# Introduction to Audio Content Analysis

## Module 9.7: Music Structure Detection

alexander lerch

Georgia Tech Center for Music Technology
College of Design

## 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

## 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 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 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

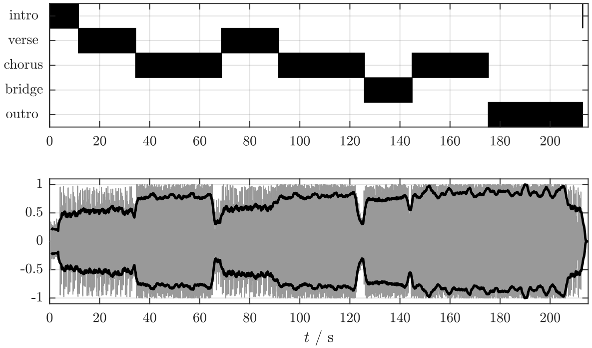
- **objective**
  - reveal structural properties and relationships
  - generate a list of parts and repetitions

- typical **processing steps**
  1. feature extraction
  2. Self Distance Matrix (SDM) or Self Similarity Matrix (SSM)
  3. detect segments
     - novelty
     - homogeneity
     - repetition

# music structure analysis

## example





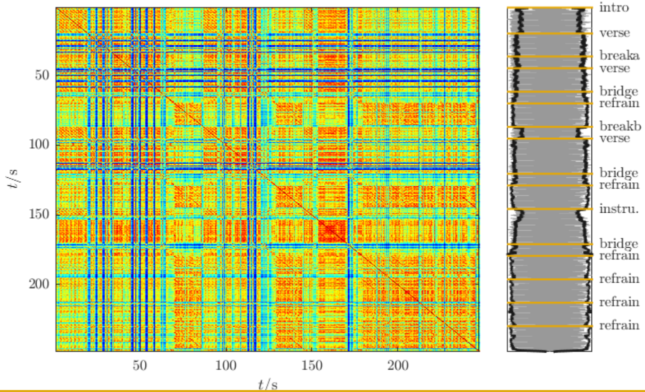

matlab source: plotStructure.m

- 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



$$S(n_\mathrm{A}, n_\mathrm{B}) = \mathrm{s}\big(v(n_\mathrm{A}), v(n_\mathrm{B})\big)$$

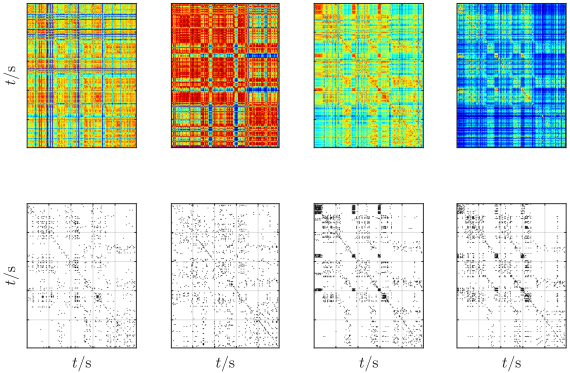



matlab source: plotSelfSimilarityMatrix.m

# music structure analysis

## feature dependency of similarity

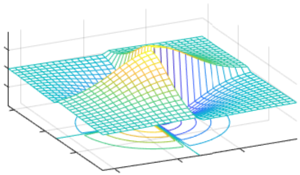


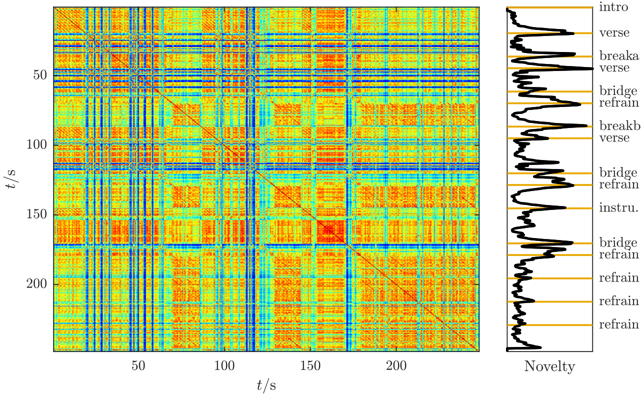



matlab source: plotSsmFeatures.m

# music structure analysis
## novelty analysis





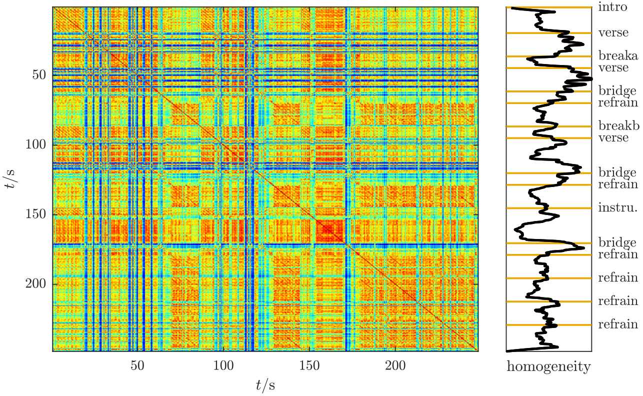



matlab source: plotCheckerBoard.m

# music structure analysis

## homogeneity analysis 1/2

Georgia Tech | Center for Music Technology College of Design




matlab source: plotSsmLowPass.m

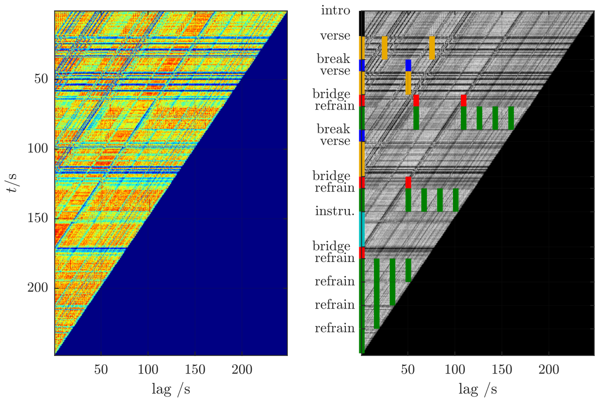
# music structure analysis
## homogeneity analysis 2/2



- 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**

$\Rightarrow$ 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

- 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
  - $F = 50 \ldots 70\%$

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

<table>
<tr><td>ann 1</td><td>intro</td><td colspan="4">A</td><td colspan="4">A</td><td>outro</td></tr>
<tr><td>ann 2</td><td>intro</td><td colspan="2">verse</td><td colspan="2">chorus</td><td colspan="2">verse</td><td colspan="2">chorus</td><td>outro</td></tr>
<tr><td>ann 3</td><td>intro</td><td>$V_1$</td><td>$V_2$</td><td>$C_1$</td><td>$C_2$</td><td>$V_1$</td><td>$V_2$</td><td>$C_1$</td><td>$C_2$</td><td>outro</td></tr>
</table>

- *method and metric*
  - boundary matching
  - frame level, e.g., pairwise match

- typical range of results
  - $F = 50 \ldots 70\%$

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

<table>
<tr><td>ann 1</td><td>intro</td><td colspan="4">A</td><td colspan="4">A</td><td>outro</td></tr>
<tr><td>ann 2</td><td>intro</td><td colspan="2">verse</td><td colspan="2">chorus</td><td colspan="2">verse</td><td colspan="2">chorus</td><td>outro</td></tr>
<tr><td>ann 3</td><td>intro</td><td>$V_1$</td><td>$V_2$</td><td>$C_1$</td><td>$C_2$</td><td>$V_1$</td><td>$V_2$</td><td>$C_1$</td><td>$C_2$</td><td>outro</td></tr>
</table>

- *method and metric*
  - boundary matching
  - frame level, e.g., pairwise match

- typical range of results
  - $F = 50 \ldots 70\%$

# summary
## lecture content



- **self similarity/distance matrices**
  - shows pairwise similarities/distances
  - depends on input features

- **structure detection**
  1. novelty
  2. homogeneity
  3. repetitions