

Lecture

Information Retrieval for Music and Motion

Meinard Müller

Summer Term 2008

Audio Structure Analysis



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, symphony, concerto)
- Detection of other hidden relationships

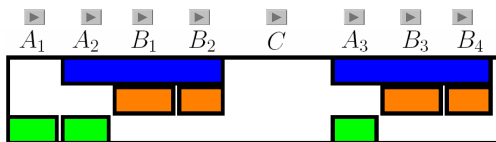
2

Audio Structure Analysis

Given: CD recording

Goal: Automatic extraction of the **repetitive structure**
(or of the **musical form**)

Example: Brahms Hungarian Dance No. 5 (Ormandy)



3

Audio Structure Analysis

- 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. Multimedia 2005)
- Goto (IEEE Trans. Audio 2006)
- Müller/Kurth (EURASIP 2007)
- Rhodes/Casey (ISMIR 2007)
- Peeters (ISMIR 2007)

4

Audio Structure Analysis

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

5

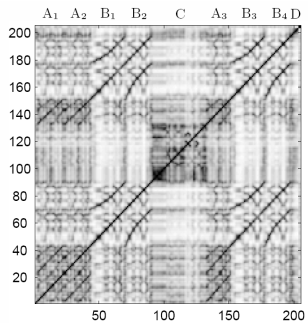
Audio Structure Analysis

- Audio $\rightsquigarrow V := (v^1, v^2, \dots, v^N)$
- $v^n = 12$ -dimensional normalized chroma vector
- Local cost measure $c : \mathbb{R}^{12} \times \mathbb{R}^{12} \rightarrow \mathbb{R}$
 $c(v^n, w^m) := 1 - \langle v^n, w^m \rangle$
- $N \times N$ cost matrix $C(n, m) := c(v^n, w^m)$
 \rightsquigarrow quadratic self-similarity matrix

6

Audio Structure Analysis

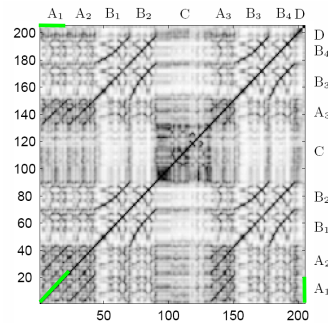
Self-similarity matrix



7

Audio Structure Analysis

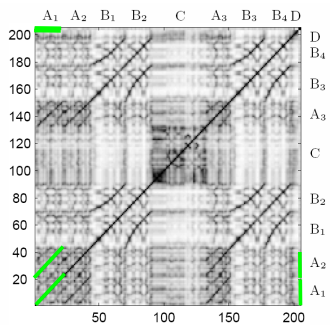
Self-similarity matrix



8

Audio Structure Analysis

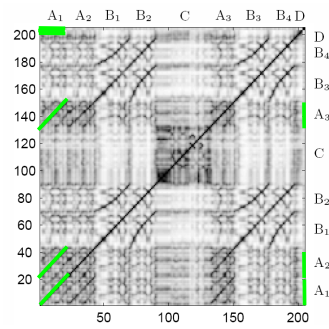
Self-similarity matrix



9

Audio Structure Analysis

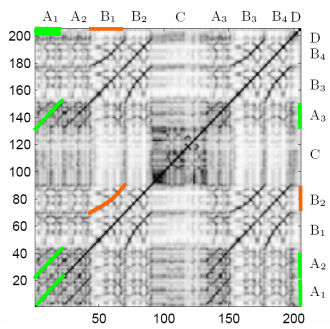
Self-similarity matrix



10

Audio Structure Analysis

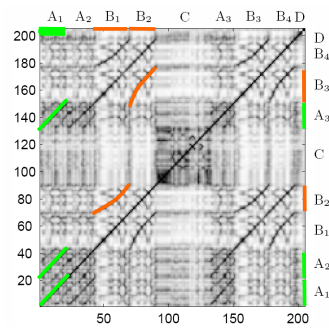
Self-similarity matrix



11

Audio Structure Analysis

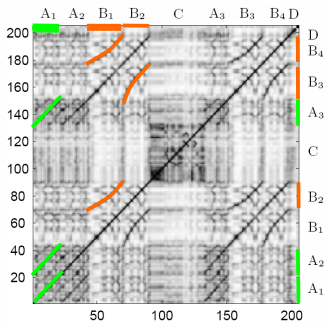
Self-similarity matrix



12

Audio Structure Analysis

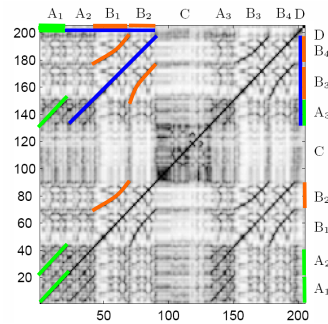
Self-similarity matrix



13

Audio Structure Analysis

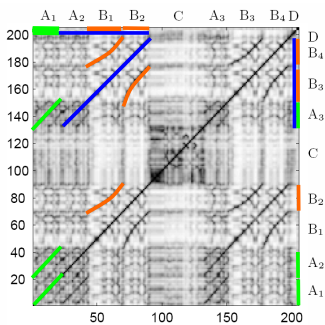
Self-similarity matrix



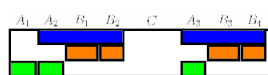
14

Audio Structure Analysis

Self-similarity matrix



Similarity cluster



15

Matrix Enhancement

Challenge: Presence of musical variations

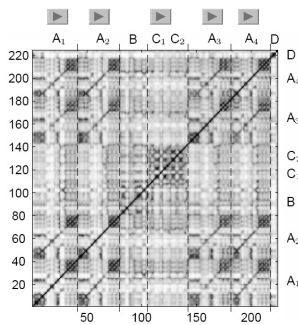
- Fragmented paths and gaps
- Paths of poor quality
- Regions of constant (low) cost
- Curved paths

Idea: Enhancement of path structure

16

Matrix Enhancement

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



17

Matrix Enhancement

Idea: Usage of contextual information (Foote 1999)

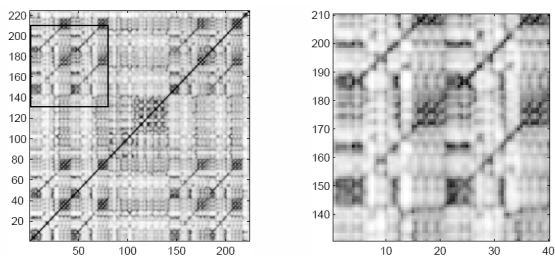
$$C_L(n, m) := \frac{1}{L} \sum_{\ell=0}^{L-1} c(v_{n+\ell}, v_{m+\ell})$$

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

↪ smoothing effect

18

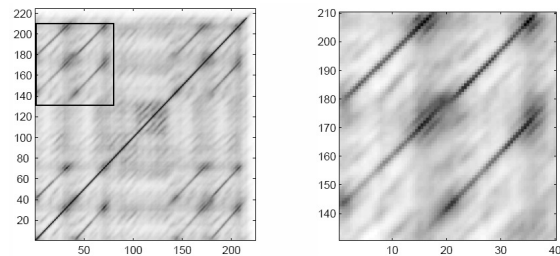
Matrix Enhancement (Shostakovich)



Cost matrix C

19

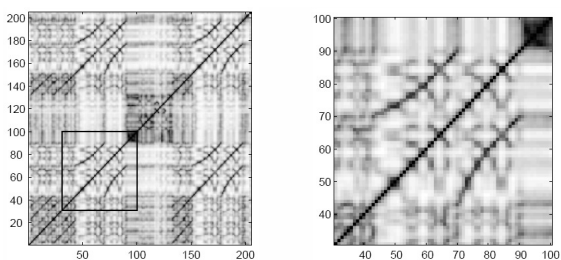
Matrix Enhancement (Shostakovich)



Enhanced cost matrix C_L

20

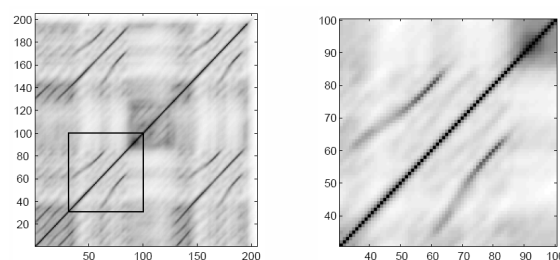
Matrix Enhancement (Brahms)



Cost matrix C

21

Matrix Enhancement (Brahms)



Enhanced cost matrix C_L

Problem: Relative tempo differences are smoothed out

22

Matrix Enhancement

Idea: Smoothing along various directions and minimizing over all directions

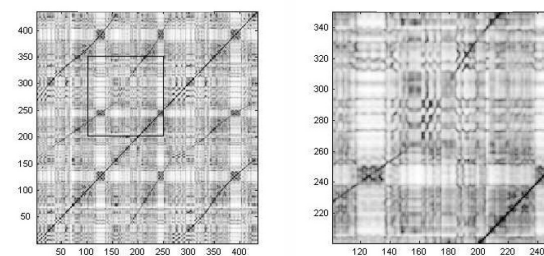
$$C_L^{\min}(n, m) := \min_k C_L^{\text{slope}_k}(n, m)$$

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

↔ tempo changes of -30 to +40 percent

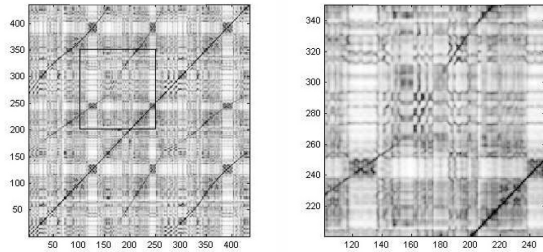
23

Matrix Enhancement



24

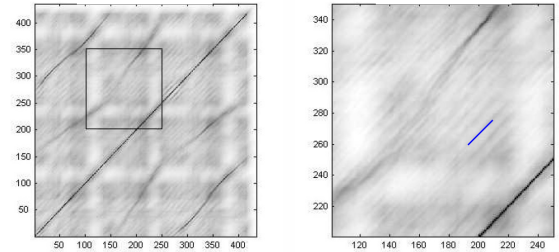
Matrix Enhancement



Cost matrix C

25

Matrix Enhancement

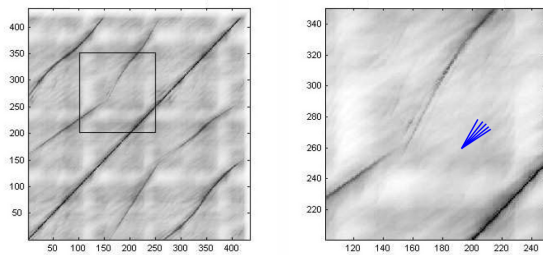


Cost matrix C_L with $L = 20$

Filtering along main diagonal

26

Matrix Enhancement

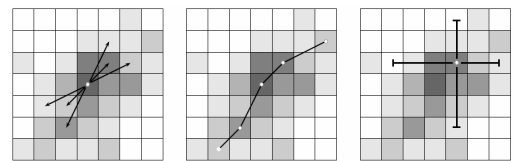


Cost matrix C_L^{\min} with $L = 20$

Filtering along 8 different directions and minimizing

27

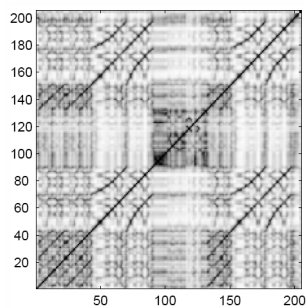
Path Extraction



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

28

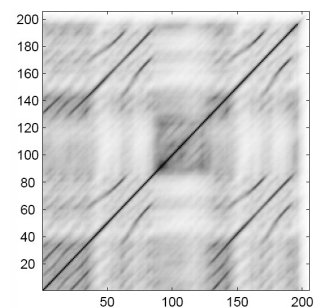
Path Extraction



Cost matrix C

29

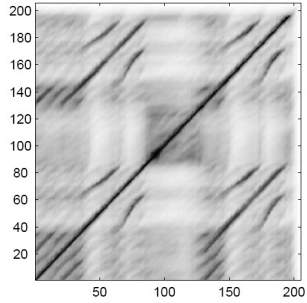
Path Extraction



Enhanced cost matrix C_L

30

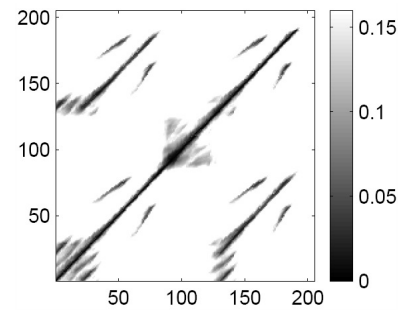
Path Extraction



Enhanced cost matrix C_L^{\min}

31

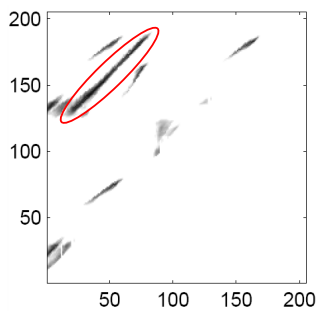
Path Extraction



Thresholded C_L^{\min}

32

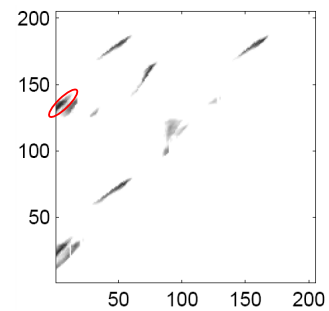
Path Extraction



Thresholded C_L^{\min} , upper left

33

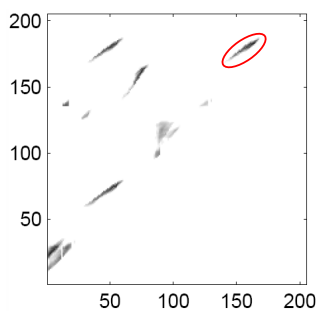
Path Extraction



Path removal

34

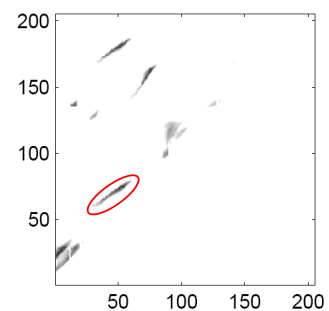
Path Extraction



Path removal

35

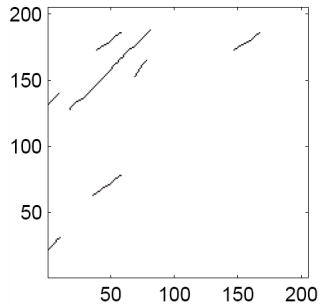
Path Extraction



Path removal

36

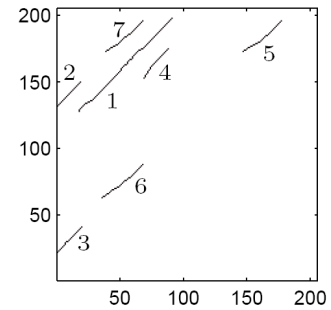
Path Extraction



Extracted paths

37

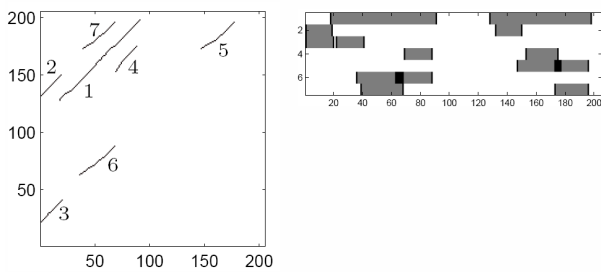
Path Extraction



Extracted paths after postprocessing

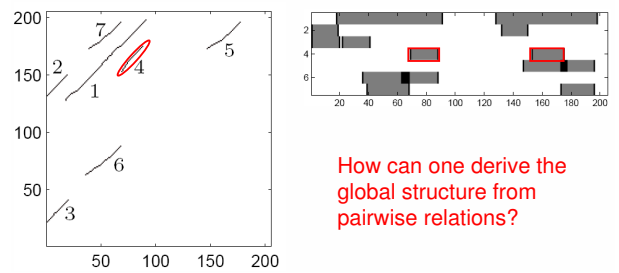
38

Global Structure



39

Global Structure



How can one derive the global structure from pairwise relations?

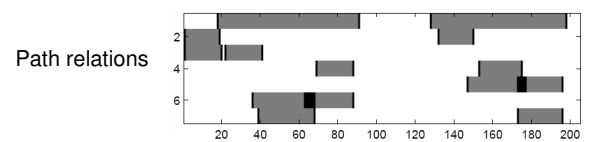
40

Global Structure

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

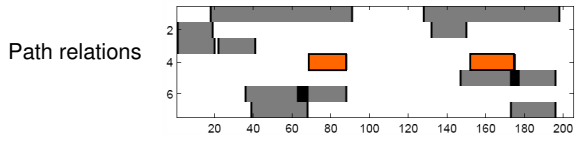
41

Global Structure



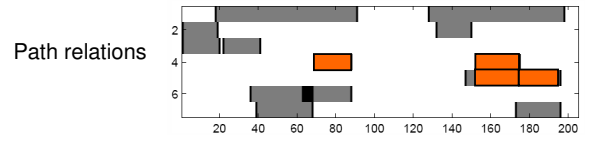
42

Global Structure



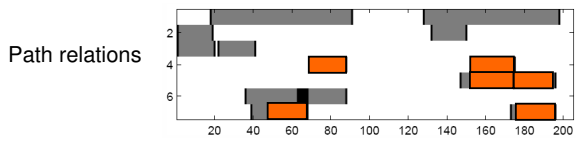
43

Global Structure



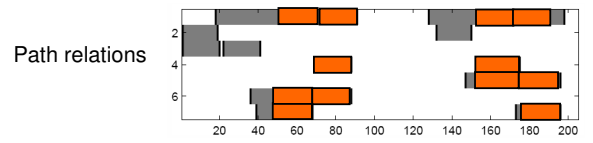
44

Global Structure



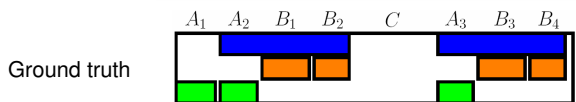
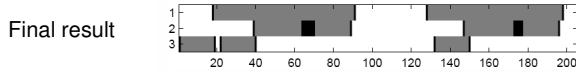
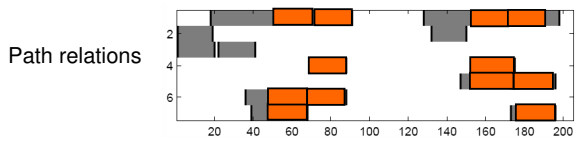
45

Global Structure



46

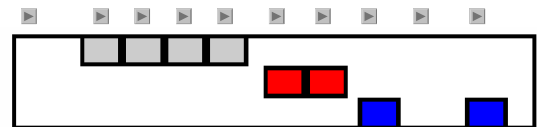
Global Structure



47

Transposition Invariance

Example: Zager & Evans "In The Year 2525"



48

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

49

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

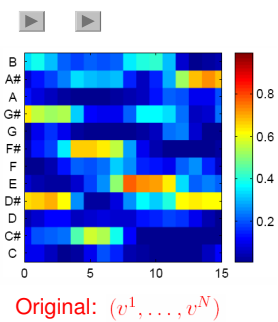
Müller/Clausen (ISMIR 2007)

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

50

Transposition Invariance

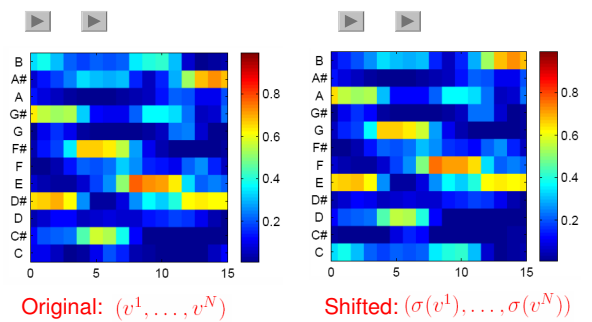
Example: Zager & Evans "In The Year 2525"



51

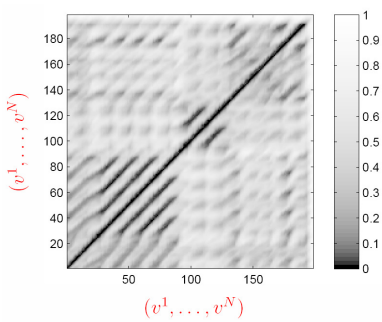
Transposition Invariance

Example: Zager & Evans "In The Year 2525"



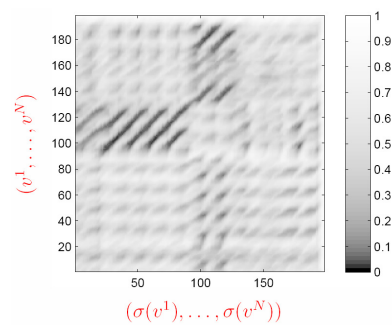
52

Transposition Invariance



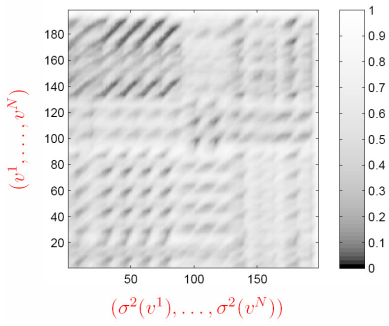
53

Transposition Invariance



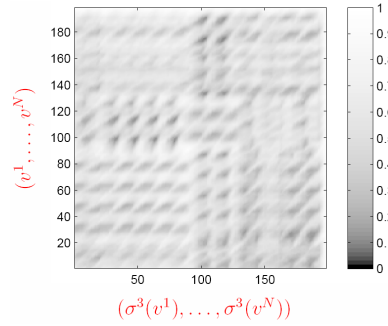
54

Transposition Invariance



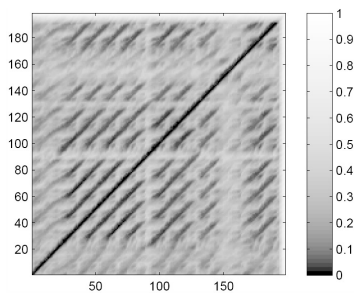
55

Transposition Invariance



56

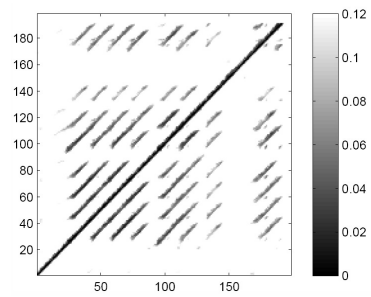
Transposition Invariance



Minimize over all twelve matrices

57

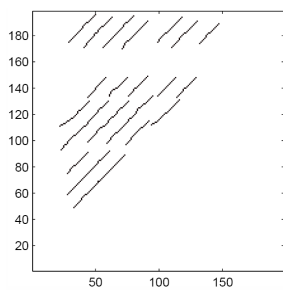
Transposition Invariance



Thresholded self-similarity matrix

58

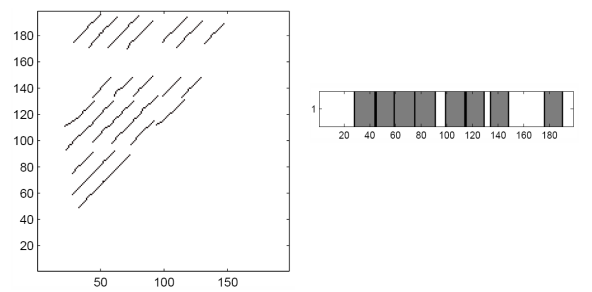
Transposition Invariance



Path extraction

59

Transposition Invariance



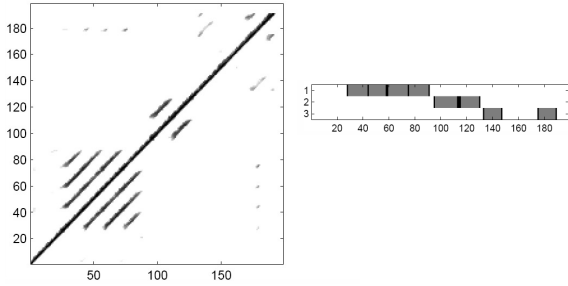
Path extraction

Computation of similarity clusters

60

Transposition Invariance

Stabilizing effect

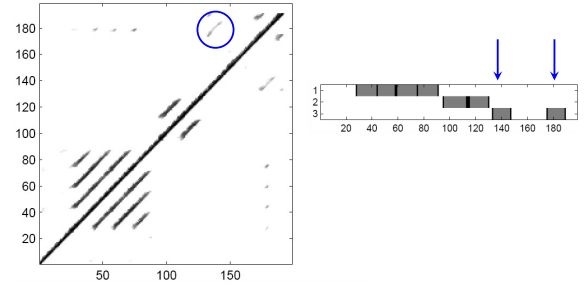


Self-similarity matrix (thresholded)

61

Transposition Invariance

Stabilizing effect

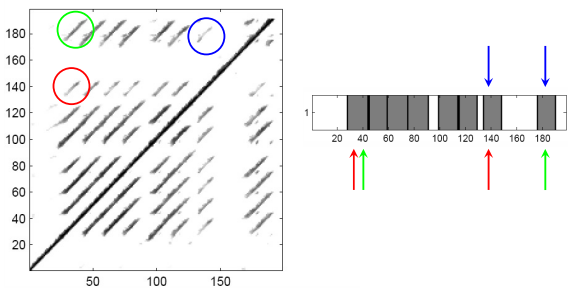


Self-similarity matrix (thresholded)

62

Transposition Invariance

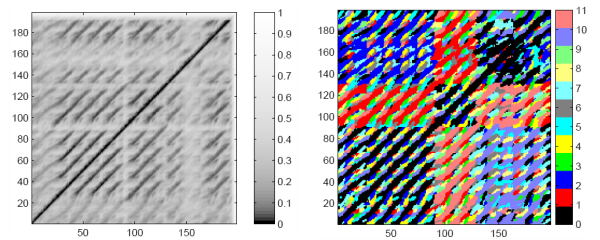
Stabilizing effect



Transposition-invariant self-similarity matrix (thresholded)

63

Transposition Invariance

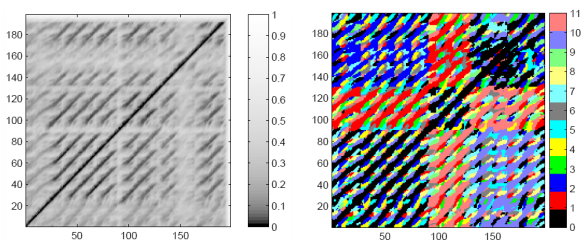


Transposition-invariant matrix

Minimizing shift index

64

Transposition Invariance

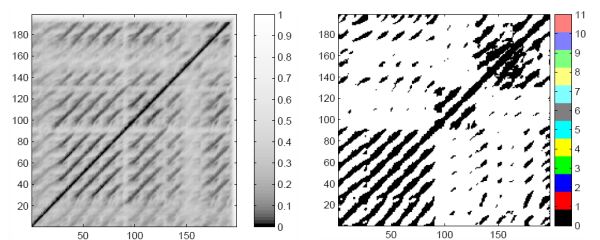


Transposition-invariant matrix

Minimizing shift index

65

Transposition Invariance

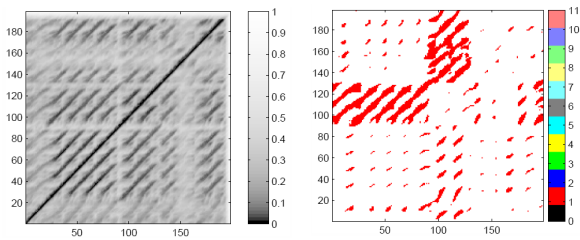


Transposition-invariant matrix

Minimizing shift index = 0

66

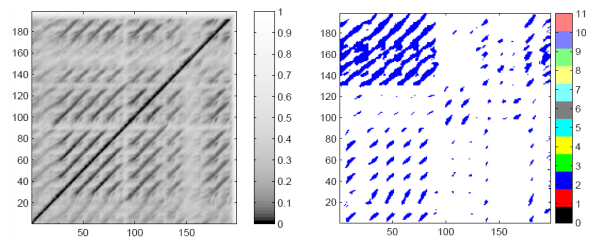
Transposition Invariance



Transposition-invariant matrix Minimizing shift index = 1

67

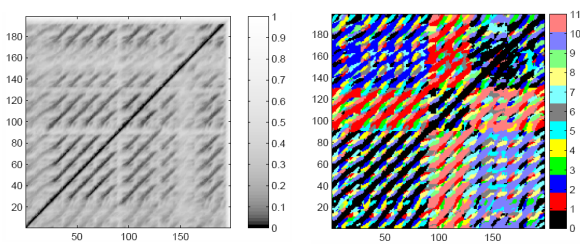
Transposition Invariance



Transposition-invariant matrix Minimizing shift index = 2

68

Transposition Invariance

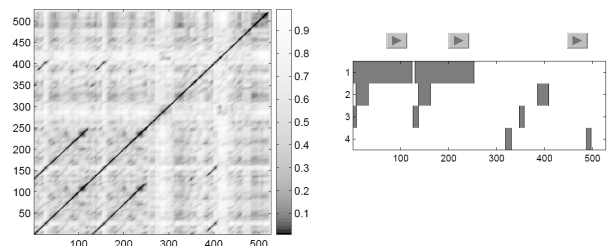


Serra/Gomez (ICASSP 2008): Used for Cover Song ID
Discrete structure \rightsquigarrow suitable for indexing?

69

Transposition Invariance

Example: Beethoven "Tempest"

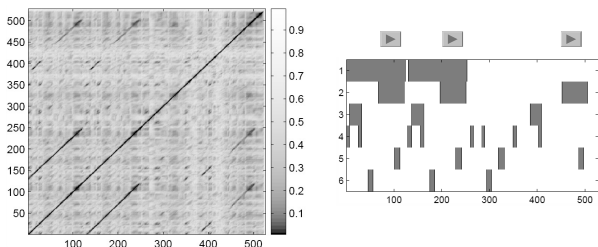


Self-similarity matrix

70

Transposition Invariance

Example: Beethoven "Tempest"



Transposition-invariant self-similarity matrix

71

Conclusions: Audio Structure Analysis

Challenge: Musical variations

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

72

Conclusions: Audio Structure Analysis

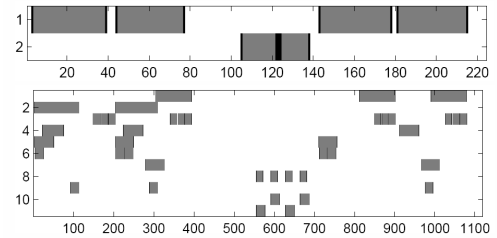
Strategy: Matrix enhancement

- 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)

73

Conclusions: Audio Structure Analysis

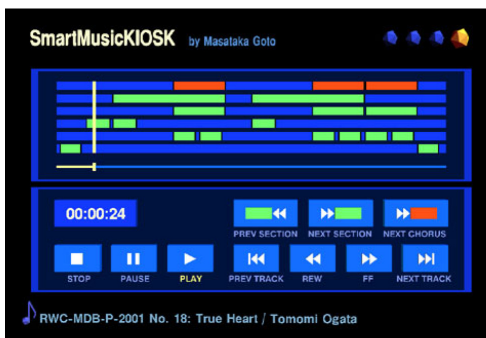
Challenge: Hierarchical structure of music



Rhodes/Casey (ISMIR 2007)

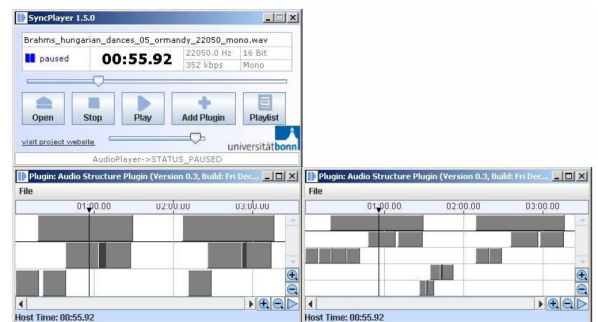
74

System: SmartMusicKiosk (Goto)



75

System: SyncPlayer/AudioStructure



76