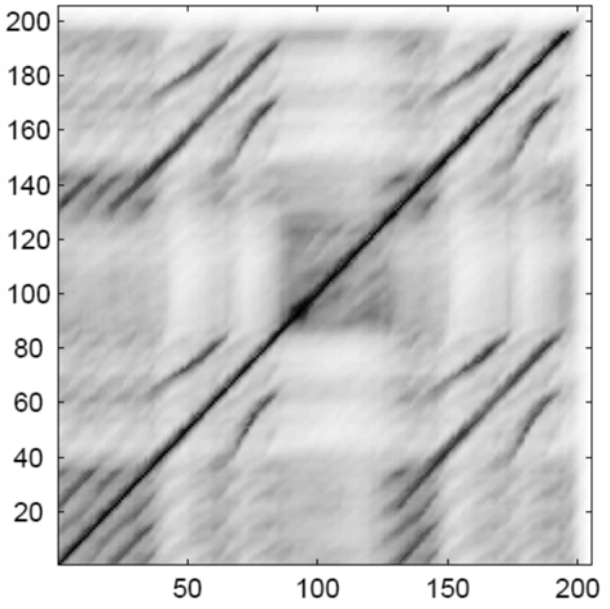
Cost matrix C

Path Extraction



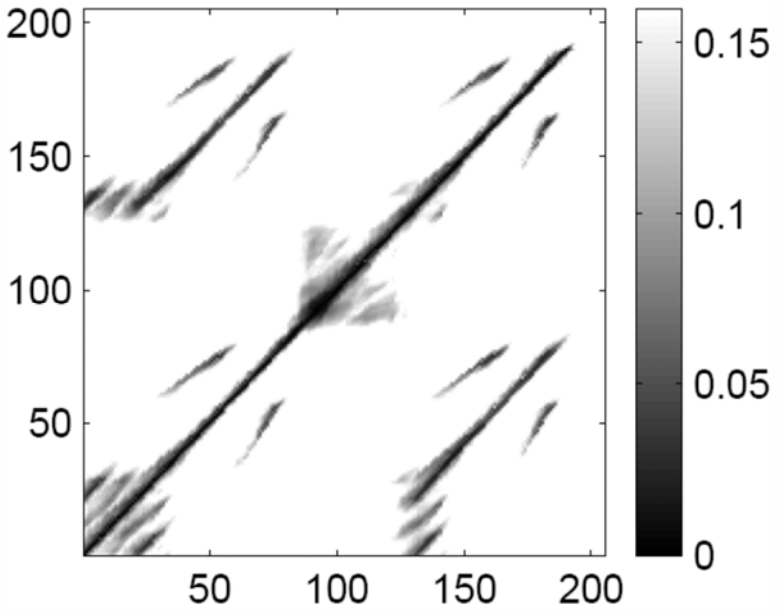
Enhanced cost matrix C_L

Path Extraction



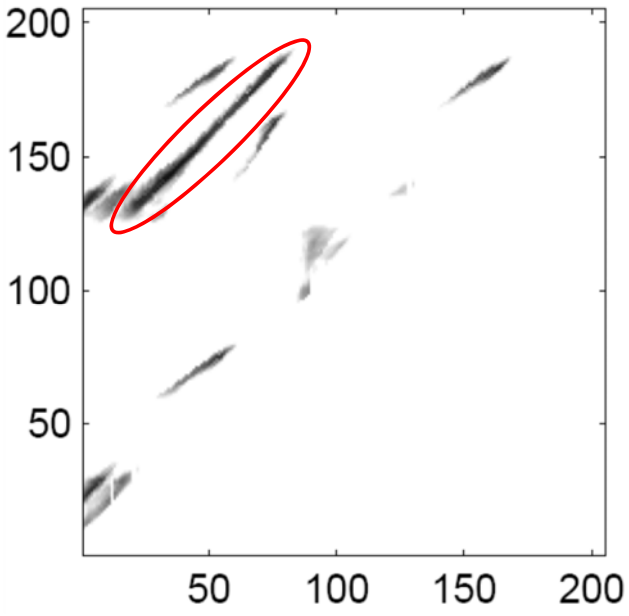
Enhanced cost matrix $C_L^{\min}$

Path Extraction



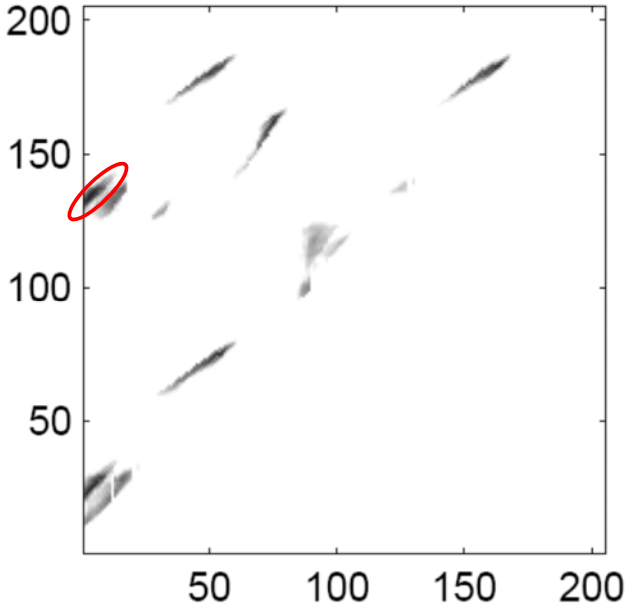
Thresholded $C_L^{\min}$

Path Extraction



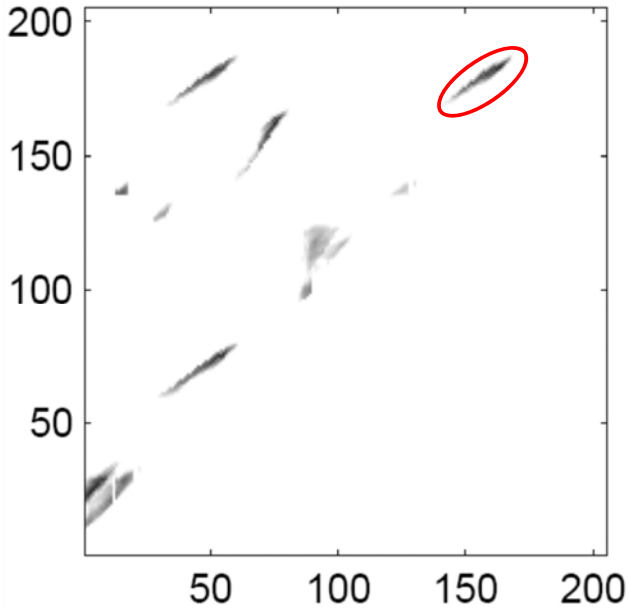
Thresholded $C_L^{\min}$, upper left

Path Extraction



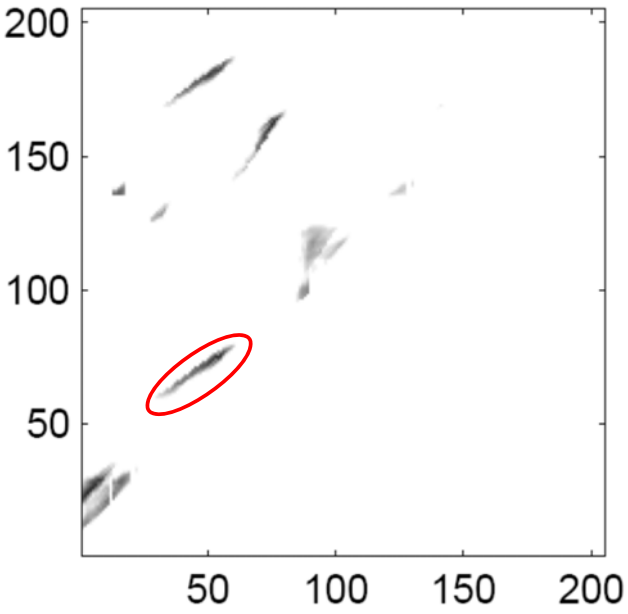
Path removal

Path Extraction



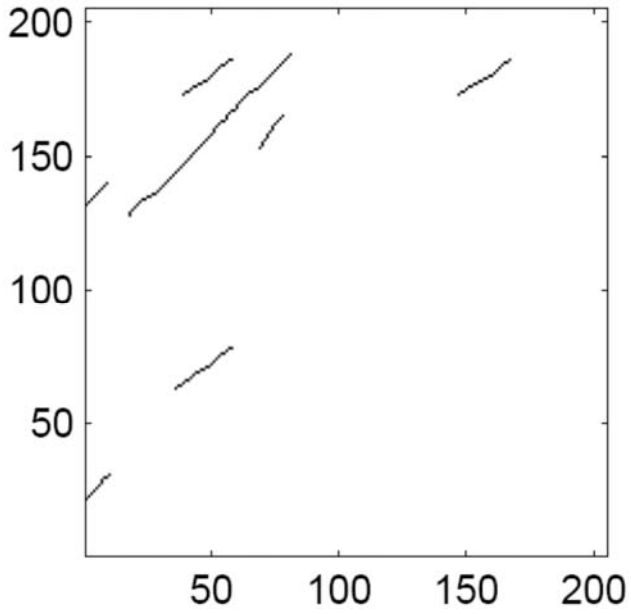
Path removal

Path Extraction



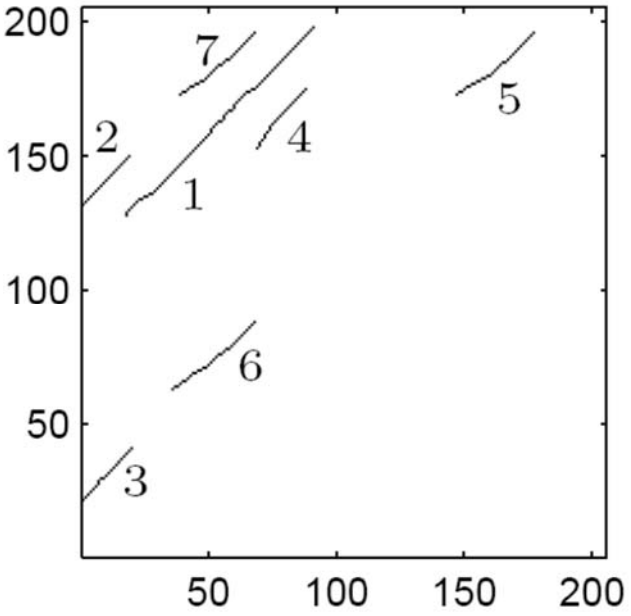
Path removal

Path Extraction



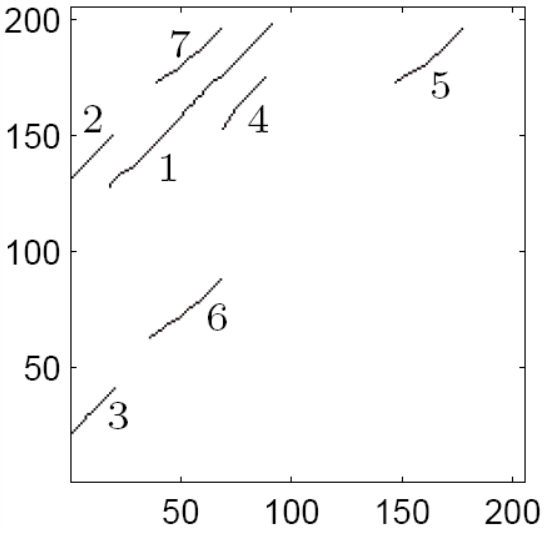
Extracted paths

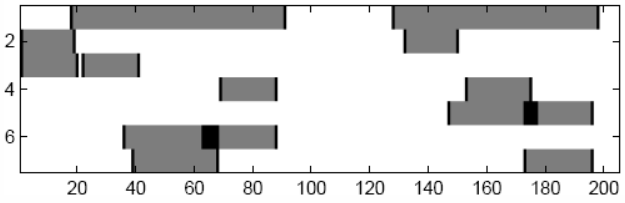
Path Extraction



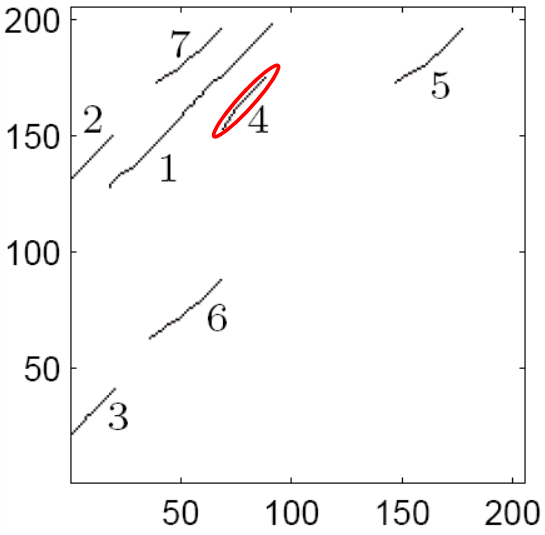
Extracted paths after postprocessing

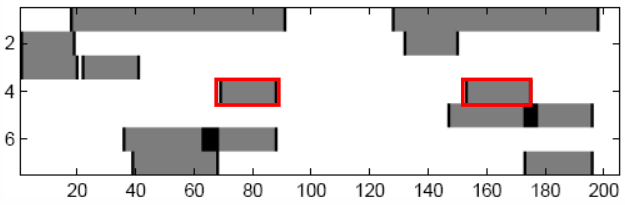
Global Structure





Global Structure





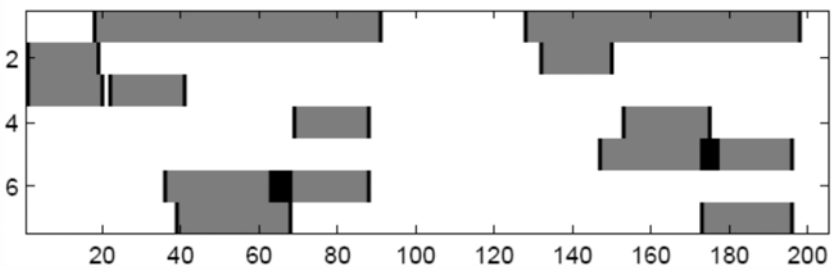
How can one derive the global structure from pairwise relations?

Global Structure

- Taks: Computation of similarity clusters
- Problem: Missing and inconsistent path relations
- Strategy: Approximate “transitive hull”

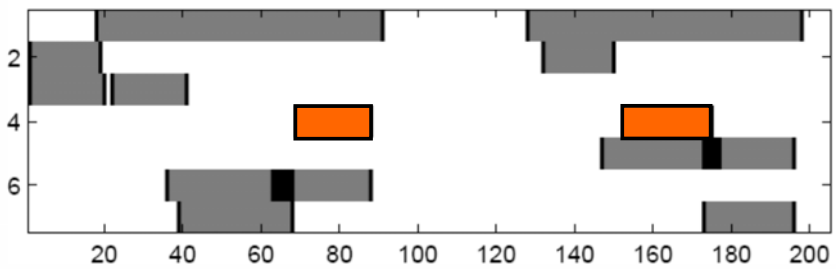
Global Structure

Path relations



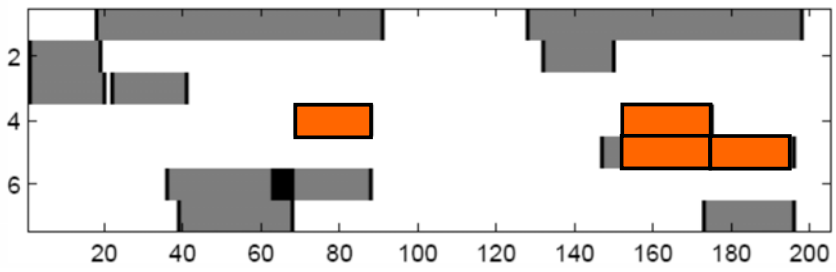
Global Structure

Path relations



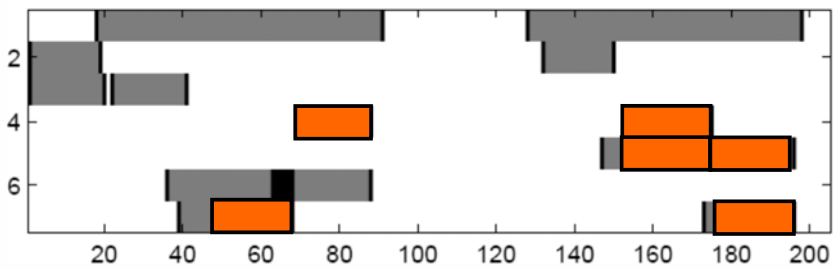
Global Structure

Path relations



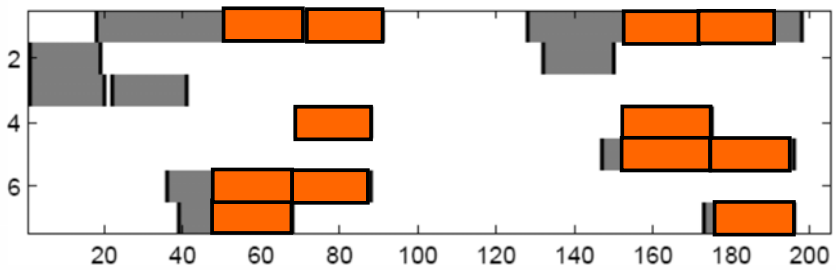
Global Structure

Path relations



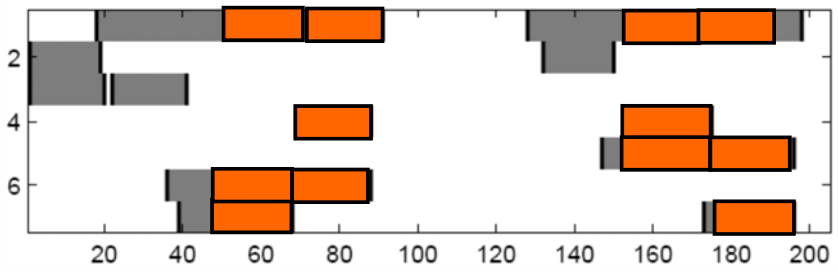
Global Structure

Path relations

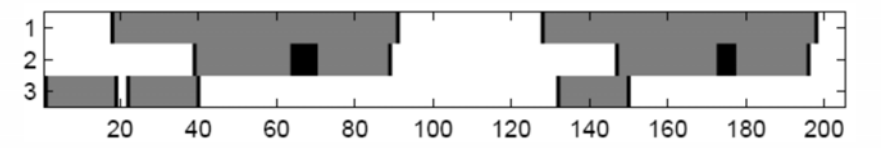


Global Structure

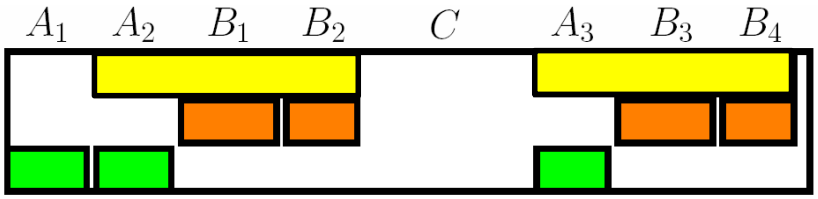
Path relations



Final result



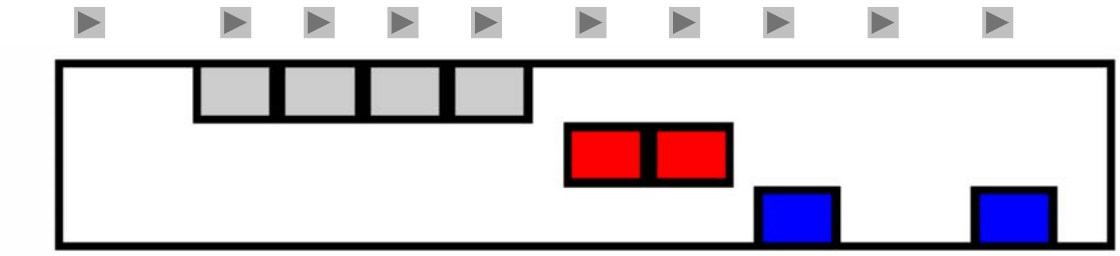
Ground truth



Time (seconds)

Transposition Invariance

Example: Zager & Evans “In The Year 2525”



Transposition Invariance

Goto (ICASSP 2003)

- Cyclically shift chroma vectors in one sequence
- Compare shifted sequence with original sequence
- Perform for each of the twelve shifts a separate structure analysis
- Combine the results

Transposition Invariance

Goto (ICASSP 2003)

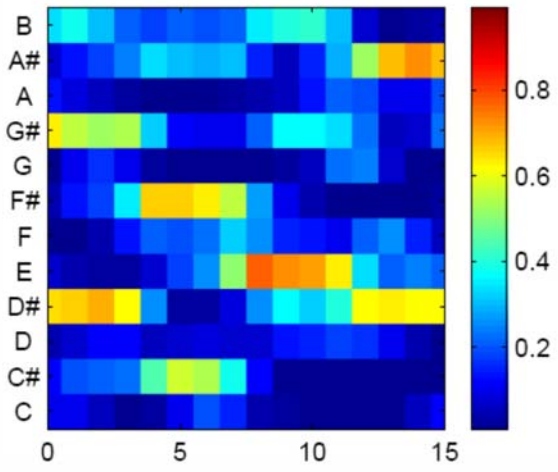
- Cyclically shift chroma vectors in one sequence
- Compare shifted sequence with original sequence
- Perform for each of the twelve shifts a separate structure analysis
- Combine the results

Müller/Clausen (ISMIR 2007)

- Integrate all cyclic information in one transposition-invariant self-similarity matrix
- Perform one joint structure analysis

Transposition Invariance

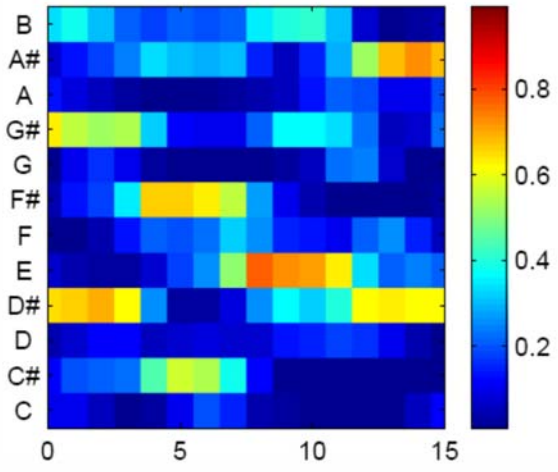
Example: Zager & Evans “In The Year 2525”



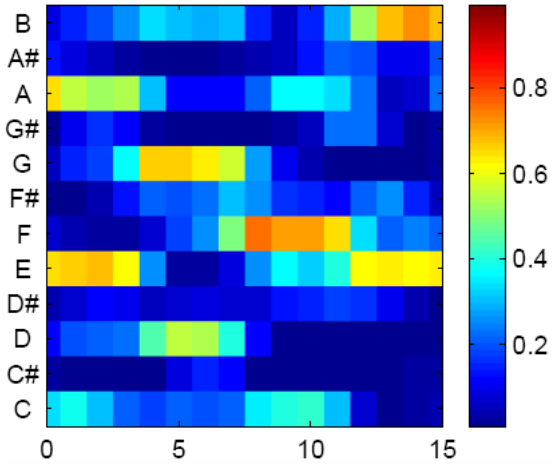
Original: $(v^1, \ldots, v^N)$

Transposition Invariance

Example: Zager & Evans “In The Year 2525”

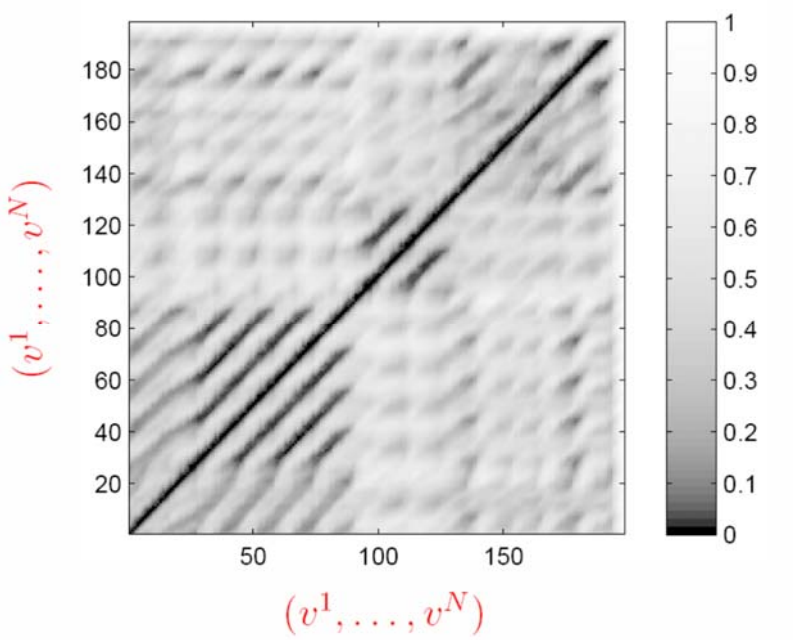


Original: $(v^1, \ldots, v^N)$

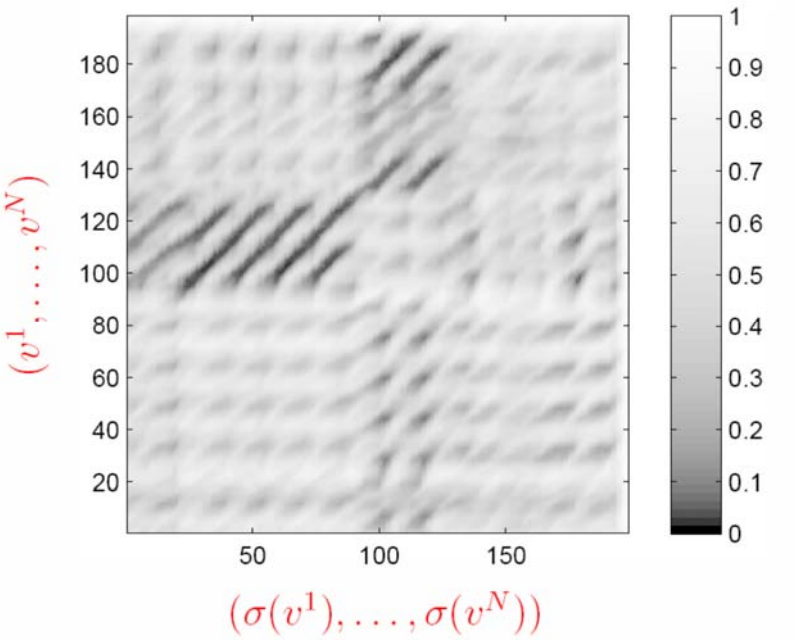


Shifted: $(\sigma(v^1), \ldots, \sigma(v^N))$

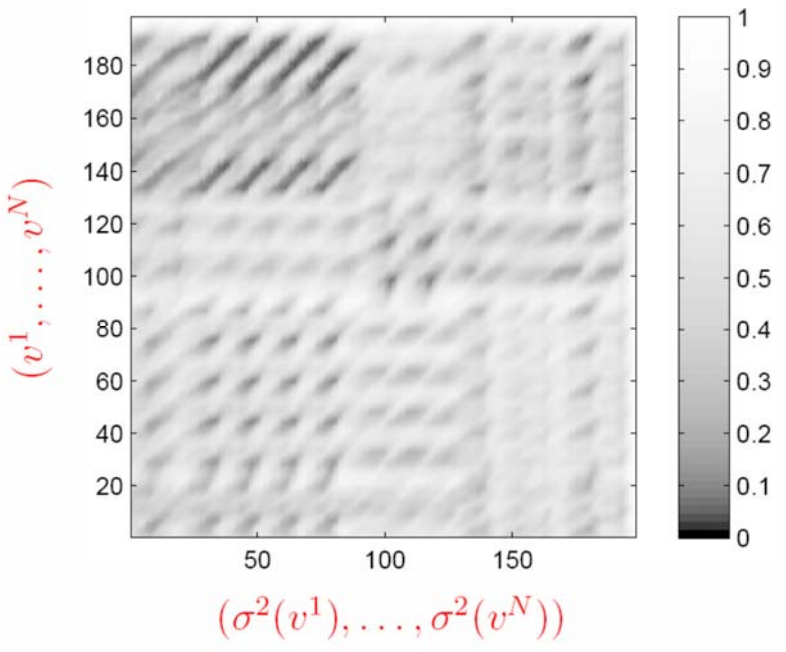
Transposition Invariance



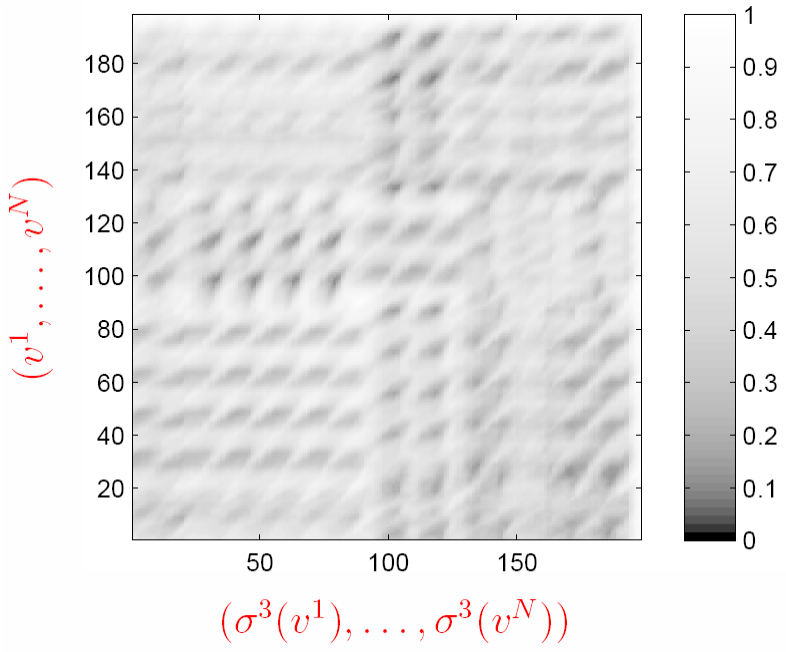
Transposition Invariance



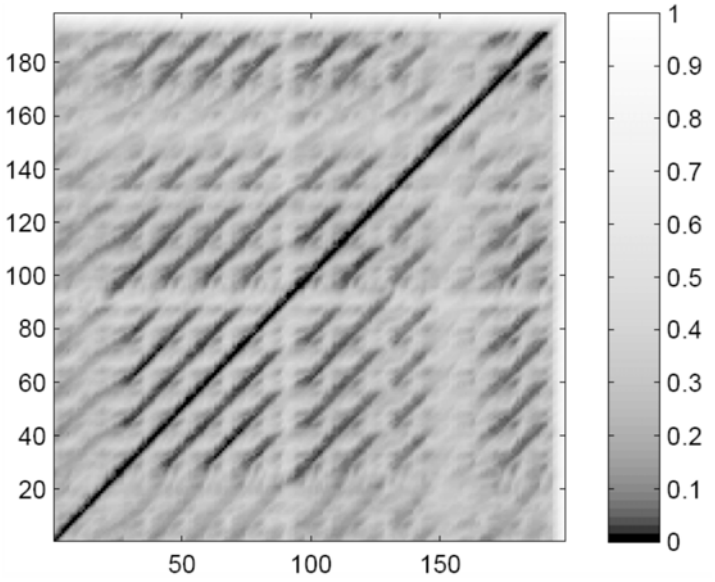
Transposition Invariance



Transposition Invariance

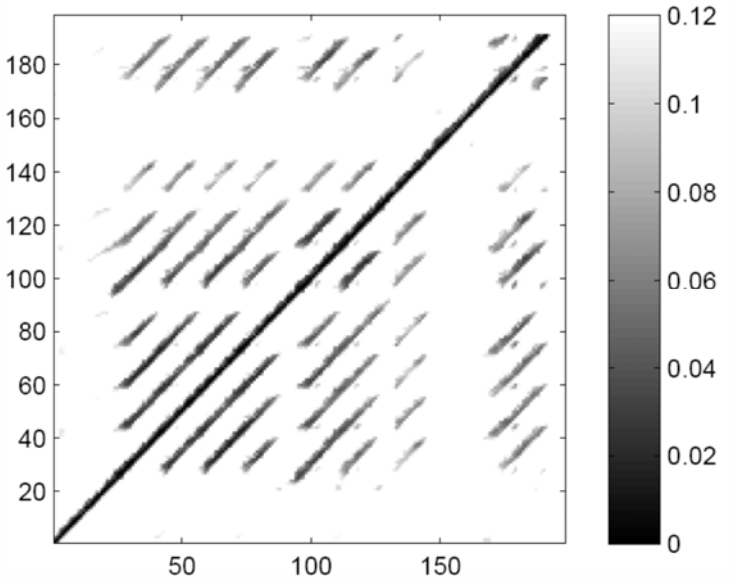


Transposition Invariance



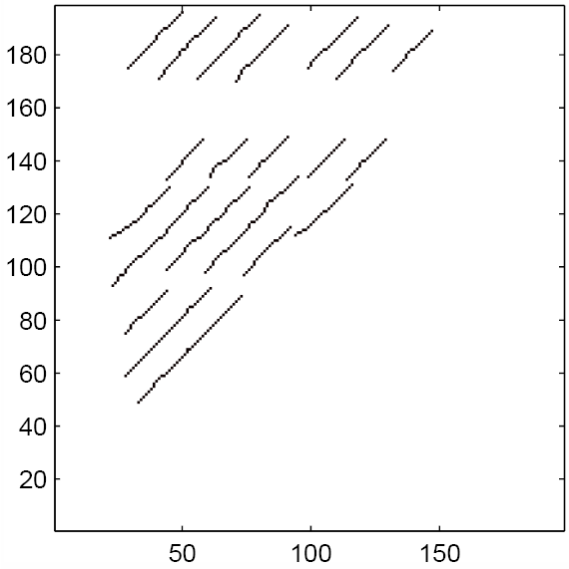
Minimize over all twelve matrices

Transposition Invariance



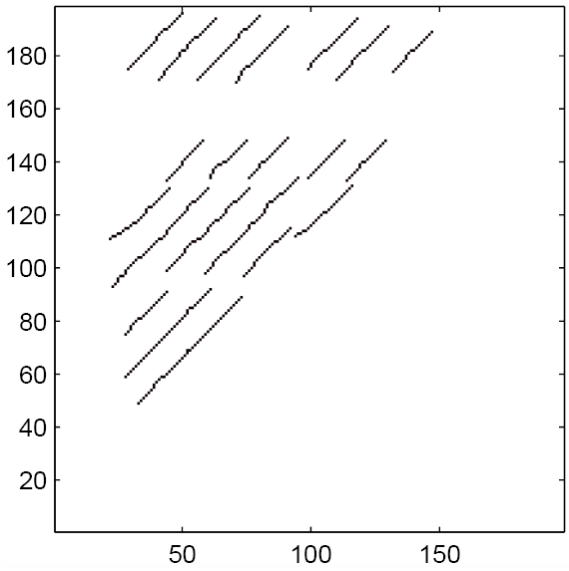
Thresholded self-similarity matrix

Transposition Invariance

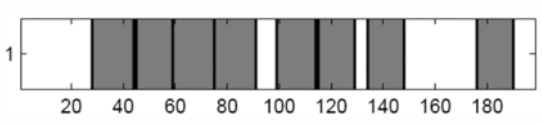


Path extraction

Transposition Invariance



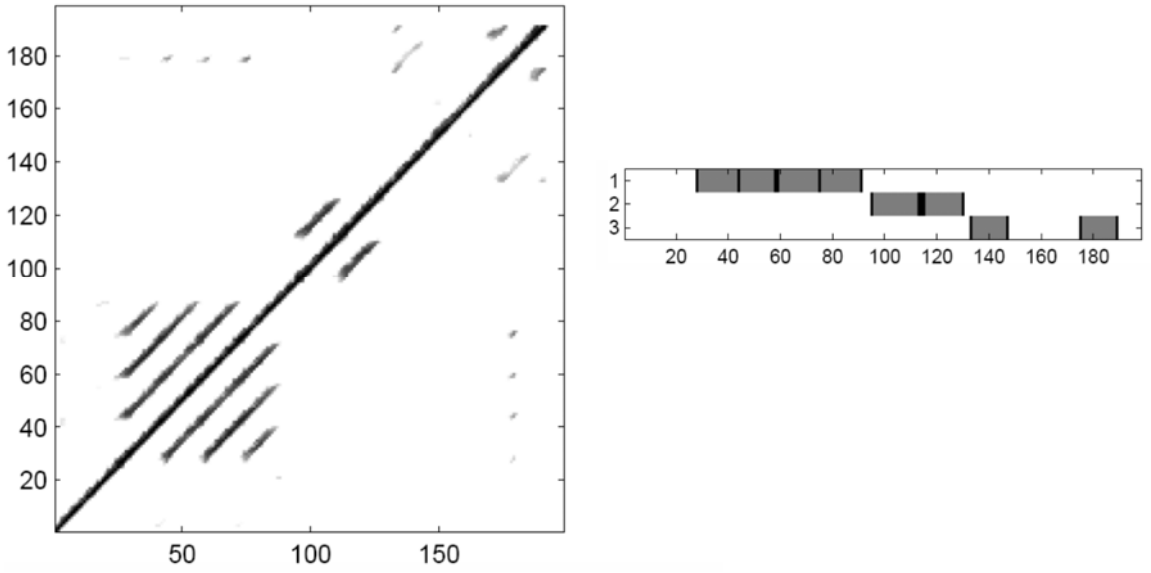
Path extraction



Computation of similarity clusters

Transposition Invariance

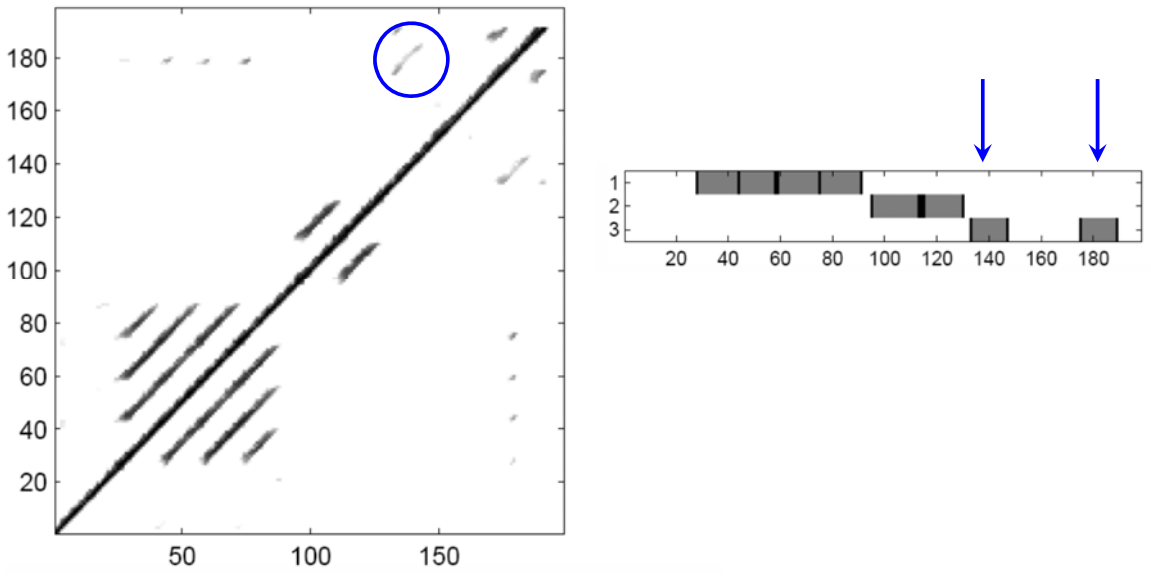
Stabilizing effect



Self-similarity matrix (thresholded)

Transposition Invariance

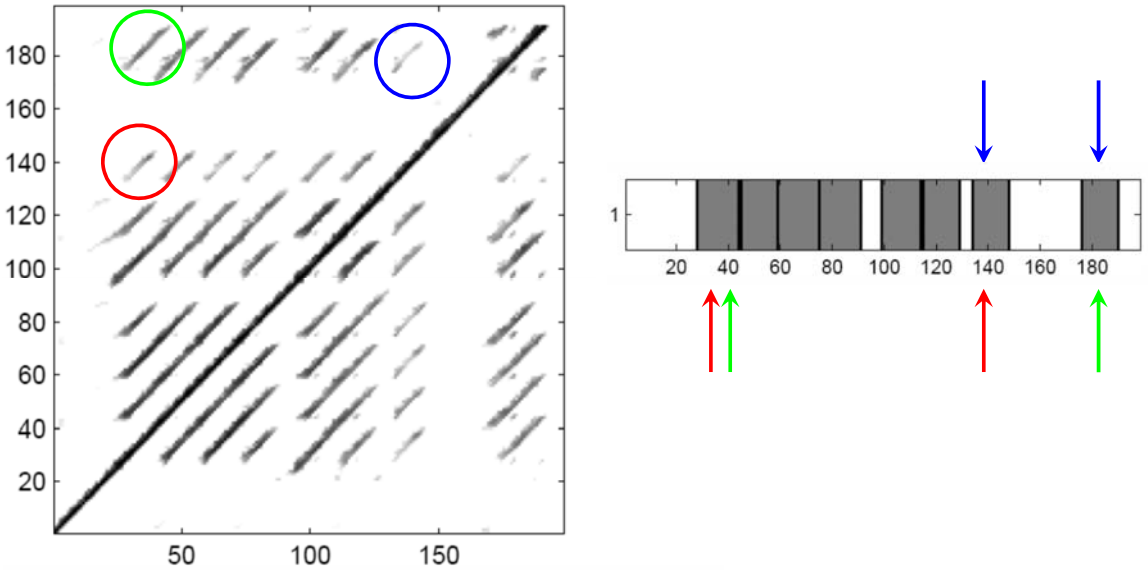
Stabilizing effect



Self-similarity matrix (thresholded)

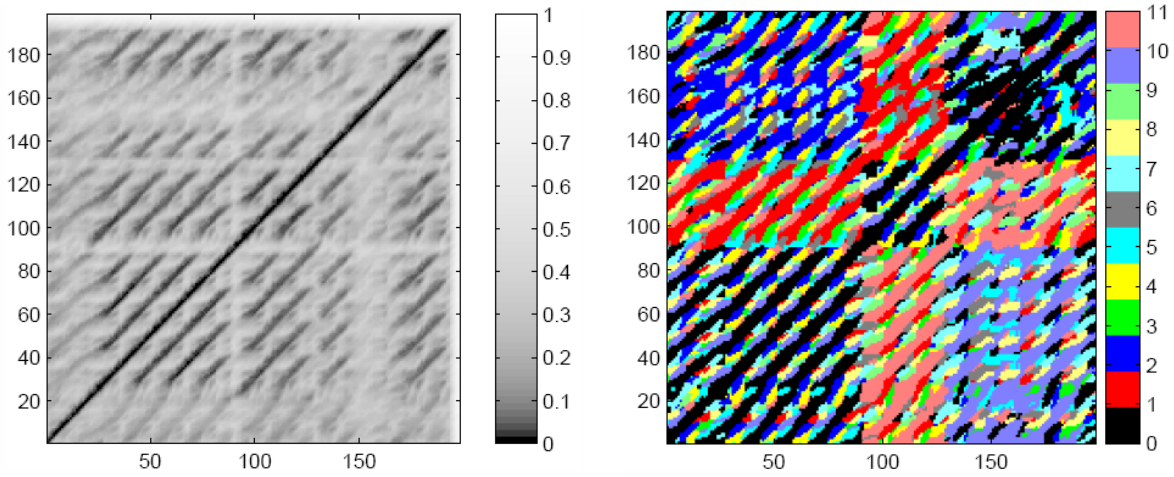
Transposition Invariance

Stabilizing effect



Transposition-invariant self-similarity matrix (thresholded)

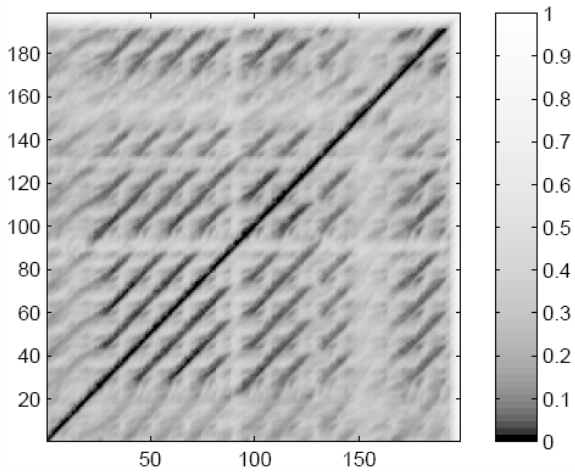
Transposition Invariance



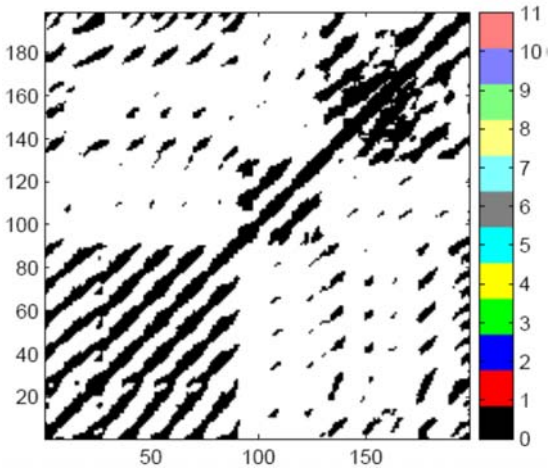
Transposition-invariant matrix

Minimizing shift index

Transposition Invariance

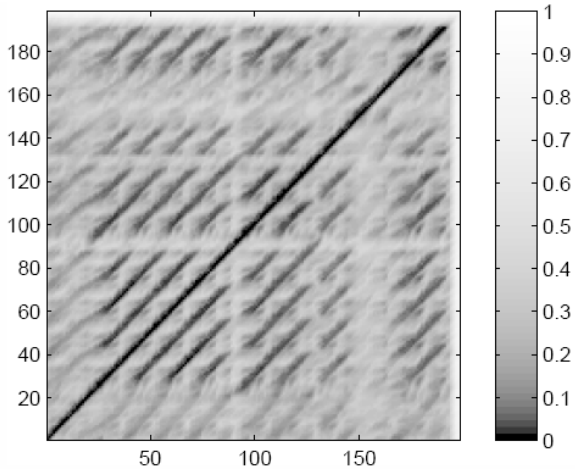


Transposition-invariant matrix

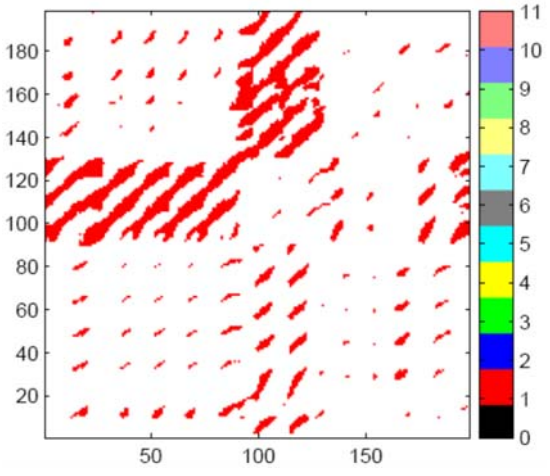


Minimizing shift index = 0

Transposition Invariance

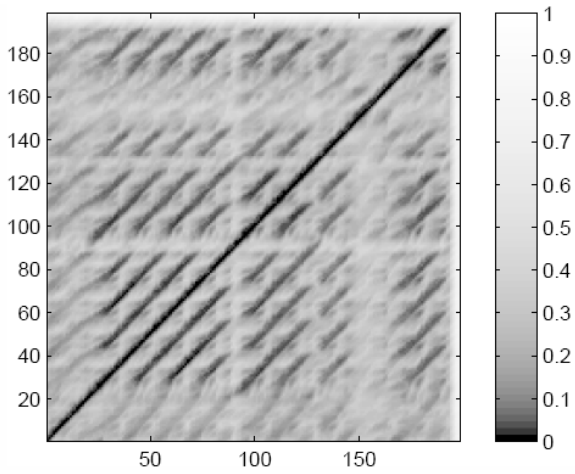


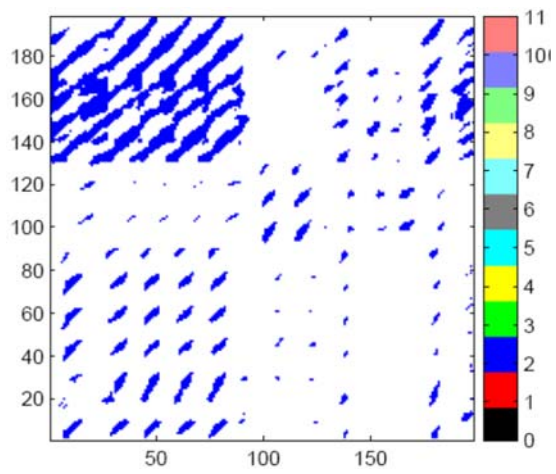
Transposition-invariant matrix



Minimizing shift index = 1

Transposition Invariance

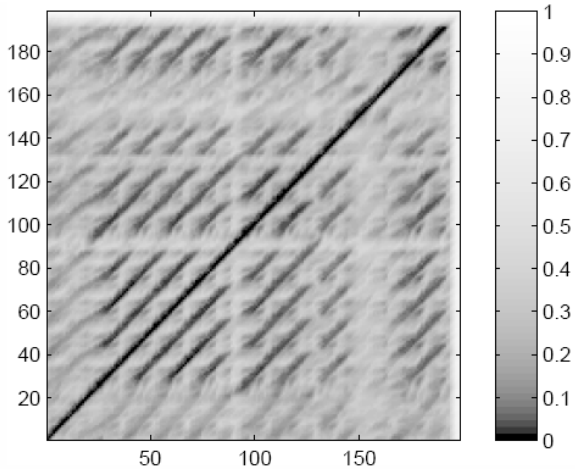


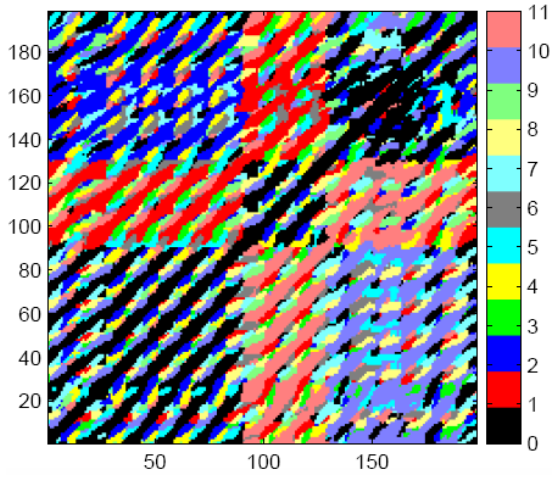


Transposition-invariant matrix

Minimizing shift index = 2

Transposition Invariance



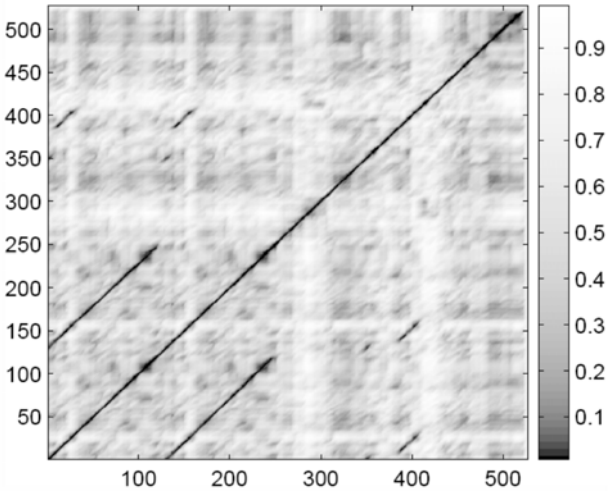


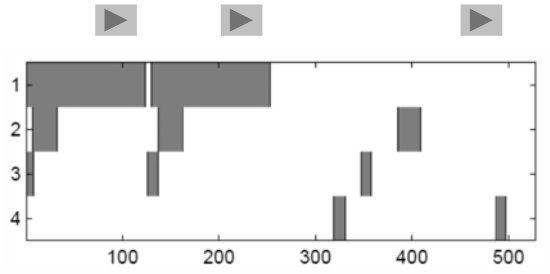
Serra/Gomez (ICASSP 2008): Used for Cover Song ID

Discrete structure $\rightsquigarrow$ suitable for indexing?

Transposition Invariance

Example: Beethoven “Tempest”

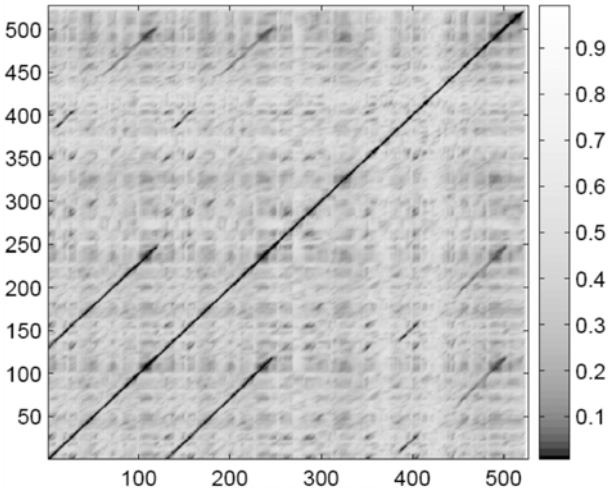


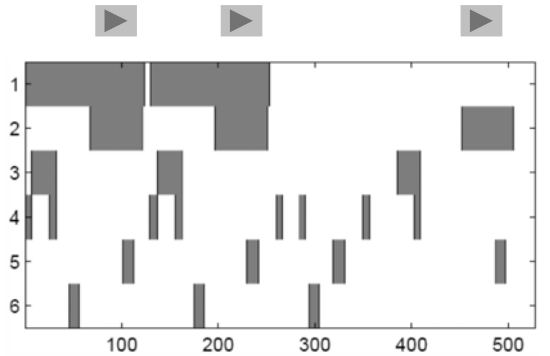


Self-similarity matrix

Transposition Invariance

Example: Beethoven “Tempest”





Transposition-invariant self-similarity matrix

Conclusions: Audio Structure Analysis

Challenge: Musical variations

- Timbre, dynamics, tempo
- Musical key ⇝ cyclic chroma shifts
- Major/minor
- Differences at note level / improvisations

Conclusions: Audio Structure Analysis

Strategy: Matrix enhancement

- Filtering techniques / contextual information
  - Cooper/Foote (ISMIR 2002)
  - Müller/Kurth (ICASSP 2006)
- Transposition-invariant similarity matrices
  - Goto (ICASSP 2003)
  - Müller/Clausen (ISMIR 2007)
- Higher-order similarity matrices
  - Peeters (ISMIR 2007)

Novel Approach for Audio Thumbnailing

Original approach: Two steps

1. Path extraction
   - Paths of poor quality (fragmented, gaps)
   - Regions of constant (low) cost
   - Curved paths
2. Grouping:
   - Noisy relations (missing, distorted, overlapping)
   - Transitivity computation difficult

Both steps are problematic!

Our main idea: Do both, path extraction and grouping, jointly

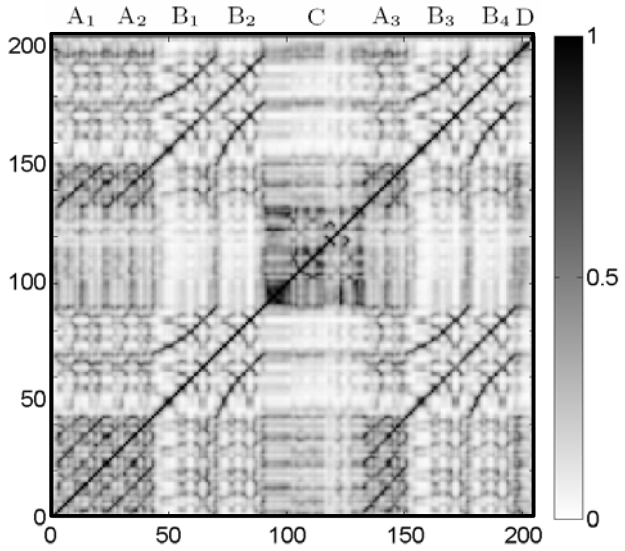
- One optimization scheme for both steps
- Stabilizing effect
- Efficient

Novel Approach for Audio Thumbnailing

Our main idea: Do both path extraction and grouping jointly

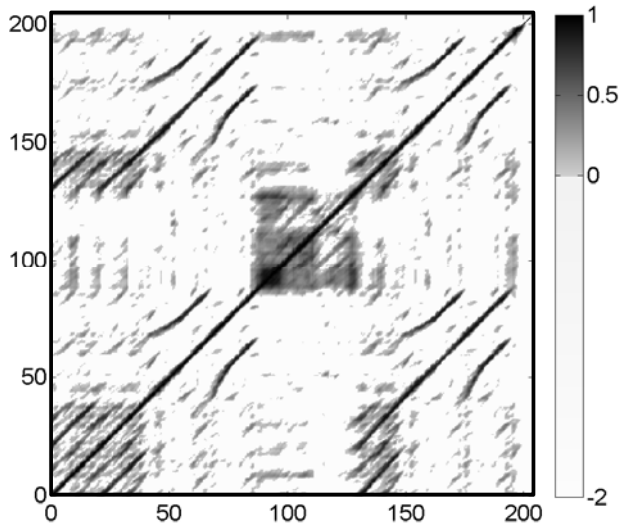
- For each audio segment we define a fitness value
- This fitness value expresses “how well” the segment explains the entire audio recording
- The segment with the highest fitness value is considered to be the thumbnail
- As main technical concept we introduce the notion of a path family

Fitness Measure



Self-similarity matrix

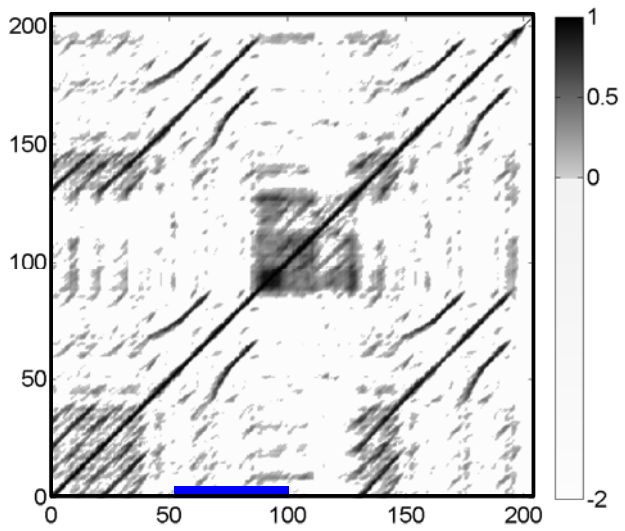
Fitness Measure



Self-similarity matrix

- Smoothing
- Transposition-Invariance
- Normalization
- Thresholding
- Negative score

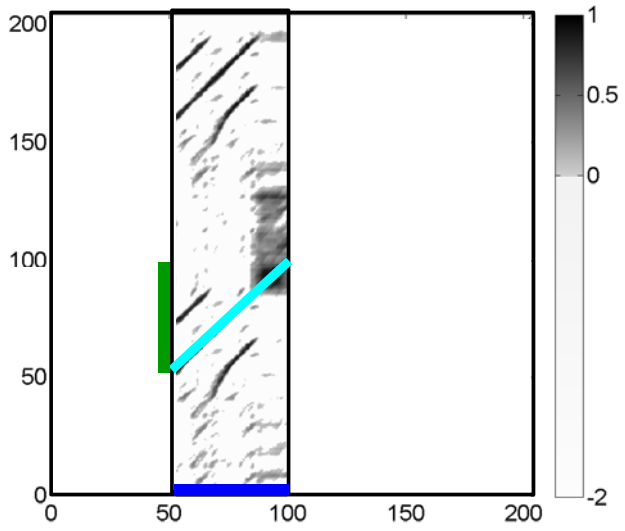
Fitness Measure



Path over segment

- Consider a fixed segment

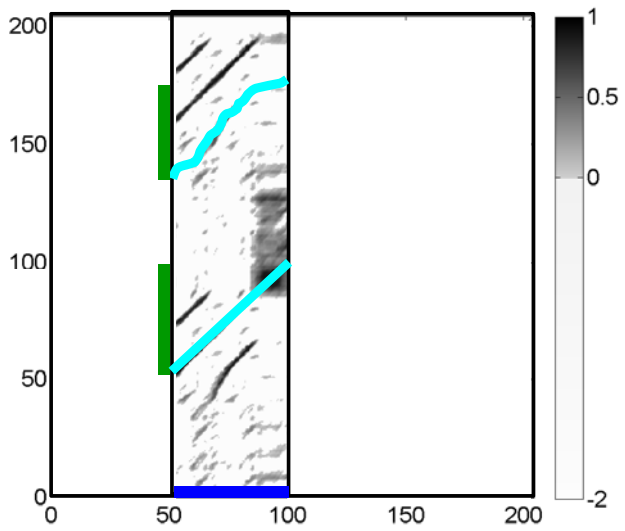
Fitness Measure



Path over segment

- Consider a fixed segment
- Path over segment
- Induced segment
- Score is high

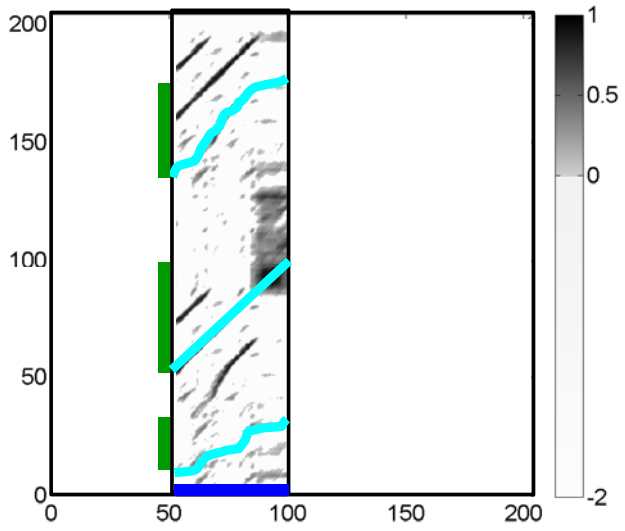
Fitness Measure



Path over segment

- Consider a fixed segment
- Path over segment
- Induced segment
- Score is high
- A second path over segment
- Induced segment
- Score is not so high

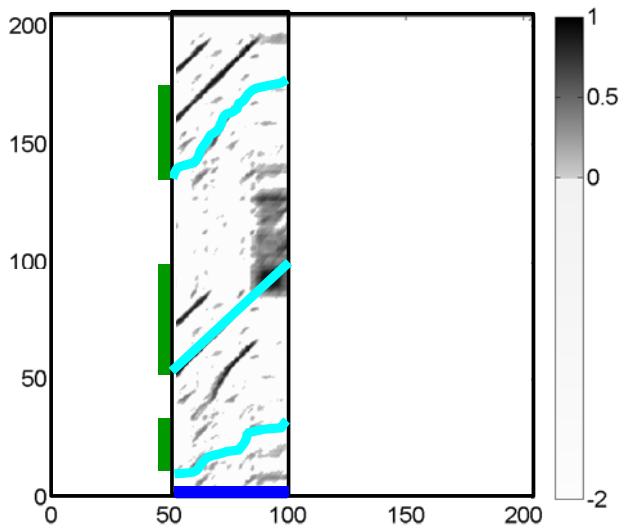
Fitness Measure



Path over segment

- Consider a fixed segment
- Path over segment
- Induced segment
- Score is high
- A second path over segment
- Induced segment
- Score is not so high
- A third path over segment
- Induced segment
- Score is very low

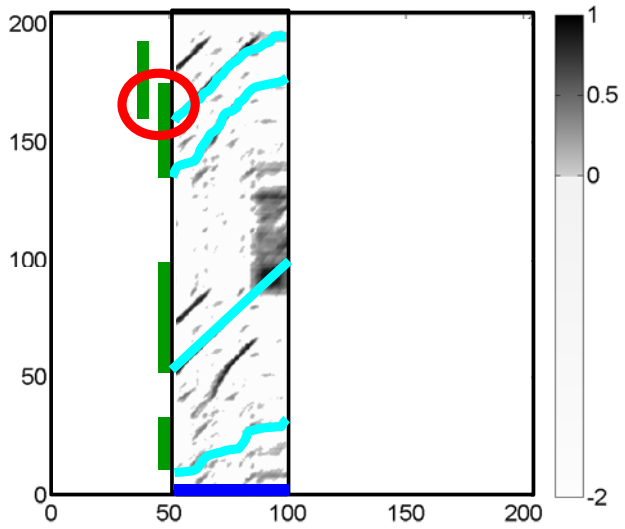
Fitness Measure



Path family

- Consider a fixed segment
- A path family over a segment is a family of paths such that the induced segments do not overlap.

Fitness Measure

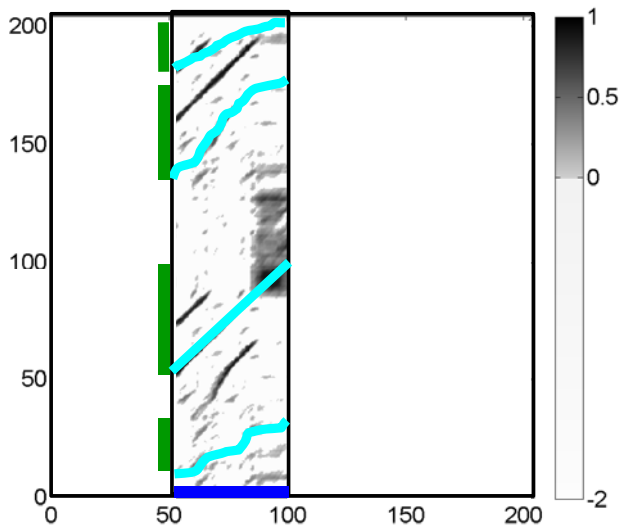


Path family

- Consider a fixed segment
- A path family over a segment is a family of paths such that the induced segments do not overlap.

This is not a path family!

Fitness Measure



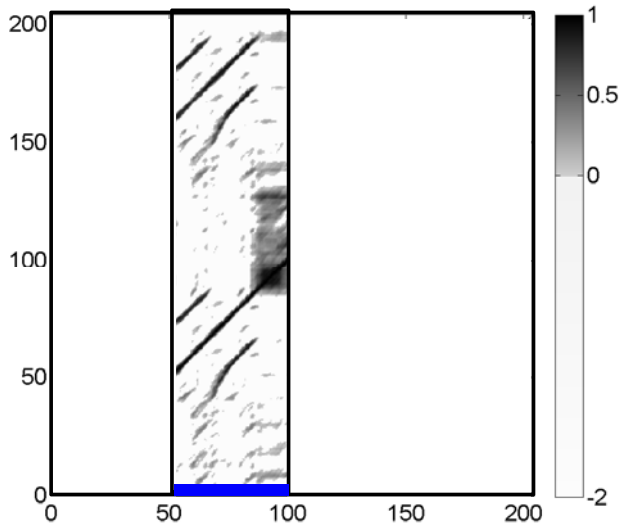
Path family

- Consider a fixed segment
- A path family over a segment is a family of paths such that the induced segments do not overlap.

This is a path family!

(Even though not a good one)

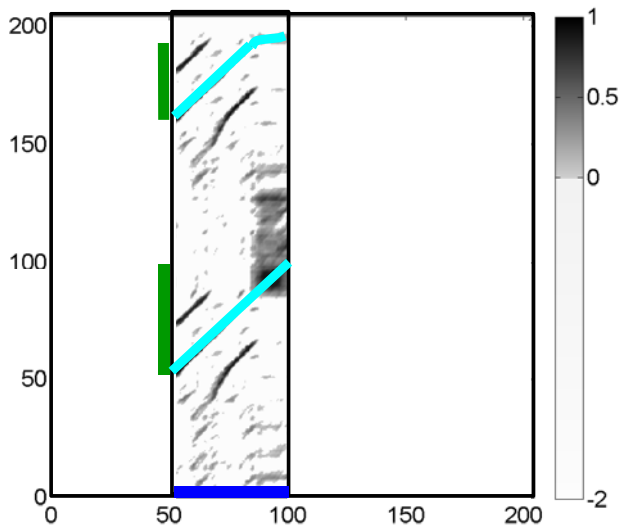
Fitness Measure



Optimal path family

- Consider a fixed segment

Fitness Measure



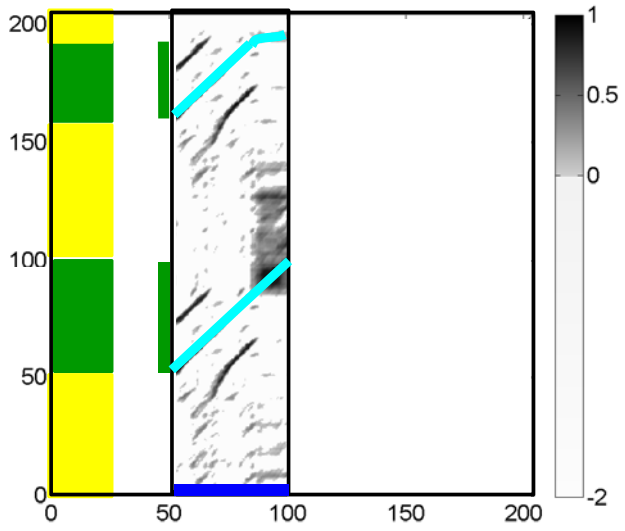
Optimal path family

- Consider a fixed segment
- Consider over the segment the optimal path family, i.e., the path family having maximal overall score.
- Call this value:

  Score(segment)

Note: This optimal path family can be computed using dynamic programming.

Fitness Measure



Optimal path family

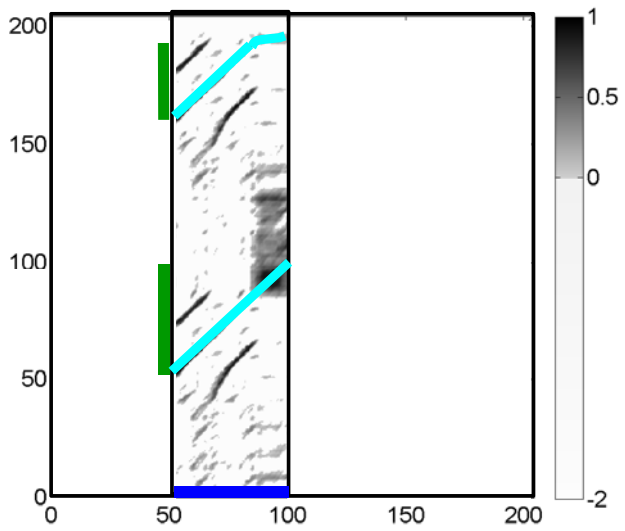
- Consider a fixed segment
- Consider over the segment the optimal path family, i.e., the path family having maximal overall score.
- Call this value:

  Score(segment)

- Furthermore consider the amount covered by the induced segments.
- Call this value:

  Coverage(segment)

Fitness Measure



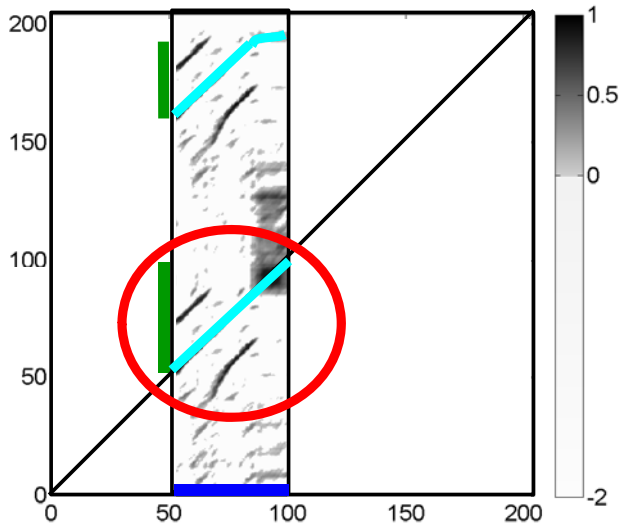
P := Score(segment)

R := Coverage(segment)

Fitness

- Consider a fixed segment

Fitness Measure



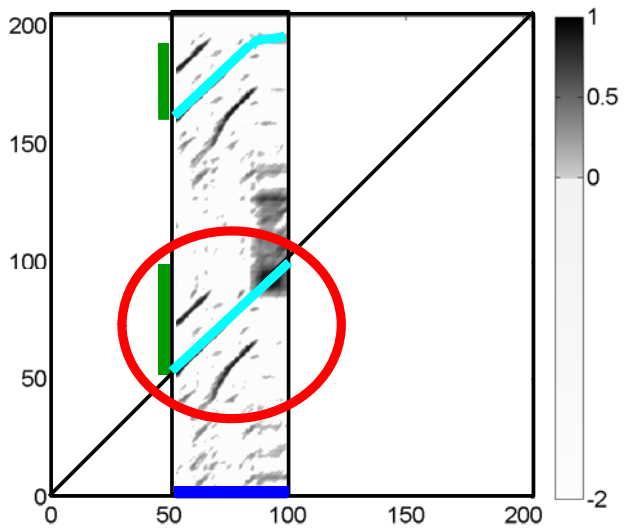
P := Score(segment)

R := Coverage(segment)

Fitness

- Consider a fixed segment
- Self-explanation are trivial!

Fitness Measure



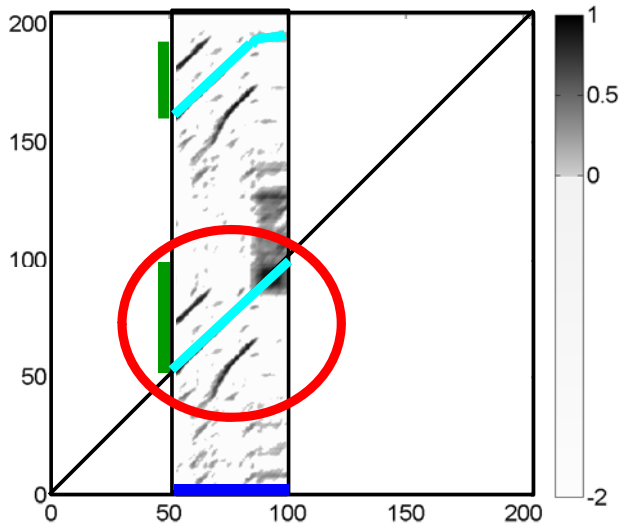
Fitness

- Consider a fixed segment
- Self-explanation are trivial!
- Substract length of segment

P := Score(segment) - length(segment)

R := Coverage(segment) - length(segment)

Fitness Measure



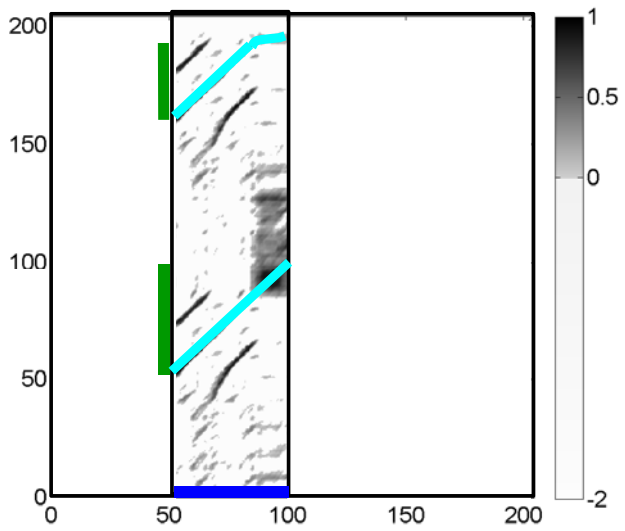
Fitness

- Consider a fixed segment
- Self-explanation are trivial!
- Substract length of segment
- Normalization

P := Normalize(Score(segment) - length(segment)) $\in [0,1]$

R := Normalize(Coverage(segment) - length(segment)) $\in [0,1]$

Fitness Measure



Fitness

- Consider a fixed segment

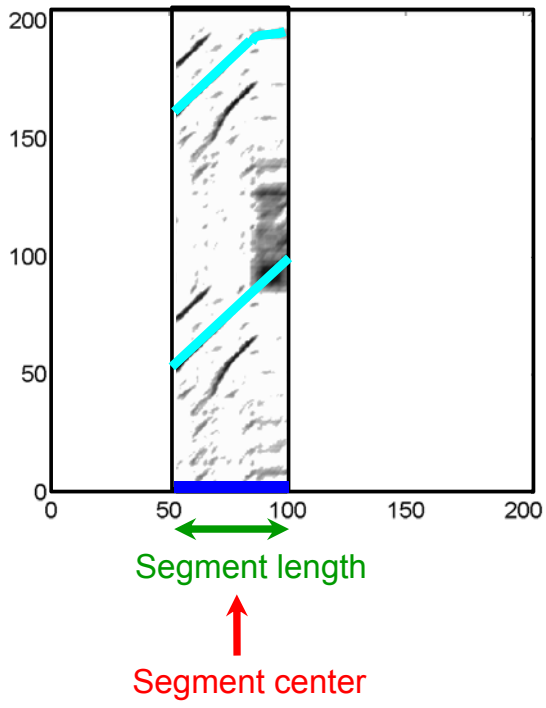
Fitness(segment)

F := 2 • P • R / (P + R)

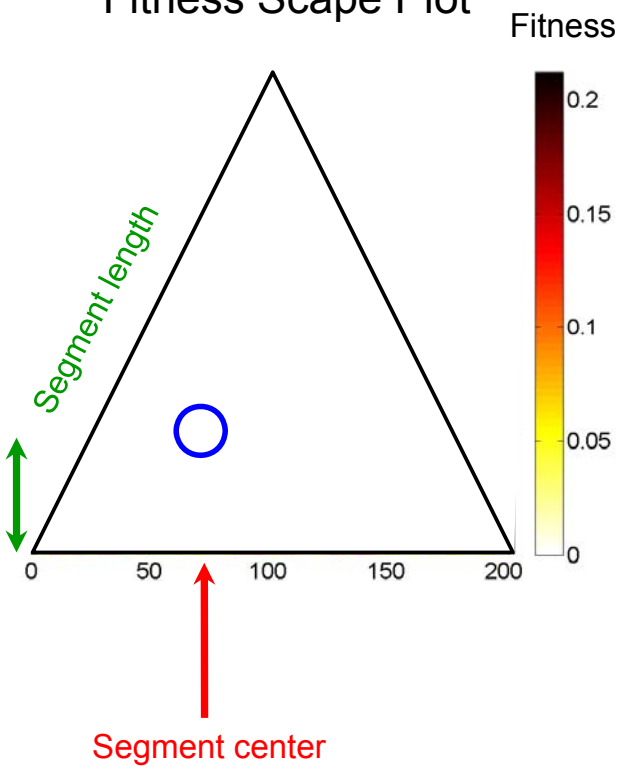
P := Normalize(Score(segment) - length(segment)) $\in [0,1]$

R := Normalize(Coverage(segment) - length(segment)) $\in [0,1]$

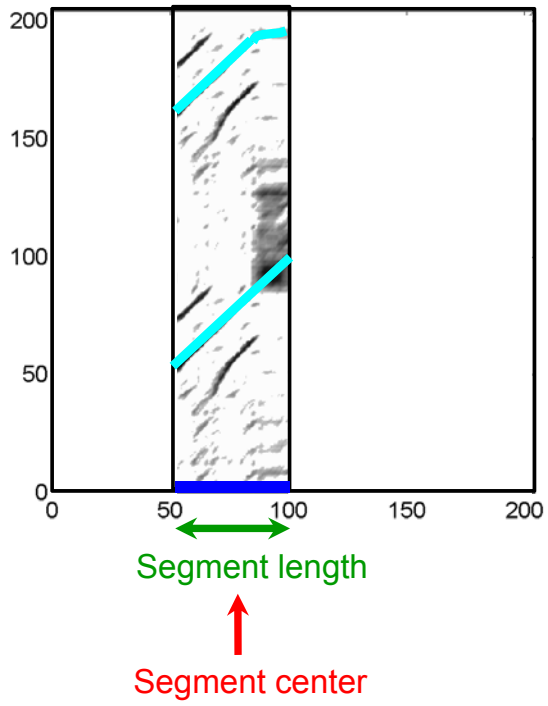
Thumbnail

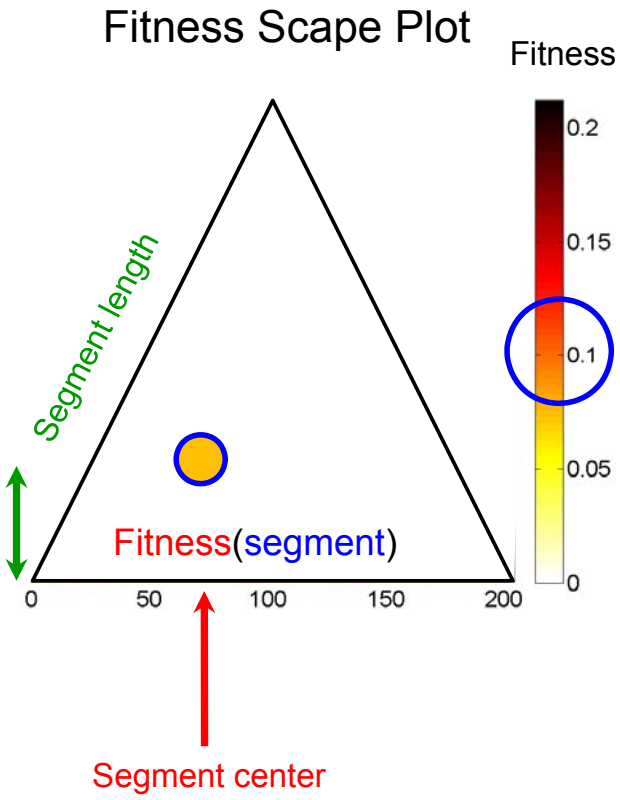


Fitness Scape Plot

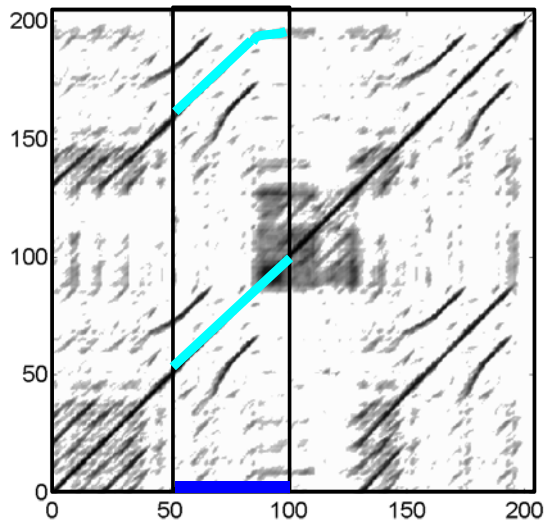


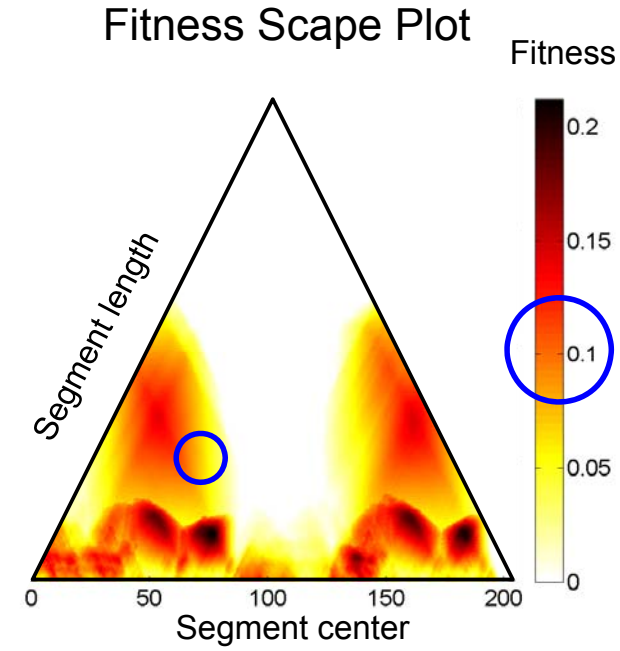
Thumbnail



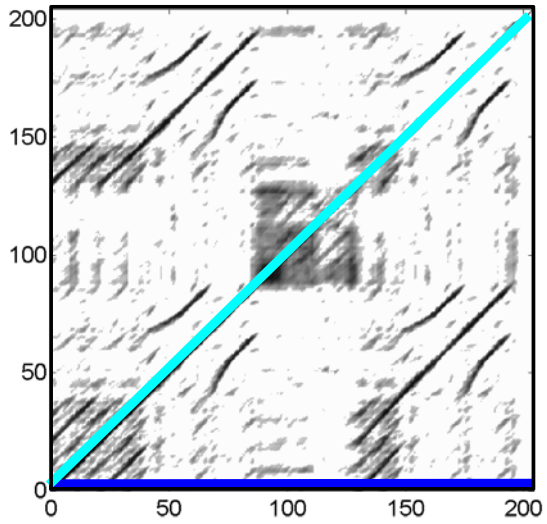


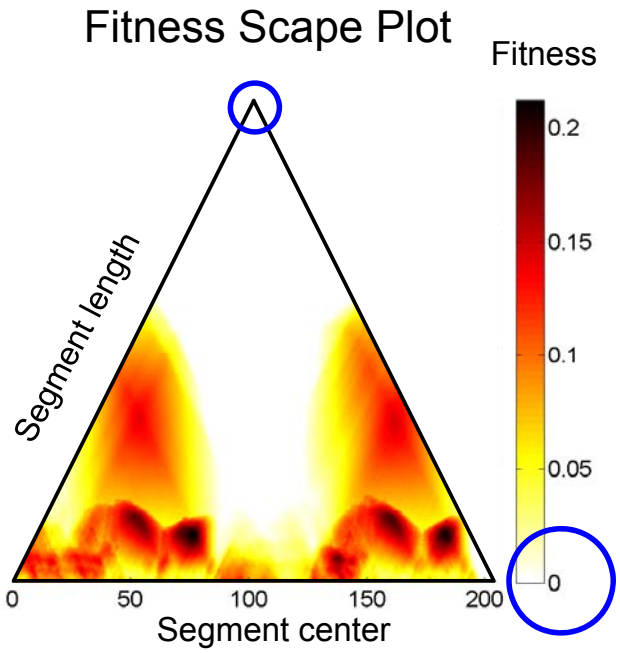
Thumbnail





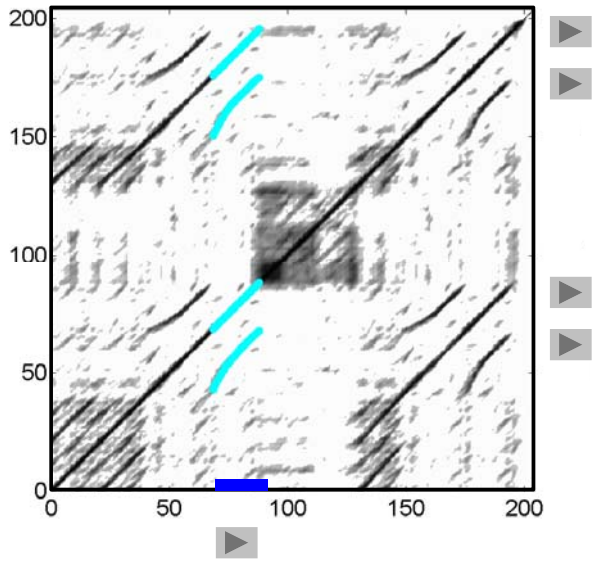
Thumbnail

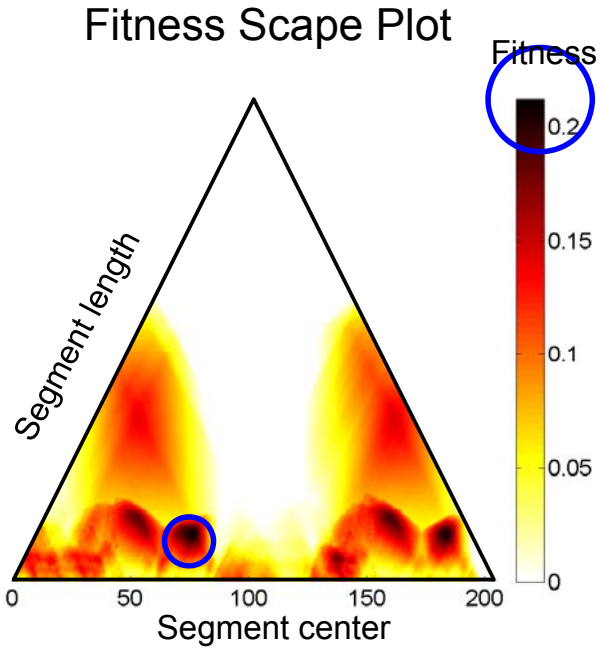




Note: Self-explanations are ignored → fitness is zero

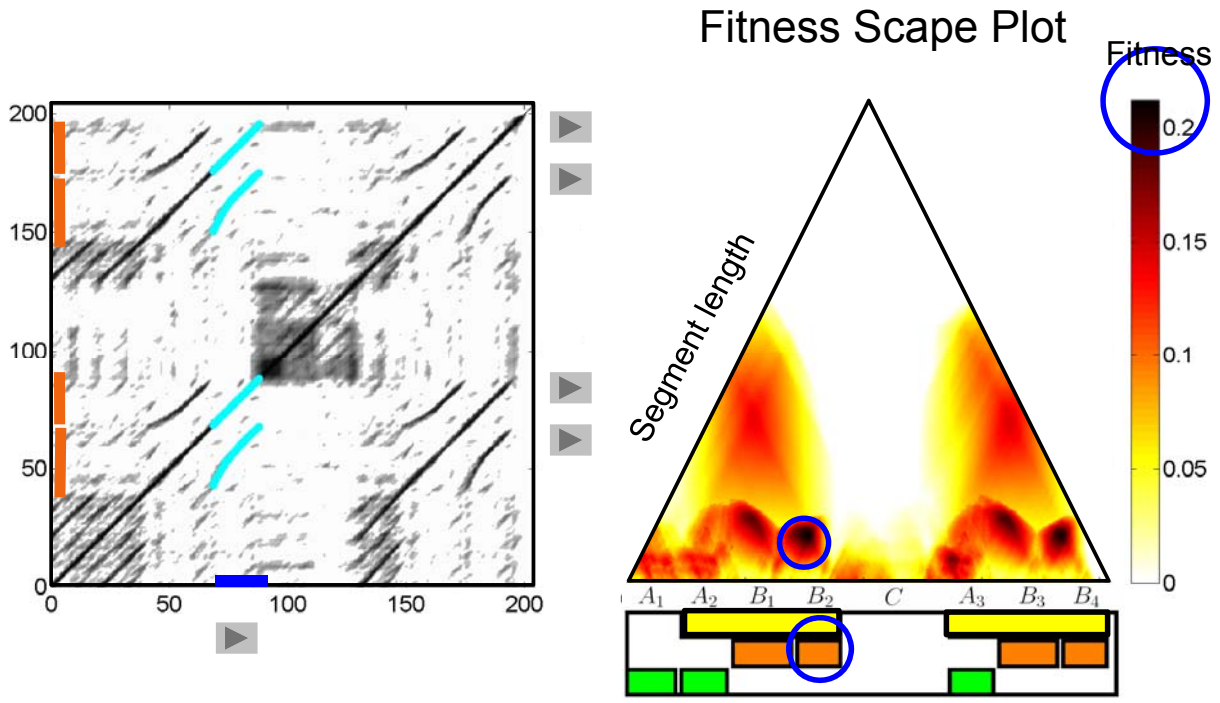
Thumbnail





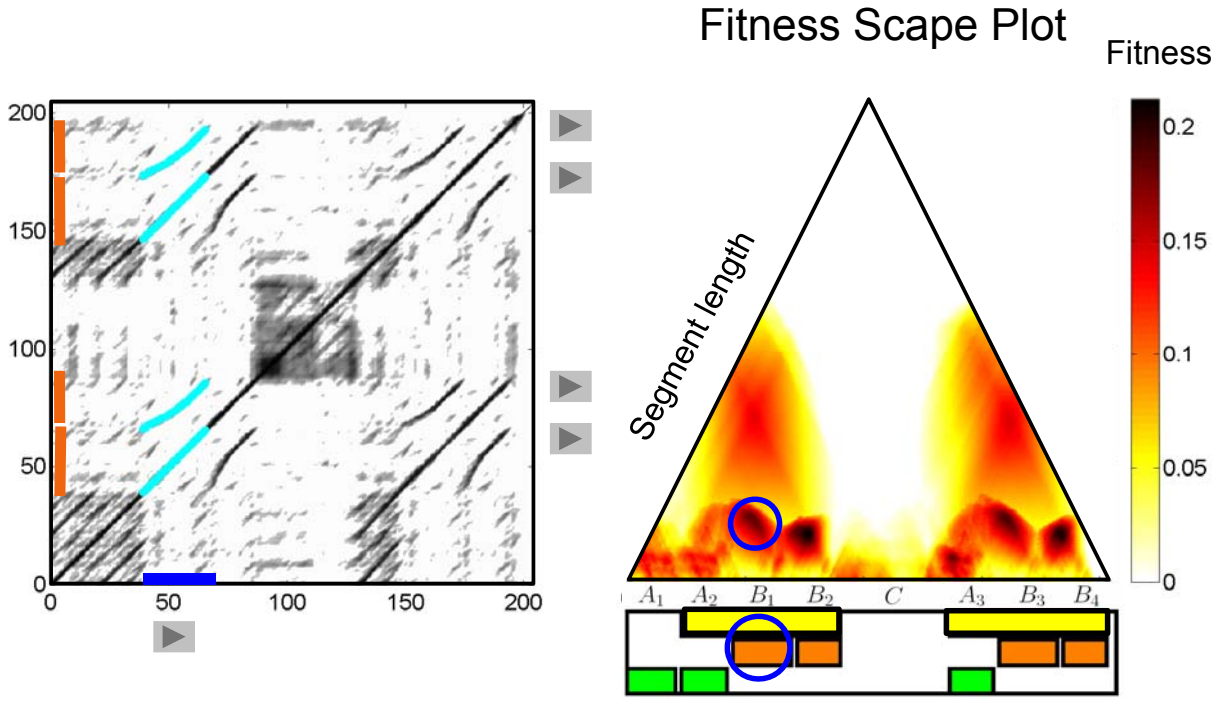
Thumbnail := segment having the highest fitness

Thumbnail



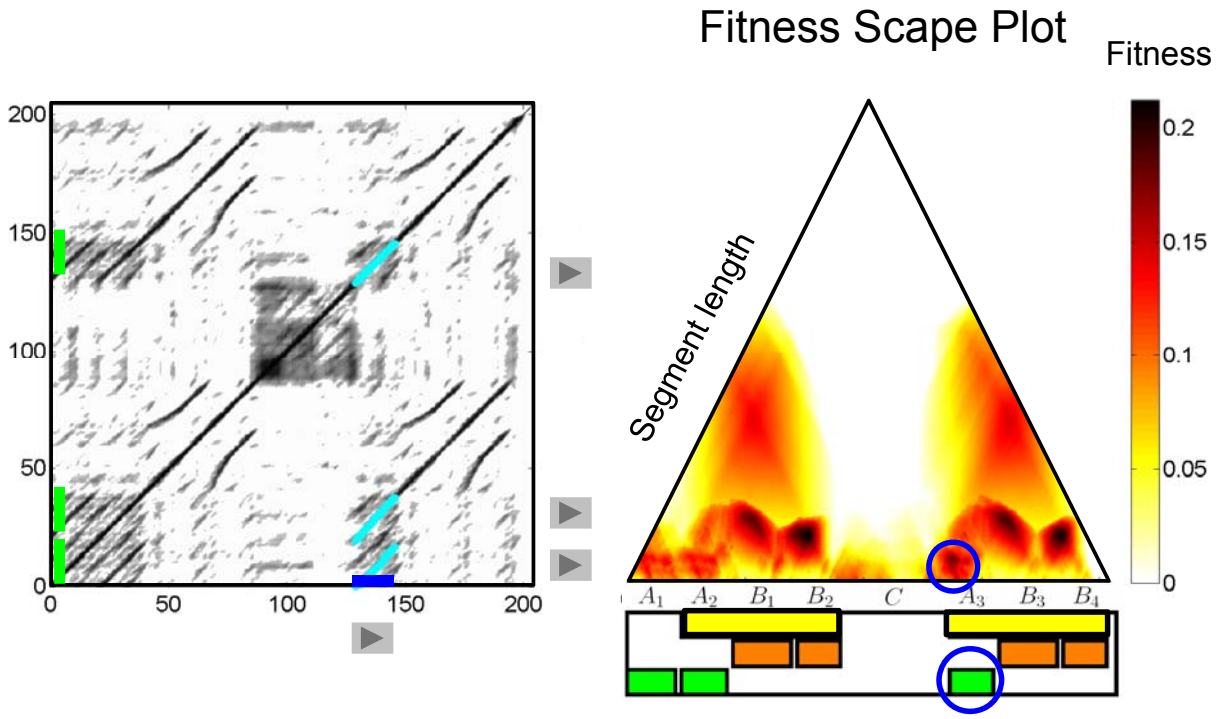
Example: Brahms Hungarian Dance No. 5 (Ormandy)

Thumbnail



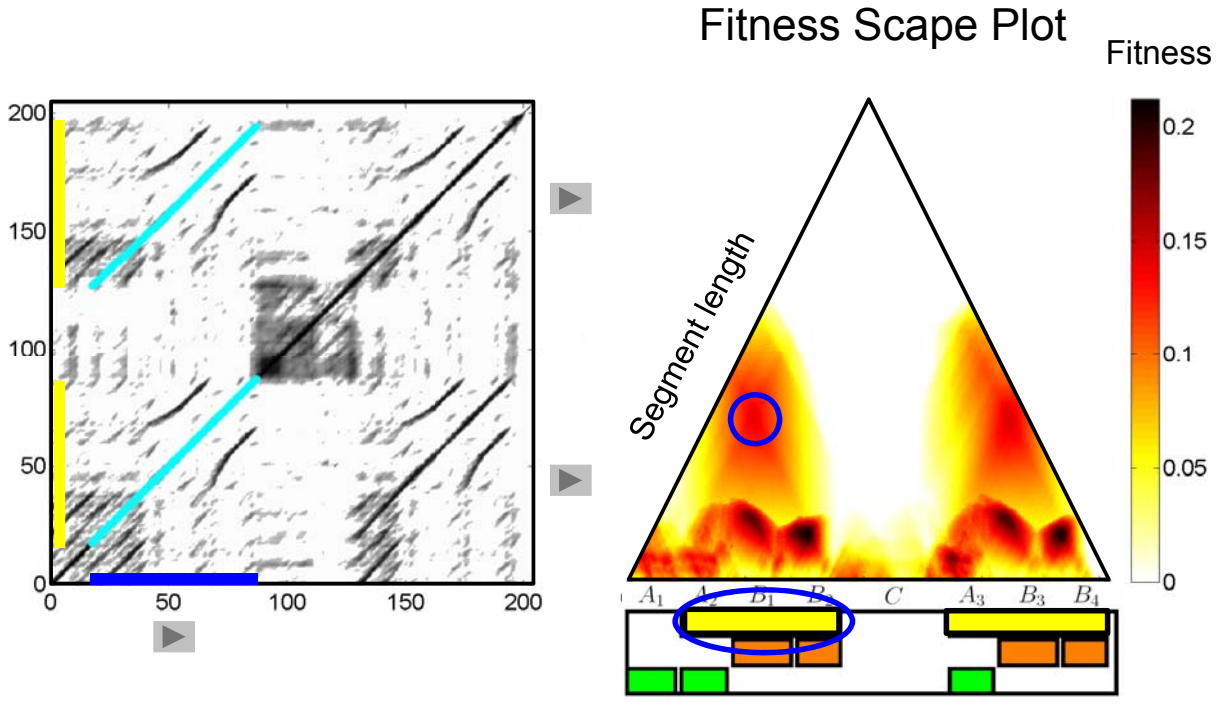
Example: Brahms Hungarian Dance No. 5 (Ormandy)

Thumbnail



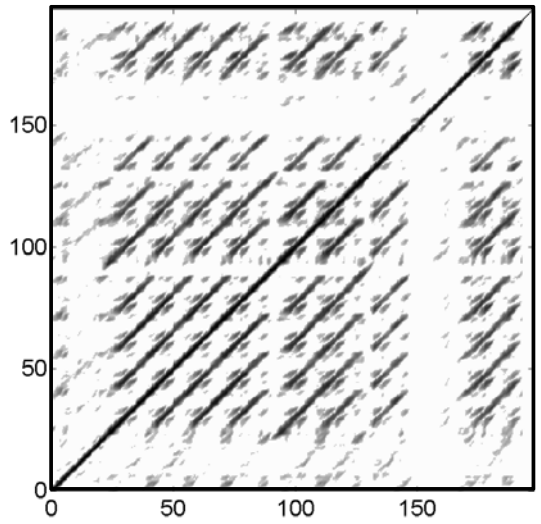
Example: Brahms Hungarian Dance No. 5 (Ormandy)

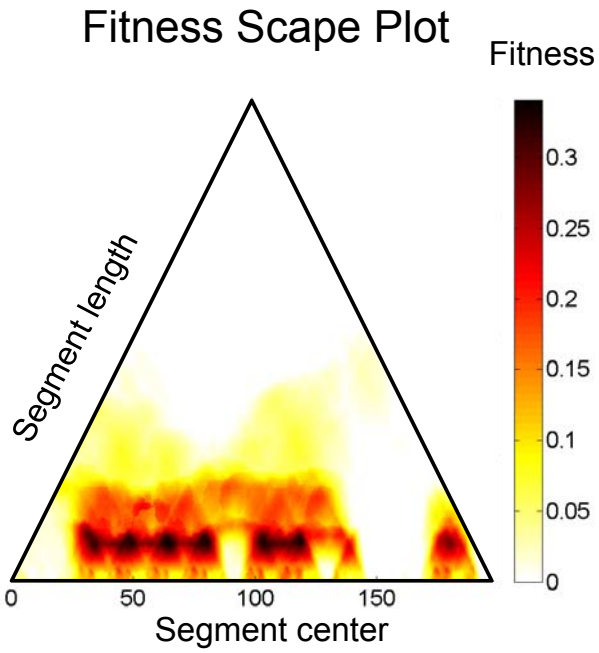
Thumbnail



Example: Brahms Hungarian Dance No. 5 (Ormandy)

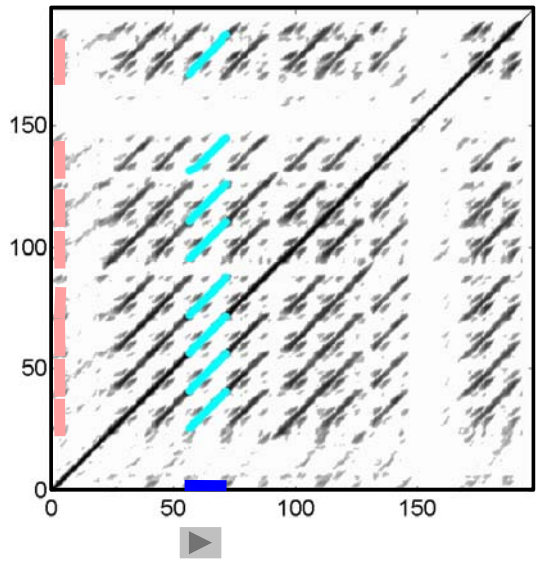
Thumbnail

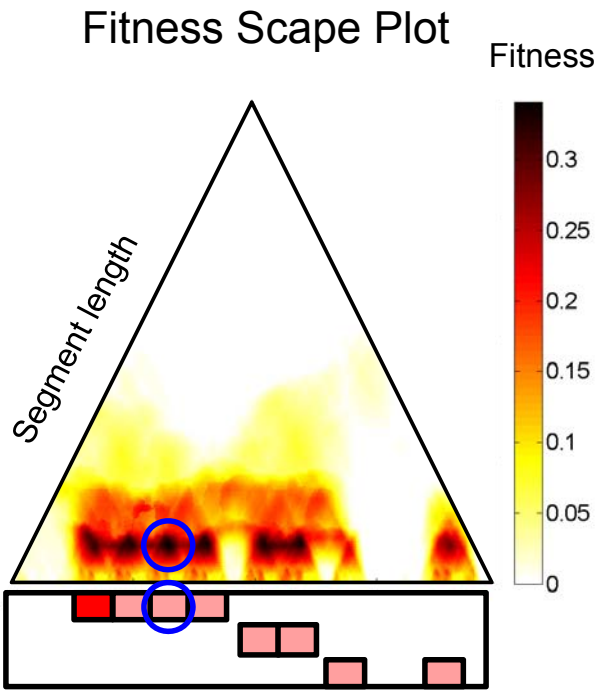




Example: Zager & Evans “In The Year 2525”

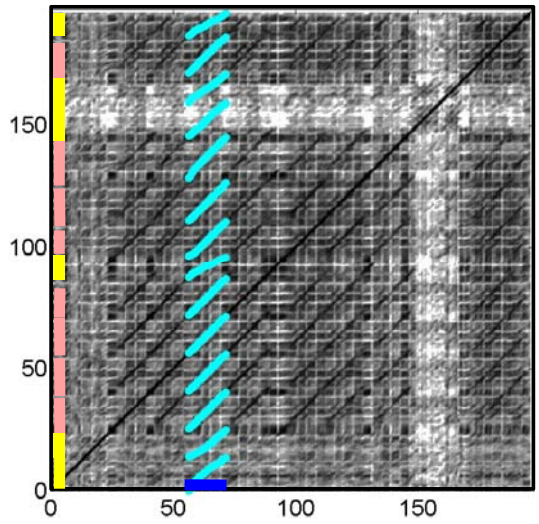
Thumbnail

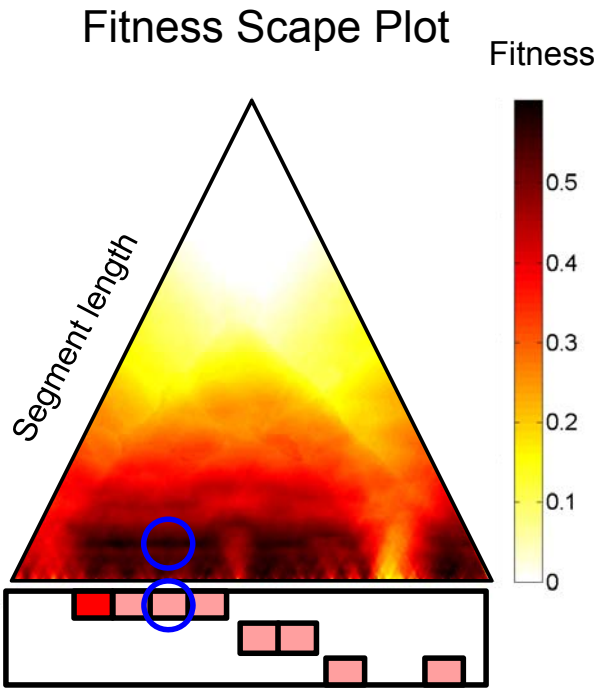




Example: Zager & Evans “In The Year 2525”

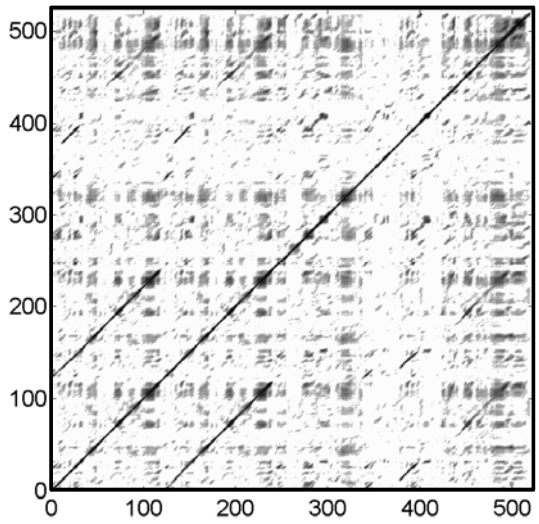
Thumbnail

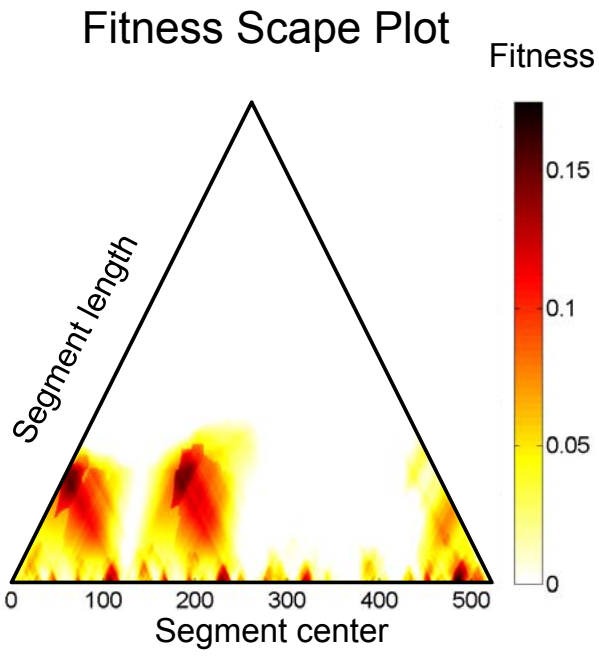




Example: Zager & Evans “In The Year 2525”

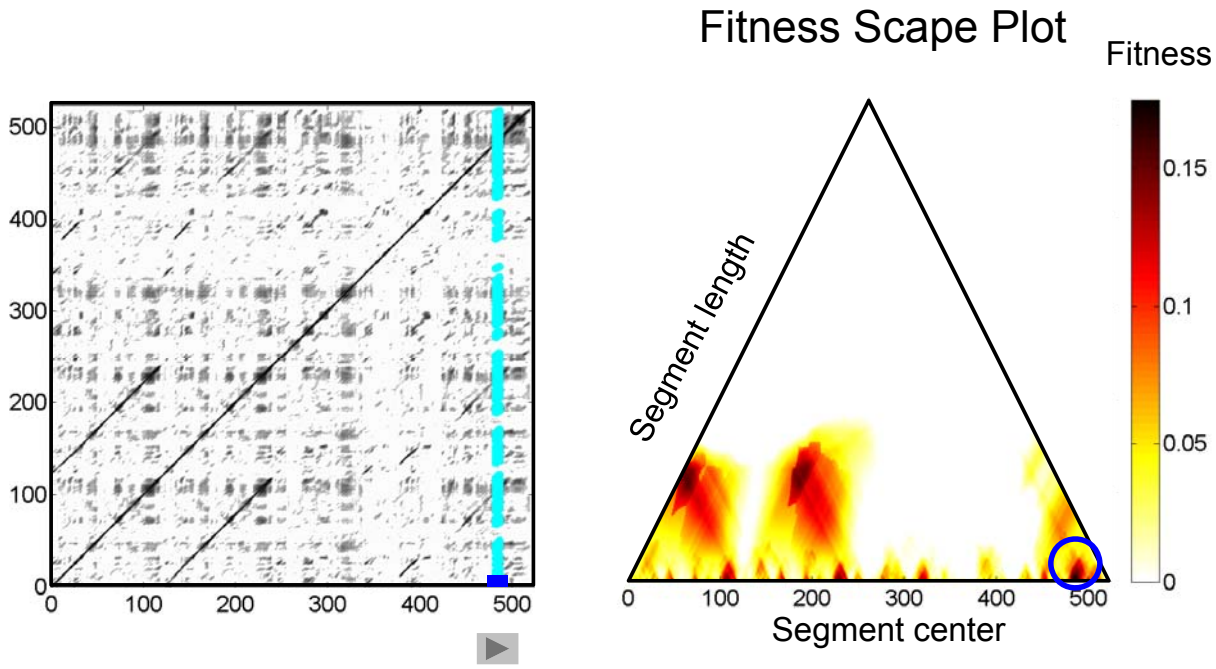
Thumbnail





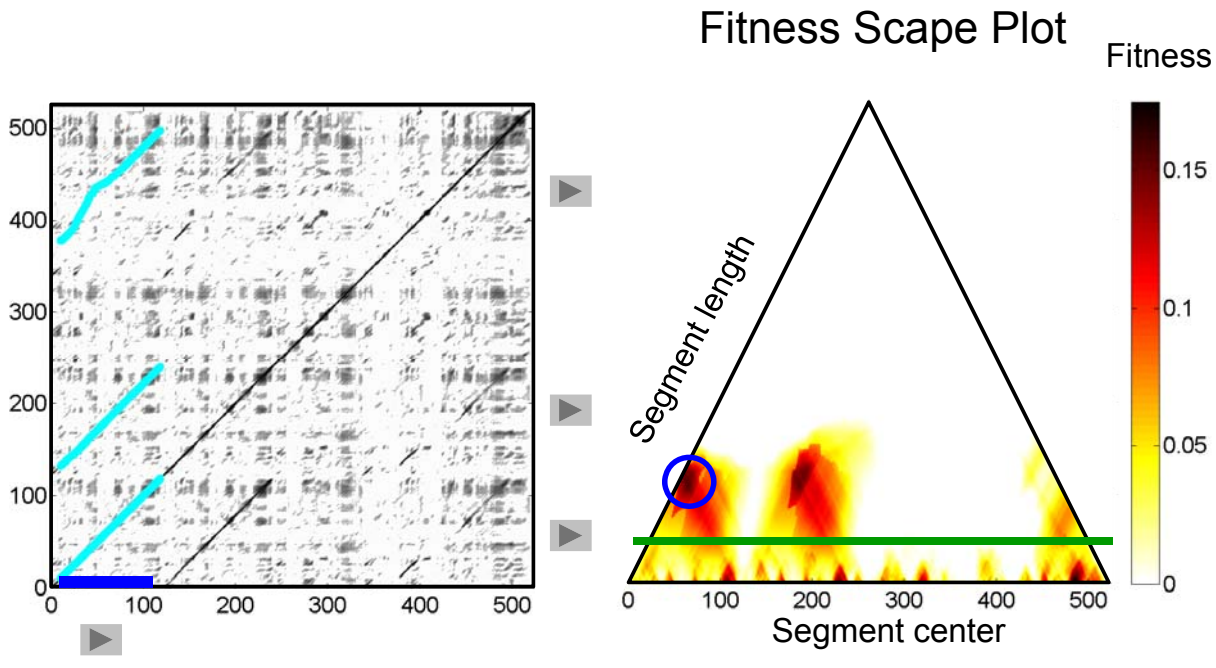
Example: Beethoven “Tempest”, Pollini

Thumbnail



Example: Beethoven “Tempest”, Pollini

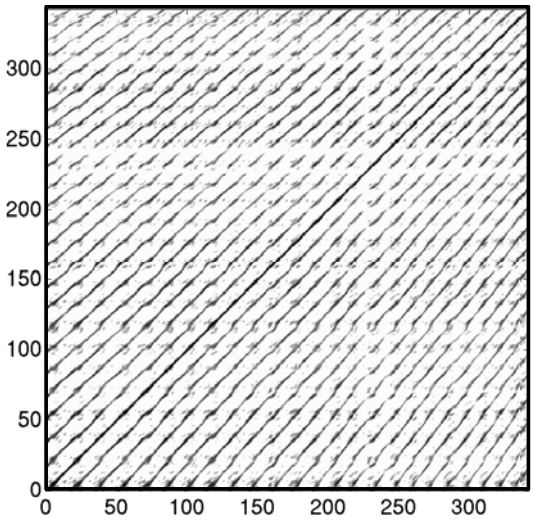
Thumbnail

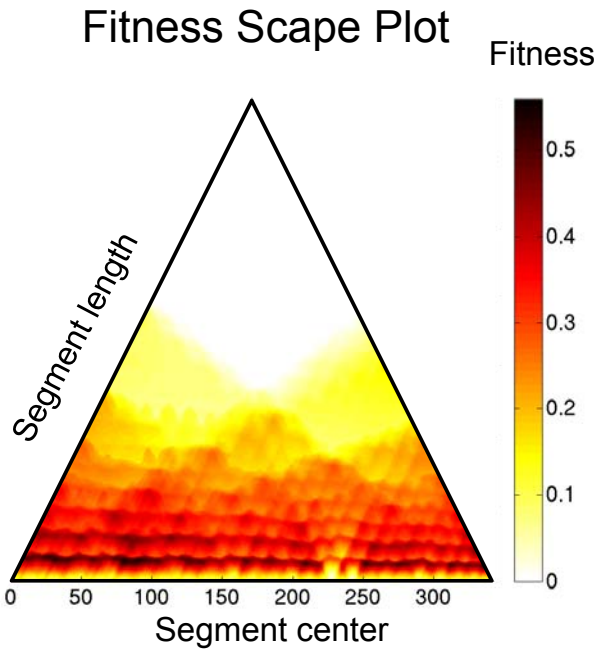


Example: Beethoven “Tempest”, Pollini

Musical knowledge: Minimum length for thumbnail

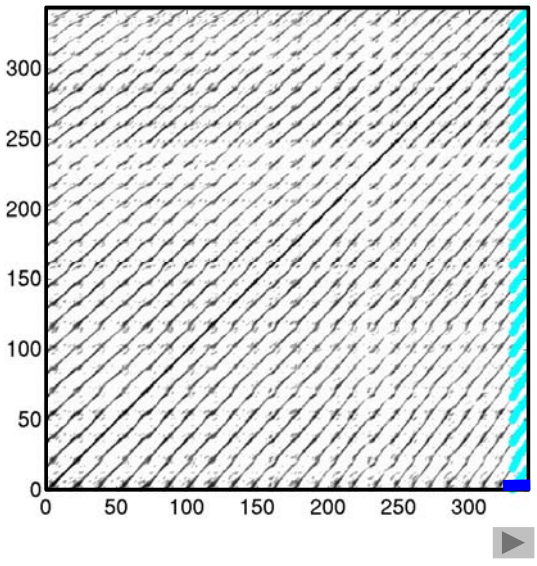
Thumbnail

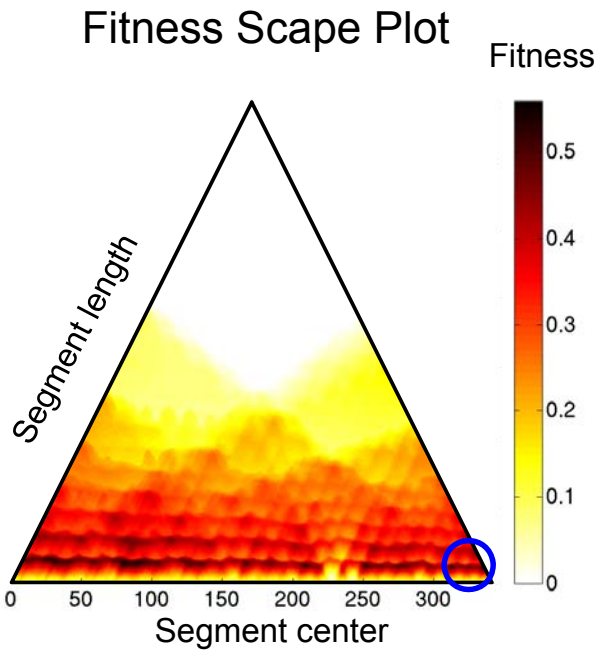




Example: NLB72246

Thumbnail

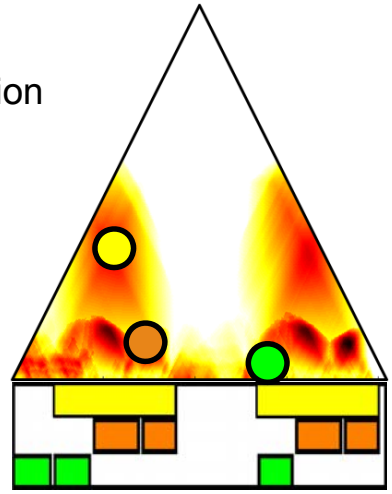




Example: NLB72246

Conclusions

- Path family: Couples path extraction and grouping
- Fitness: Quality of segment in context of entire recording
  - Combination of score and coverage
  - Trivial self-explanation are disregarded
- Thumbnail: Segment of maximal fitness
- Fintness scape plot: Global structure visualization



Future work:

- Multiscale approach
- Combination with novelty detection
- Interface for structure navigation