



Tutorial T1
Fundamentals of Music Processing:
An Introduction using Python and Jupyter Notebooks

Music Representations and Retrieval

Meinard Müller, Frank Zalkow

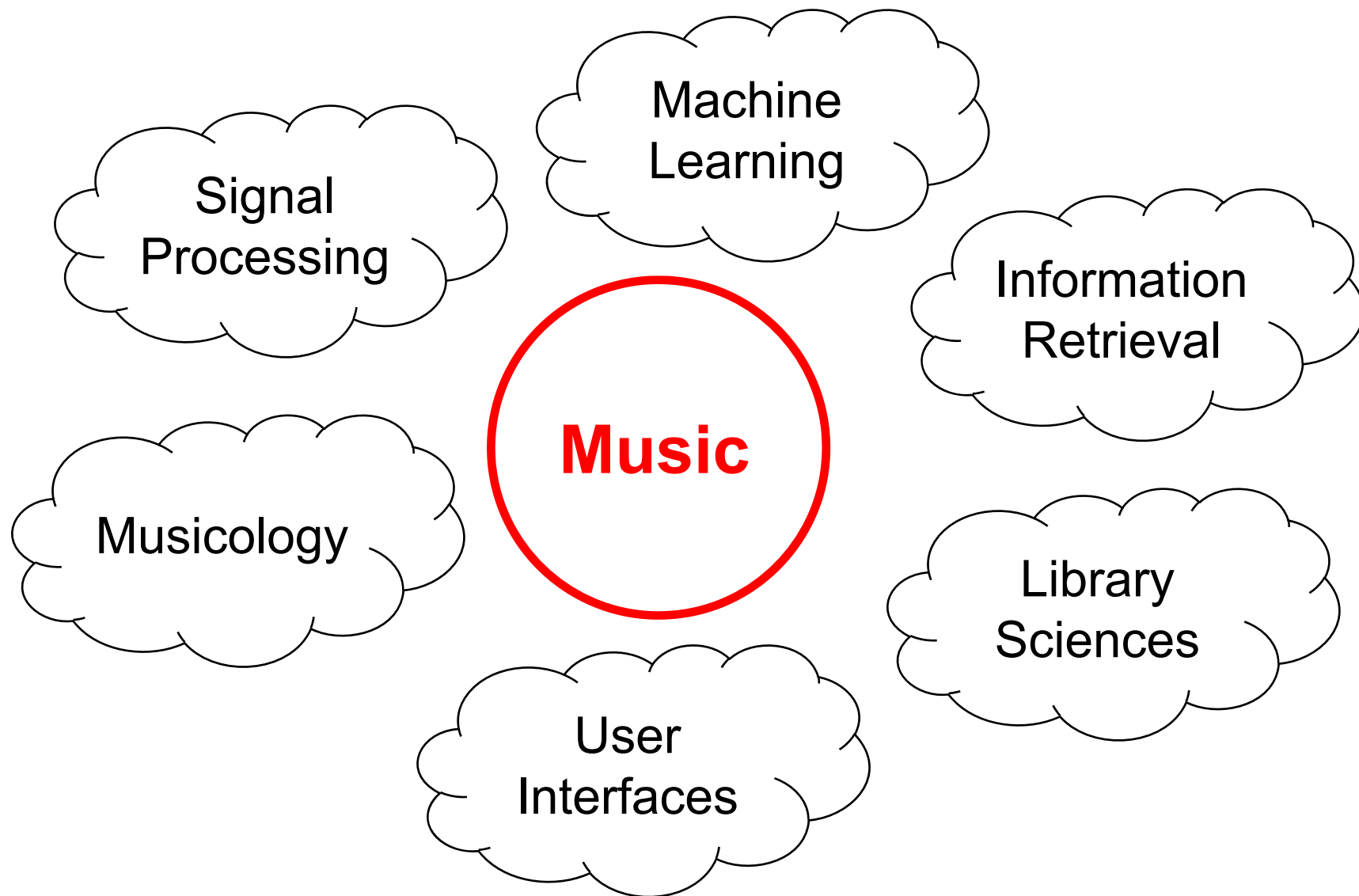
International Audio Laboratories Erlangen

meinard.mueller@audiolabs-erlangen.de, frank.zalkow@audiolabs-erlangen.de

Music



Music Information Retrieval (MIR)

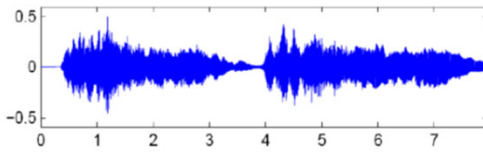


Music Information Retrieval (MIR)

Sheet Music (Image)



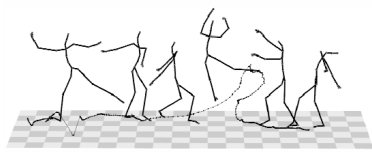
CD / MP3 (Audio)



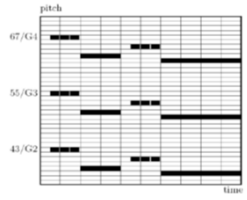
MusicXML (Text)

```
<note>  
  <pitch>  
    <step>E</step>  
    <alter>-1</alter>  
    <octave>4</octave>  
  </pitch>  
  <duration>2</duration>  
  <type>half</type>  
</note>
```

Dance / Motion (Mocap)



MIDI



Singing / Voice (Audio)



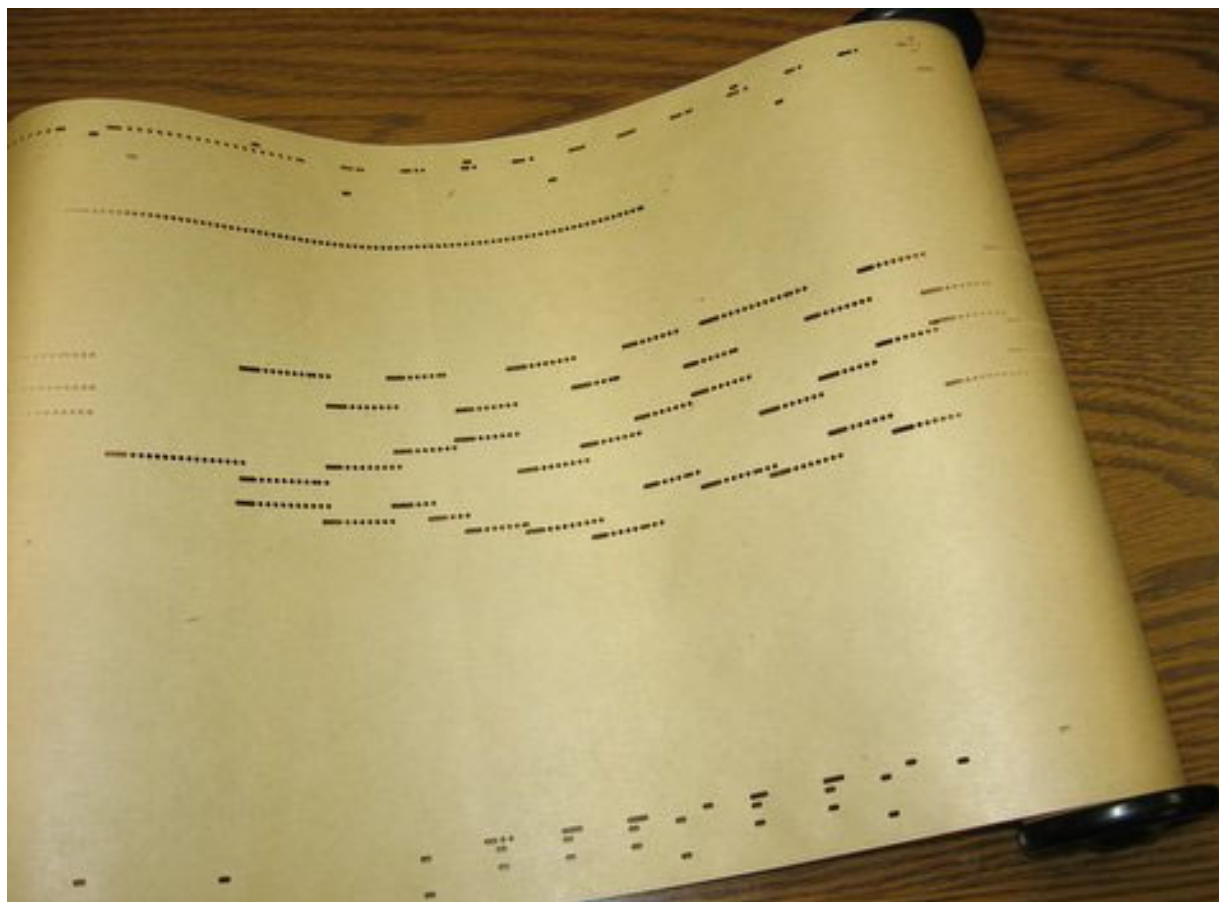
Music Film (Video)



Music Literature (Text)



Piano Roll Representation



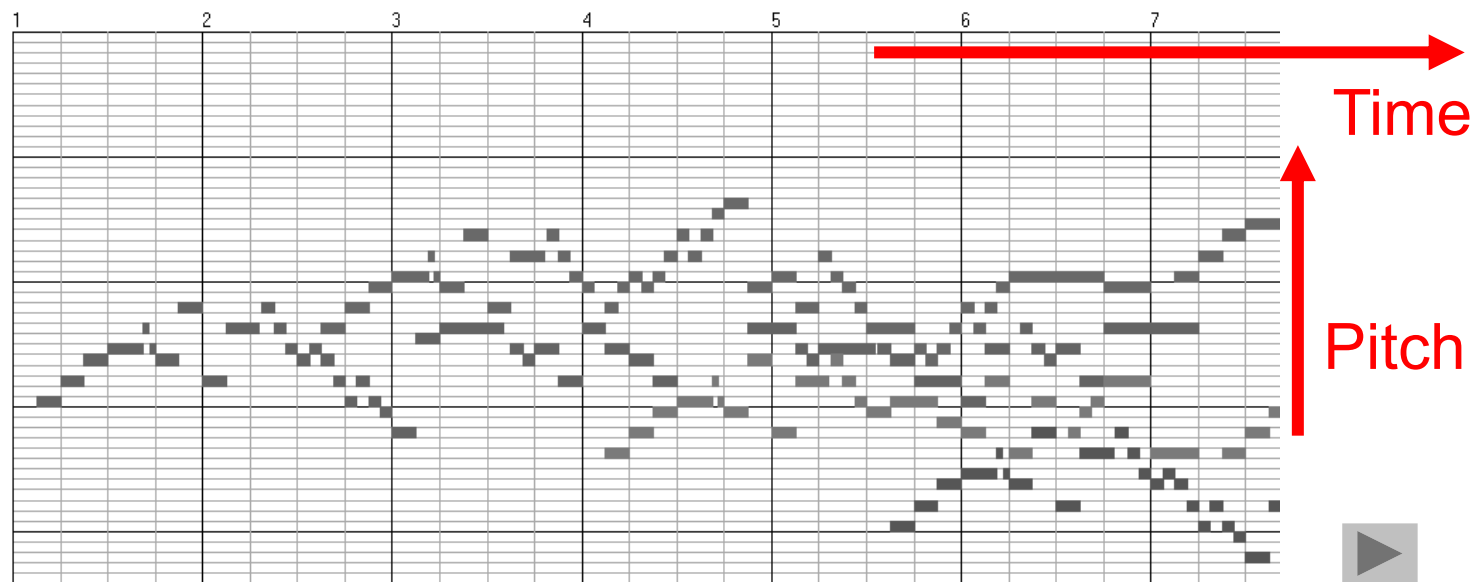
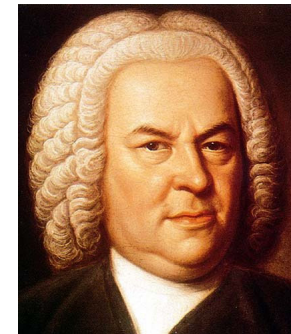
Player Piano (1900)



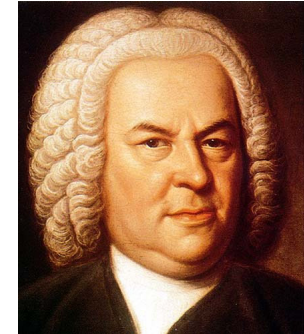
Piano Roll Representation (MIDI)

J.S. Bach, C-Major Fuge

(Well Tempered Piano, BWV 846)



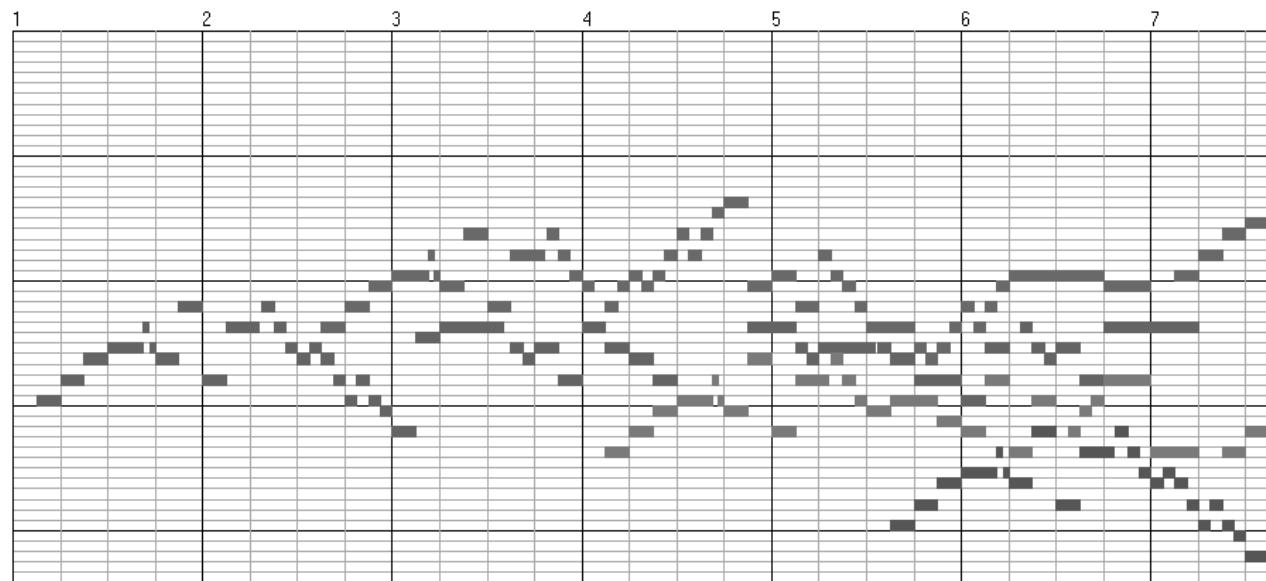
Piano Roll Representation (MIDI)



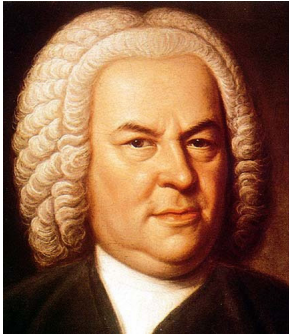
Query:



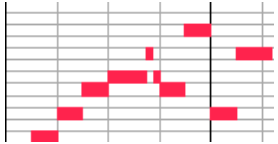
Goal: Find all occurrences of the query



Piano Roll Representation (MIDI)

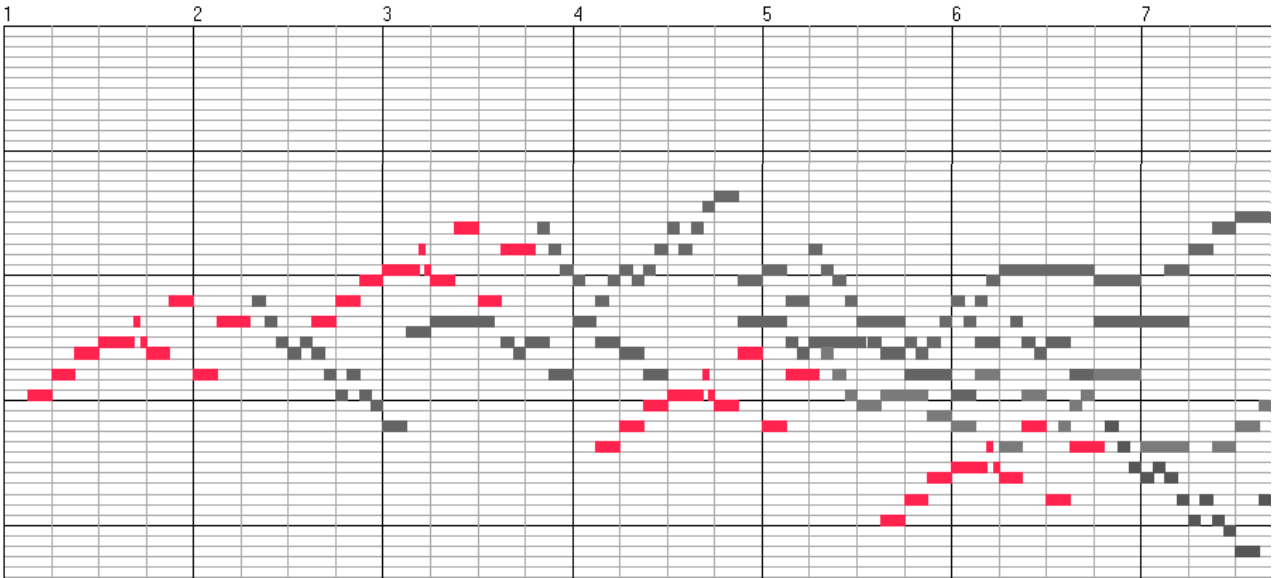


Query:



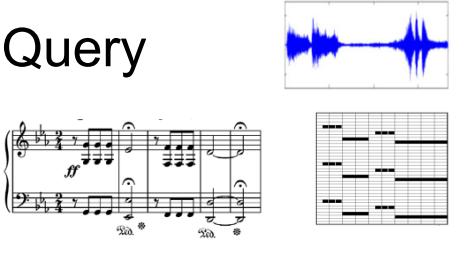
Goal: Find all occurrences of the query

Matches:

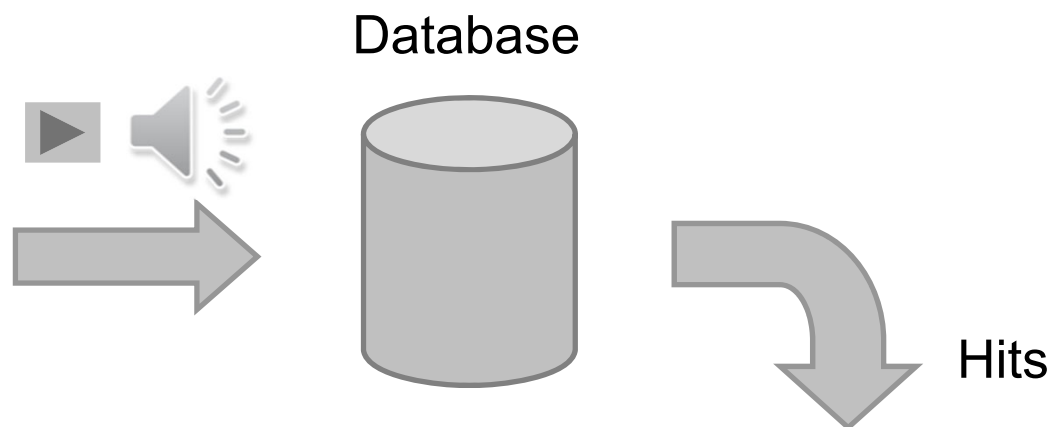


Music Retrieval

Query



The query box contains three visual representations of a musical piece: a musical score with treble and bass clefs, a blue waveform, and a piano roll.



Retrieval tasks:

Audio identification

Audio matching

Version identification

Category-based music retrieval

Bernstein (1962)
Beethoven, Symphony No. 5

Beethoven, Symphony No. 5:

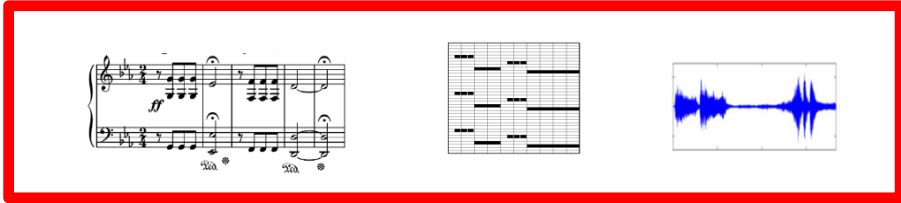
- Bernstein (1962)
- Karajan (1982)
- Gould (1992)

- Beethoven, Symphony No. 9
- Beethoven, Symphony No. 3
- Haydn Symphony No. 94



Music Retrieval

Modalities



Retrieval tasks:

Audio identification

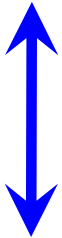
Audio matching

Version identification

Category-based music retrieval

Specificity

High specificity



Low specificity

Granularity

Fragment-based retrieval

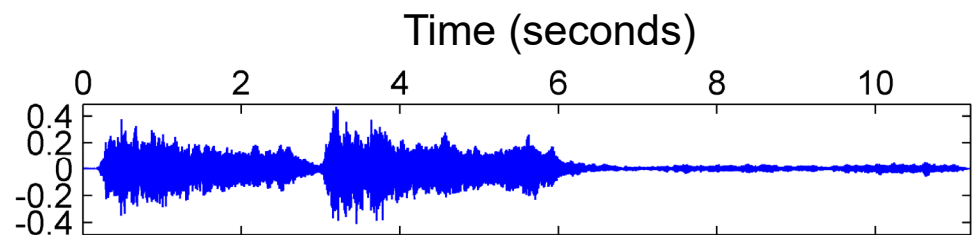


Document-based retrieval

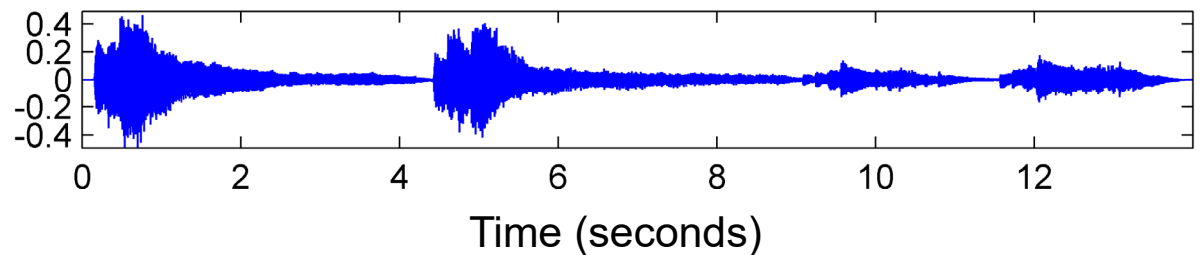
Music Synchronization: Audio-Audio

Beethoven's Fifth

Karajan 



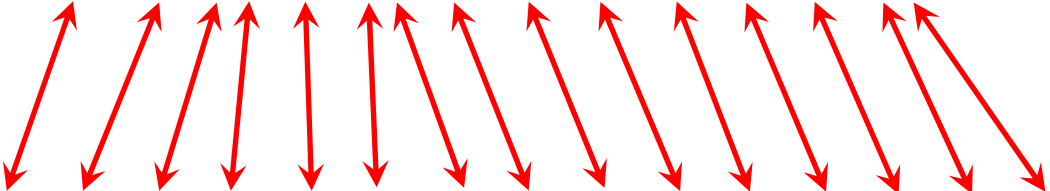
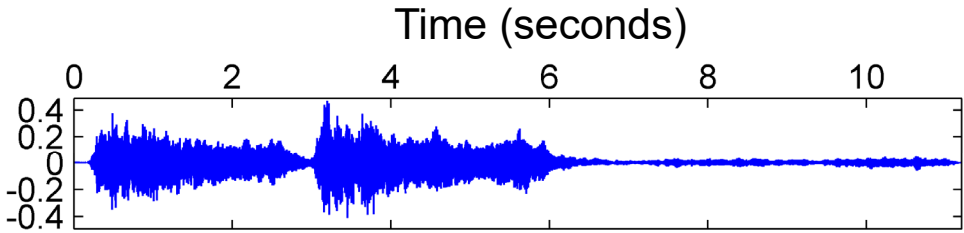
Gould 



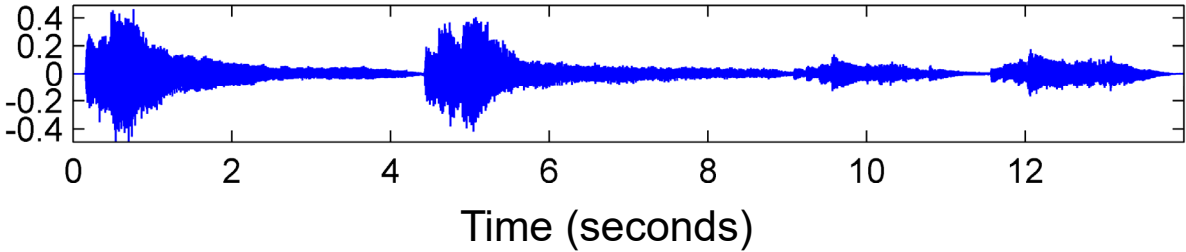
Music Synchronization: Audio-Audio

Beethoven's Fifth

Karajan 



Gould 



Music Synchronization: Audio-Audio

Task

Given: Two different audio recordings (two versions) of the same underlying piece of music.

Goal: Find for each position in one audio recording the **musically** corresponding position in the other audio recording.

Music Synchronization: Audio-Audio

Two main steps:

1.) Feature extraction

- Robust to variations (e.g., instrumentation, timbre, dynamics)
- Discriminative (e.g., capturing harmonic, melodic, tonal aspects)

➡ **Chroma features**

2.) Temporal alignment

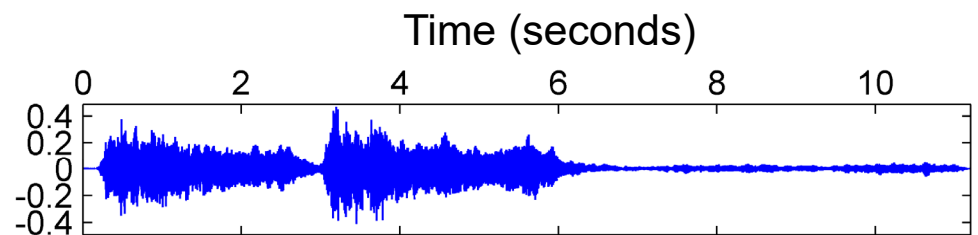
- Capturing local and global tempo variations
- Trade-off: Robustness vs. accuracy
- Efficiency

➡ **Dynamic time warping (DTW)**

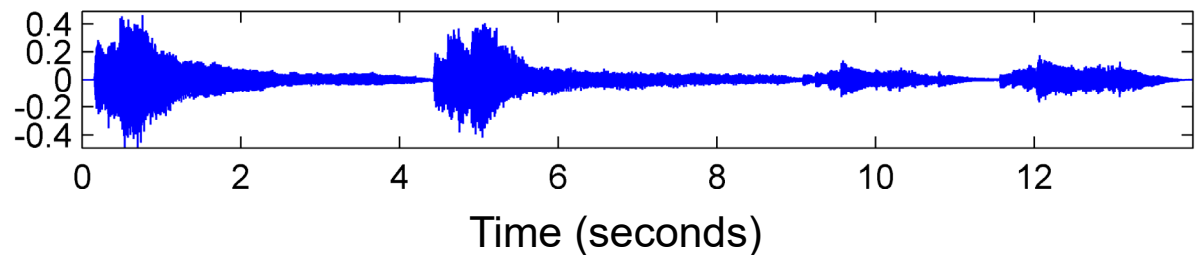
Music Synchronization: Audio-Audio

Beethoven's Fifth

Karajan 



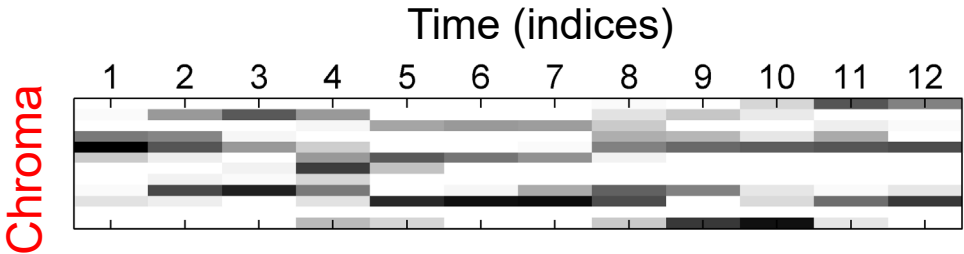
Gould 



Music Synchronization: Audio-Audio

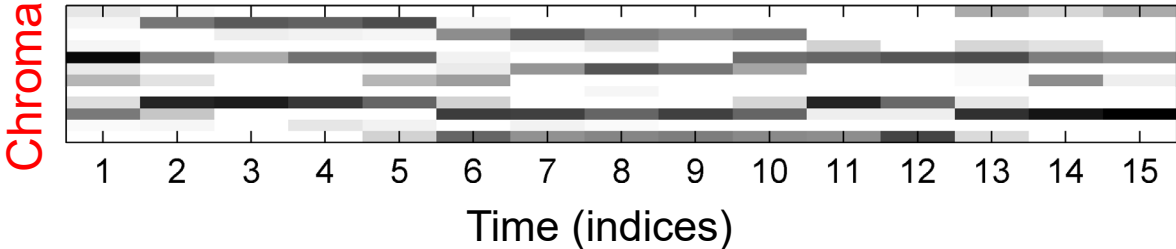
Beethoven's Fifth

Karajan





Time–chroma representations

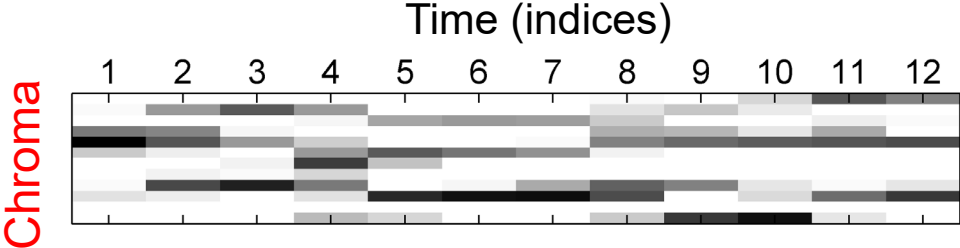
Gould



Music Synchronization: Audio-Audio

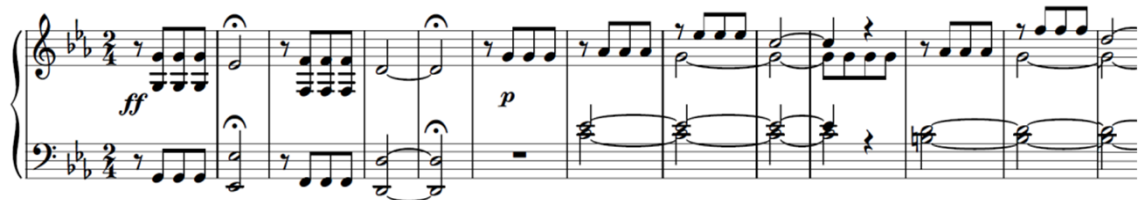
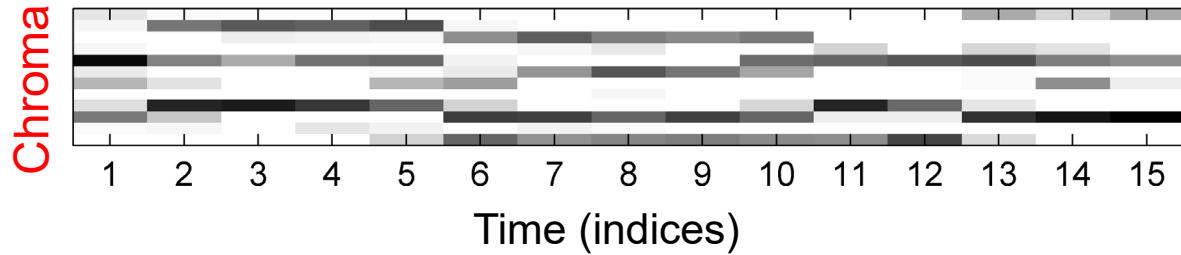
Beethoven's Fifth

Karajan 




Time–chroma representations

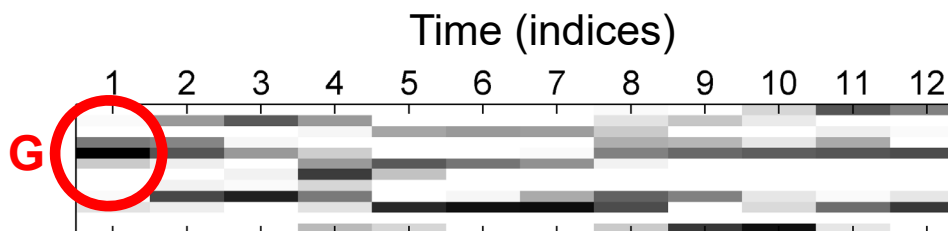
Gould 

Music Synchronization: Audio-Audio

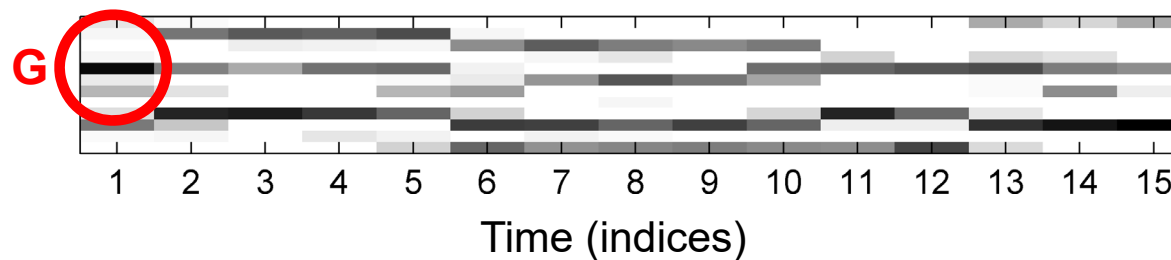
Beethoven's Fifth

Karajan





Time–chroma representations

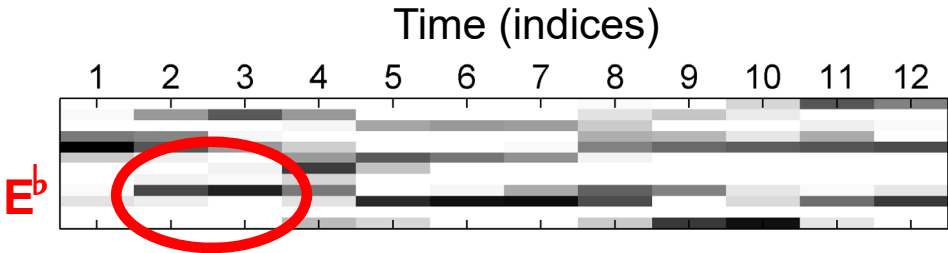
Gould



Music Synchronization: Audio-Audio

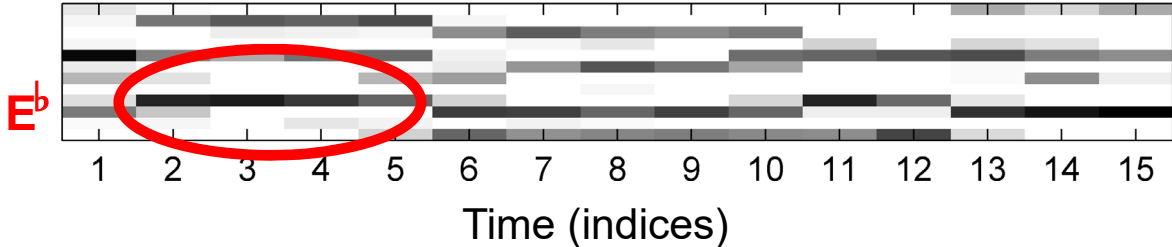
Beethoven's Fifth

Karajan 


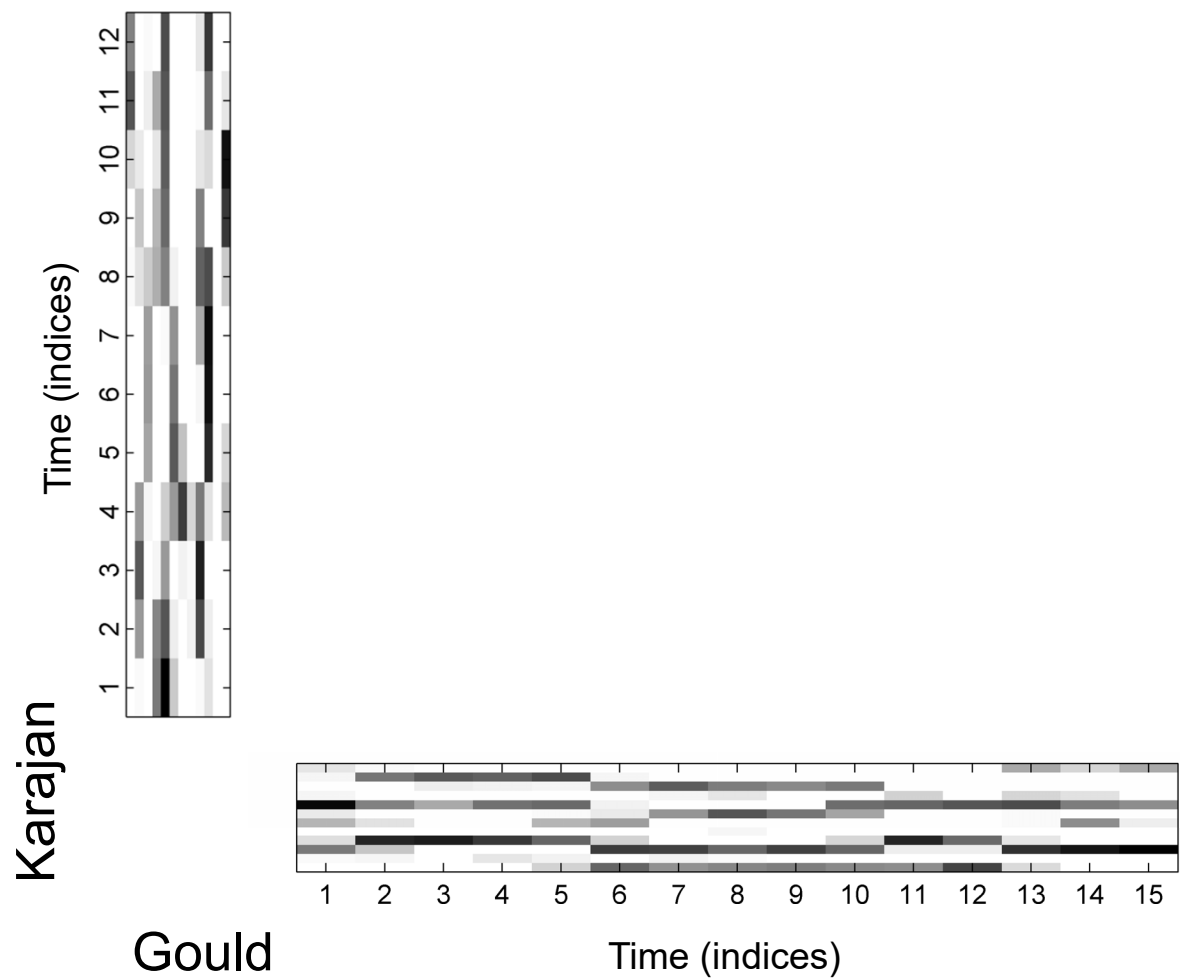


Time-chroma representations

Gould 

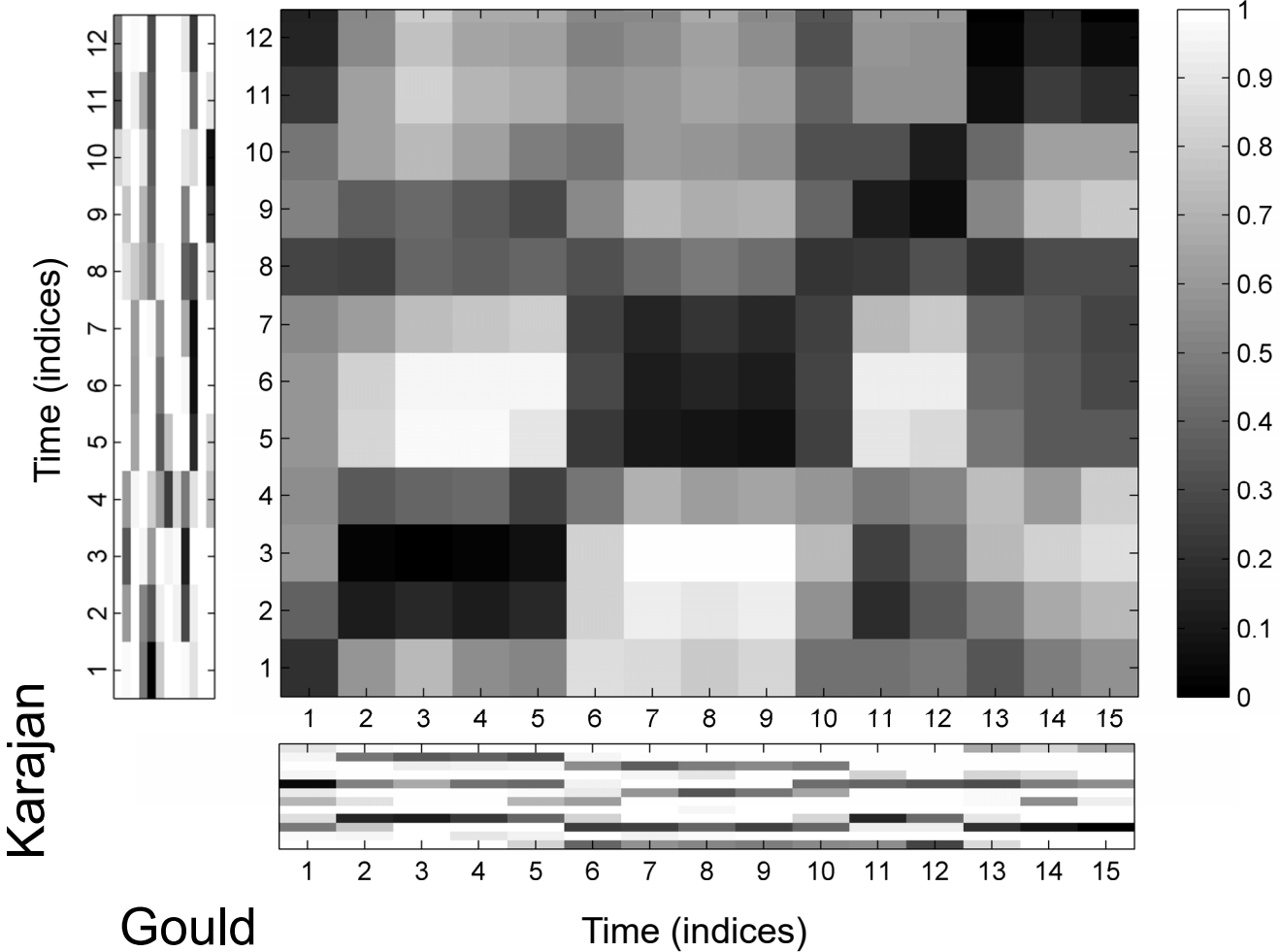



Music Synchronization: Audio-Audio



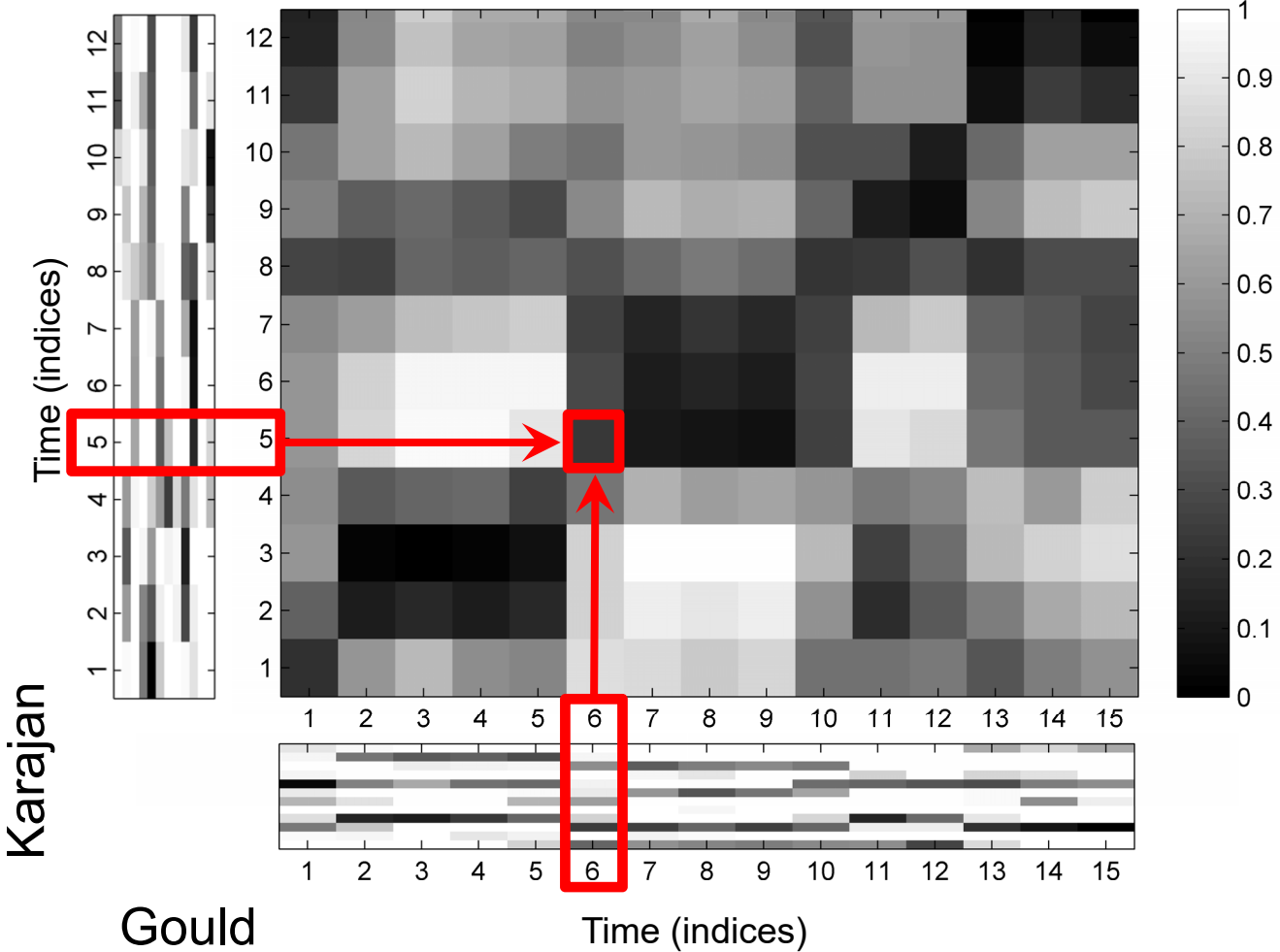
Music Synchronization: Audio-Audio

Cost matrix



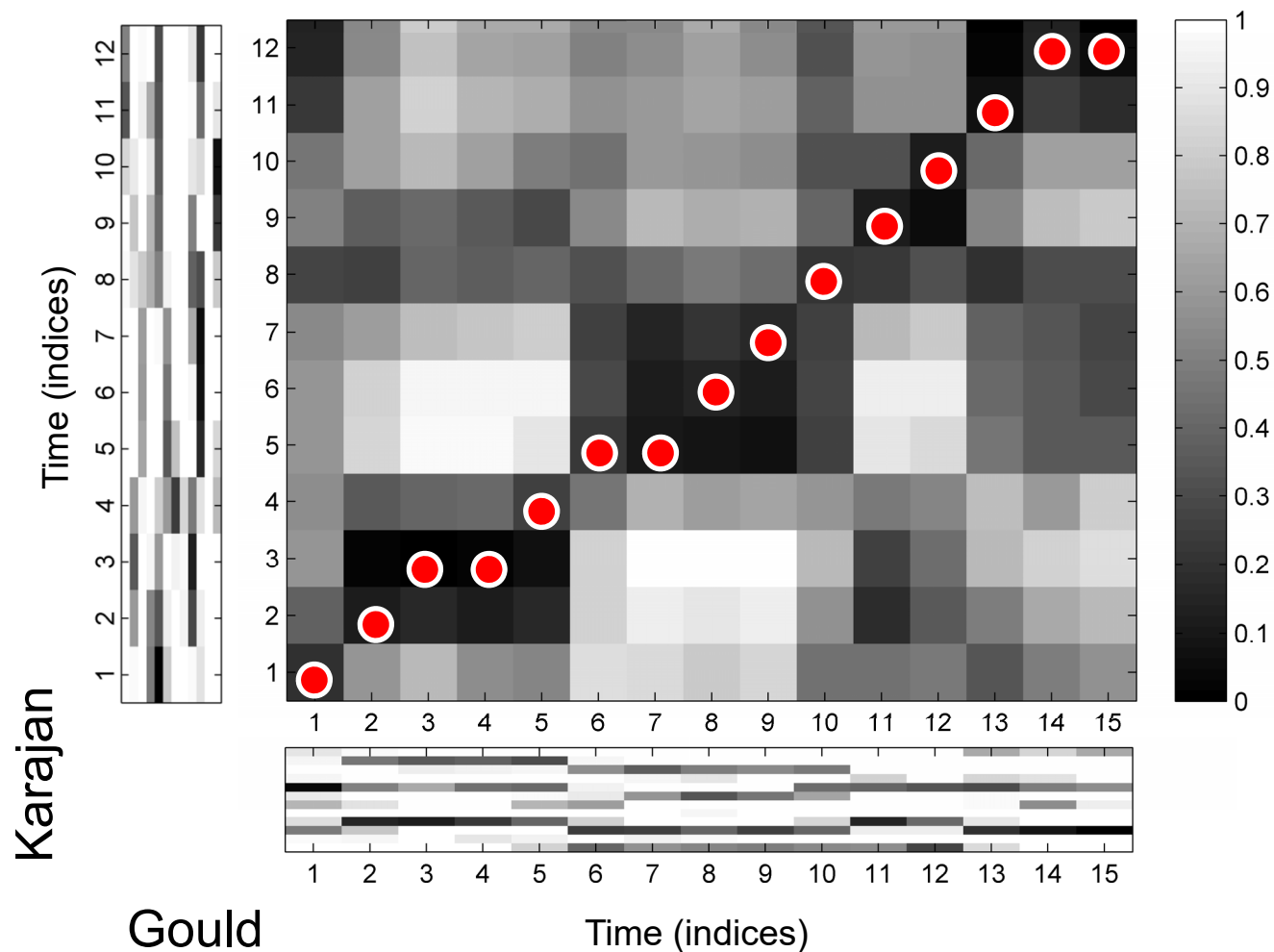
Music Synchronization: Audio-Audio

Cost matrix



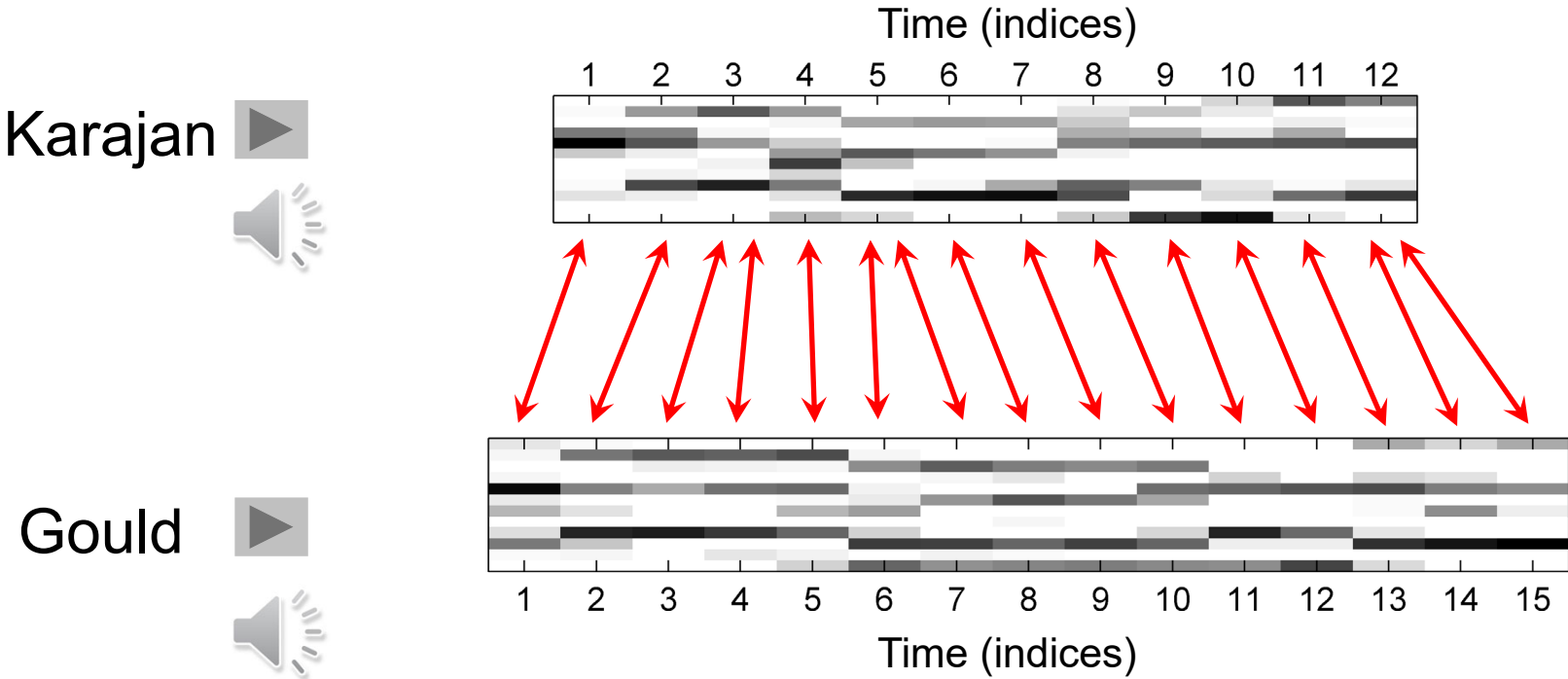
Music Synchronization: Audio-Audio

Cost-minimizing warping path



Music Synchronization: Audio-Audio

Optimal alignment (cost-minimizing warping path)



Application: Interpretation Switcher

Interpretation Switcher
Beethoven, Op067-1_Symphony5

midi 00:00.00

Bernstein 00:00.00

Sawallisch 00:00.00

Scherbakov 00:00.00

- midi
- Bernstein
- Sawallisch
- Scherbakov

Deselect all

Absolute Reference
Relative Reference

Movement selection Interval Repeat **Info**



Music Synchronization: Image-Audio

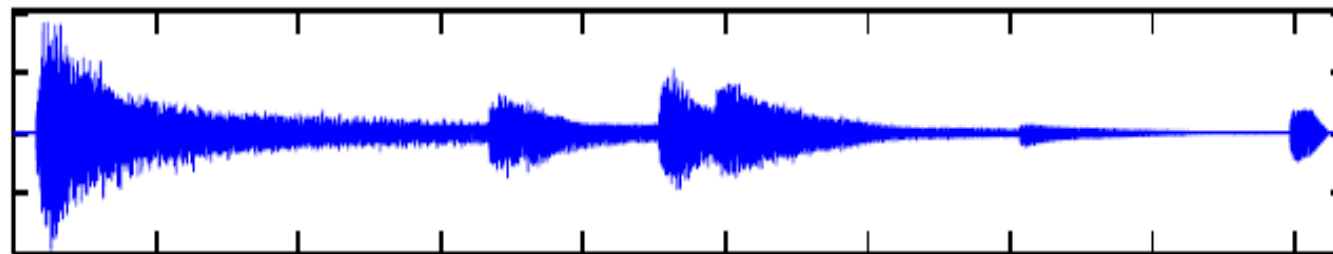
Image

Grave.

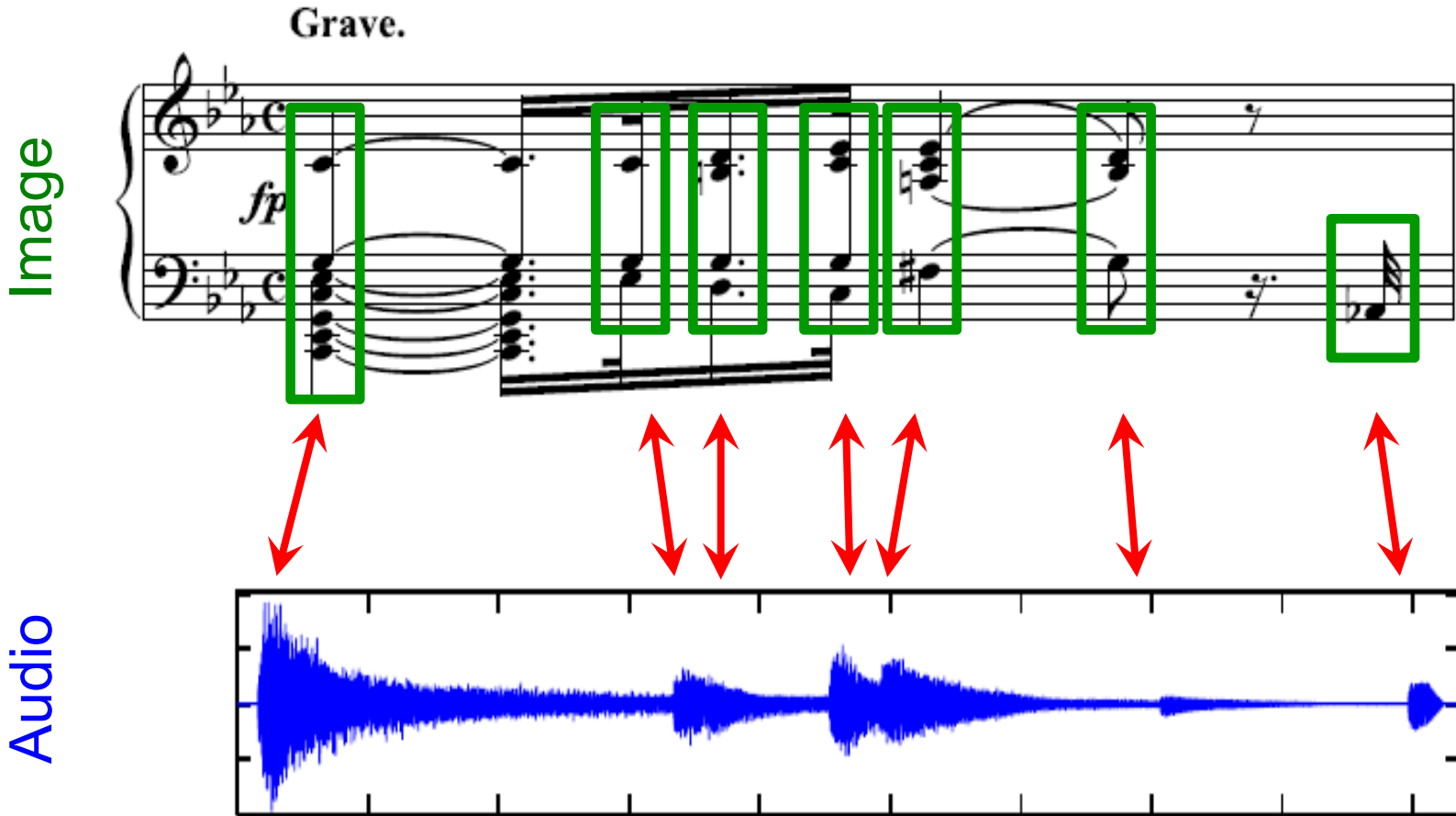


The image shows a musical score for piano, marked "Grave." and "fp". The score is written in G major (one sharp) and common time (C). It consists of two staves: a treble clef staff and a bass clef staff. The music features a slow, somber tempo with a focus on sustained chords and melodic lines. The first staff begins with a treble clef, a key signature of one sharp (F#), and a common time signature (C). The second staff begins with a bass clef, the same key signature, and a common time signature. The music is characterized by a slow, steady pace, with a focus on sustained chords and melodic lines. The first staff begins with a treble clef, a key signature of one sharp (F#), and a common time signature (C). The second staff begins with a bass clef, the same key signature, and a common time signature. The music is characterized by a slow, steady pace, with a focus on sustained chords and melodic lines.

Audio



Music Synchronization: Image-Audio

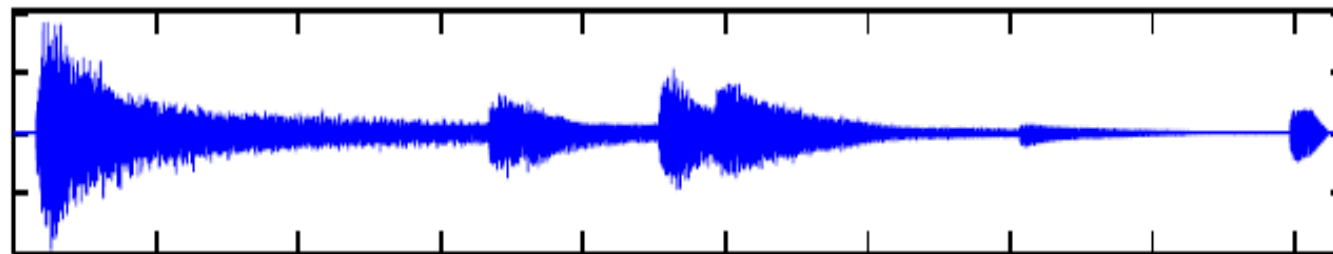


How to make the data comparable?

Image



Audio



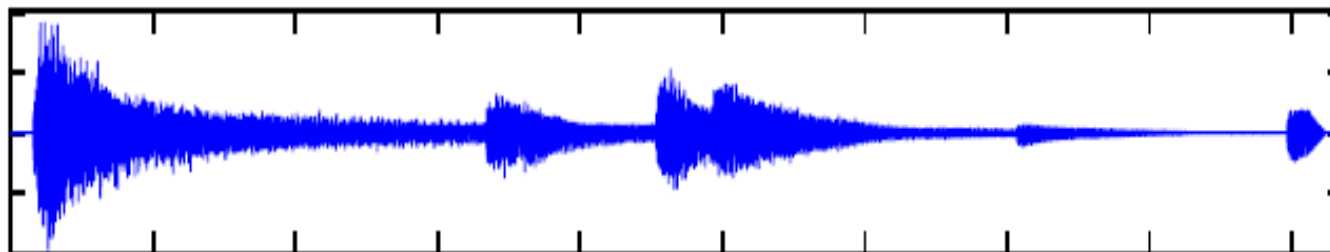
How to make the data comparable?

Image Processing: Optical Music Recognition

Image



Audio



How to make the data comparable?

Image Processing: Optical Music Recognition

Image



Audio



Audio Processing: Fourier Analysis



How to make the data comparable?

Image Processing: Optical Music Recognition

Image



Audio



Audio Processing: Fourier Analysis



Application: Score Viewer

The screenshot displays two windows from a music application. The top window, titled "ScoreViewer", shows a musical score for "Beethoven - Klaviersonaten Band 1 - Henle". The score is for "Sonata no.8 in C minor, op.13 'Pathétique' / Rondo (Allegro)". The score is displayed in a multi-staff format. A yellow highlight is visible on the first staff. Below the score, there are navigation controls: "Track: 29 / 54", "Bar: 1 / 211", and "Page: 159 / 285". There are also "Play" and "Stop" buttons, and a "Score Following On" indicator.

The bottom window, titled "AudioViewer", shows a track list for "Beethoven - Piano Sonatas-Alfred Brendel". The track list is as follows:

Track	Time
03 Sonata no.1 in F minor, op.2 no.1 / Menuetto (Allegretto)	3:24
04 Sonata no.1 in F minor, op.2 no.1 / Prestissimo	5:32
05 Sonata no.2 in A major, op.2 no.2 / Allegro vivace	7:15
06 Sonata no.2 in A major, op.2 no.2 / Largo appassionato	6:28
07 Sonata no.2 in A major, op.2 no.2 / Scherzo (Allegretto)	3:30
08 Sonata no.2 in A major, op.2 no.2 / Rondo (Grazioso)	7:03
09 Sonata no.8 in C minor, op.13 "Pathétique" / Allegro di molto e con brio	9:40
10 Sonata no.8 in C minor, op.13 "Pathétique" / Adagio cantabile	5:17
11 Sonata no.8 in C minor, op.13 "Pathétique" / Rondo (Allegro)	4:20

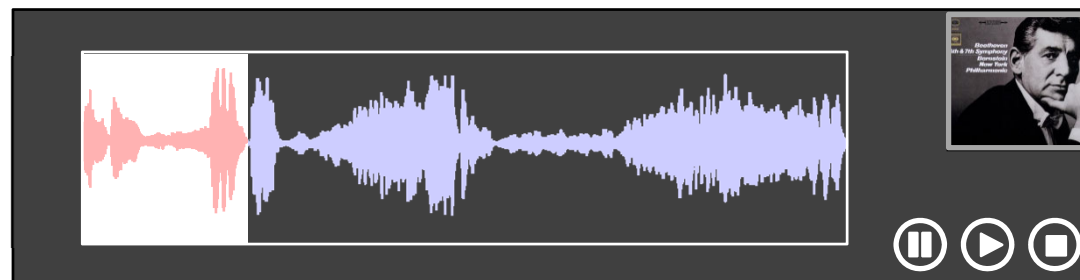
Below the track list, there is a waveform visualization. At the bottom of the window, there are navigation controls: "Disc: 1 / 11", "Track: 11 / 11", and "Time: 00:00.00 / 4:30.35". There are also "Play" and "Stop" buttons.



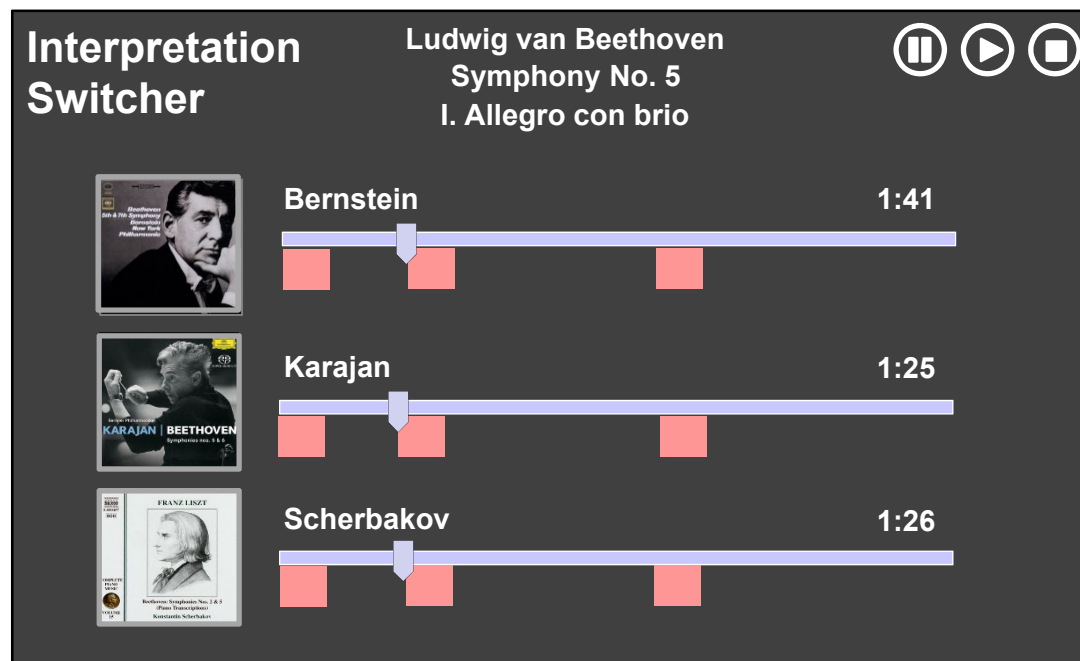
Audio Matching

Task

Query:



Database: Matches

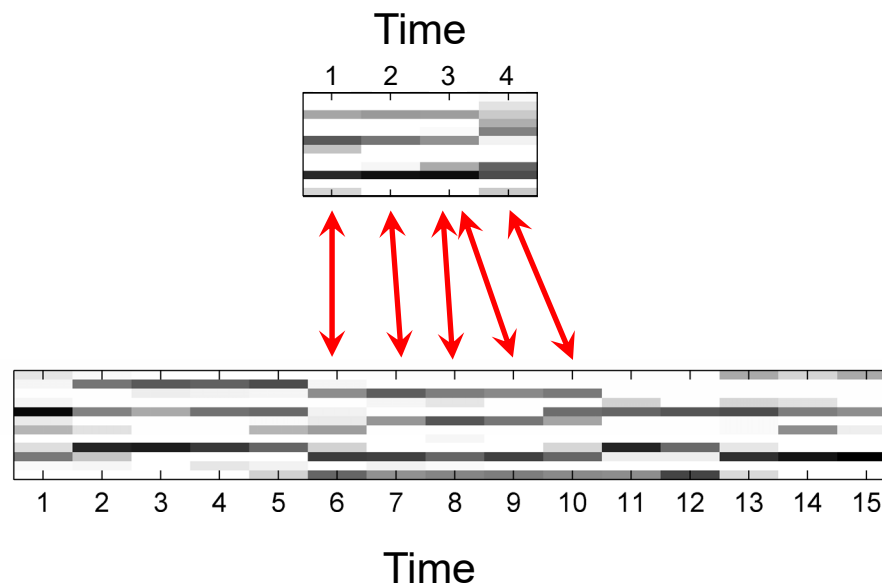


Audio Matching

Task

Query: Sequence X

Database: Sequence Y

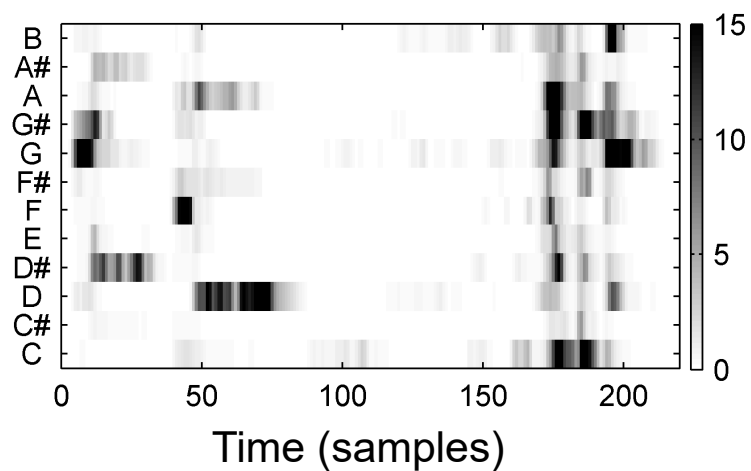


Subsequence matching

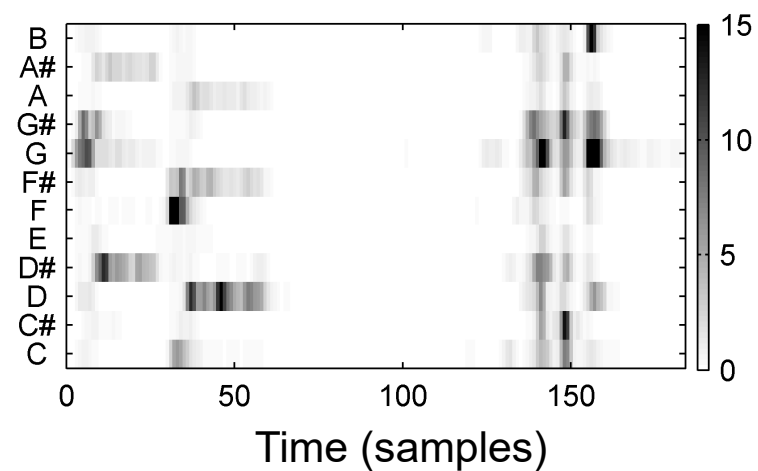
Audio Features

Example: Beethoven's Fifth

Bernstein



Karajan

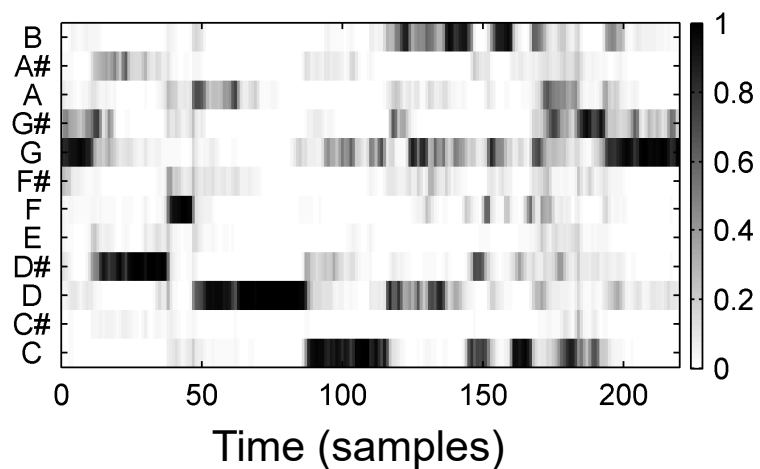


Chroma representation (10 Hz)

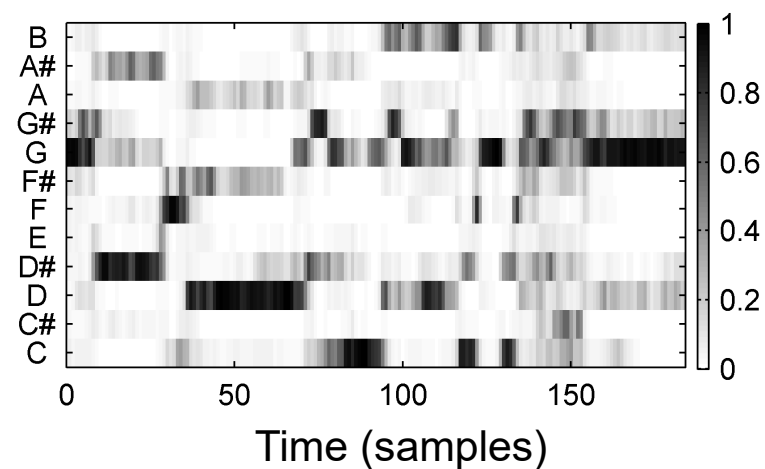
Audio Features

Example: Beethoven's Fifth

Bernstein



Karajan



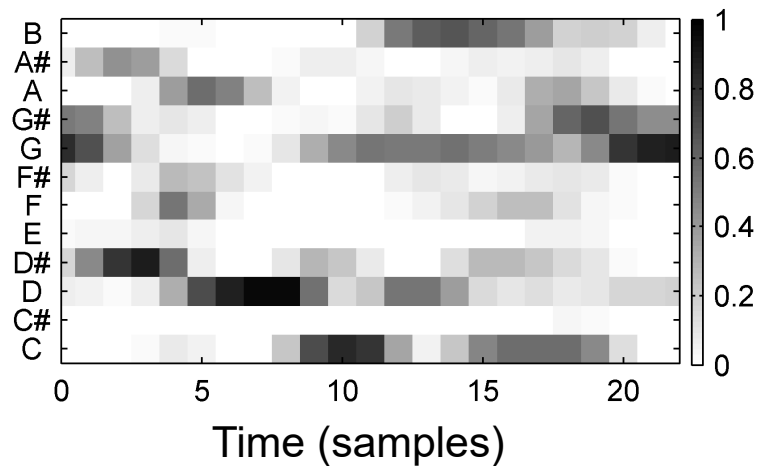
Chroma representation (10 Hz)

- Normalization

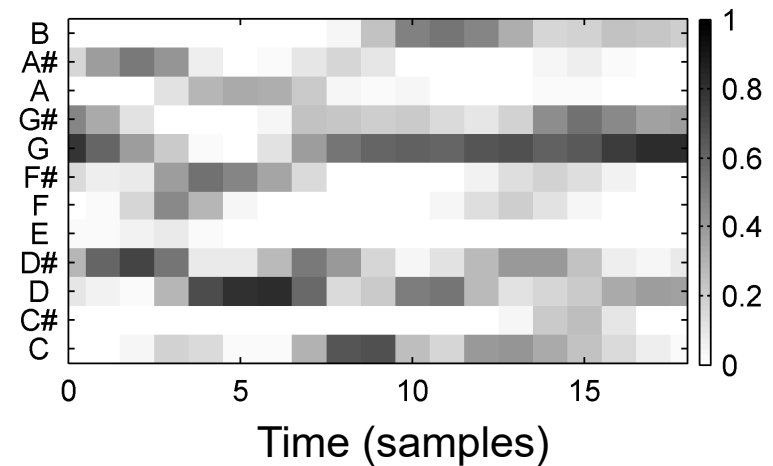
Audio Features

Example: Beethoven's Fifth

Bernstein



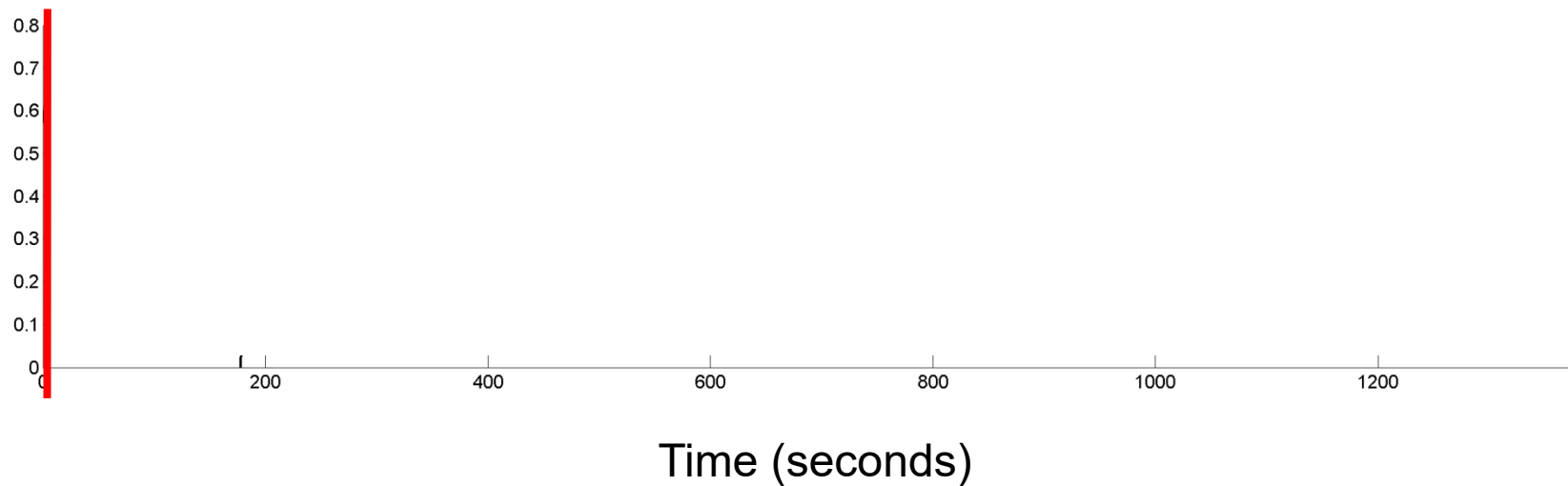
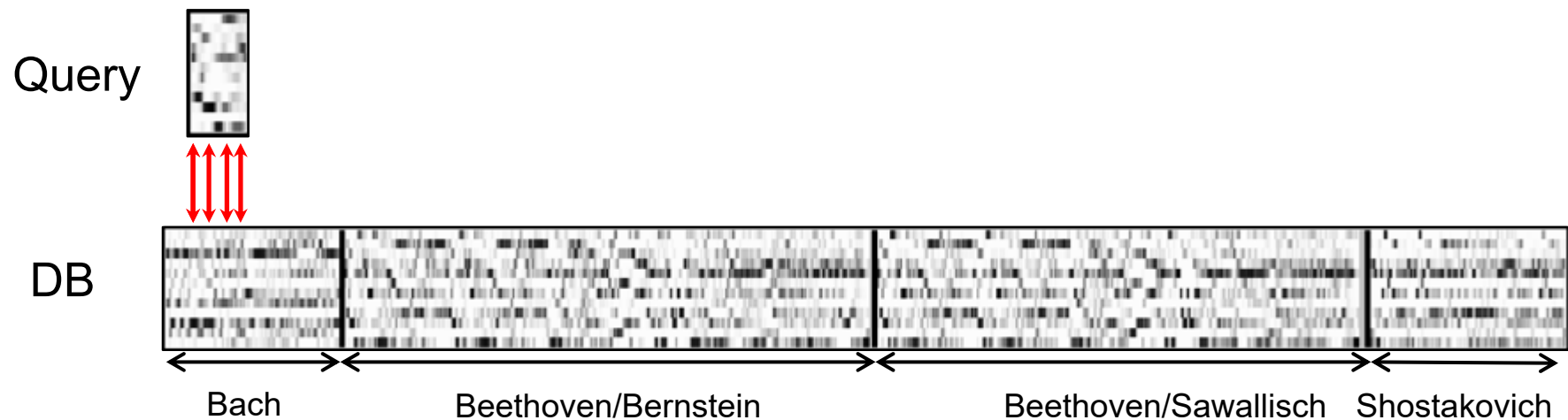
Karajan



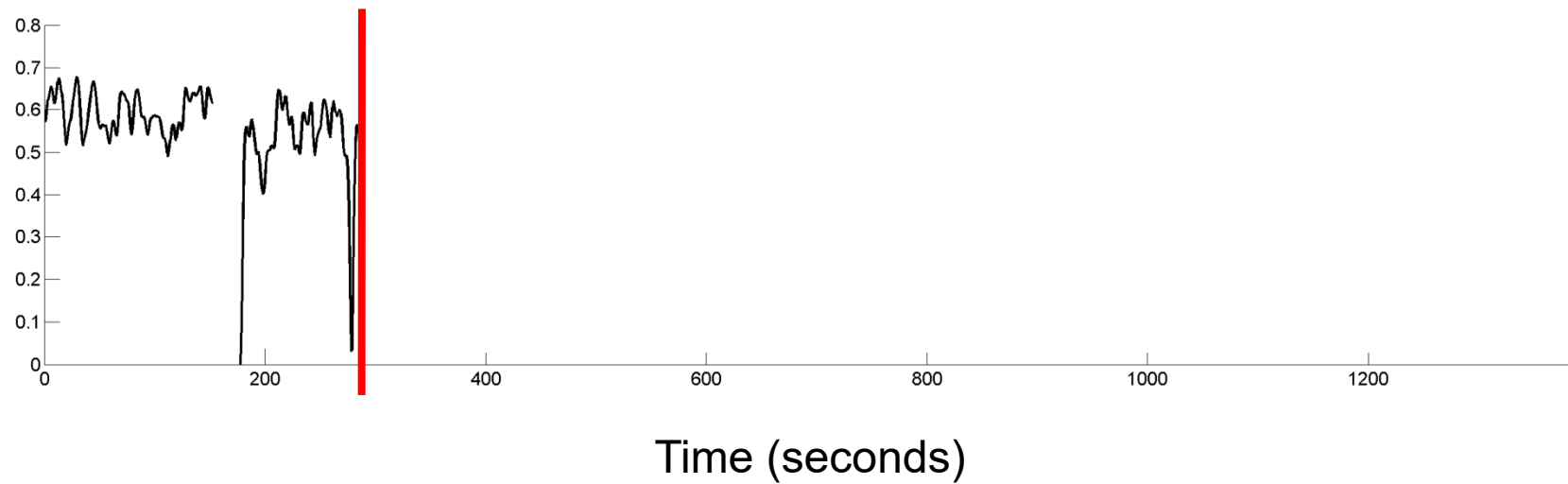
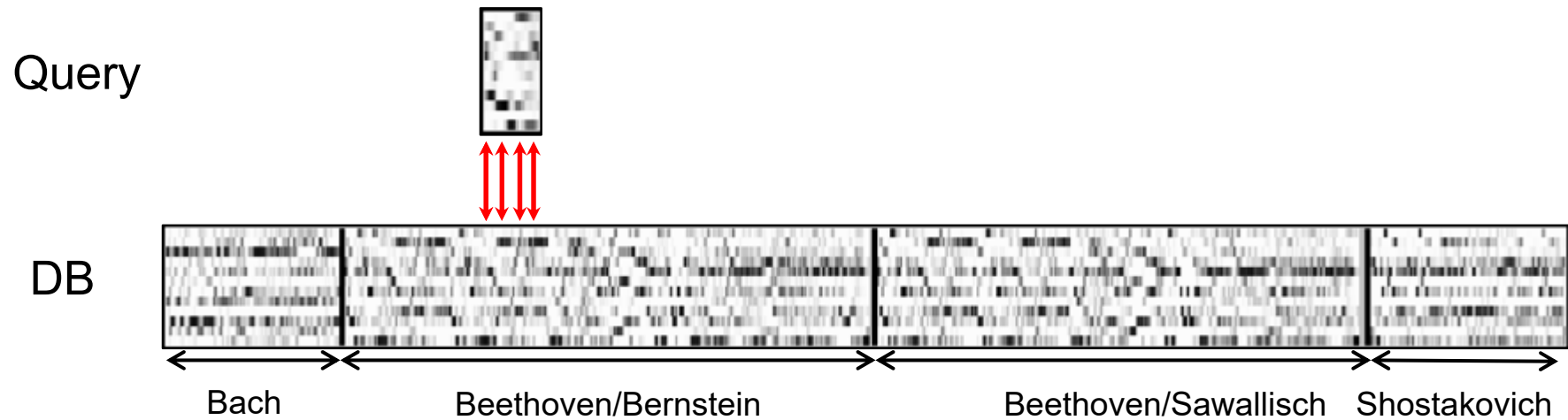
Chroma representation (1 Hz)

- Normalization
- Smoothing & downsampling

Matching Procedure



Matching Procedure

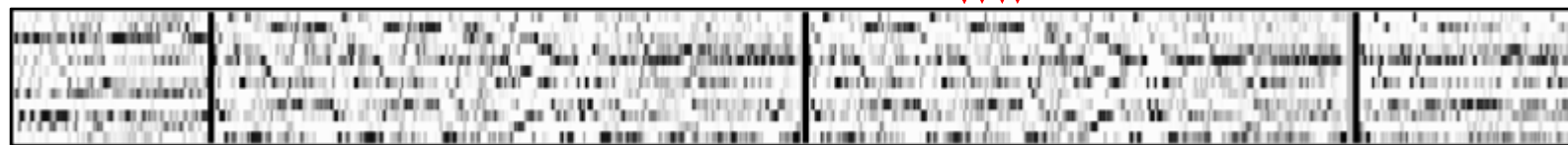


Matching Procedure

Query



DB

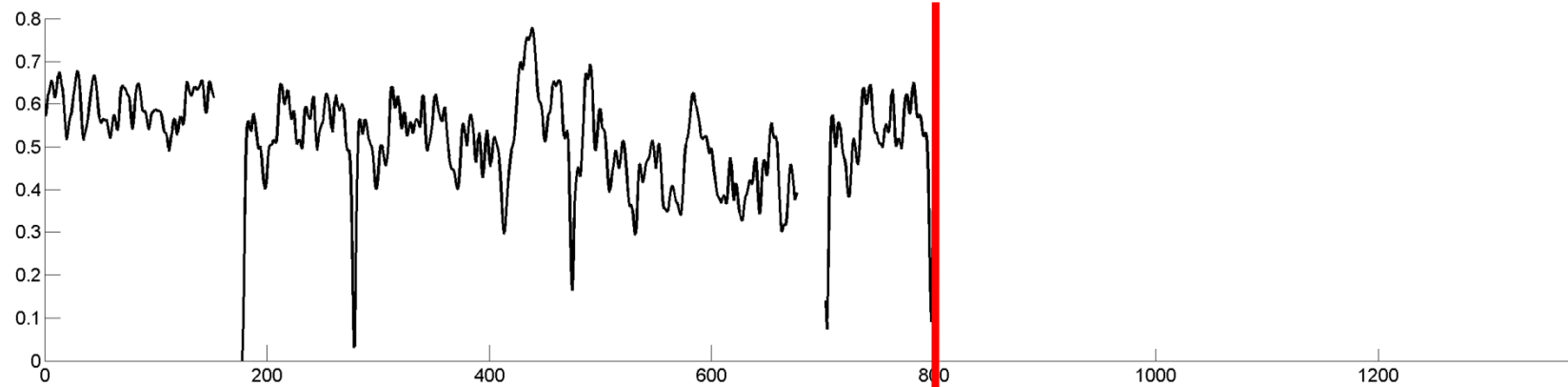


Bach

Beethoven/Bernstein

Beethoven/Sawallisch

Shostakovich



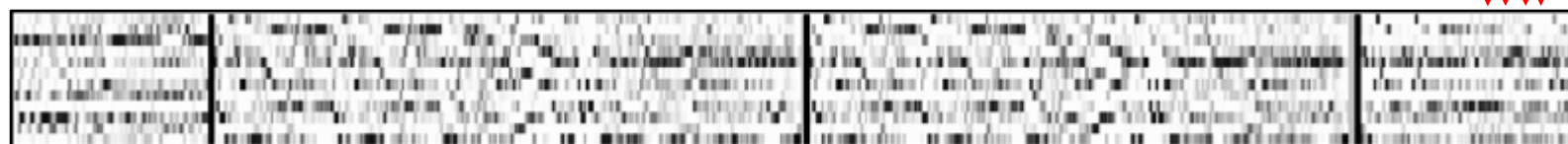
Time (seconds)

Matching Procedure

Query



DB

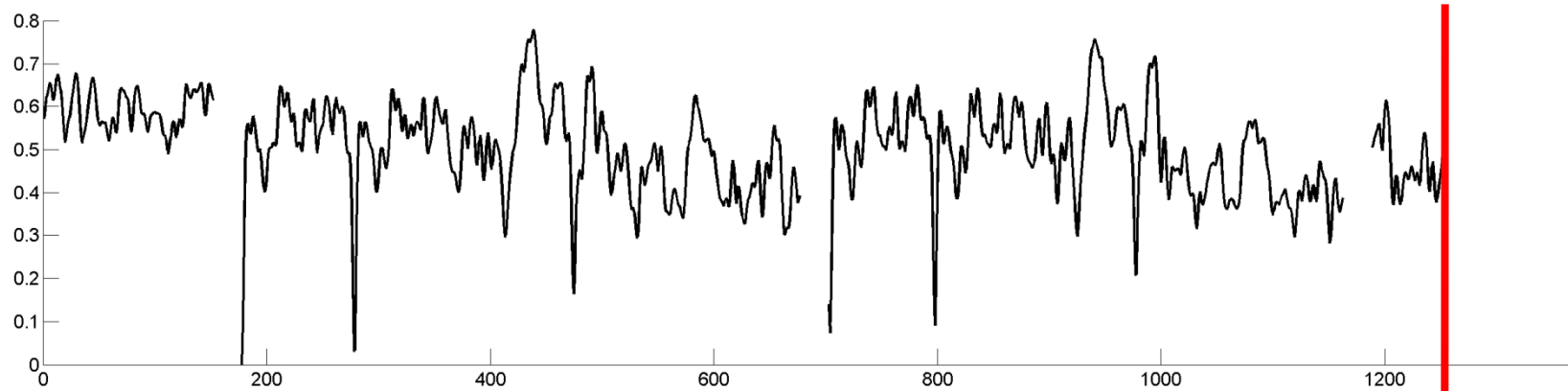


Bach

Beethoven/Bernstein

Beethoven/Sawallisch

Shostakovich

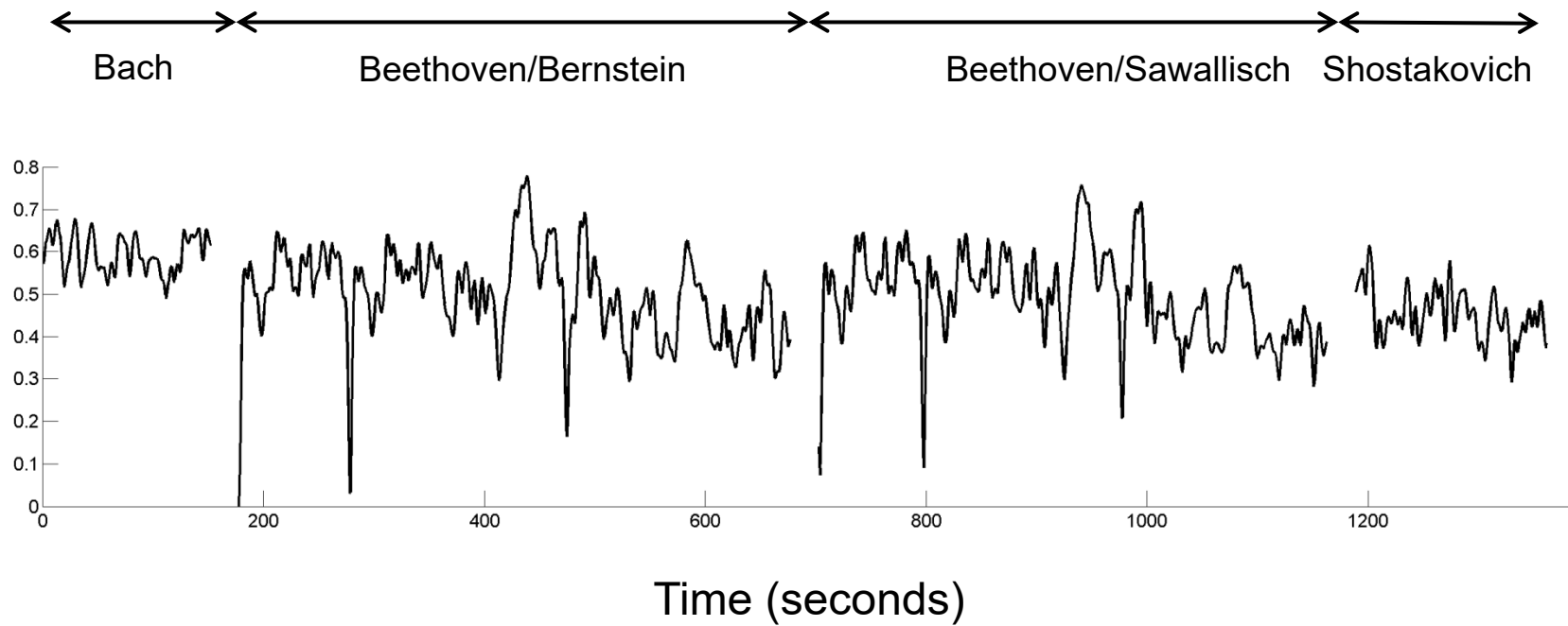


Time (seconds)

Matching Procedure

Matching curve

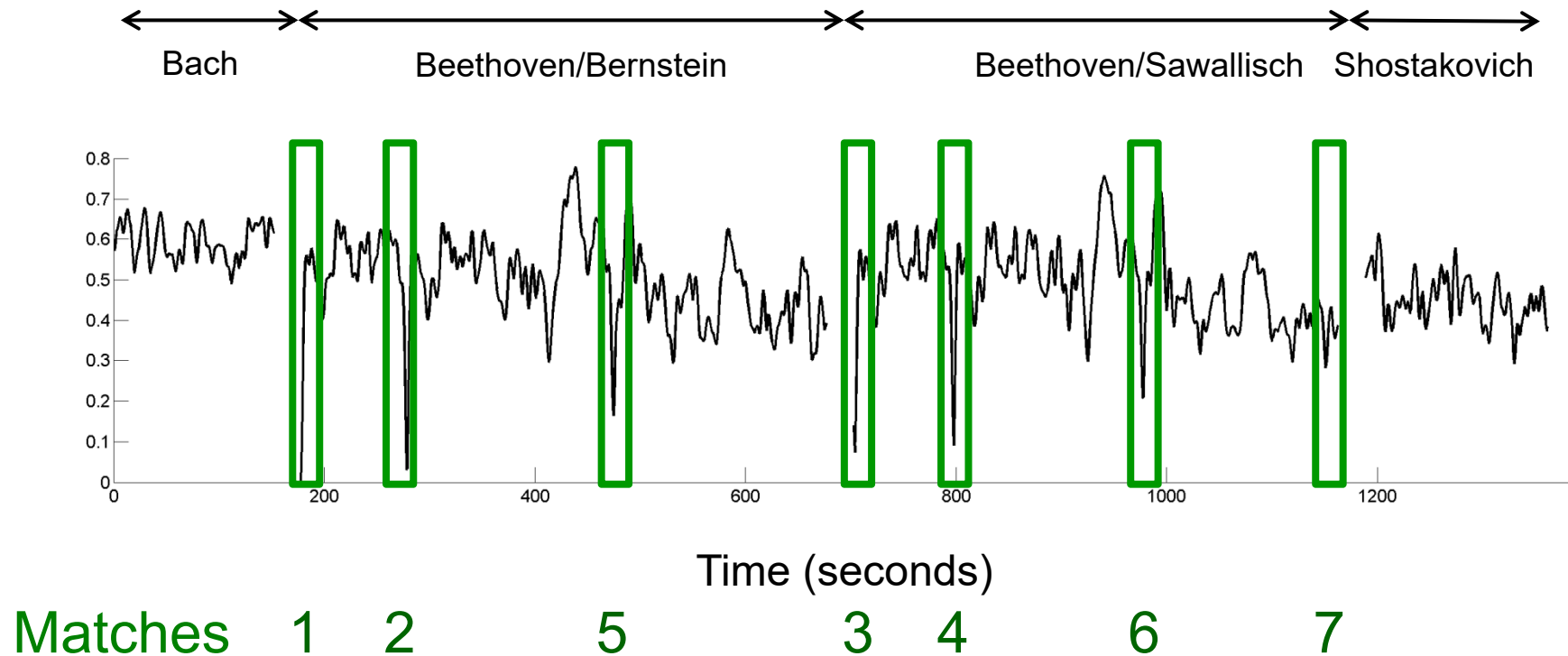
Query: Beethoven's Fifth / Bernstein (first 20 seconds)



Matching Procedure

Matching curve

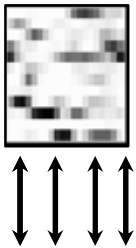
Query: Beethoven's Fifth / Bernstein (first 20 seconds)



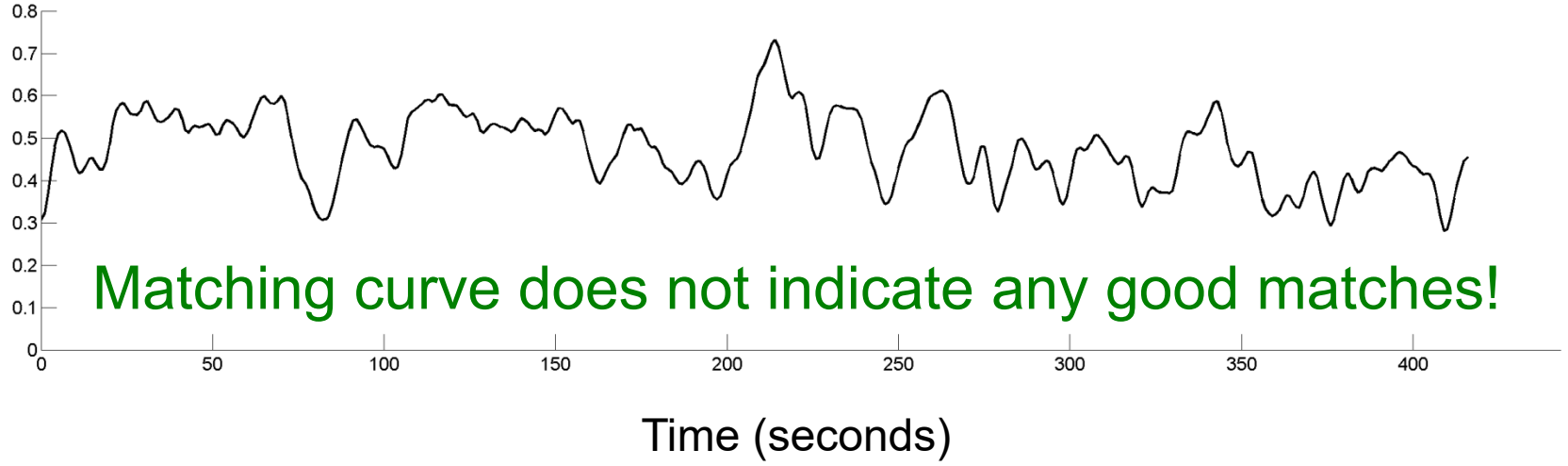
Matching Procedure

Problem: How to deal with tempo differences?

Karajan is much faster than Bernstein!



Beethoven/Karajan



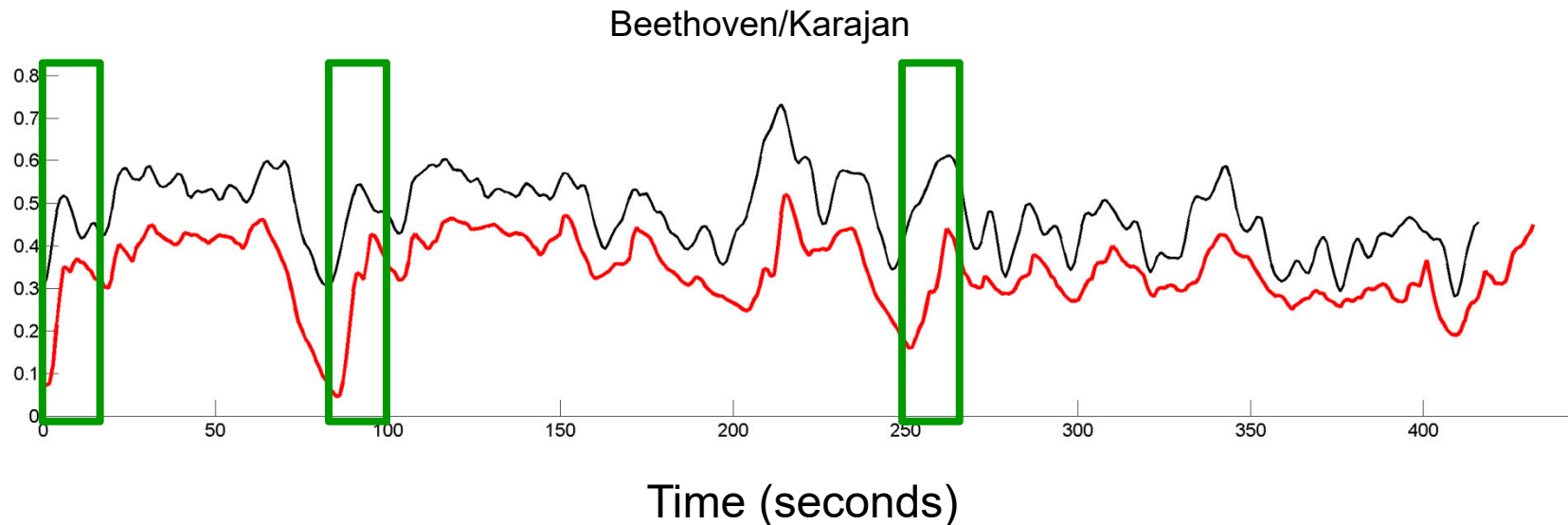
Matching Procedure

1. Strategy: Usage of local warping

Karajan is much faster than Bernstein!

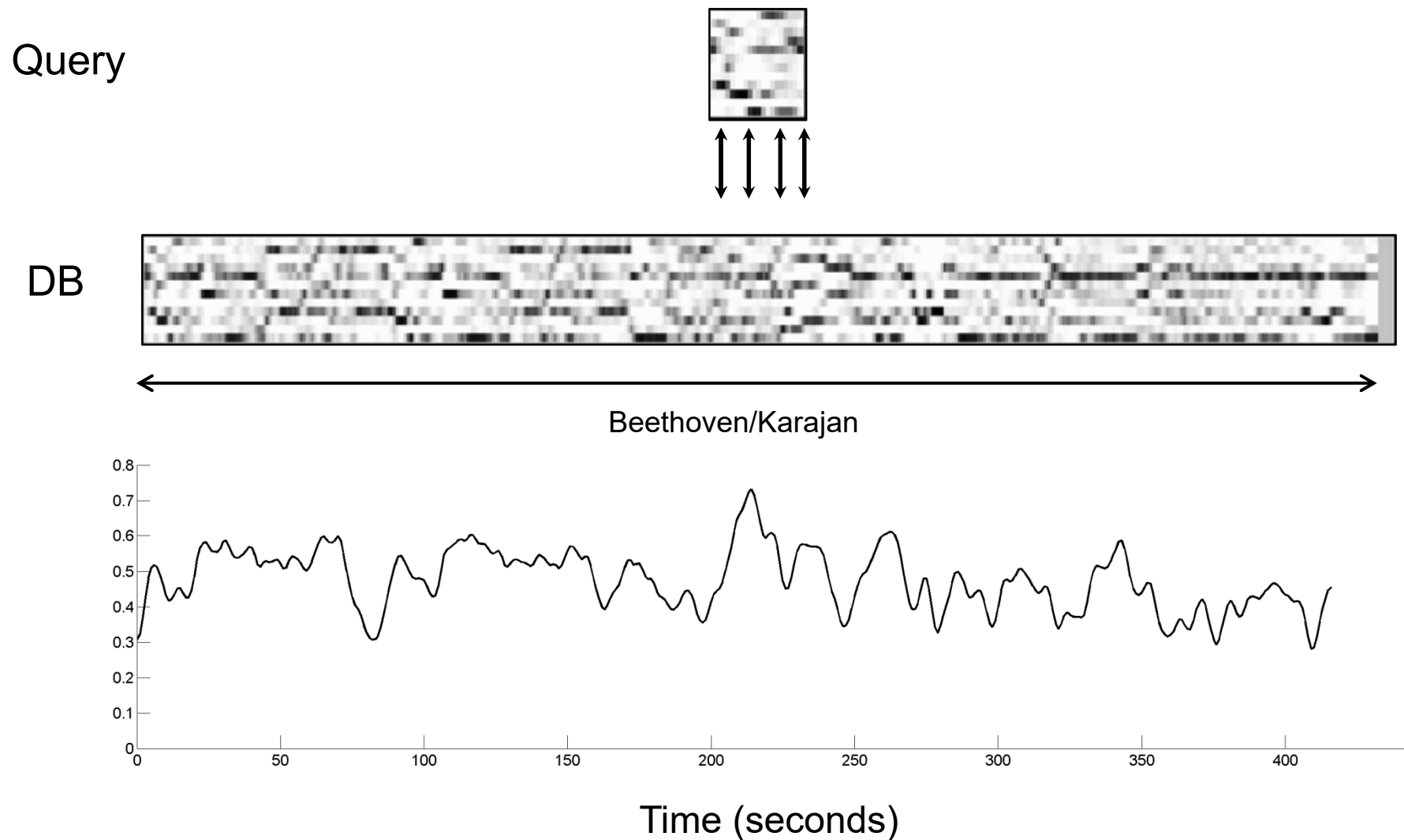


Warping strategies are computationally expensive and hard for indexing.



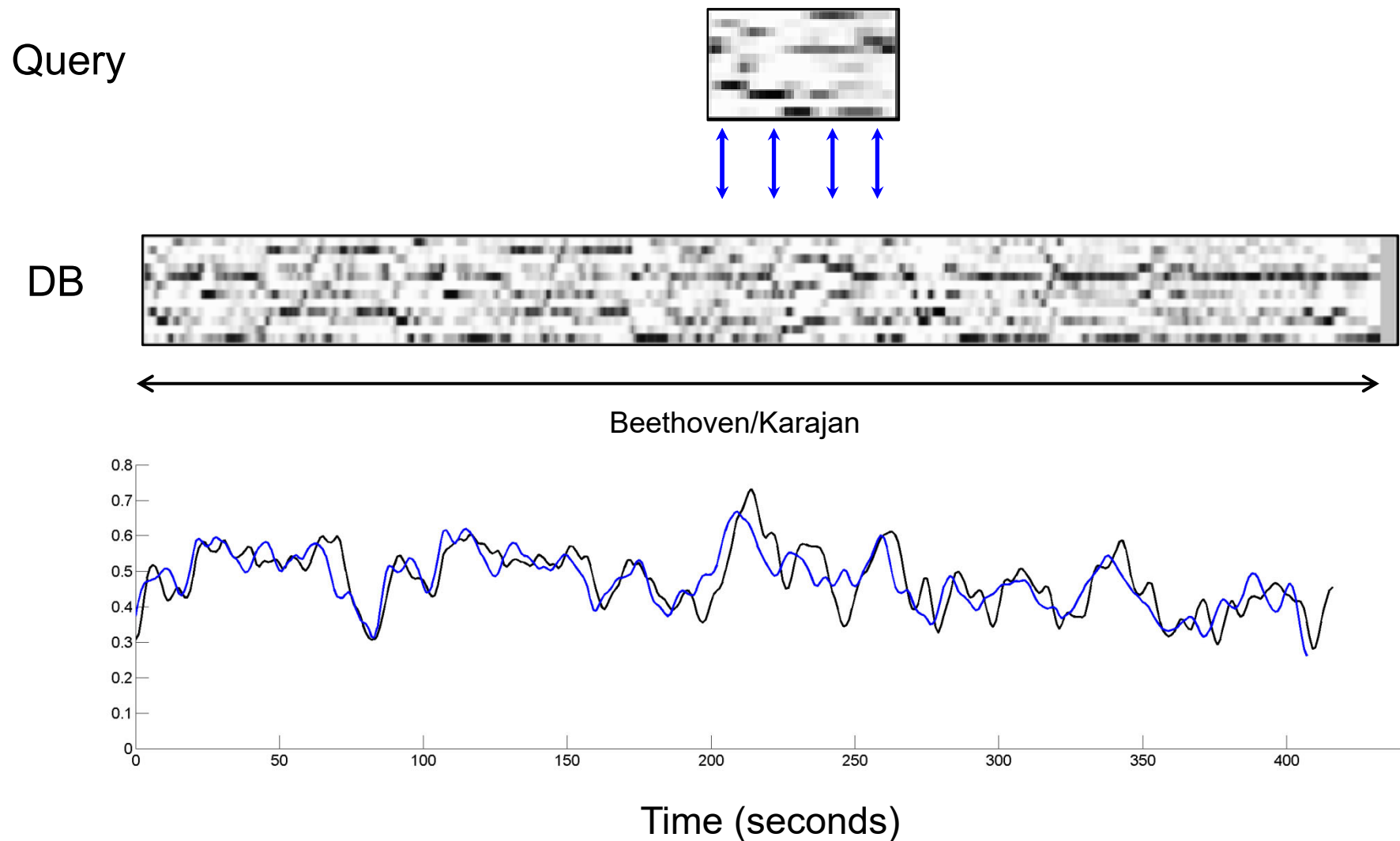
Matching Procedure

2. Strategy: Usage of multiple scaling



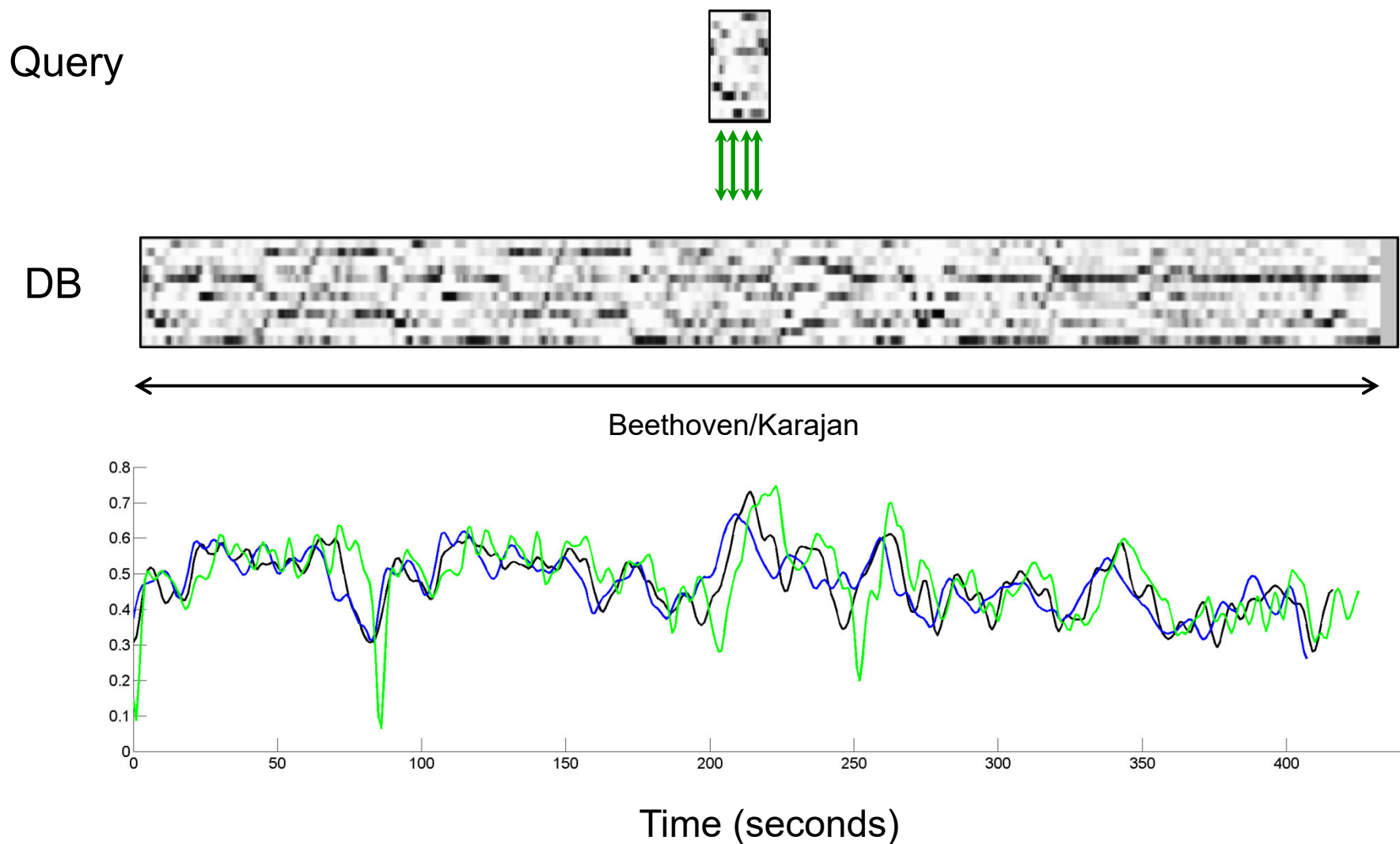
Matching Procedure

2. Strategy: Usage of multiple scaling



Matching Procedure

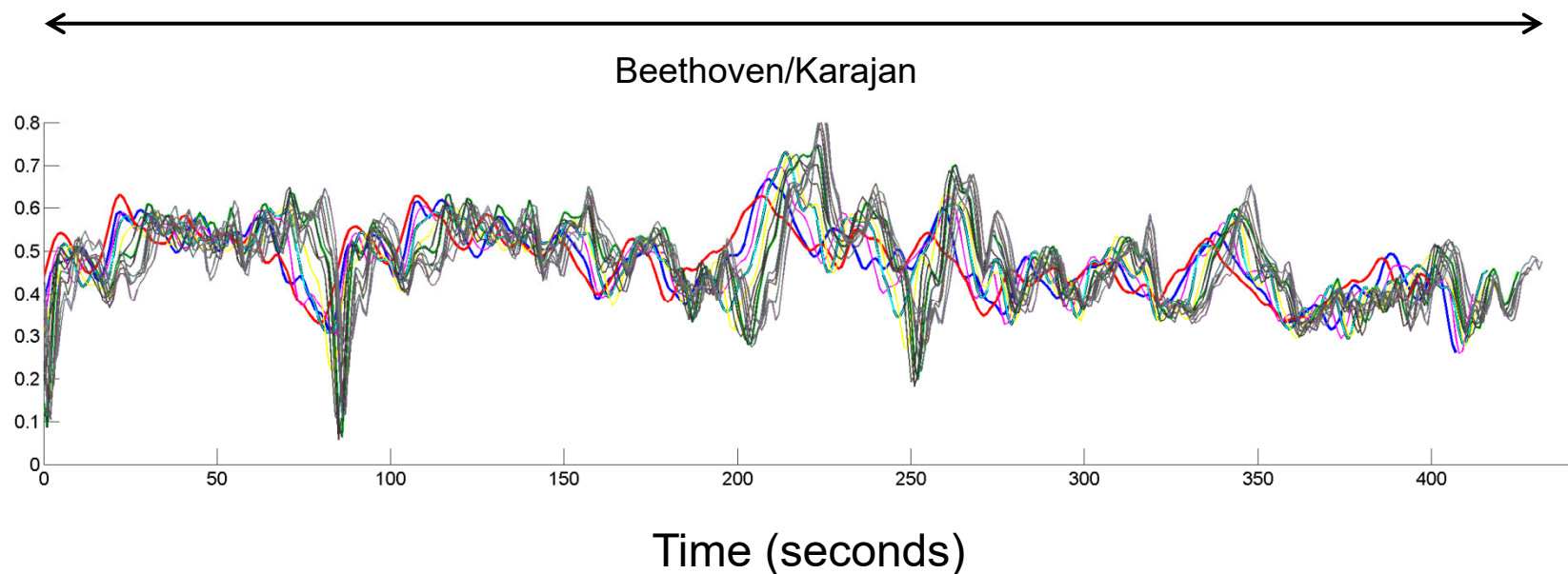
2. Strategy: Usage of multiple scaling



Matching Procedure

2. Strategy: Usage of multiple scaling

Query resampling simulates tempo changes

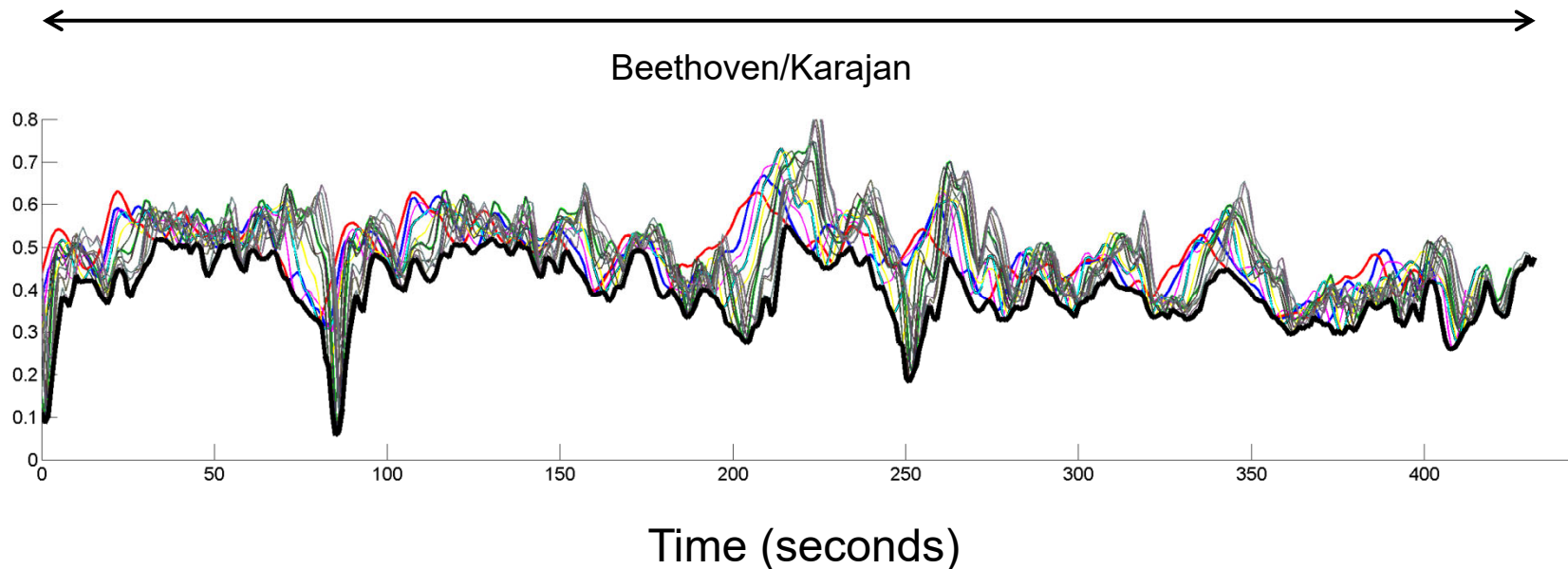


Matching Procedure

2. Strategy: Usage of multiple scaling

Query resampling simulates tempo changes

Minimize over all curves



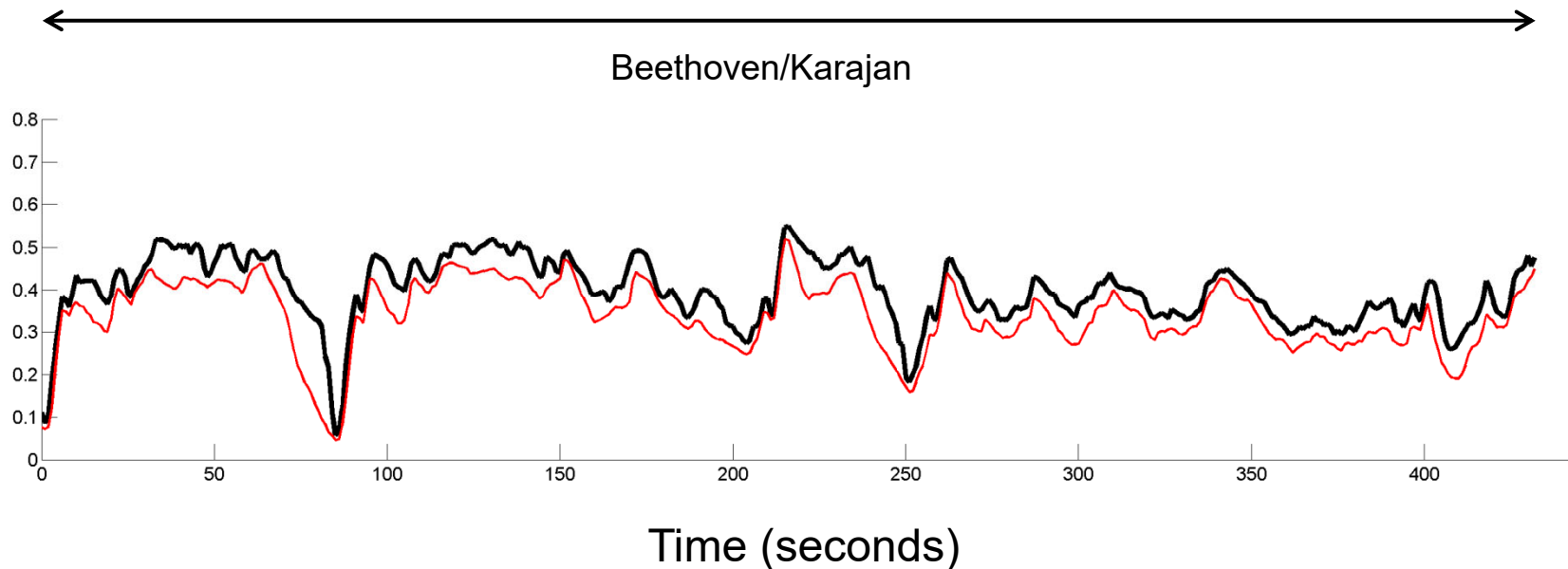
Matching Procedure

2. Strategy: Usage of multiple scaling

Query resampling simulates tempo changes









Minimize over all curves

Resulting curve is similar to **warping curve**



Audio Matching

Query: Beethoven's Fifth / Bernstein (first 20 seconds)

Rank	Piece	Position	
1	Beethoven's Fifth/Bernstein	0 - 21	
2	Beethoven's Fifth/Bernstein	101- 122	
3	Beethoven's Fifth/Karajan	86 - 103	
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
10	Beethoven's Fifth/Karajan	252 - 271	
11	Beethoven's Fifth/Scherbakov	0 - 19	
12	Beethoven's Fifth/Sawallisch	275 - 296	
13	Beethoven's Fifth/Scherbakov	86 - 103	
14	Schumann Op. 97,1/Levine	28 - 43	

Audio Matching: Conclusions

Strategy: Handle variations at various levels

- Chroma → invariance to timbre
- Normalization → invariance to dynamics
- Smoothing → invariance to local time deviations
- Multiple queries → invariance to global tempo

Audio Matching: Conclusions

Strategy: Handle variations at various levels

- Chroma → invariance to timbre
- Normalization → invariance to dynamics
- Smoothing → invariance to local time deviations
- Multiple queries → invariance to global tempo

Notes:

- There is no “standard” chroma feature.
→ Variants can make a huge difference!
- Learn invariance from examples
→ “Deep Chroma” [Korzeniowski, Widmer; ISMIR 2016]
- Temporal warping makes problem hard
- Efficiency