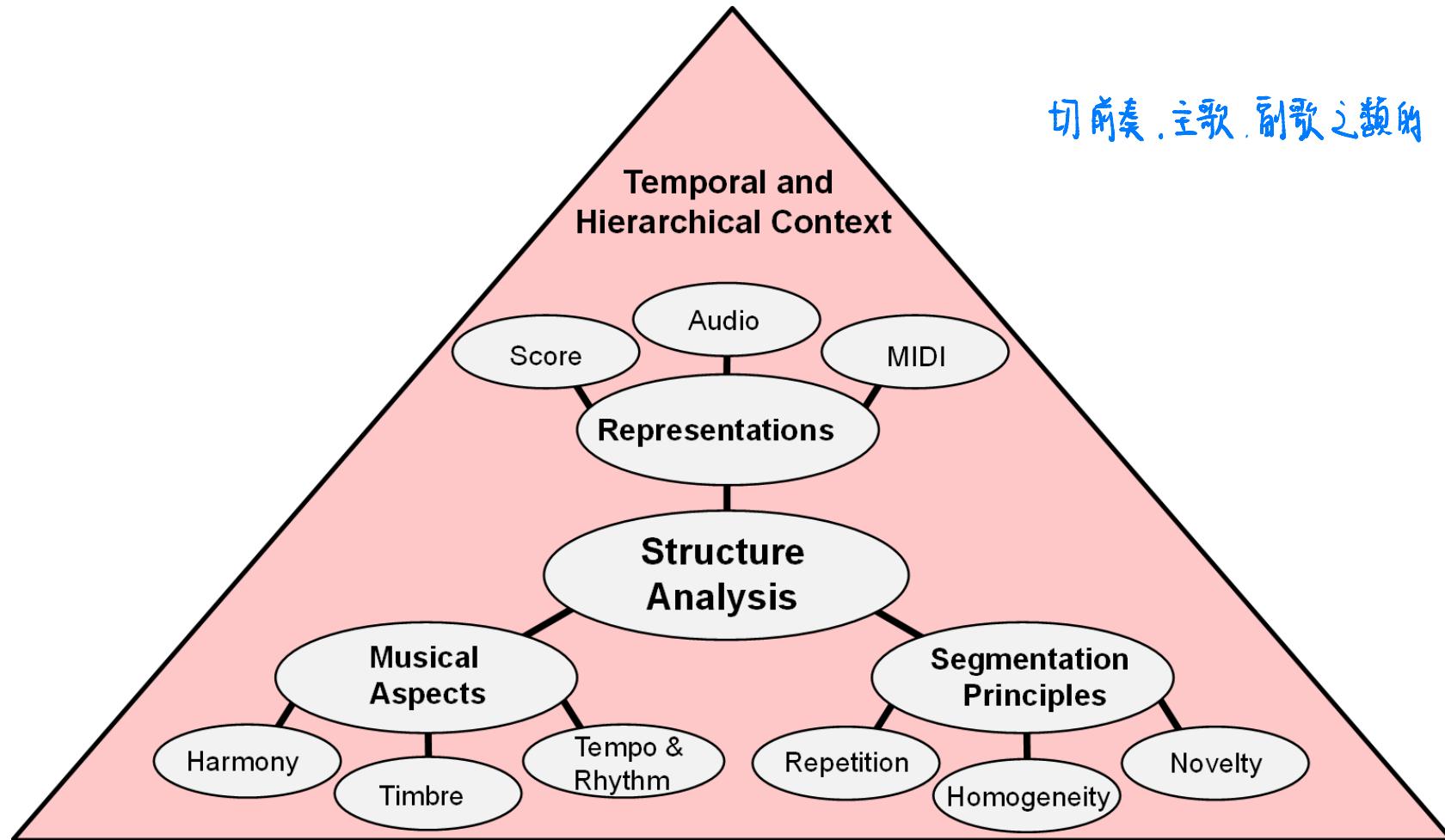


Structural segmentation

Li Su

2019/05/07

The structure of music 音樂結構分析



Introduction

- The structure of Brahms' Hungarian Dance no.5

Two hierarchies

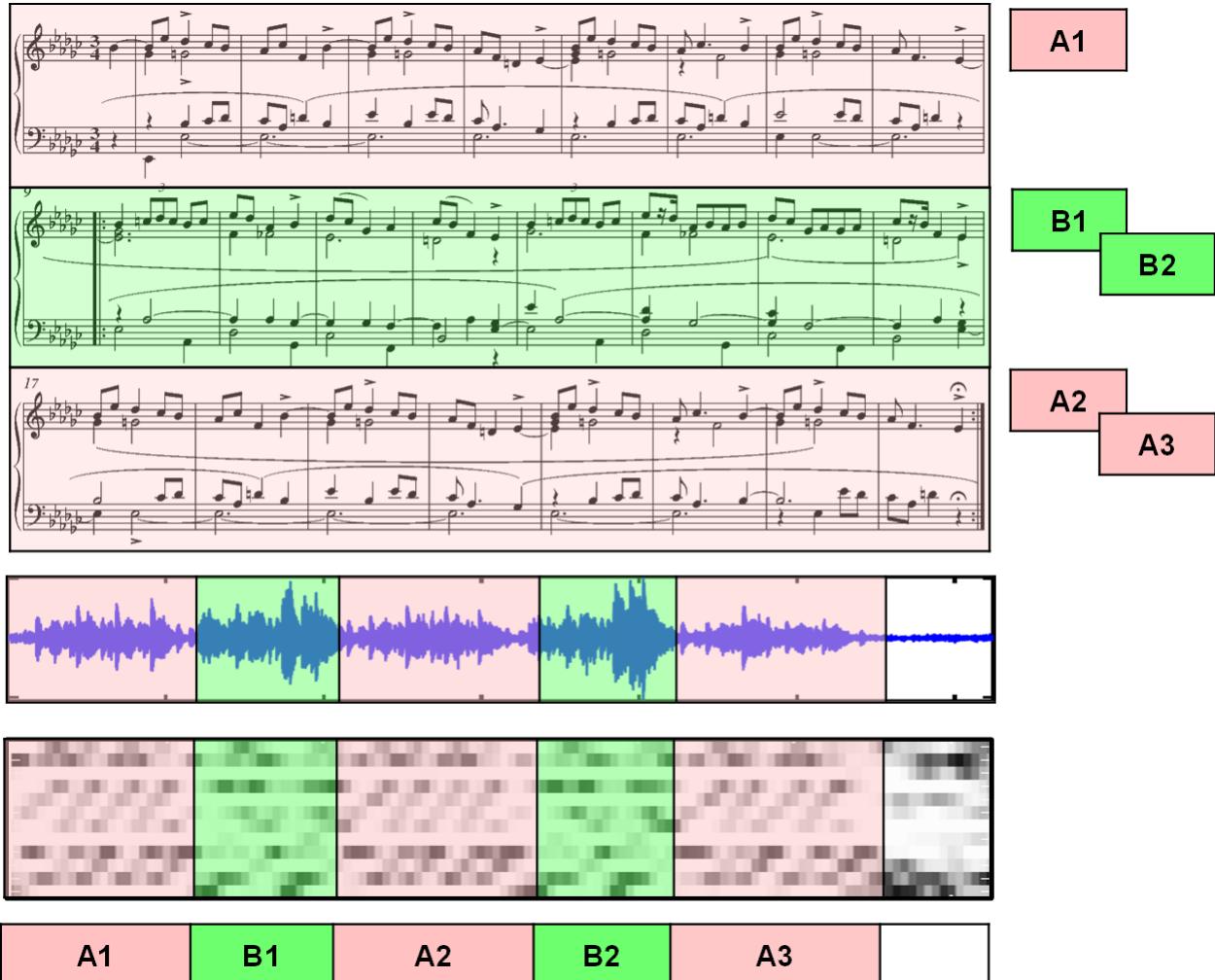
The image shows a musical score with two main sections: **Allegro.** and **Vivace.** The **Allegro.** section starts with a 2/4 time signature and transitions to 3/4. The **Vivace.** section follows with a 2/2 time signature. The score is annotated with various performance instructions and boxes:

- Allegro.:**
 - Measures 1-4: No specific box.
 - Measures 5-8: Box A1 (pink).
 - Measures 9-12: Box A2 (pink).
 - Measures 13-16: Box B1 (green).
 - Measures 17-20: Box B2 (green).
 - Measures 21-24: Box C (light blue).
 - Measures 25-28: Box A3 (pink).
 - Measures 29-32: Box B3 (green).
 - Measures 33-36: Box B4 (green).
 - Measures 37-40: Box D (yellow).
- Vivace.:**
 - Measures 1-4: Box A1 (pink).
 - Measures 5-8: Box A2 (pink).
 - Measures 9-12: Box B1 (green).
 - Measures 13-16: Box B2 (green).
 - Measures 17-20: Box C (light blue).
 - Measures 21-24: Box A3 (pink).
 - Measures 25-28: Box B3 (green).
 - Measures 29-32: Box B4 (green).
 - Measures 33-36: Box D (yellow).

Performance instructions visible in the score include: **marc.**, **in tempo**, **poco ritard.**, **poco rit.**, **poco ritard.**, **p in tempo**, **poco ritard.**, **p legg.**, **p in tempo**, **poco ritard.**, **poco rit.**, **Din tempo**, **E marcato**, and **p poco ritard.**.

Introduction

- Musical structure and signals



Structure segmentation

- General goal:
 - Divide an audio recording into temporal segments
切出有意義的分段點 corresponding to musical parts 實作上容易切太細：切成兩三。約
 - Group these segments into musically meaningful categories
歸類 / clustering
- Examples
 - *Intro, verse, chorus, bridge, outro* of a pop song
 - *Exposition, development, recapitulation, coda* of a sonata 遇轉調專處理 chroma feature

Types of musical structures

MIREX 有 structure segmentation 比賽 (目前做得不好)

Strophic form



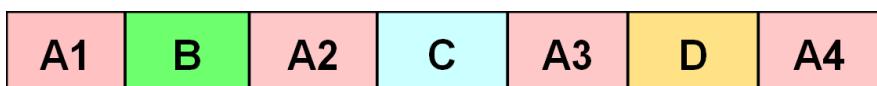
Sonata form 奏鳴曲式



Chain form



Rondo form 輪旋曲



“Tell Me Why” by Beatles



“Yesterday” by Beatles



Music segmentation (1)

中型結構

- Motifs, riffs, phrases, themes, and variations...

- A motif is a recurring pattern in a writing work that helps to strengthen the theme
- A riff is a short, repeated, memorable musical phrase (or melody figure)

找重覆出現不難，找有差異的重覆不容易

LA CUCARACHA

The musical score for 'La Cucaracha' is shown in two parts. The left part is a simple melody with lyrics, and the right part is a more complex variation by Mozart. The left part has blue circles highlighting 'motif 1', 'exact repetition', 'motif 2 (answer)', 'motif 1a', and 'exact repetition'. The right part shows the title '12 VARIATIONEN 小星星變奏' and 'für das Pianoforte von W. A. MOZART.' Below it, the 'TEMA.' (theme) is shown in its original form, followed by 'VAR. I.' (Variation I) which includes a 'legato' instruction.

<https://tamingthesaxophone.com/composition-motif>

Music segmentation (2)

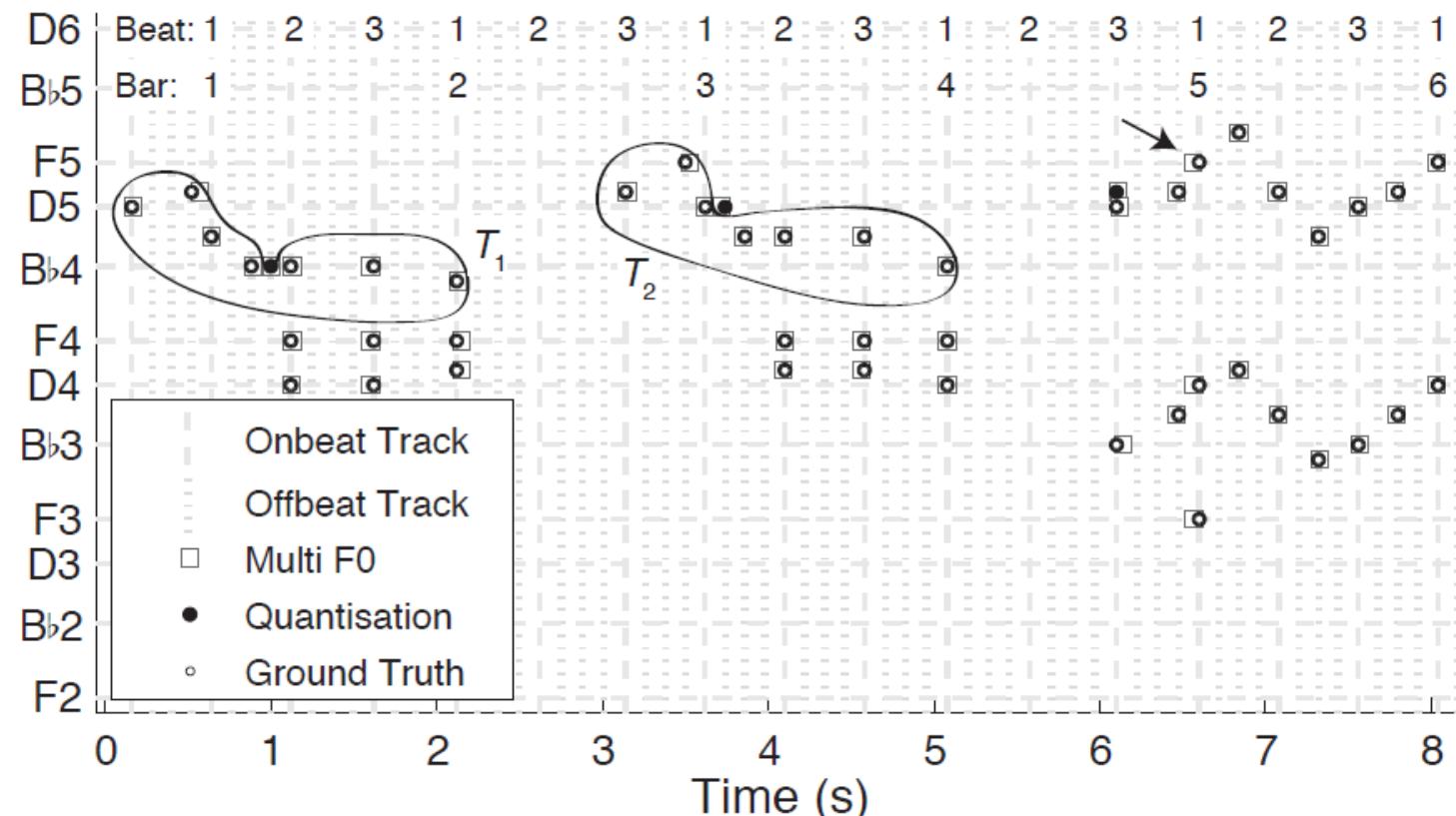
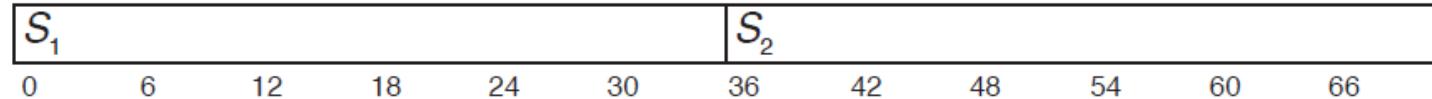
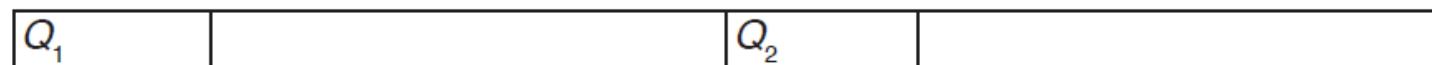
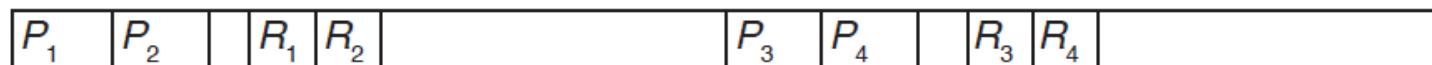
- The key to bridge the audio-to-semantic gap
- Example: Mozart's Piano Sonata no.4 in E-flat major K282 by Friedrich Gulda

Pattern discovery problem ↗ oversegmentation

The image shows a musical score for Mozart's Piano Sonata no.4 in E-flat major K282, specifically the Menuetto movement. The score is in 2/4 time with a key signature of one flat. The tempo is marked as 'Tempo di Menuetto' with a dotted quarter note followed by '120'. The score consists of two staves: treble and bass. The music is divided into measures by vertical bar lines. Above the staff, measure numbers and their corresponding total measures in parentheses are listed: 0 (36), 3 (39), 6 (42), 9 (45), 12 (48), 15 (51), 18 (54), 21 (57), 24 (60), 27 (63), 30 (66), and 33 (69). The music is annotated with several labels indicating patterns:

- P₁**:出现在第0拍。
- P₂**:出现在第3拍。
- Q₁**:出现在第6拍。
- R₁**:出现在第12拍。
- R₂**:出现在第15拍。
- S₁, S₂**:出现在第18拍。

此外，乐谱上方有一段手写标注：“Pattern discovery problem ↗ oversegmentation”，指出了自动模式识别可能存在的过度分割问题。



Challenging case

重點出現好多次，但每次呈現都差很多

- Chopin's Nocturne Op. 62, No. 2 in E major

1. Lento .889

sostenuto

This musical score excerpt shows a piano part in E major (three sharps) with a tempo of Lento (.889). The instruction 'sostenuto' is written below the staff. The melody consists of eighth-note patterns, and the harmonic progression is indicated by Roman numerals I, II, III, IV, V, VI, VII, and I.

2. .653

3 6

This musical score excerpt shows a piano part in E major (three sharps) with a tempo of .653. The instruction '3 6' is written above the staff. The melody consists of eighth-note patterns, and the harmonic progression is indicated by Roman numerals I, II, III, IV, V, VI, VII, and I.

3. .264

p

cresc.

dim.

This musical score excerpt shows a piano part in E major (three sharps) with a tempo of .264. The instruction 'p' is written below the staff, and 'cresc.' is written above the staff. The harmonic progression is indicated by Roman numerals I, II, III, IV, V, VI, VII, and I. A dynamic instruction 'dim.' is also present.

4. .458

IS:16

This musical score excerpt shows a piano part in E major (three sharps) with a tempo of .458. The instruction 'IS:16' is written below the staff. The melody consists of sixteenth-note patterns, and the harmonic progression is indicated by Roman numerals I, II, III, IV, V, VI, VII, and I.

5. 1.000

pp

cresc.

This musical score excerpt shows a piano part in E major (three sharps) with a tempo of 1.000. The instruction 'pp' is written below the staff, and 'cresc.' is written above the staff. The melody consists of eighth-note patterns, and the harmonic progression is indicated by Roman numerals I, II, III, IV, V, VI, VII, and I.

6. .944

dim.

p

PP

cresc.

This musical score excerpt shows a piano part in E major (three sharps) with a tempo of .944. The instruction 'dim.' is written below the staff, and 'p' is written above the staff. The instruction 'PP' is written below the staff, and 'cresc.' is written above the staff. The harmonic progression is indicated by Roman numerals I, II, III, IV, V, VI, VII, and I.

Collins, Tom, et al. "SIARCT-CFP: Improving Precision and the Discovery of Inexact Musical Patterns in Point-Set Representations." *ISMIR*. 2013.

Tasks

- **Segmentation**
 - The process of **dividing** a given audio stream into acoustically meaningful segments
 - Notes > Beat intervals > Sentences (4 or 8 measures) > Sections
- **Structure analysis**
 - Find the relations between the segments
- **Structure labeling**
 - Computational music analysis

MIREX competition

目前 performance 低

- Structural Segmentation 單單切太小段
- Discovery of Repeated Themes & Sections

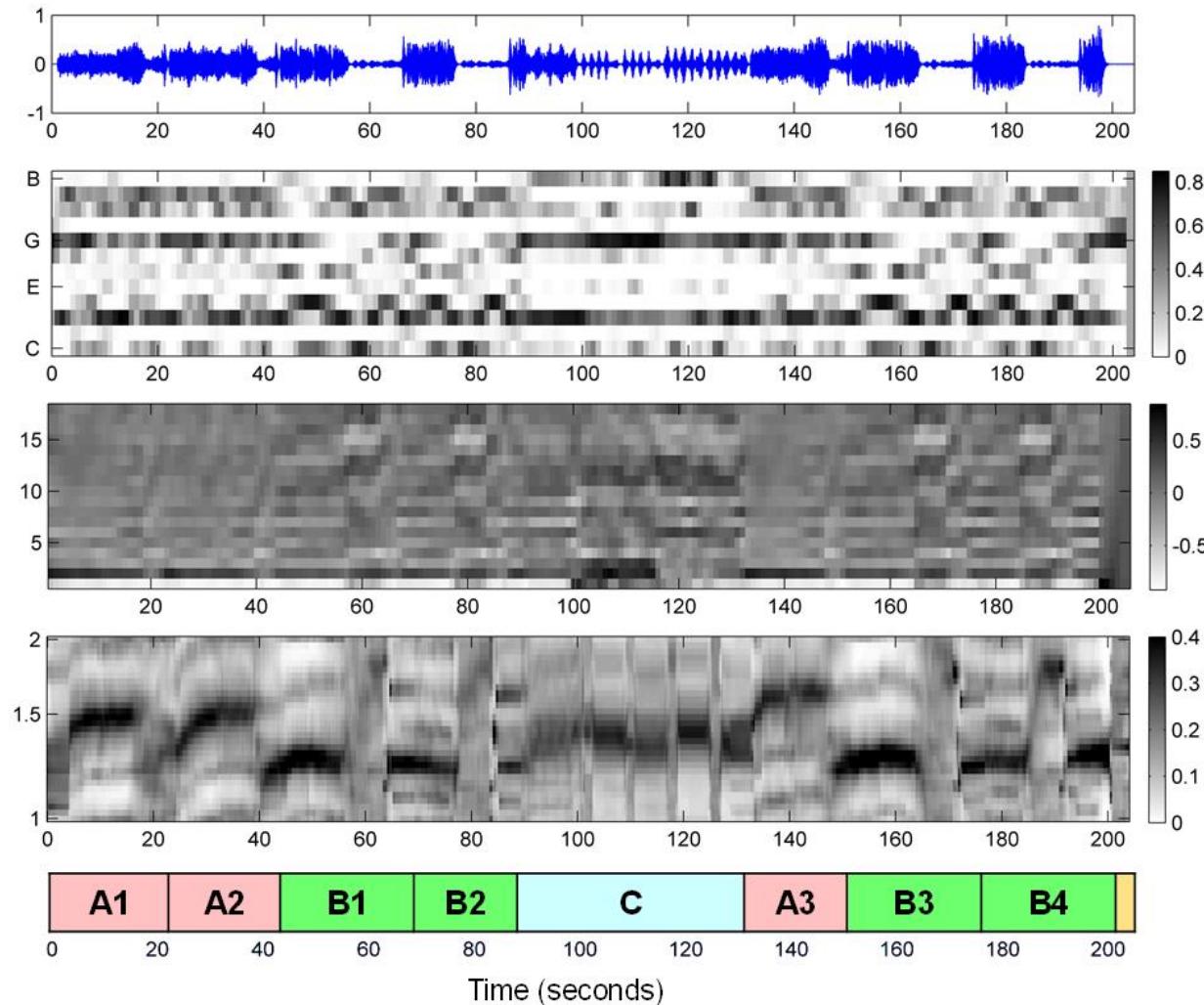
找曲子的動機 (未必週期性出現)

可思考長項嘴脣
Songle.jp 却歌問題

General principle

- In terms of signal processing, we wish to find
- 1) Novelty 訊號跟週圍不同，有新東西
- 2) Homogeneity
- 3) Repetition
- For a music signal in a macro scale

Feature representation



From: M. Mueller, *Fundamentals of Music Processing*, Chapter 4, Springer 2015

Self-similarity matrix (SSM)

- Given a sequence of feature vectors $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$
- Self-similarity matrix (SSM) $\mathbf{S} \in R^{n \times n}$:

$$\mathbf{S}(i, j) = s(\mathbf{x}_i, \mathbf{x}_j), \quad i, j \in [1: n]$$

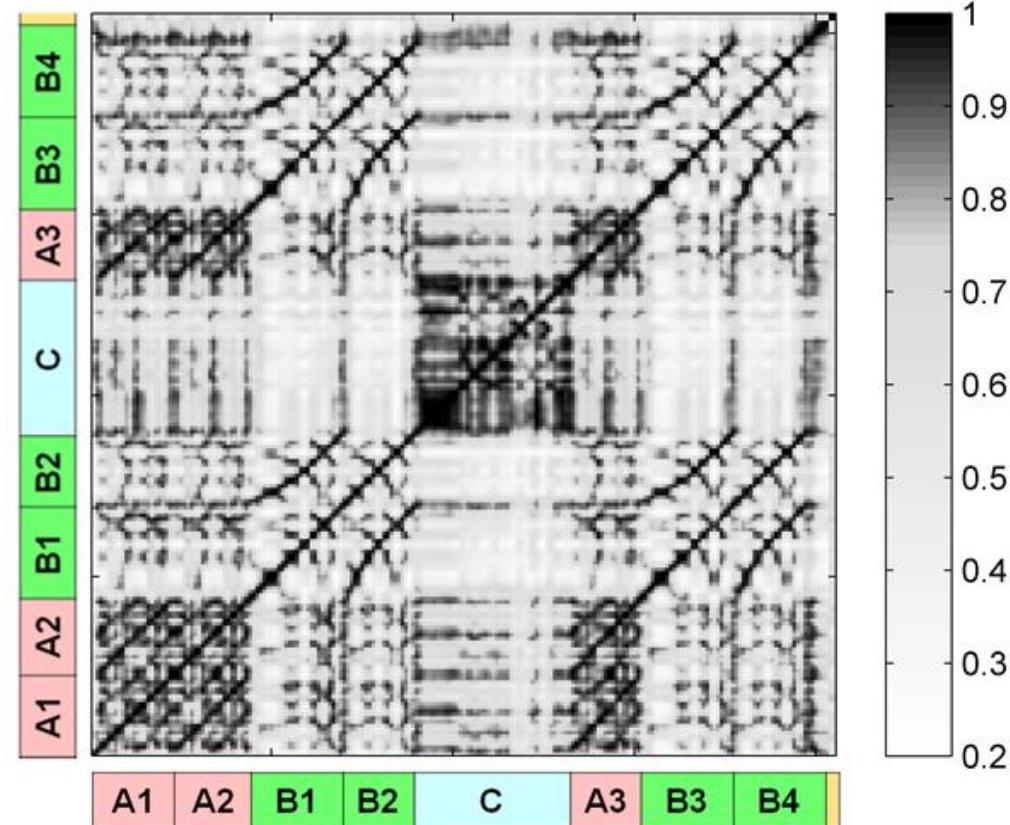
- Cosine-similarity

可試試別的 metric

$$s(\mathbf{x}_i, \mathbf{x}_j) = \frac{\mathbf{x}_i \cdot \mathbf{x}_j}{\|\mathbf{x}_i\| \|\mathbf{x}_j\|}$$

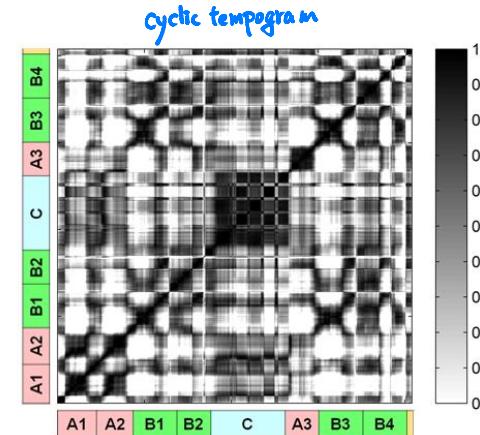
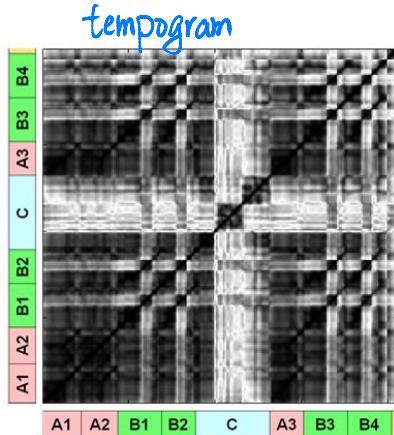
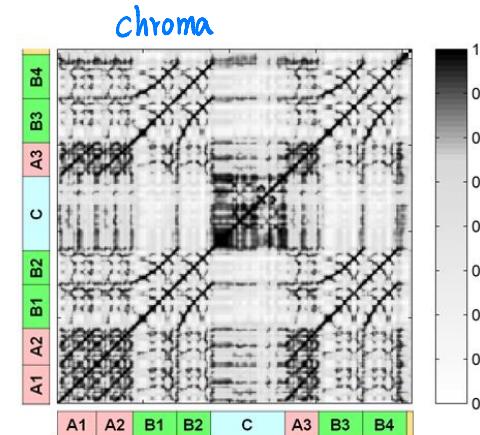
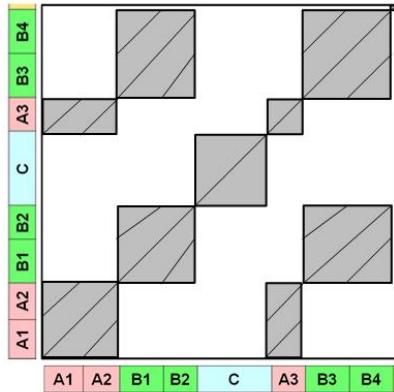
Self-similarity matrix (SSM)

- Cosine similarity of the i th and the j th frames



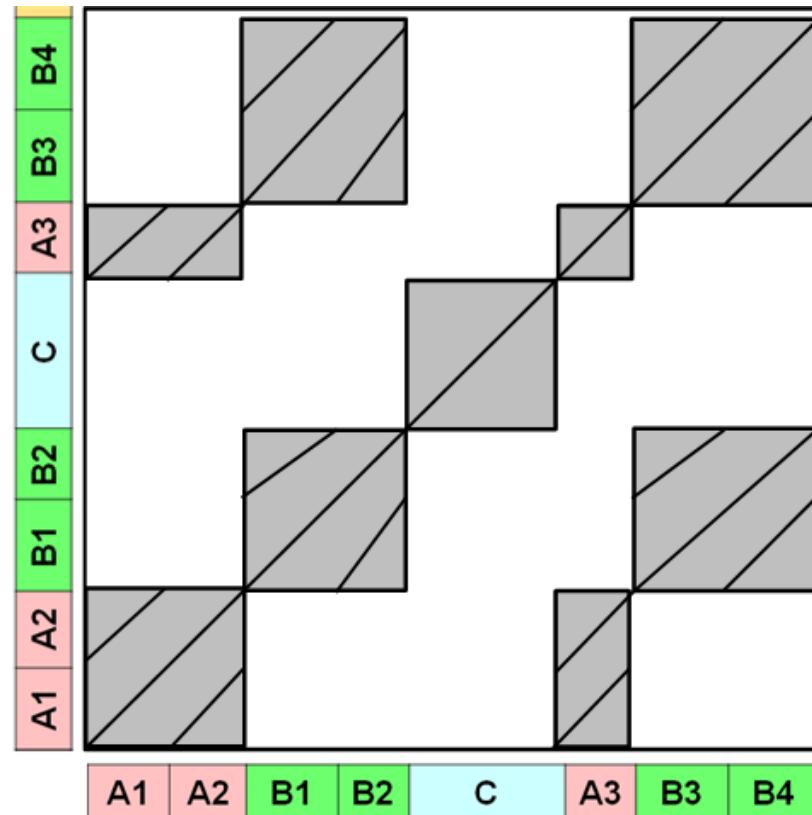
Self-similarity matrix (SSM)

也可用mfcc



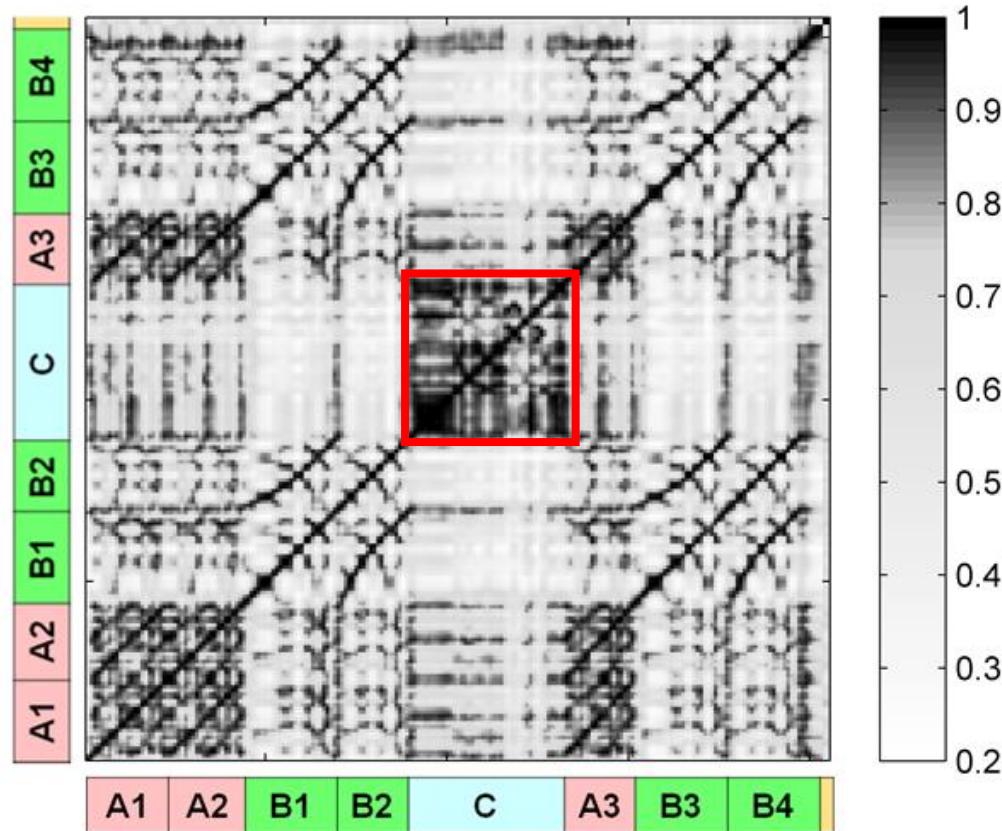
Self-similarity matrix (SSM)

- How to interpret the meaning of a SSM?

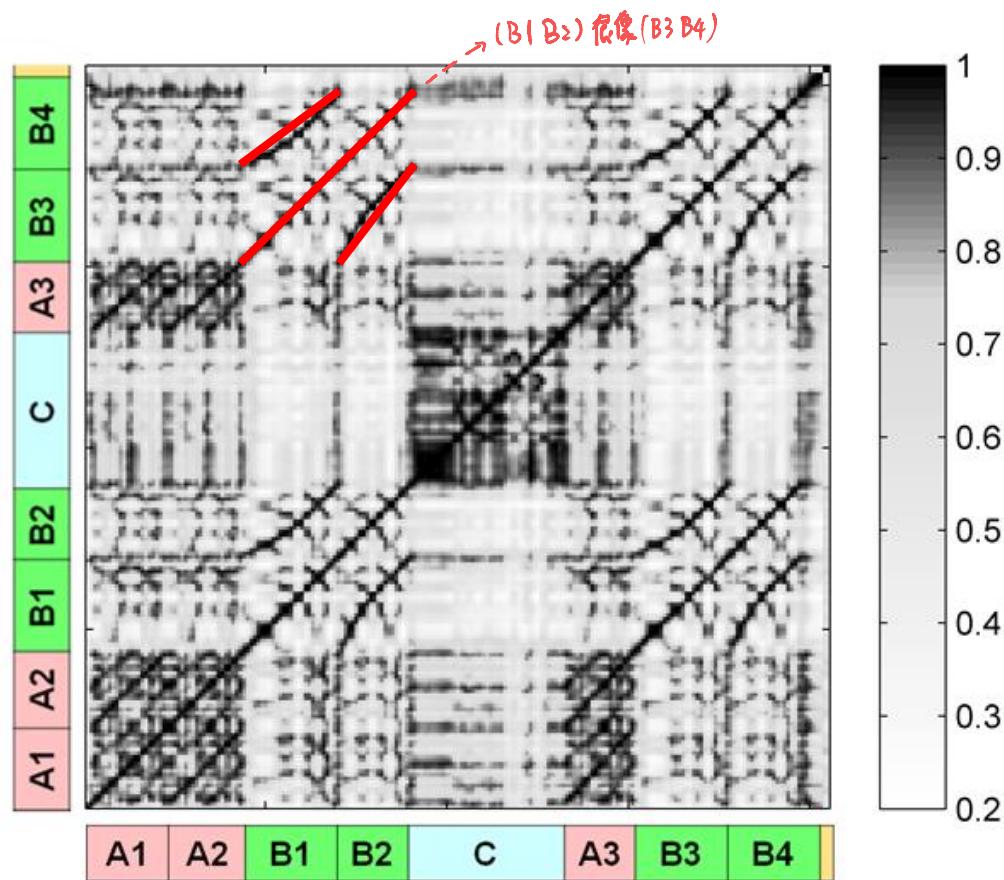


Blocks: homogeneity

自我重複，跟別人不像

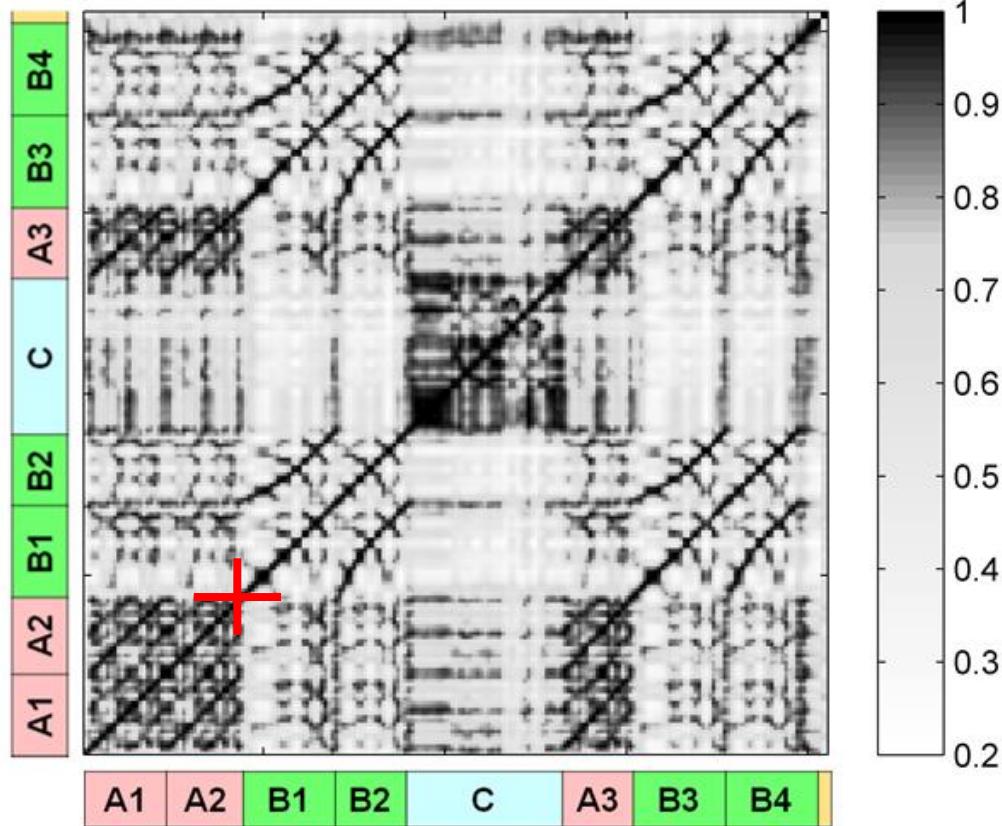


Paths: repetition



Corner: novelty

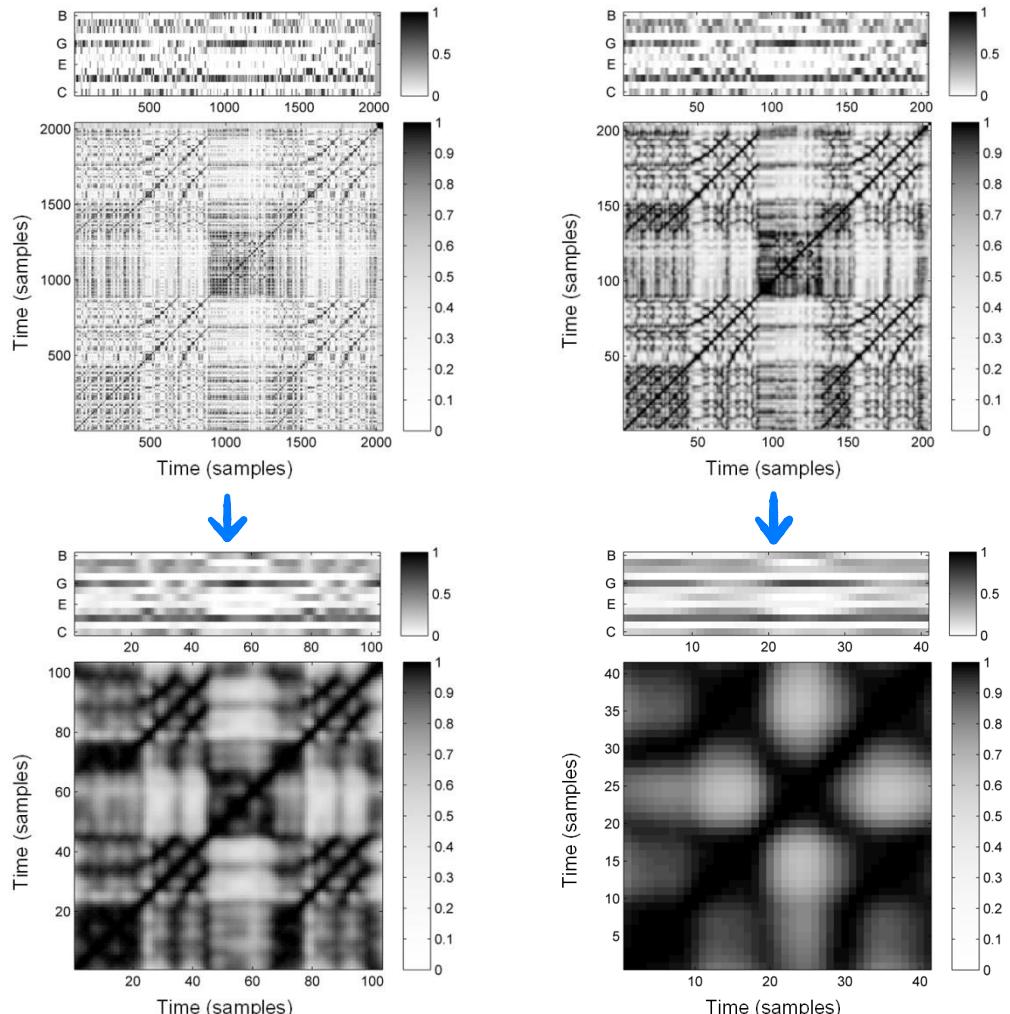
Maybe use whole column / row
filter.



Block enhancement

Smoothing on
(1) feature
(2) SSM

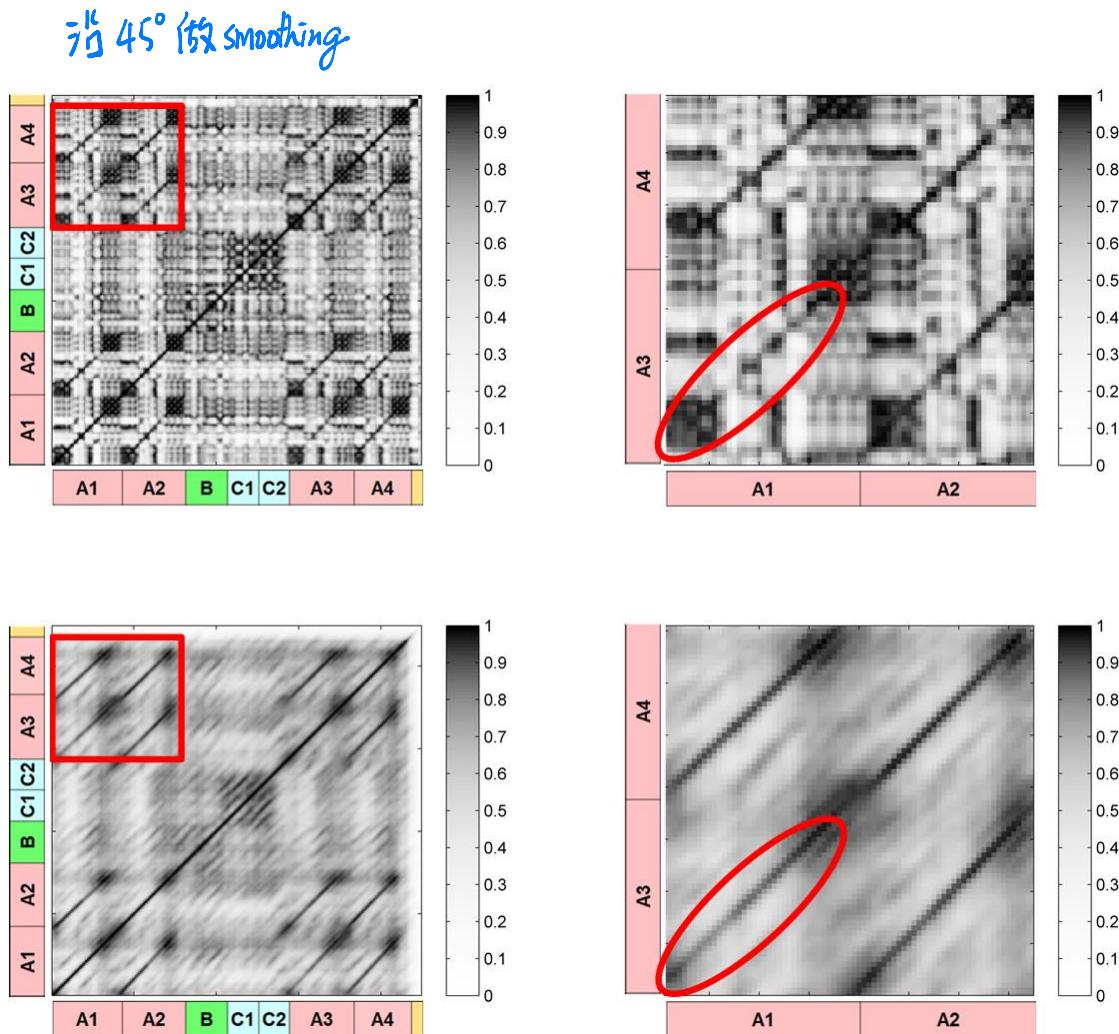
- Feature smoothing
- Coarsening



Path enhancement

- Filtering along diagonals

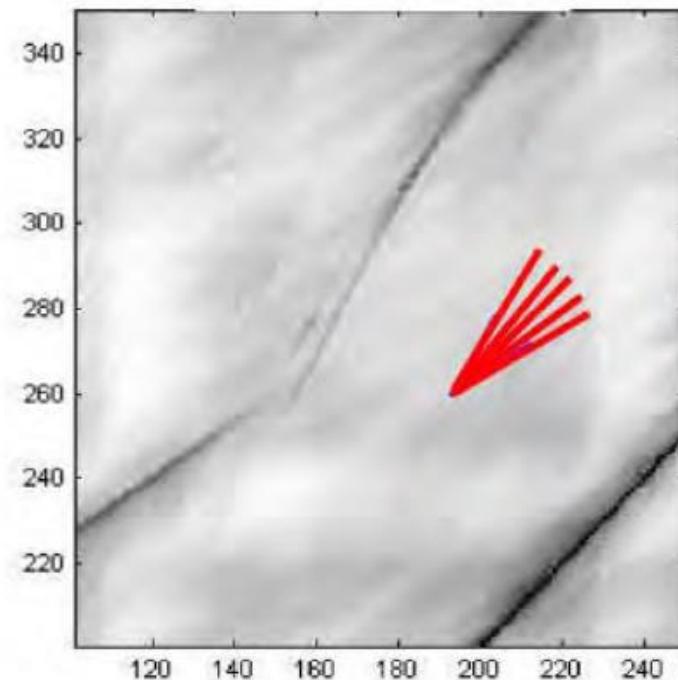
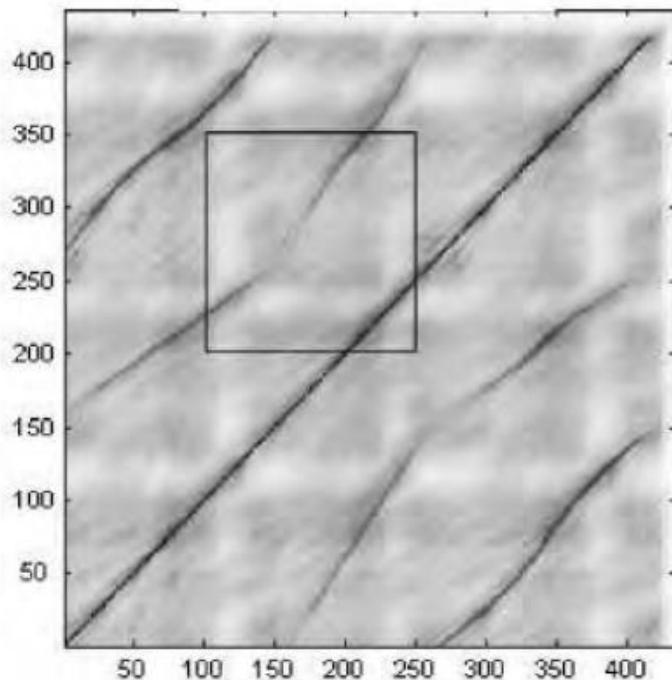
$$\mathbf{S}(i,j) = \frac{1}{L} \sum_{l=1}^L \mathbf{S}(i + l, j + l)$$



Path enhancement: multiple filtering

note 45° 未必 always true · A1 和 A2 可以不同速

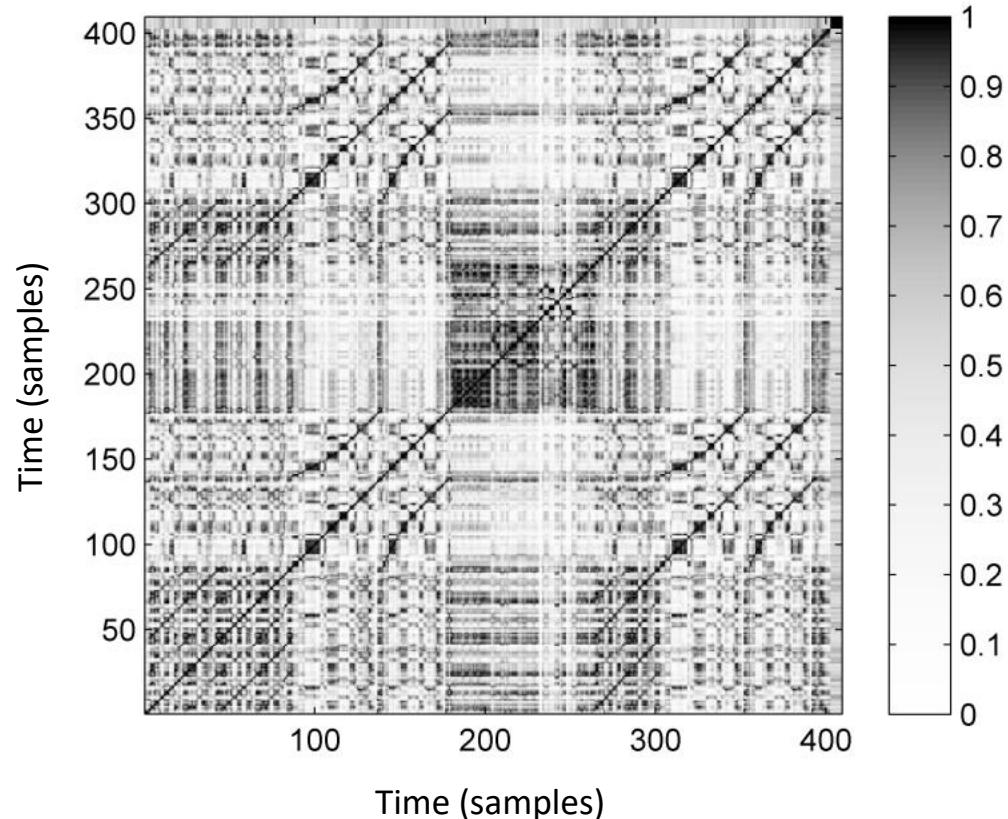
- Filtering along **different directions** and **taking the maximal value**, to account for tempo changes



Path enhancement

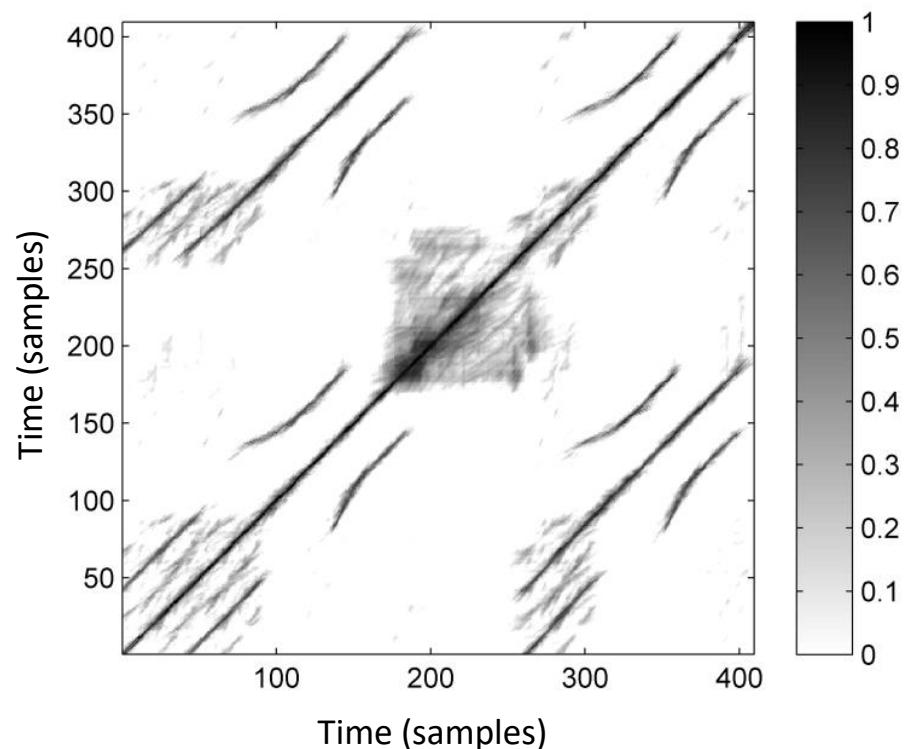
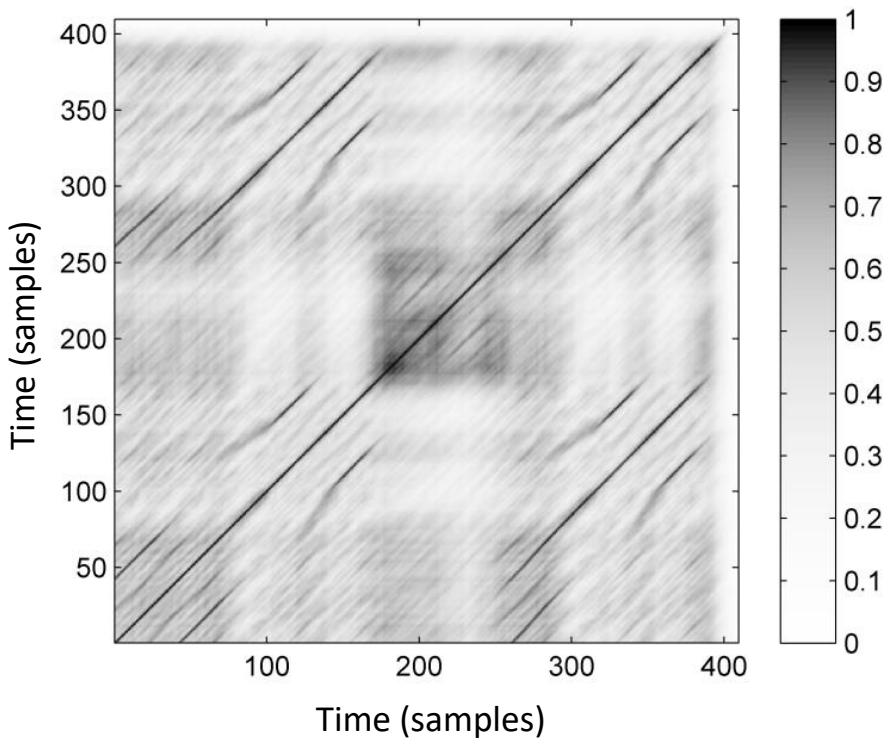
- Procedures:
 - Diagonal smoothing
 - Multiple filtering
 - Thresholding (relative)
 - Scaling and penalty
 - Path extraction
 - Transposition invariance

轉調 改奏歌曲 or 副歌升 key
要調整 chroma similarity metric



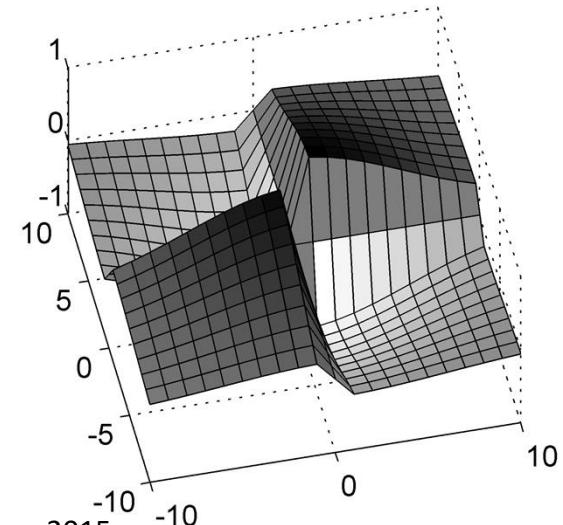
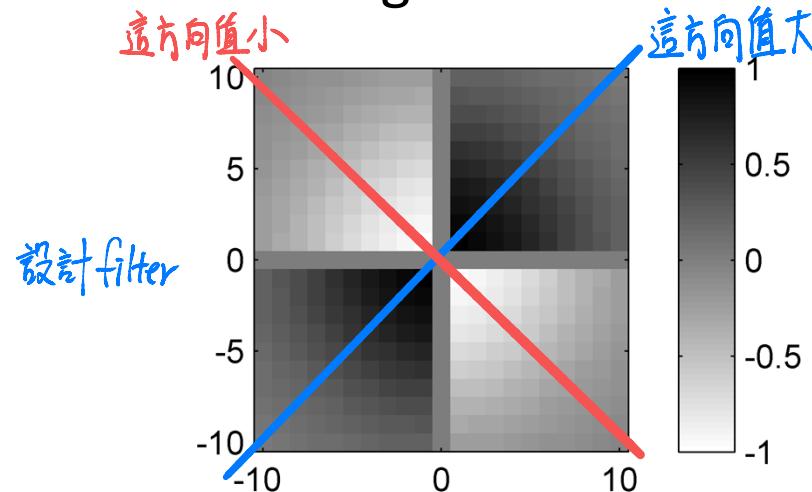
From: M. Mueller, *Fundamentals of Music Processing*, Chapter 4, Springer 2015

Filtering and thresholding



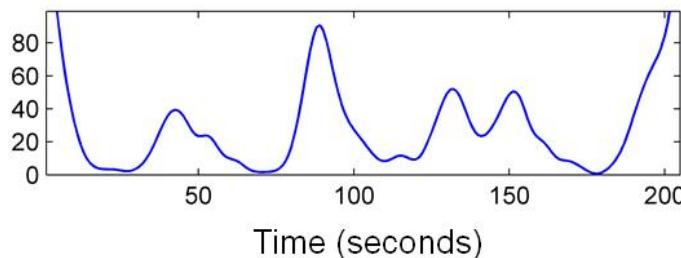
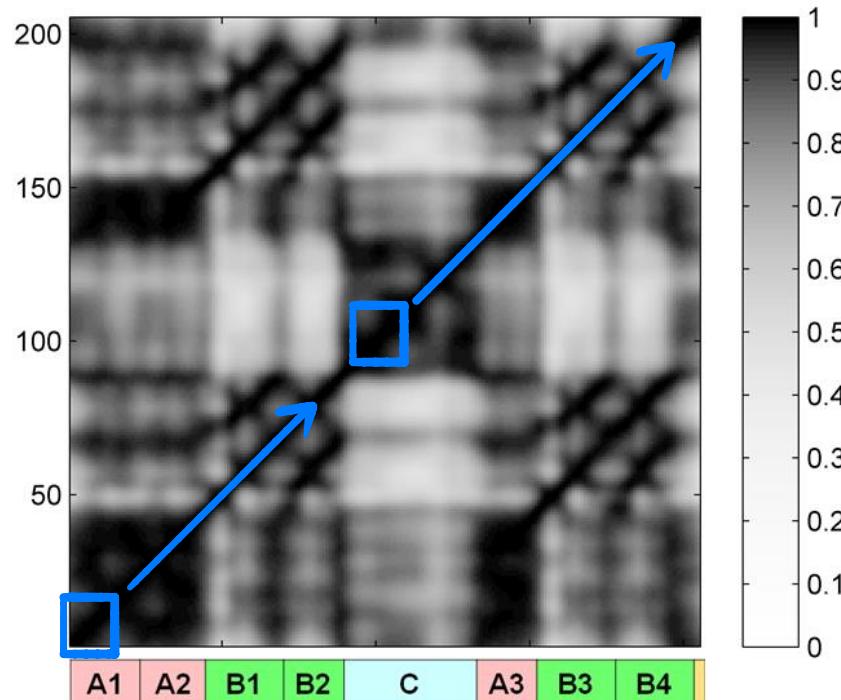
Novelty-based Segmentation

- General goals:
 - Find instances where musical changes occur
 - Find transition between subsequent musical parts
 - Main idea:
 - Use checkboard-like kernel function to detect corner points on main diagonal of SSM

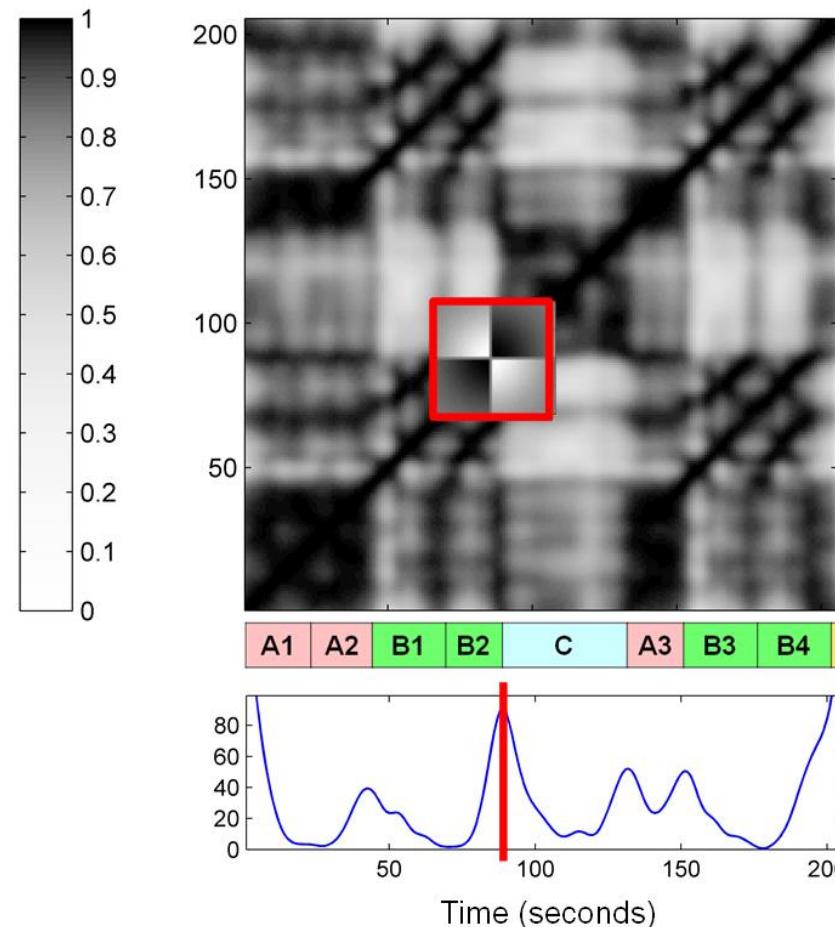
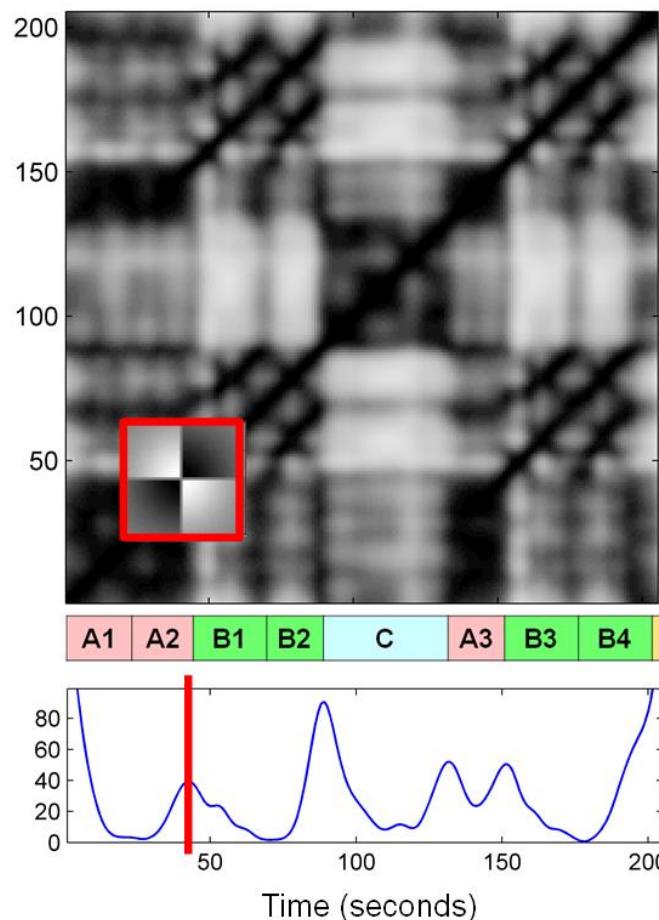


An example

From: M. Mueller, *Fundamentals of Music Processing*, Chapter 4, Springer 2015



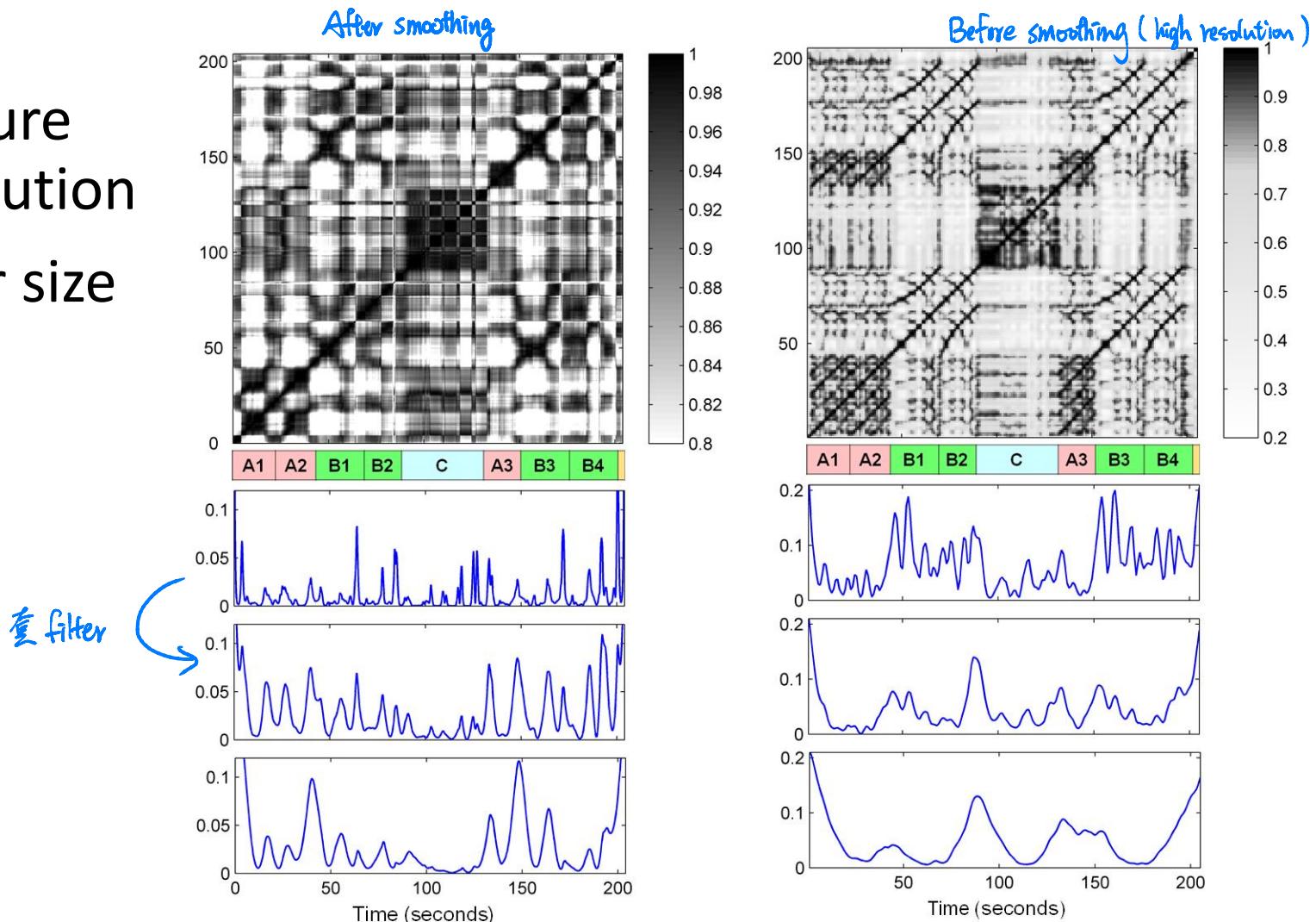
An example



From: M. Mueller, *Fundamentals of Music Processing*, Chapter 4, Springer 2015

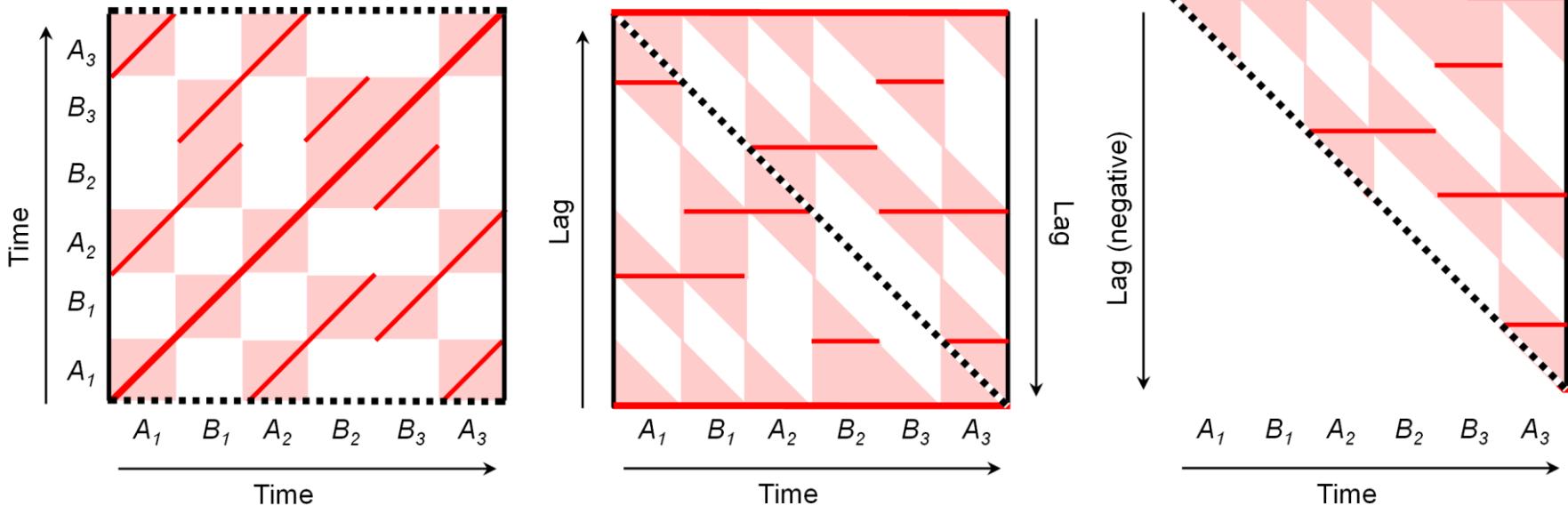
Comparison

- Feature resolution
- Filter size

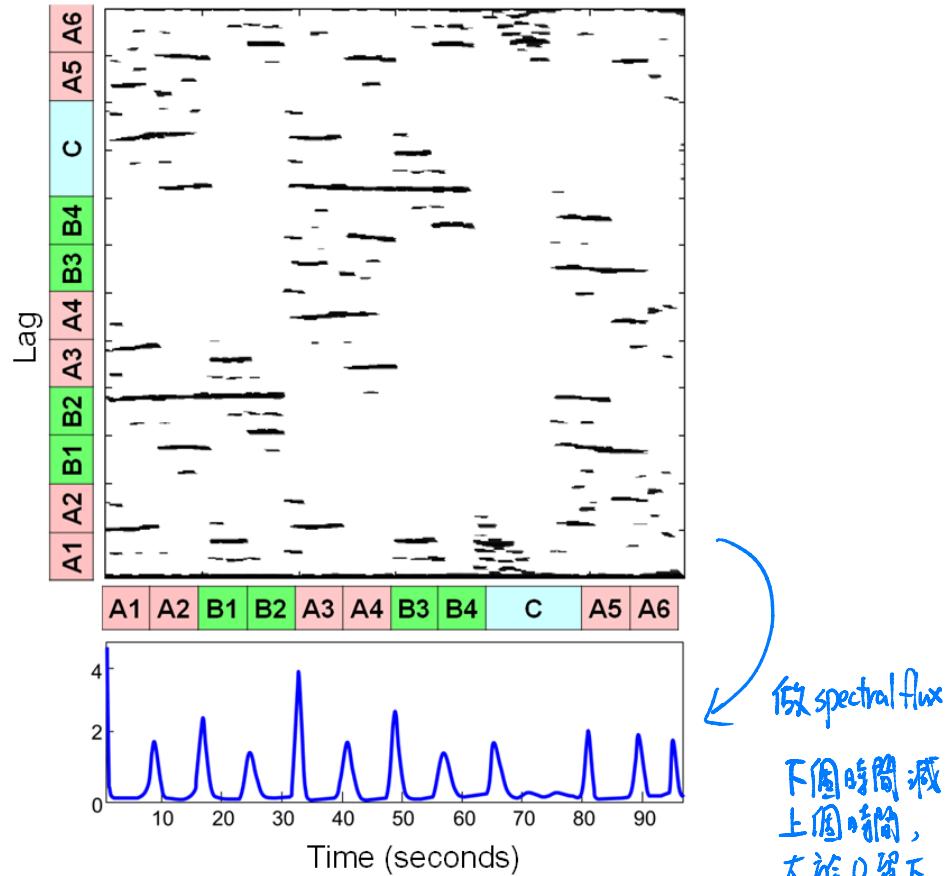
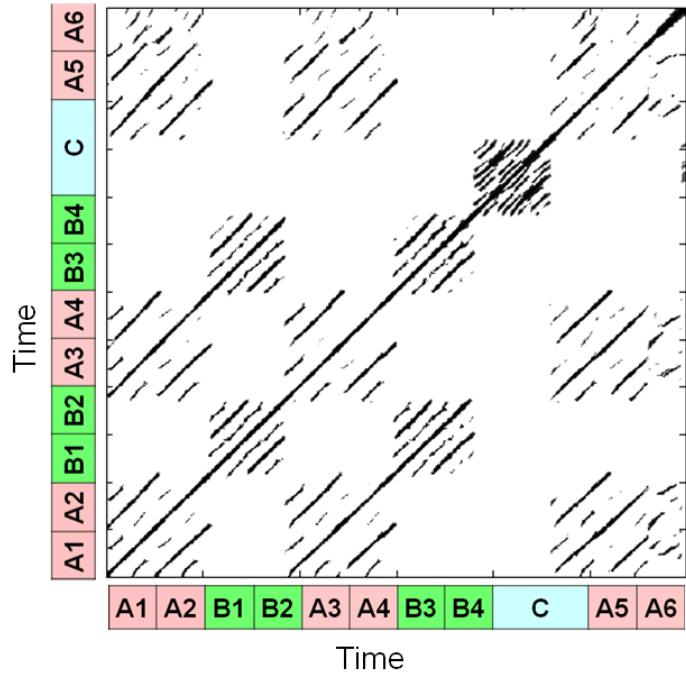


Novelty-based Segmentation: Method 2

- Structure features
 - Enhanced SSM
 - Time-lag SSM



An example



From: M. Mueller, *Fundamentals of Music Processing*, Chapter 4, Springer 2015

Further readings

- Foote, Jonathan. "Visualizing music and audio using self-similarity." *Proceedings of the seventh ACM international conference on Multimedia (Part 1)*. ACM, 1999.
- Müller, Meinard, Nanzhu Jiang, and Harald Grohganz. "SM Toolbox: MATLAB implementations for computing and enhancing similarity matrices." *Audio Engineering Society Conference: 53rd International Conference: Semantic Audio*. Audio Engineering Society, 2014.

CNN / UNET 可以

Wrap-up

- The time-frequency structure of music: what have we learned in this semester?

