



Introduction to **Audio Content Analysis**

Module 9.7: Music Structure Detection

alexander lerch

introduction

overview

corresponding textbook section

section 9.7

■ lecture content

- structure in music
- self similarity and self distance matrices
- structure detection approaches

■ learning objectives

- summarize basic difficulties in ground truth annotations of musical structure
- explain and interpret self similarity and self distance matrices
- summarize three domains for approaching music structure detection



introduction

overview

corresponding textbook section

section 9.7

■ lecture content

- structure in music
- self similarity and self distance matrices
- structure detection approaches

■ learning objectives

- summarize basic difficulties in ground truth annotations of musical structure
- explain and interpret self similarity and self distance matrices
- summarize three domains for approaching music structure detection



music structure

introduction

- **music is inherently formal**/organized/structural
- various **hierarchical structural levels**
 - *groups of notes* build rhythmic/melodic/harmonic patterns
 - *measures* group multiple events
 - *phrases* group several measures
 - *sections* contain several phrases
 - several sections can comprise *piece/movement*
 - ...
- **grouping** of musical elements/patterns is influenced by
 - 1 *contrasts & novelty*
 - ▶ rhythmic, harmonic, melodic patterns
 - 2 *similarity and repetitions*
 - ▶ rhythmic, harmonic, melodic patterns
 - 3 *homogeneity* within a section
 - ▶ instrumentation, tempo, harmony

music structure

introduction

- **music is inherently formal**/organized/structural
- various **hierarchical structural levels**
 - *groups of notes* build rhythmic/melodic/harmonic patterns
 - *measures* group multiple events
 - *phrases* group several measures
 - *sections* contain several phrases
 - several sections can comprise *piece/movement*
 - ...
- **grouping** of musical elements/patterns is influenced by
 - 1 *contrasts & novelty*
 - ▶ rhythmic, harmonic, melodic patterns
 - 2 *similarity and repetitions*
 - ▶ rhythmic, harmonic, melodic patterns
 - 3 *homogeneity* within a section
 - ▶ instrumentation, tempo, harmony

music structure analysis

introduction

■ objective

- reveal structural properties and relationships
- generate a list of parts and repetitions

■ typical processing steps

- 1 *feature extraction*
- 2 Self Distance Matrix (SDM) or *Self Similarity Matrix (SSM)* computation
- 3 *segment detection* based on
 - ▶ novelty
 - ▶ homogeneity
 - ▶ repetition

overview
o

intro
oo●

ssm
ooo

novelty
o

homogeneity
oo

repetition
oo

eval
o

summary
o

music structure analysis

example

Georgia Tech | Center for Music Technology
College of Design



music structure analysis

features 1/2

■ features from **all categories** can have impact on structure

- timbre
 - ▶ instrumentation, playing technique, effects, ...
- tonal content
 - ▶ melodic and harmonic patterns, range, ...
- rhythm content
 - ▶ tempo, rhythmic patterns, ...
- dynamics
 - ▶ loudness, range, ...

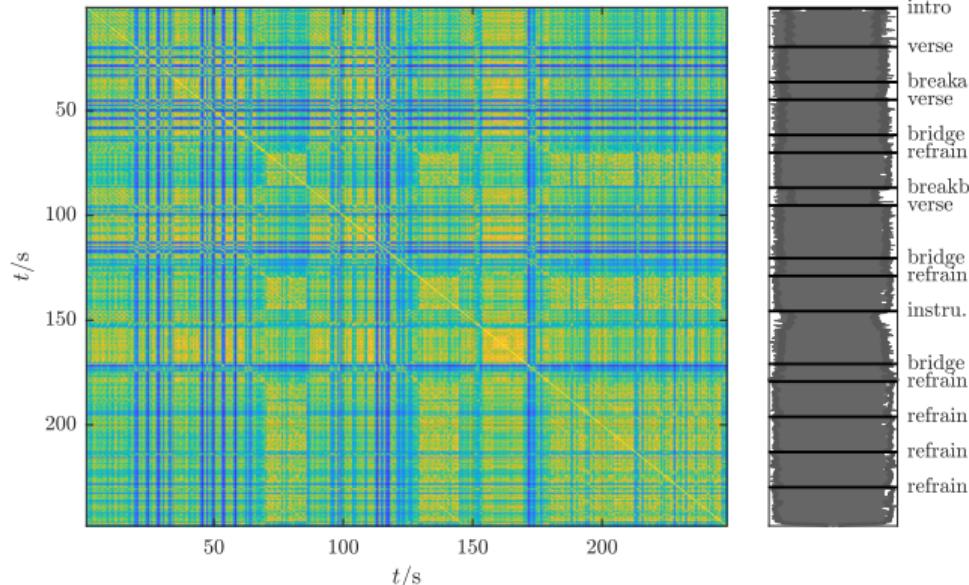
■ **feature aggregation**

- use texture window, or
- aggregate features per beat or downbeat

music structure analysis

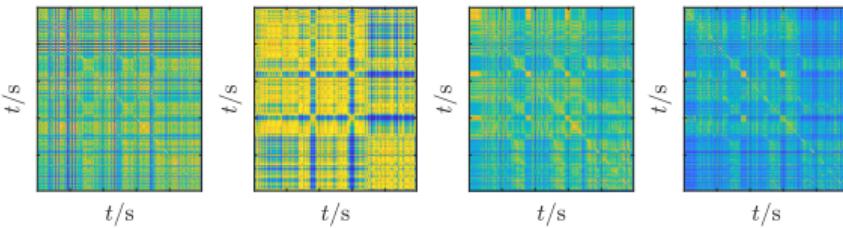
self similarity matrix

$$S(n_A, n_B) = s(v(n_A), v(n_B))$$

matlab source: [plotSSm.m](#)

music structure analysis

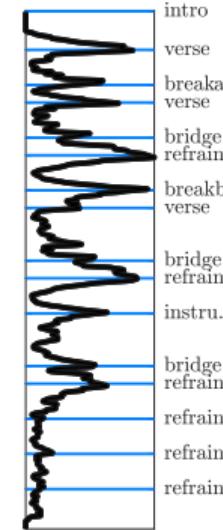
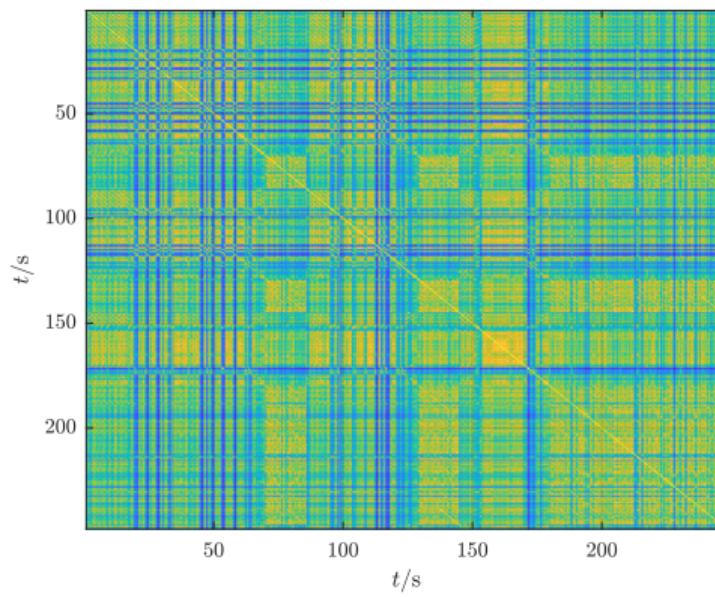
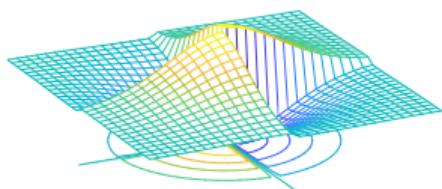
feature dependency of similarity



Features (left to right): Pitch Chroma, RMS, MFCCs, Mel Spectrogram

music structure analysis

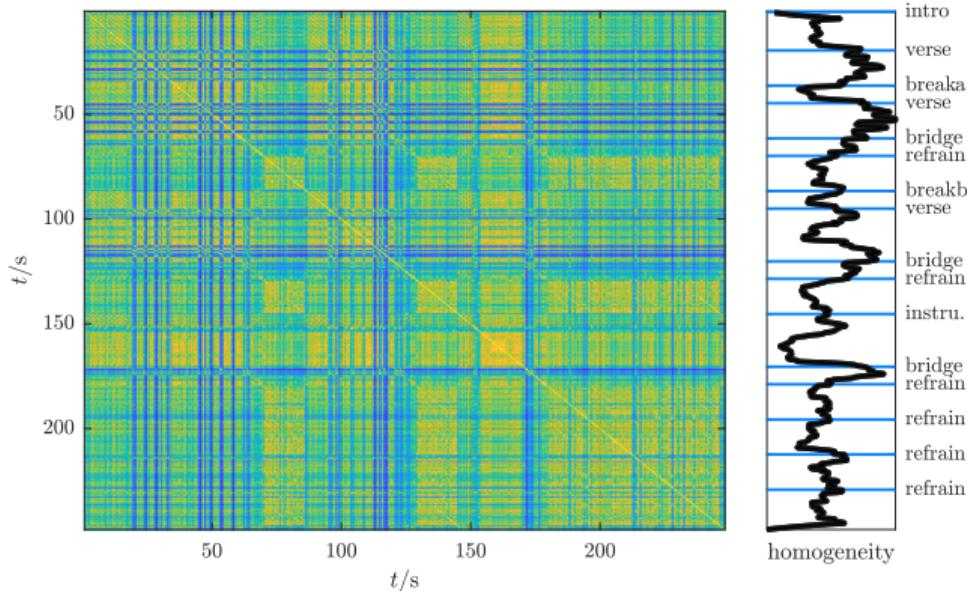
novelty analysis



Novelty

music structure analysis

homogeneity analysis 1/2



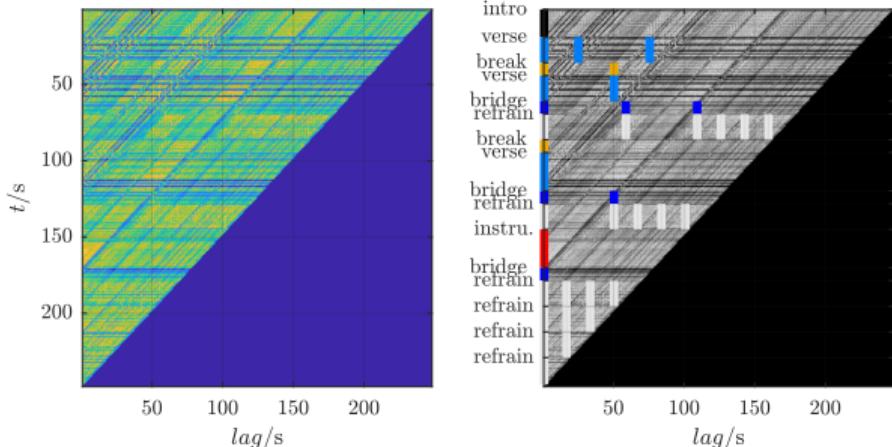
music structure analysis

homogeneity analysis 2/2

- can also be used as post-processing step after novelty-based approach, e.g.
 - 1 describe each segment with features
 - 2 cluster and see which segments are grouped together

music structure analysis

repetition analysis 1/2

matlab source: [plotSsmRotated.m](#)

music structure analysis

repetition analysis 2/2

- while in many cases it 'looks' easy, automatic extraction is **error-prone**

⇒ typical approaches for **enhancing** the distance/similarity/lag matrix

- filtering (low pass smoothing, high pass edge detection)
- use matrices with different time resolutions
- image processing methods (e.g., erosion & dilation)
- thresholding
- "path search" through probability matrix

music structure analysis

repetition analysis 2/2

- while in many cases it 'looks' easy, automatic extraction is **error-prone**

⇒ typical approaches for **enhancing** the distance/similarity/lag matrix

- filtering (low pass smoothing, high pass edge detection)
- use matrices with different time resolutions
- image processing methods (e.g., erosion & dilation)
- thresholding
- "path search" through probability matrix

music structure analysis

evaluation

■ evaluation of structure detection **challenging**

- *ground truth*
 - ▶ structure itself may be ambiguous
 - ▶ depending on annotator, varying hierarchical level of labels, e.g.

ann 1	intro	A				A				outro
ann 2	intro	verse		chorus		verse		chorus		outro
ann 3	intro	V ₁	V ₂	C ₁	C ₂	V ₁	V ₂	C ₁	C ₂	outro

■ method and metric

- boundary matching
- frame level, e.g., pairwise match

■ typical range of results

- $F = 50 \dots 70\%$

music structure analysis

evaluation

■ evaluation of structure detection **challenging**

- *ground truth*
 - ▶ structure itself may be ambiguous
 - ▶ depending on annotator, varying hierarchical level of labels, e.g.

ann 1	intro	A				A				outro
ann 2	intro	verse		chorus		verse		chorus		outro
ann 3	intro	V ₁	V ₂	C ₁	C ₂	V ₁	V ₂	C ₁	C ₂	outro

■ *method and metric*

- boundary matching
- frame level, e.g., pairwise match

■ typical range of results

- $F = 50 \dots 70\%$

music structure analysis

evaluation

■ evaluation of structure detection **challenging**

- *ground truth*
 - ▶ structure itself may be ambiguous
 - ▶ depending on annotator, varying hierarchical level of labels, e.g.

ann 1	intro	A				A				outro
ann 2	intro	verse		chorus		verse		chorus		outro
ann 3	intro	V ₁	V ₂	C ₁	C ₂	V ₁	V ₂	C ₁	C ₂	outro

■ *method and metric*

- boundary matching
- frame level, e.g., pairwise match

■ typical range of results

- $F = 50 \dots 70\%$

summary

lecture content

■ self similarity/distance matrices

- shows pairwise similarities/distances
- depends on input features

■ structure detection

- 1 novelty
- 2 homogeneity
- 3 repetitions

