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

Music Representations and Retrieval

Meinard Müller, Frank Zalkow

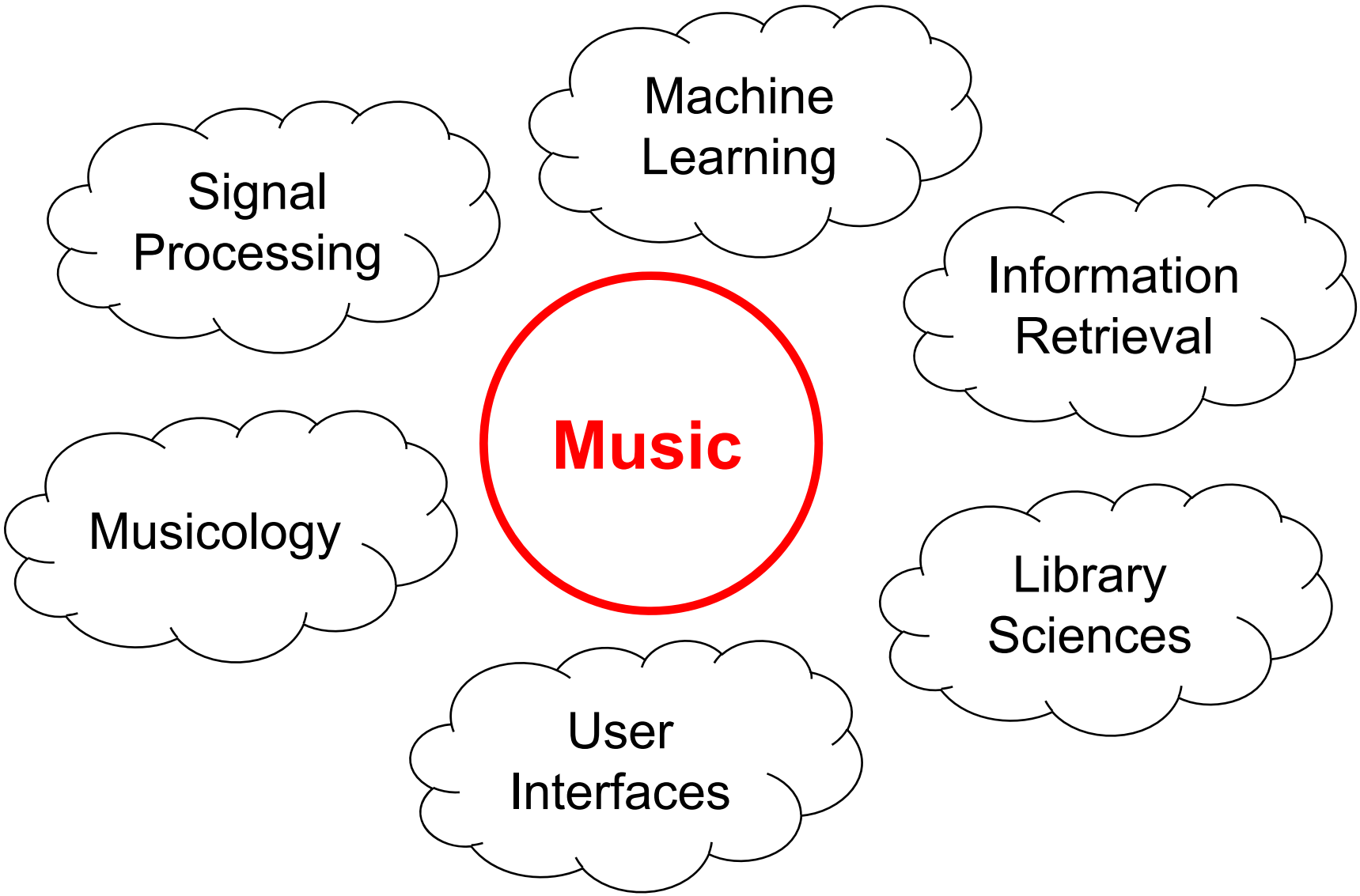
International Audio Laboratories Erlangen
meinard.mueller@audiolabs-erlangen.de, frank.zalkow@audiolabs-erlangen.de

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

Fraunhofer
IIS

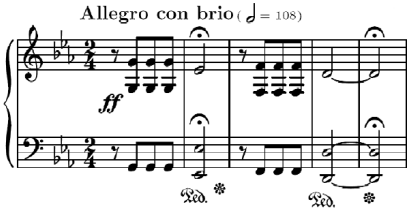
Music

Music Information Retrieval (MIR)

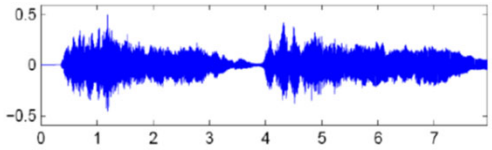


Music Information Retrieval (MIR)

Sheet Music (Image)



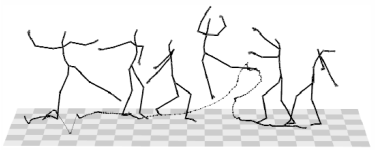
CD / MP3 (Audio)



MusicXML (Text)

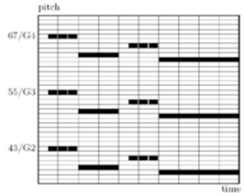
```
<note>
  <pitch>
    <step>E</step>
    <alter>-1</alter>
    <octave>4</octave>
  </pitch>
  <duration>2</duration>
  <type>half</type>
</note>
```

Dance / Motion (Mocap)



Music

MIDI

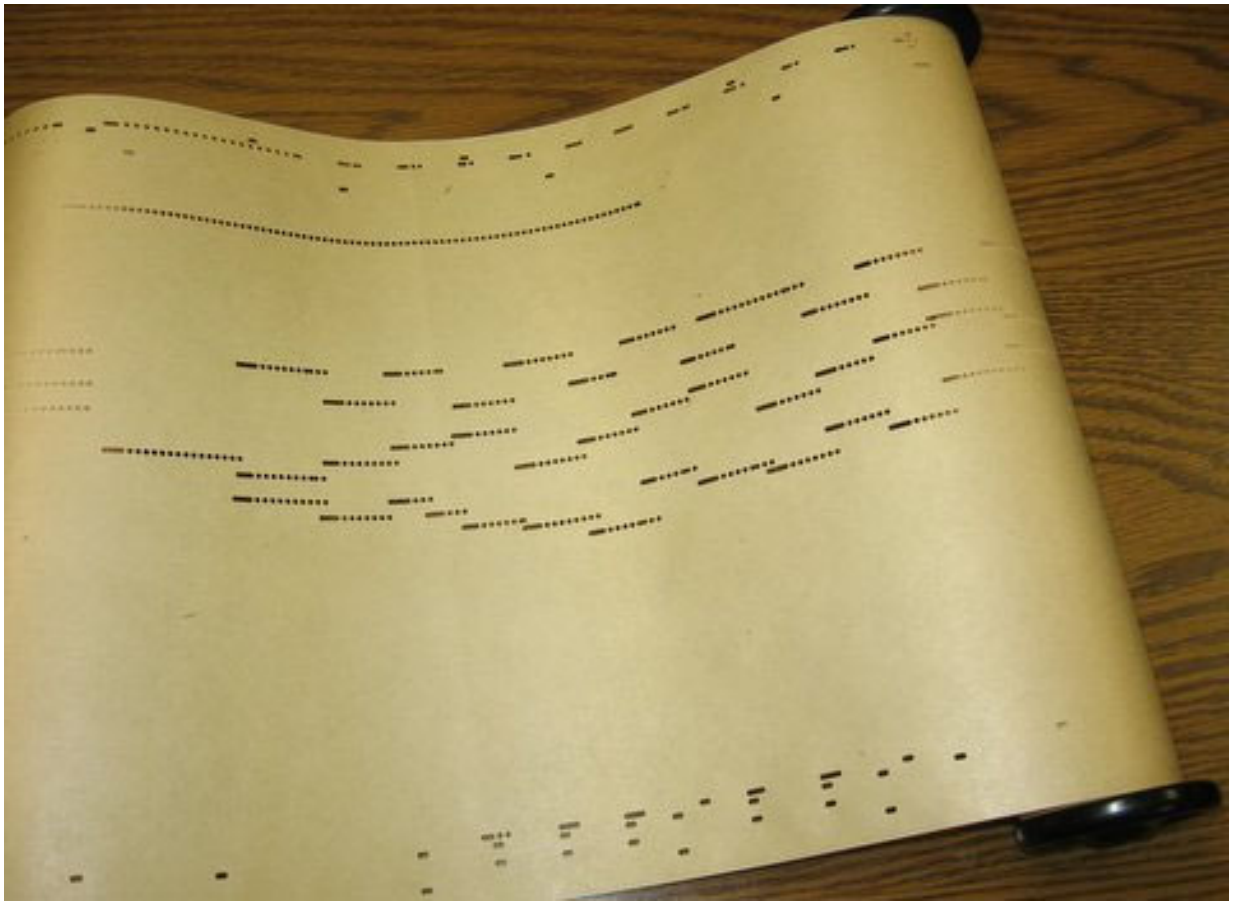


Singing / Voice (Audio)

Music Film (Video)

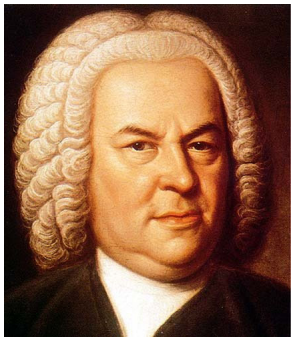
Music Literature (Text)

Piano Roll Representation



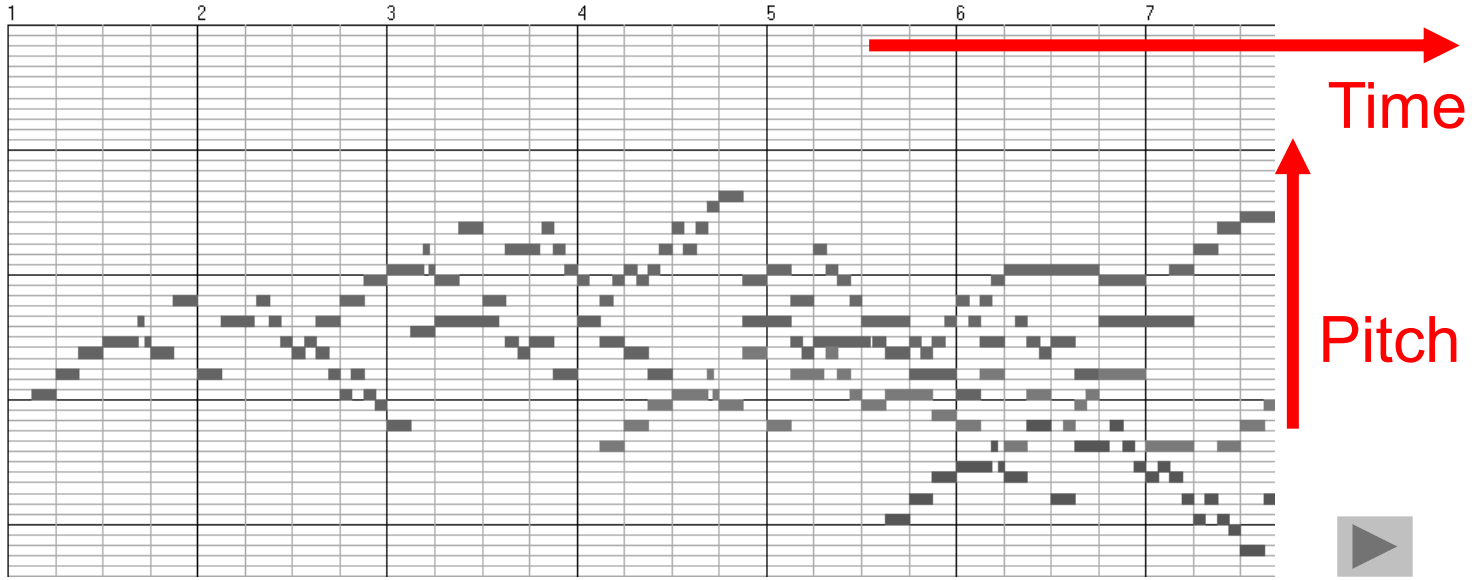
Player Piano (1900)

Piano Roll Representation (MIDI)

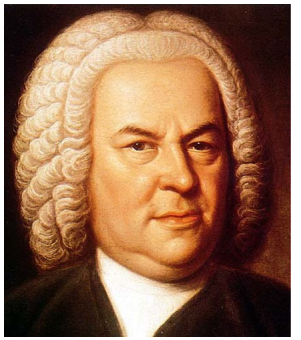


J.S. Bach, C-Major Fuge

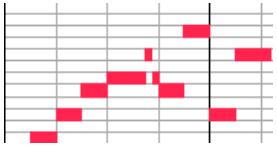
(Well Tempered Piano, BWV 846)



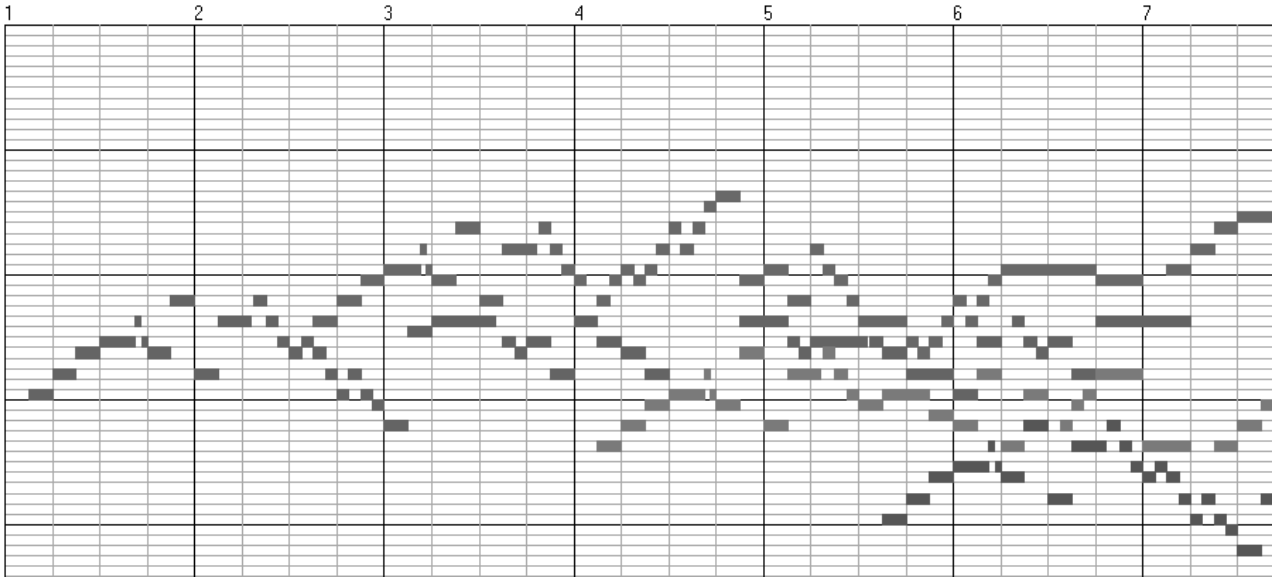
Piano Roll Representation (MIDI)



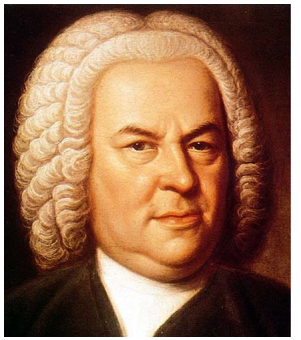
Query:



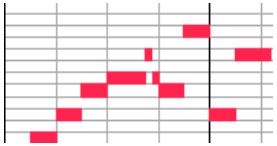
Goal: Find all occurrences of the query



Piano Roll Representation (MIDI)

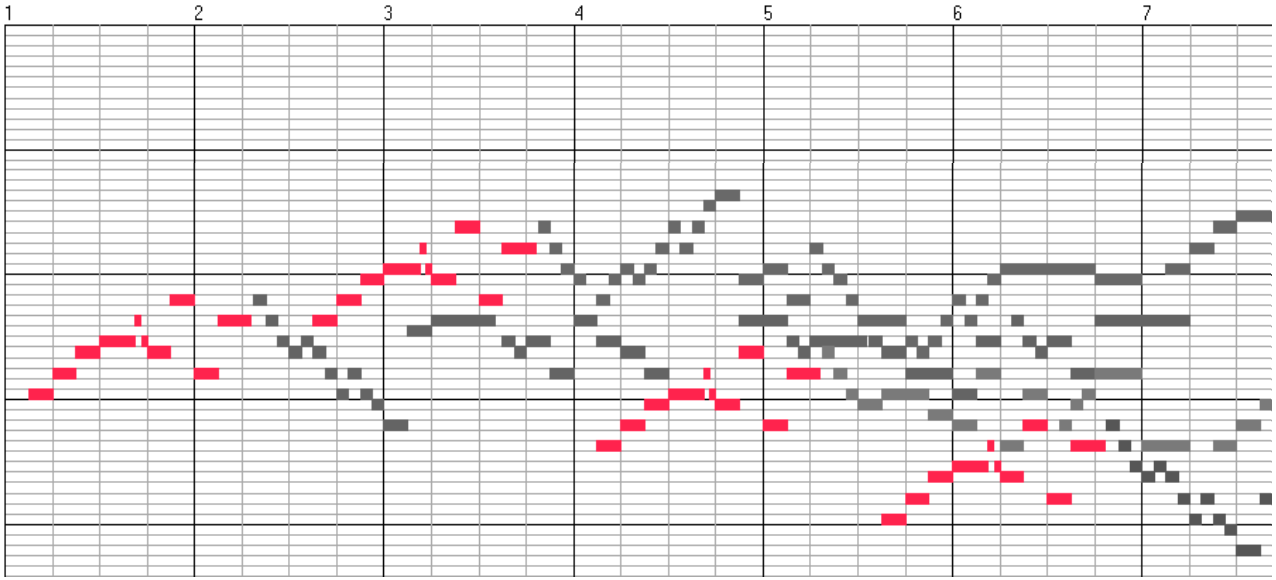


Query:

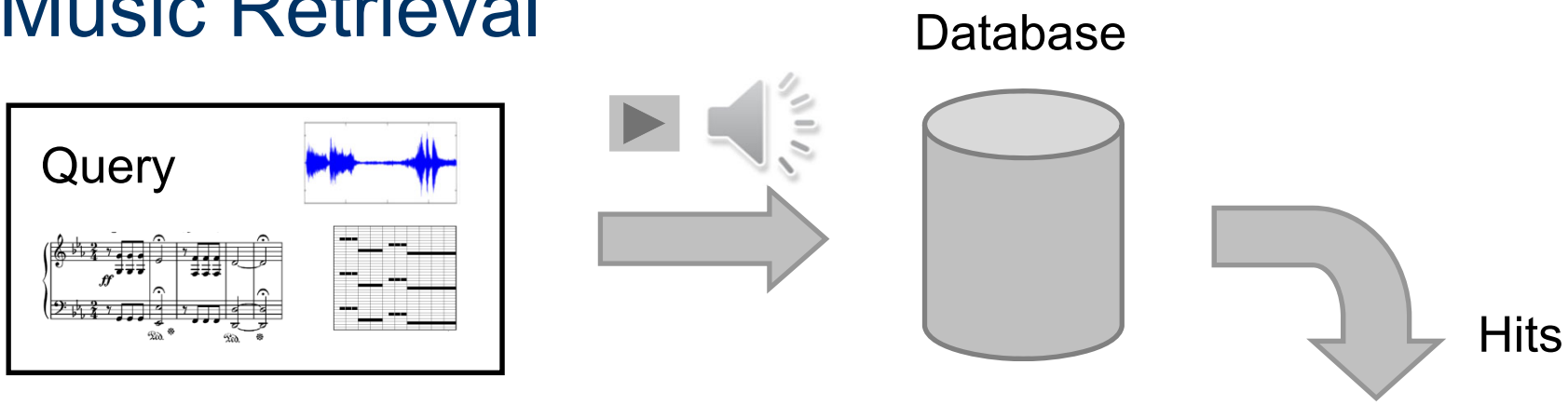


Goal: Find all occurrences of the query

Matches:



Music Retrieval



Retrieval tasks:

Audio identification

Bernstein (1962)
Beethoven, Symphony No. 5

Audio matching

Beethoven, Symphony No. 5:
- Bernstein (1962)
- Karajan (1982)
- Gould (1992)

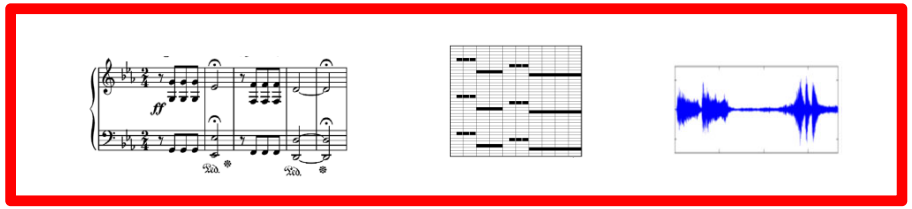
Version identification

Category-based music retrieval

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

Music Retrieval

Modalities



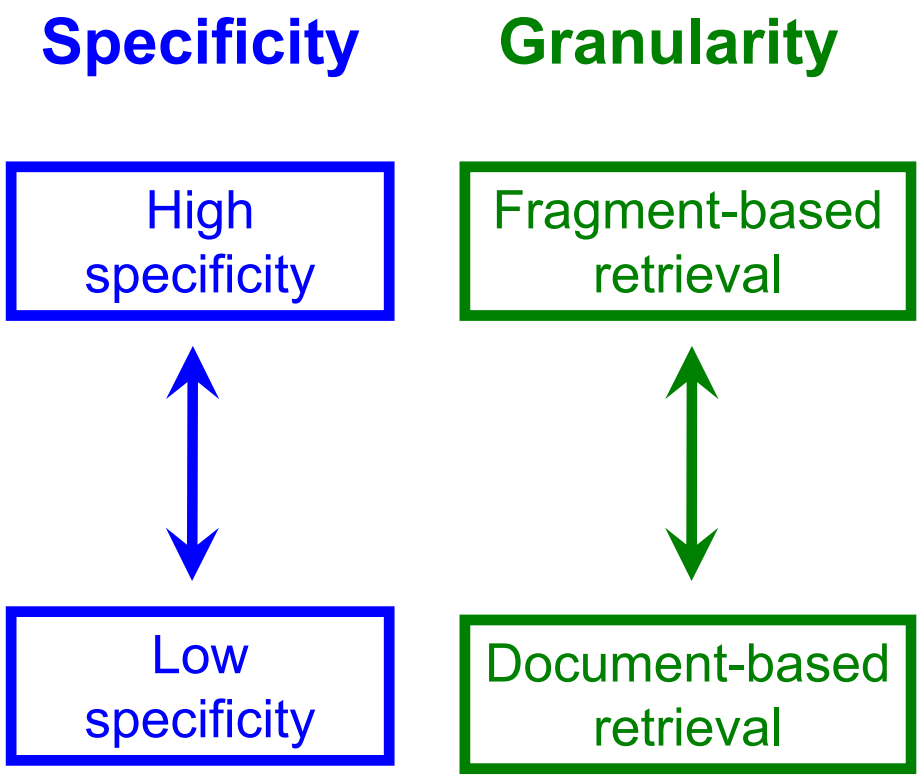
Retrieval tasks:

Audio identification

Audio matching

Version identification

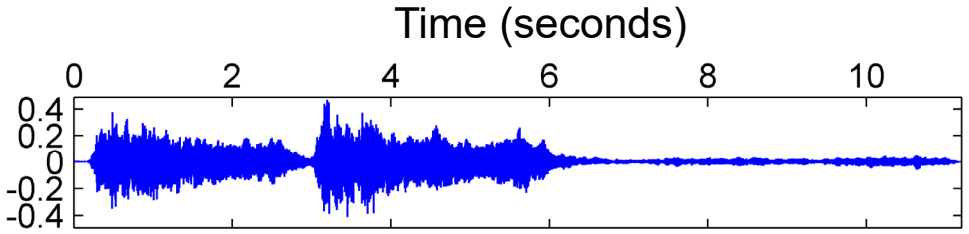
Category-based music retrieval



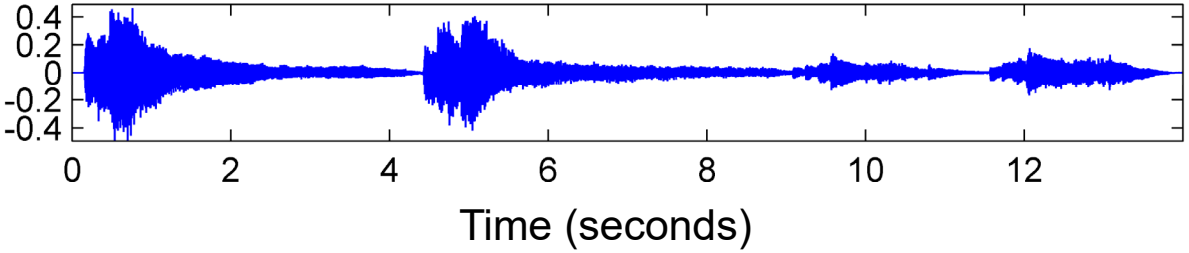
Music Synchronization: Audio-Audio

Beethoven's Fifth

Karajan



Gould

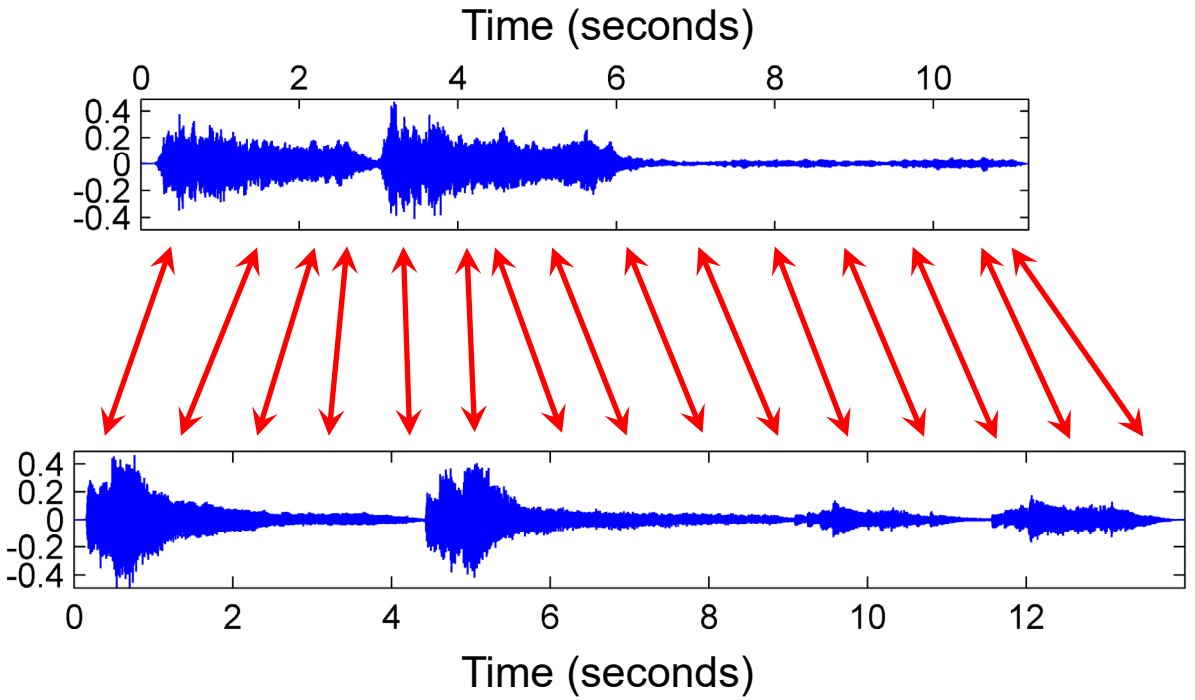


Music Synchronization: Audio-Audio

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

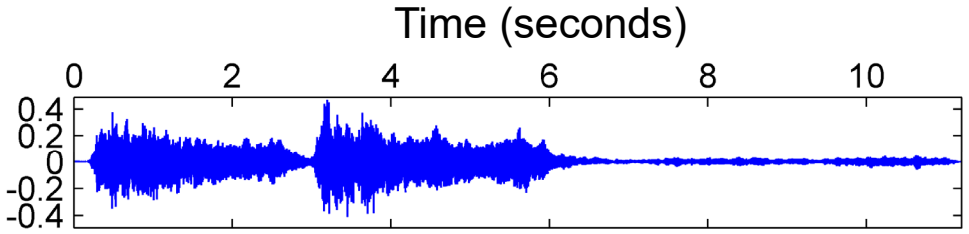
- Capturing local and global tempo variations
- Trade-off: Robustness vs. accuracy
- Efficiency

➡ **Dynamic time warping (DTW)**

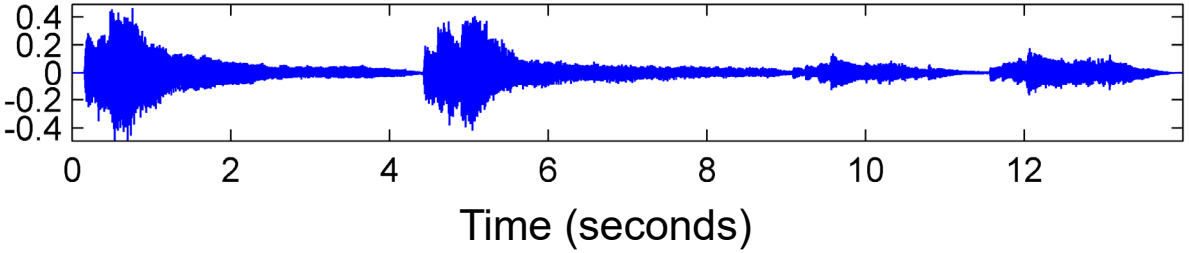
Music Synchronization: Audio-Audio

Beethoven's Fifth

Karajan



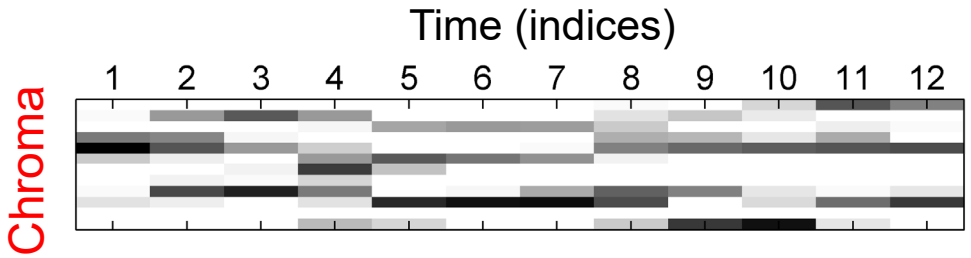
Gould



Music Synchronization: Audio-Audio

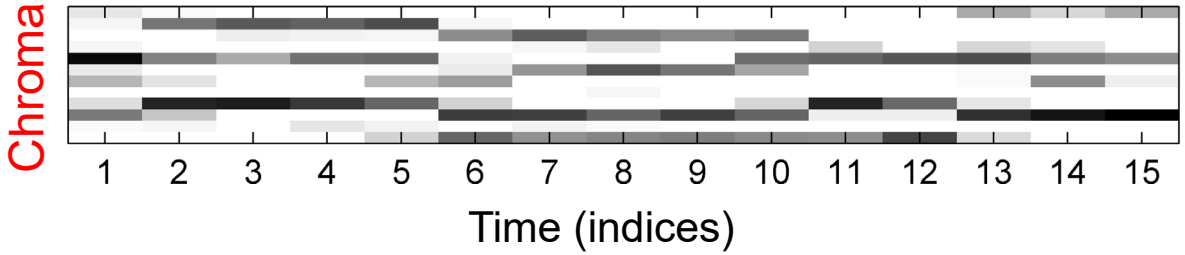
Beethoven's Fifth

Karajan



Time–chroma representations

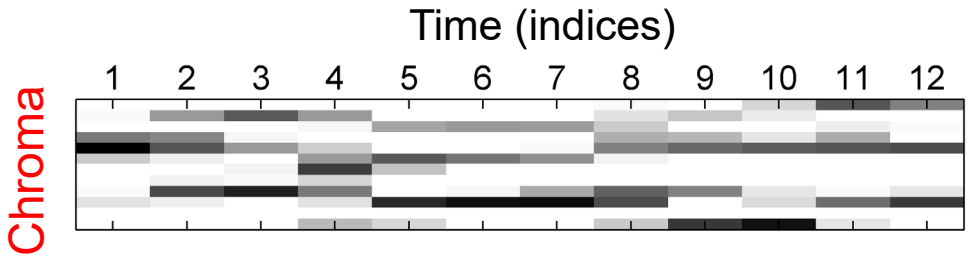
Gould



Music Synchronization: Audio-Audio

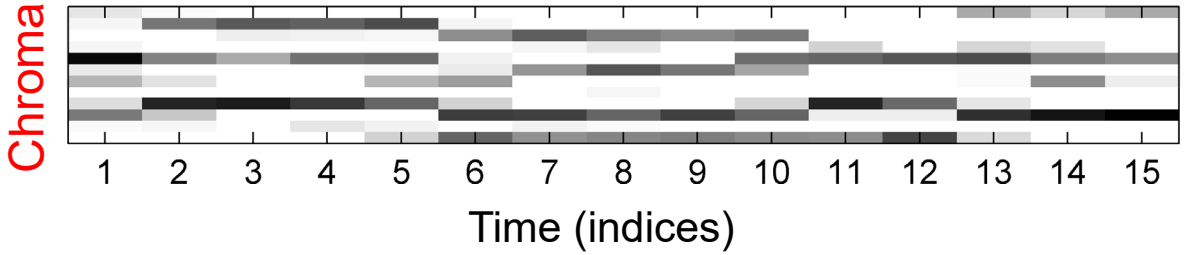
Beethoven's Fifth

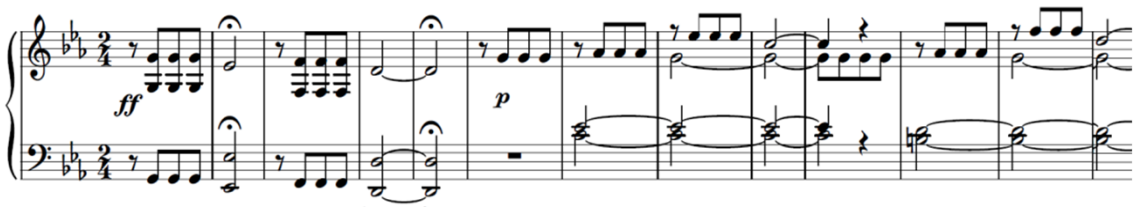
Karajan



Time–chroma representations

Gould

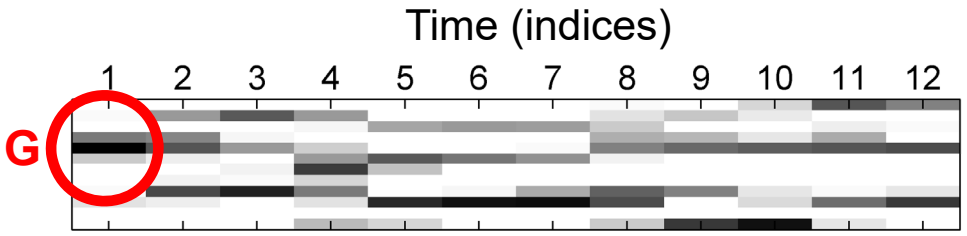




Music Synchronization: Audio-Audio

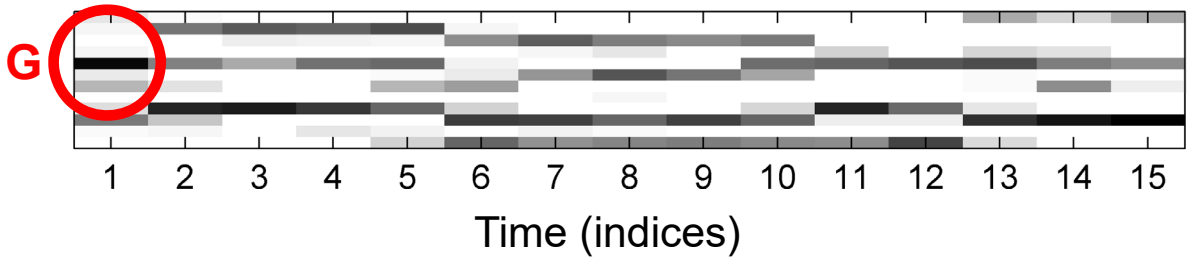
Beethoven's Fifth

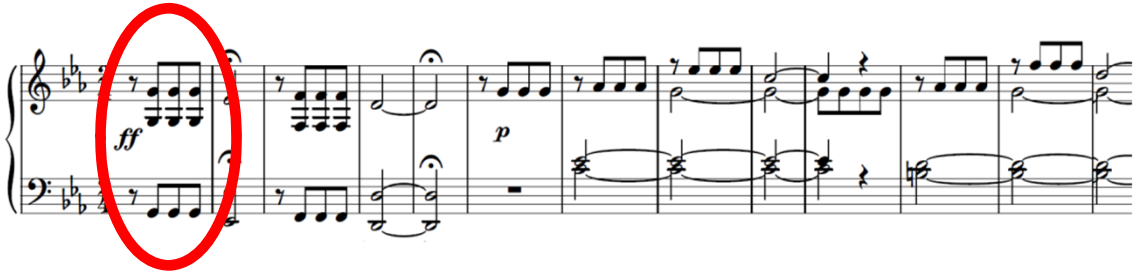
Karajan



Time–chroma representations

Gould

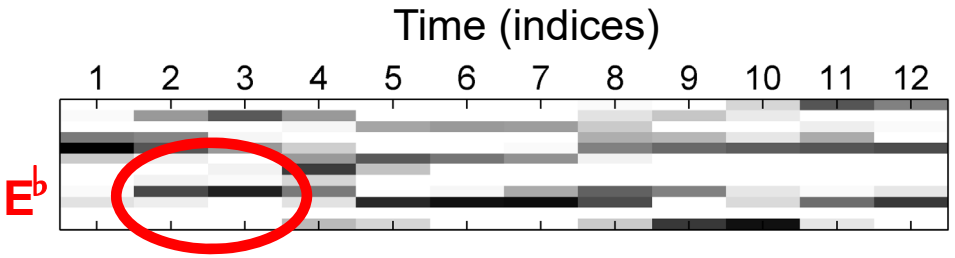




Music Synchronization: Audio-Audio

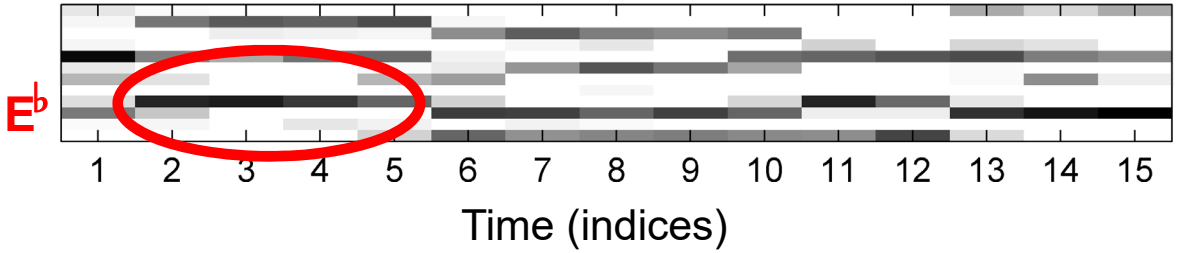
Beethoven's Fifth

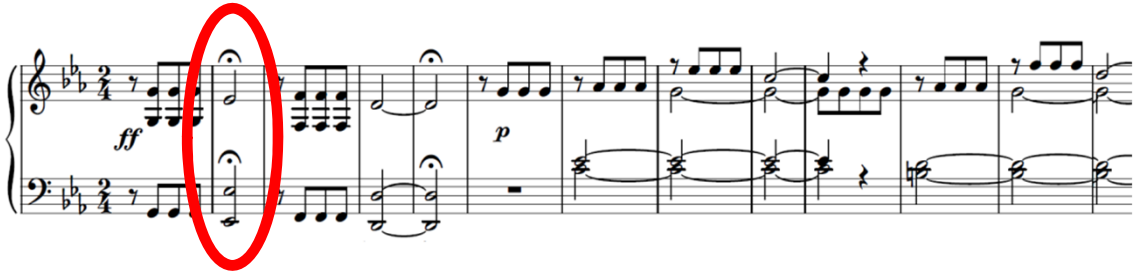
Karajan



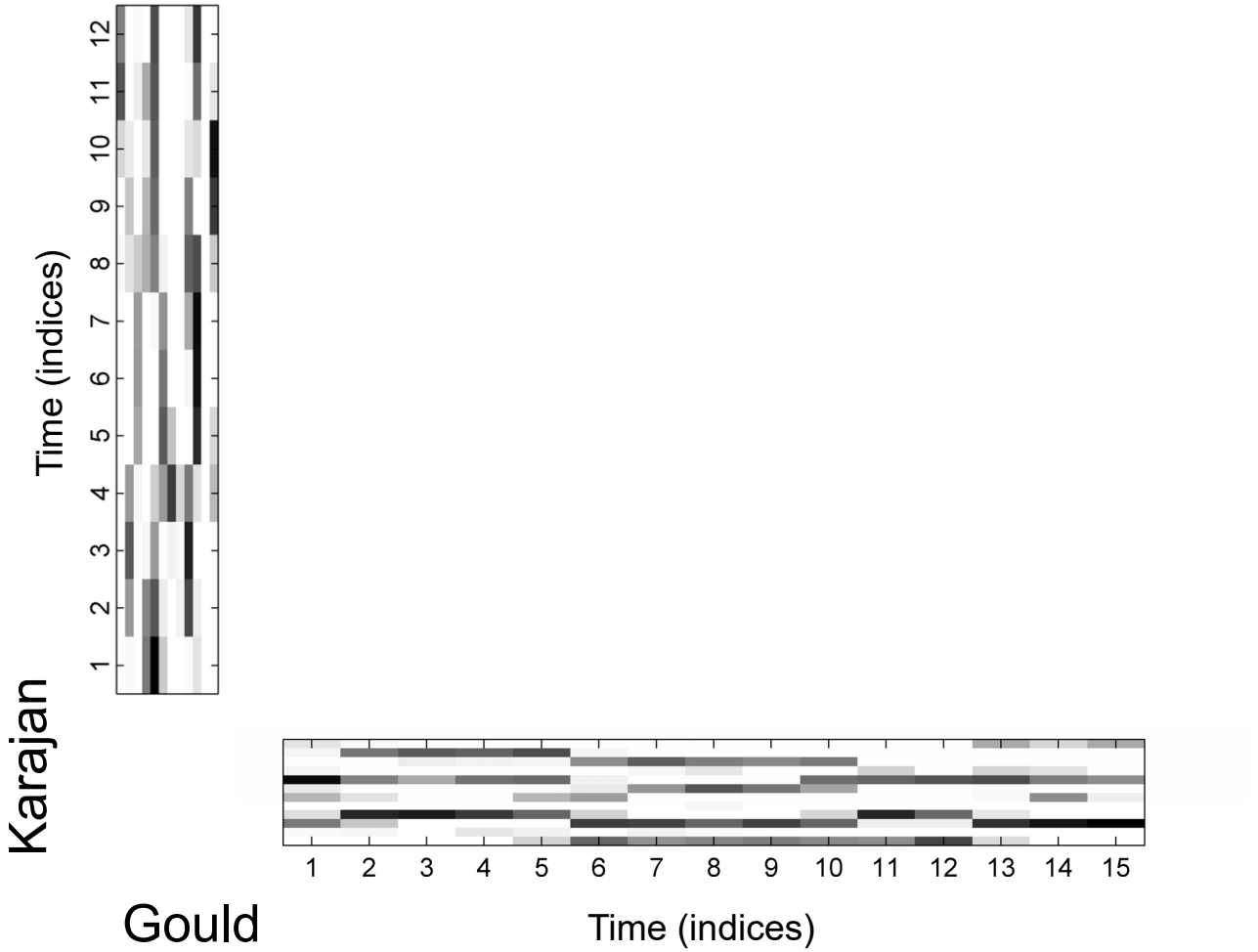
Time–chroma representations

Gould



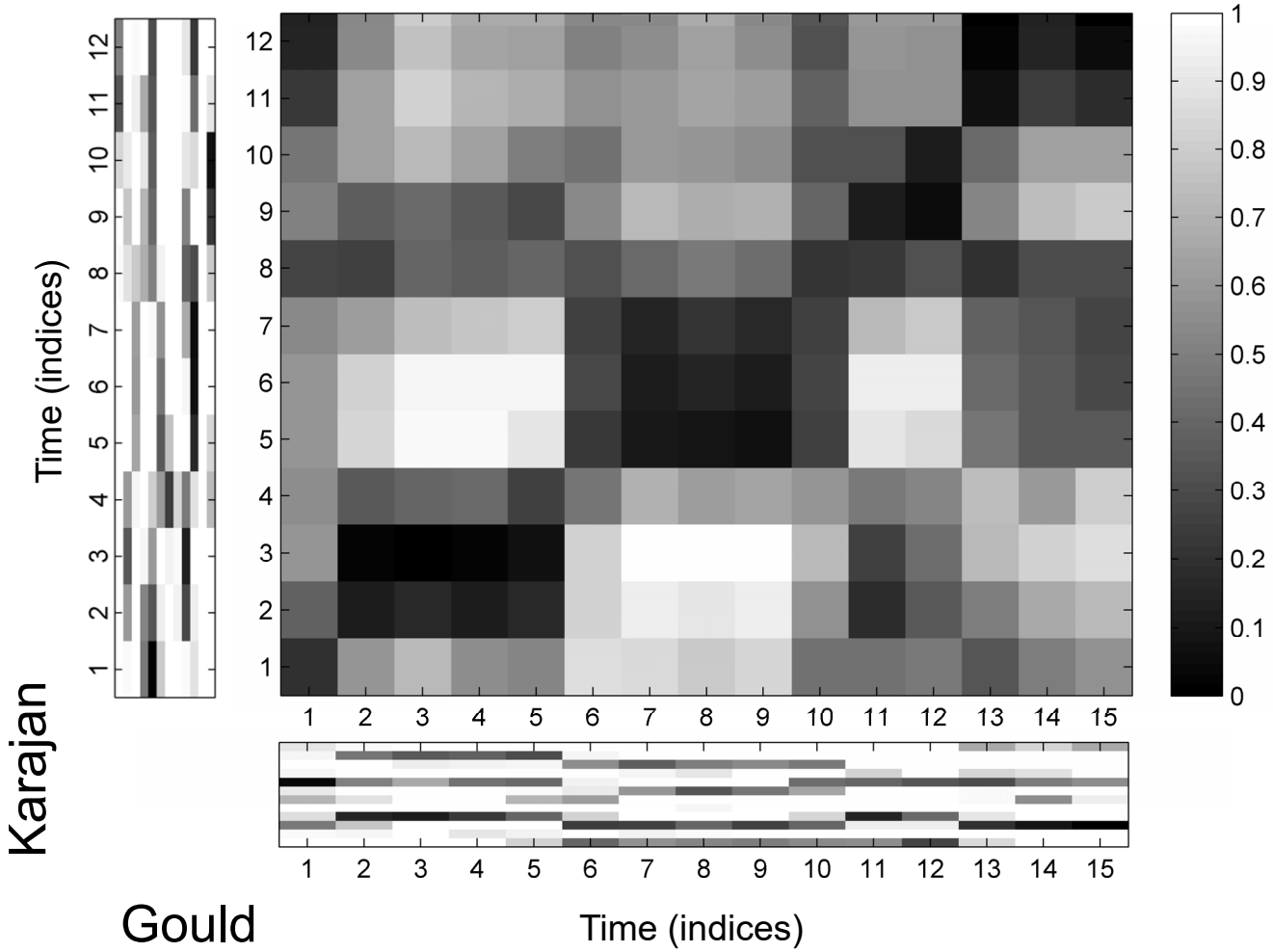


Music Synchronization: Audio-Audio



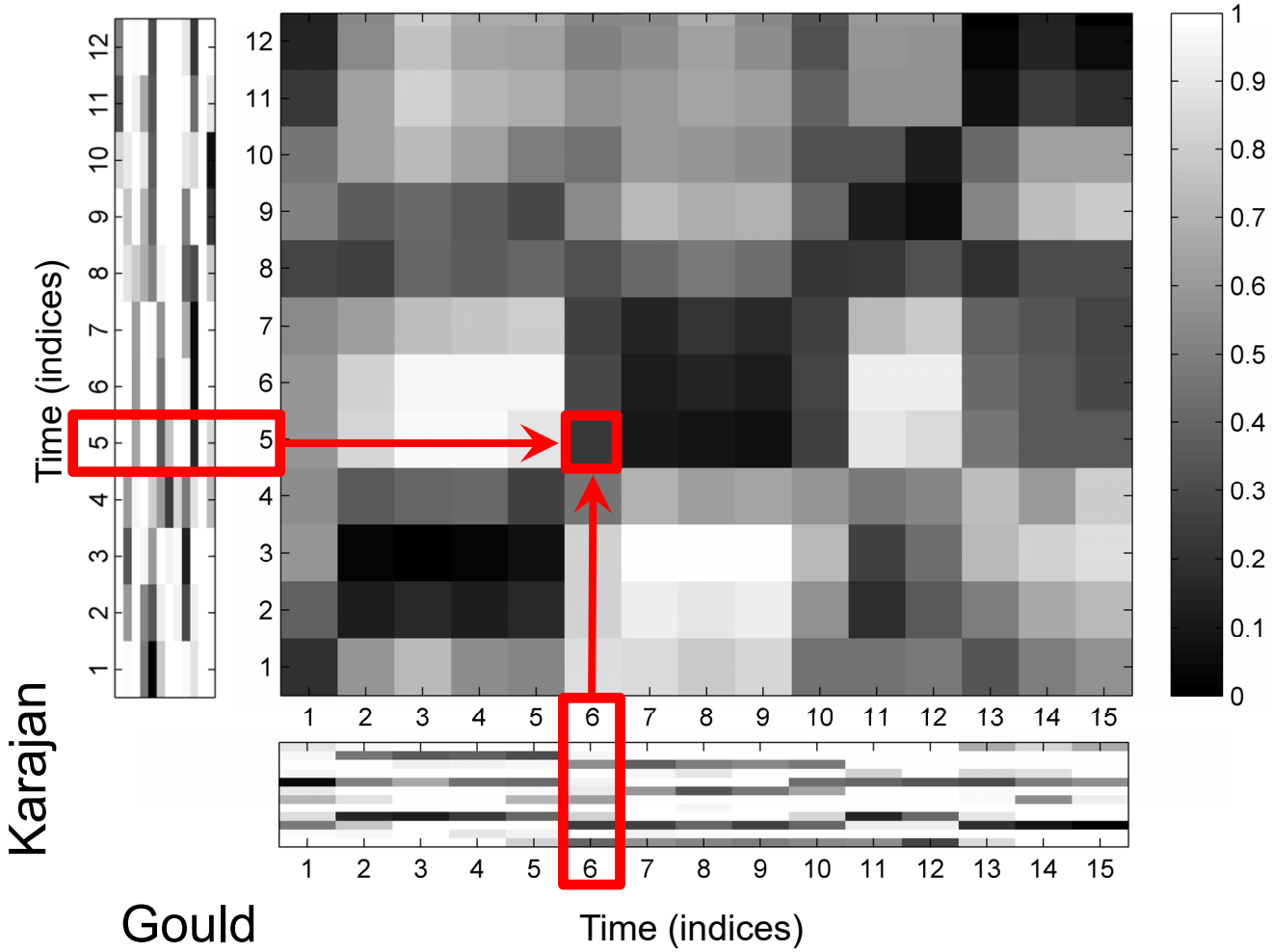
Music Synchronization: Audio-Audio

Cost matrix



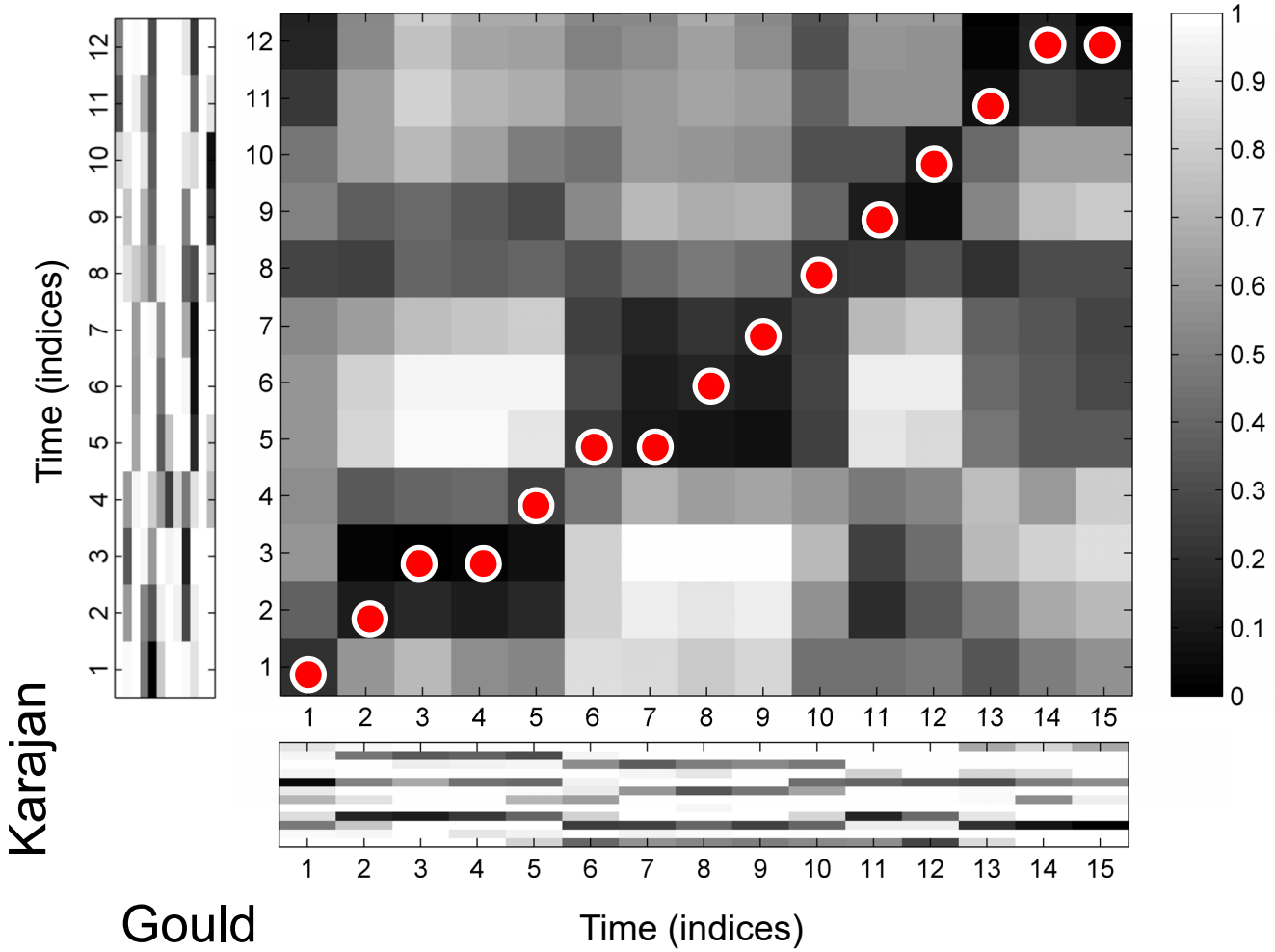
Music Synchronization: Audio-Audio

Cost matrix



Music Synchronization: Audio-Audio

Cost-minimizing warping path

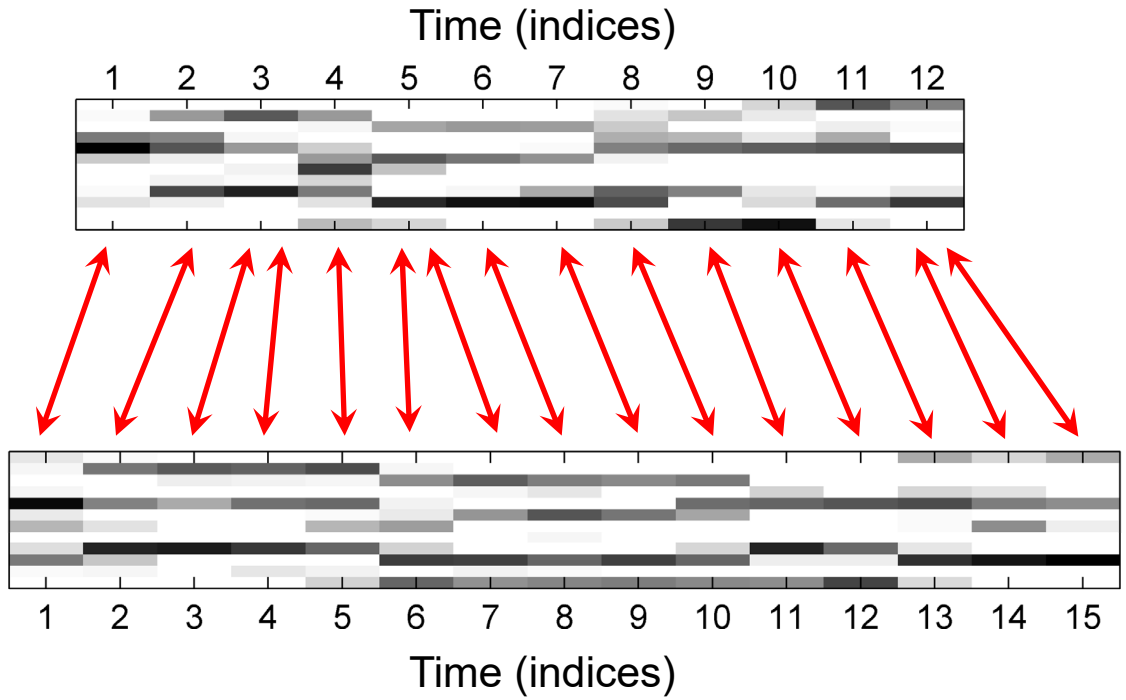


Music Synchronization: Audio-Audio

Optimal alignment (cost-minimizing warping path)

Karajan

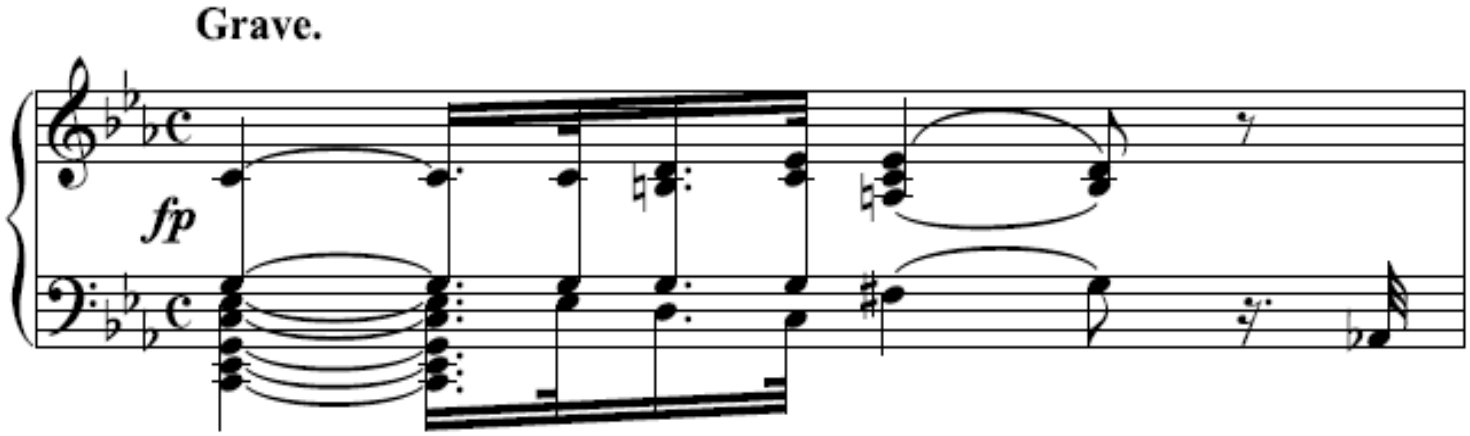
Gould



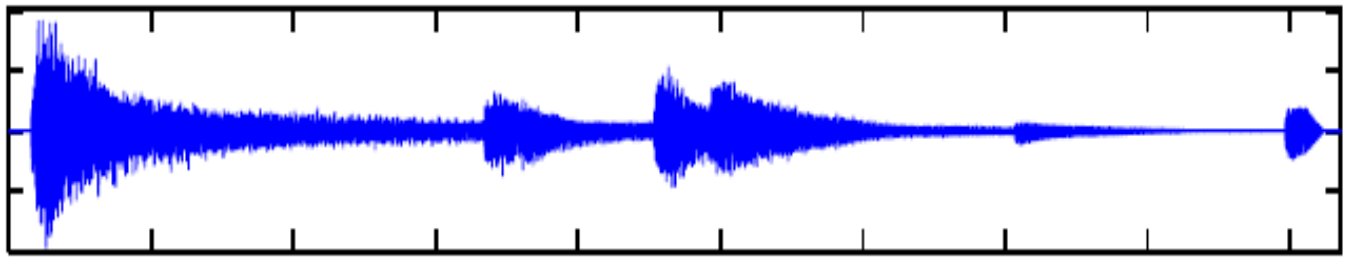
Application: Interpretation Switcher

Music Synchronization: Image-Audio

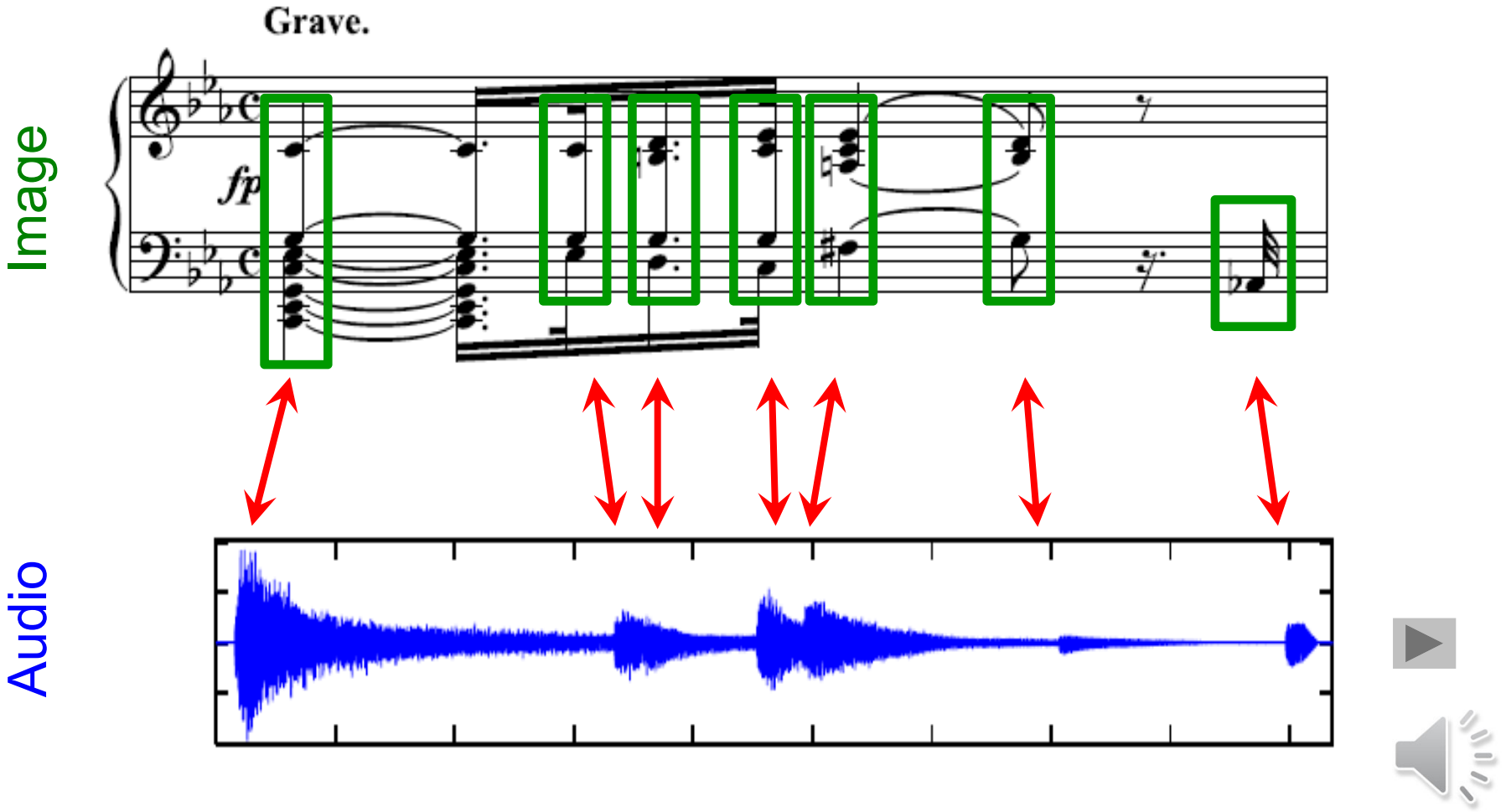
Image



Audio

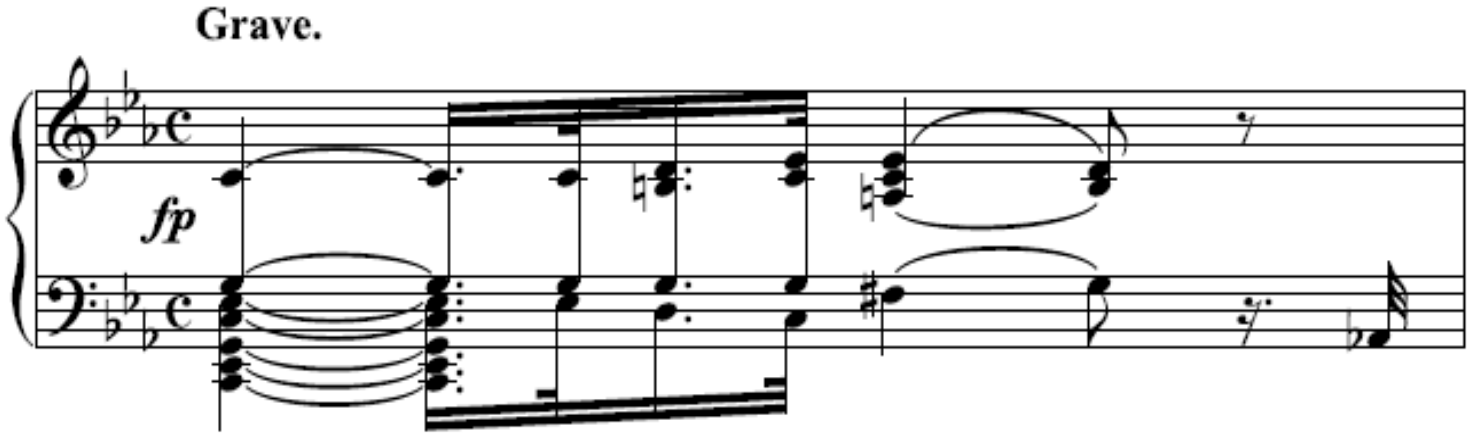


Music Synchronization: Image-Audio

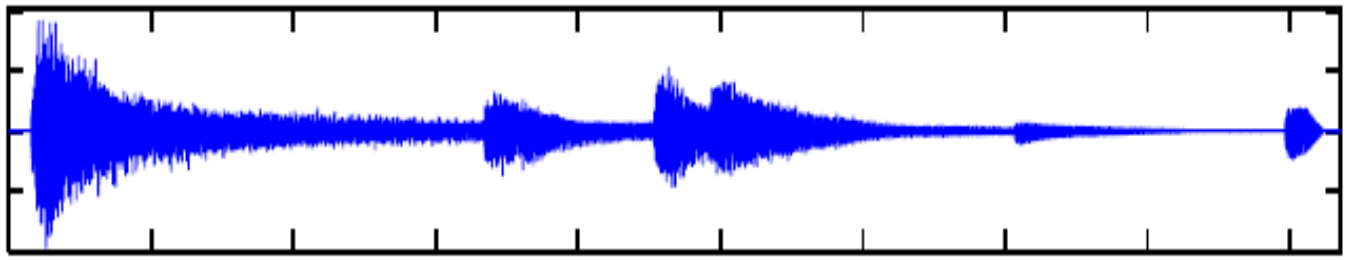


How to make the data comparable?

Image



Audio

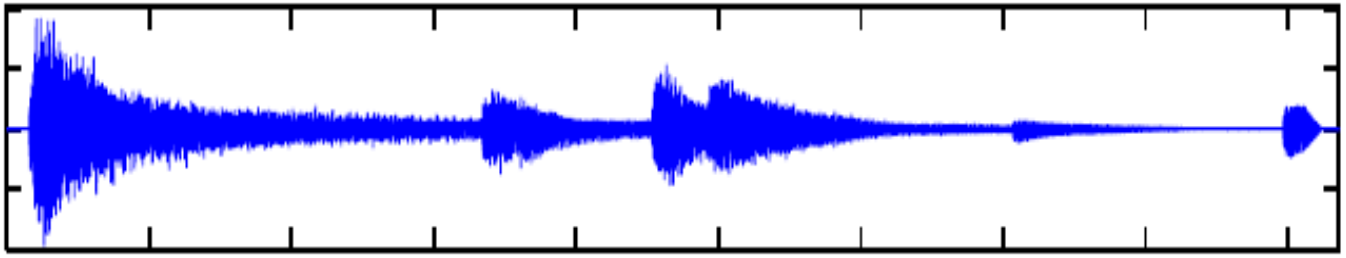


How to make the data comparable?

Image Processing: Optical Music Recognition

Image

Audio



How to make the data comparable?

Image Processing: Optical Music Recognition

Image

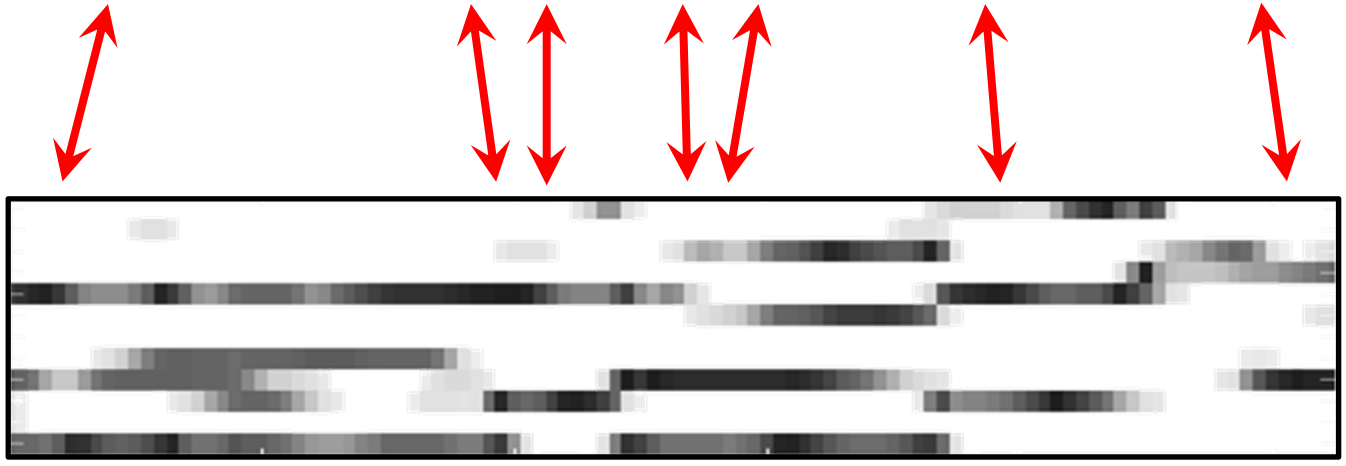
Audio

Audio Processing: Fourier Analysis

How to make the data comparable?

Image Processing: Optical Music Recognition

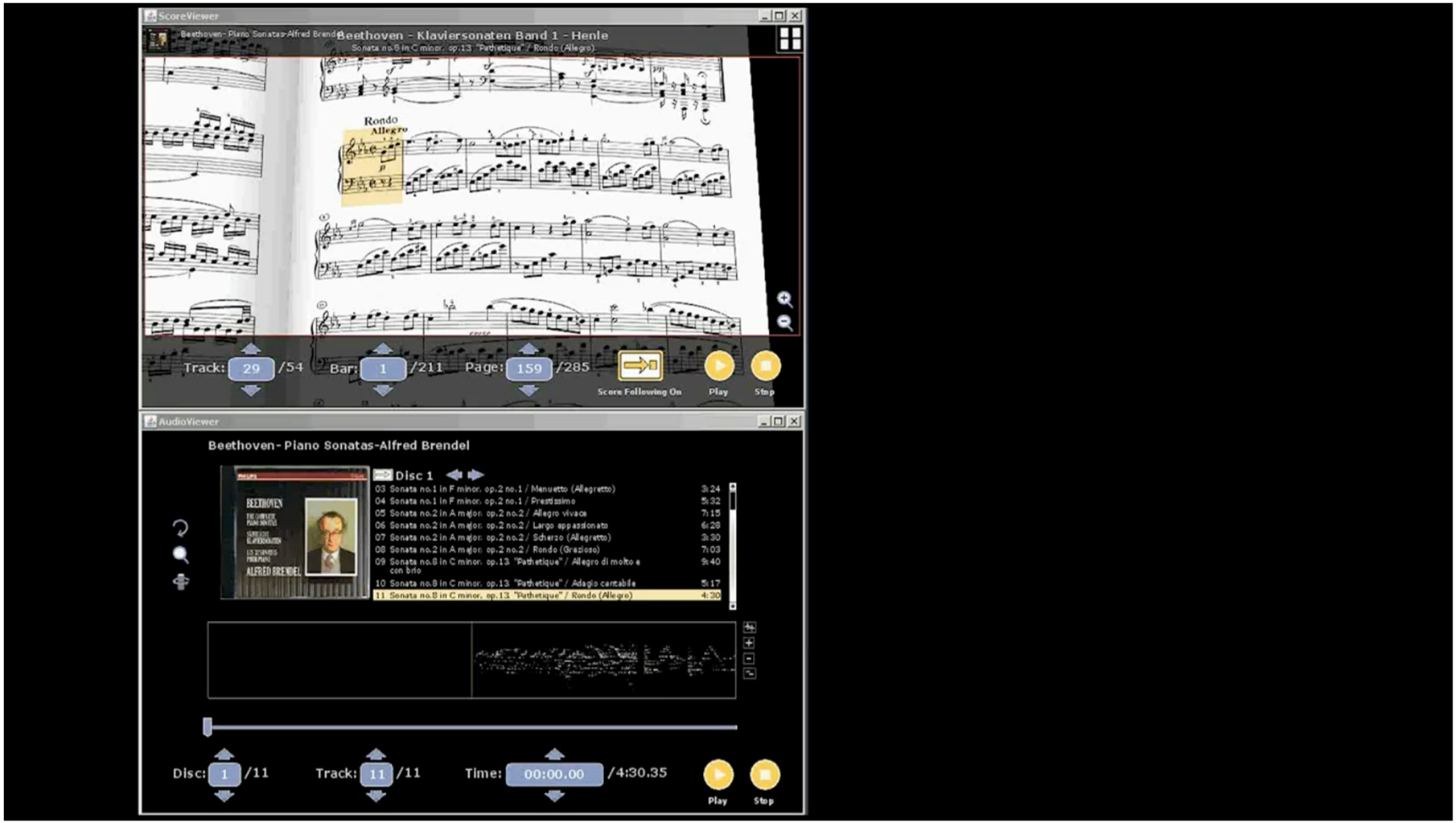
Image



Audio

Audio Processing: Fourier Analysis

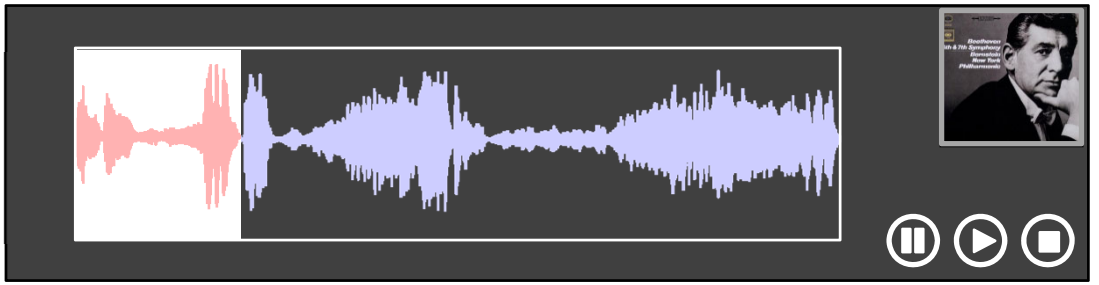
Application: Score Viewer



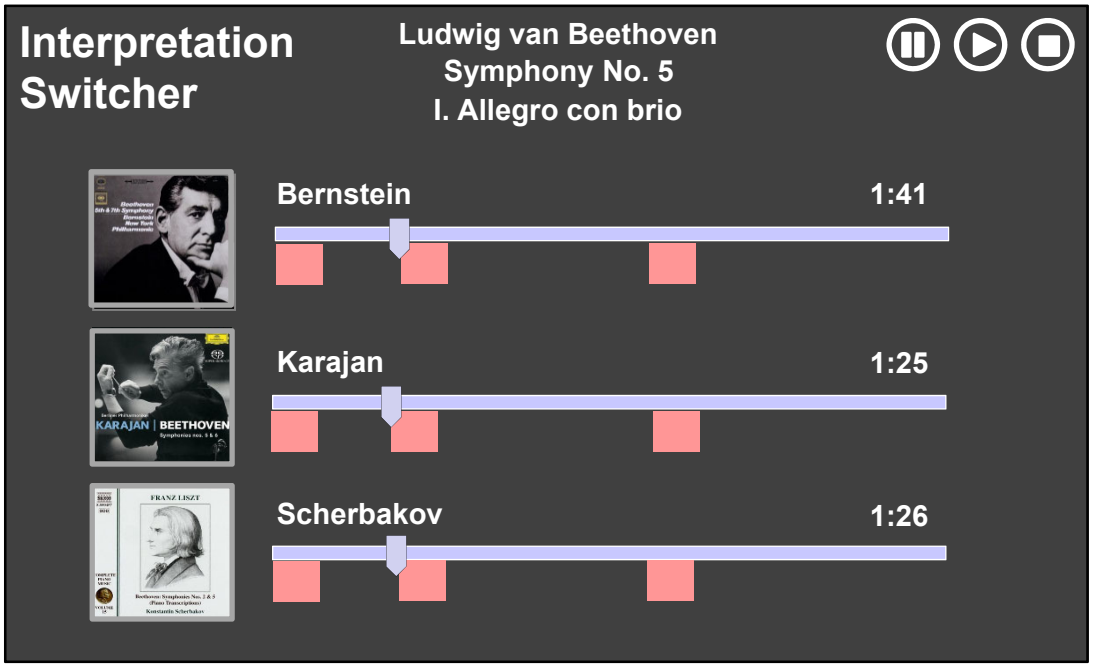
Audio Matching

Task

Query:



Database: Matches

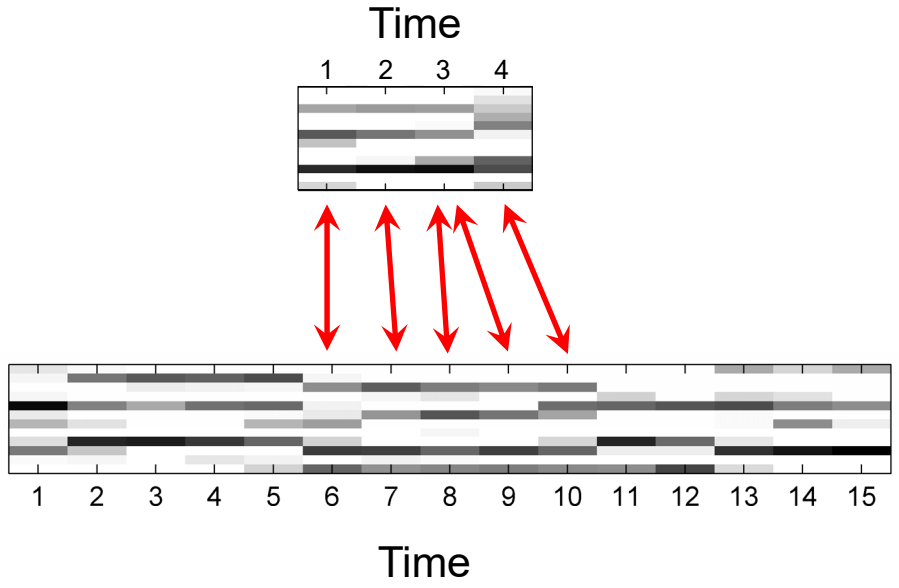


Audio Matching

Task

Query: Sequence *X*

Database: Sequence *Y*

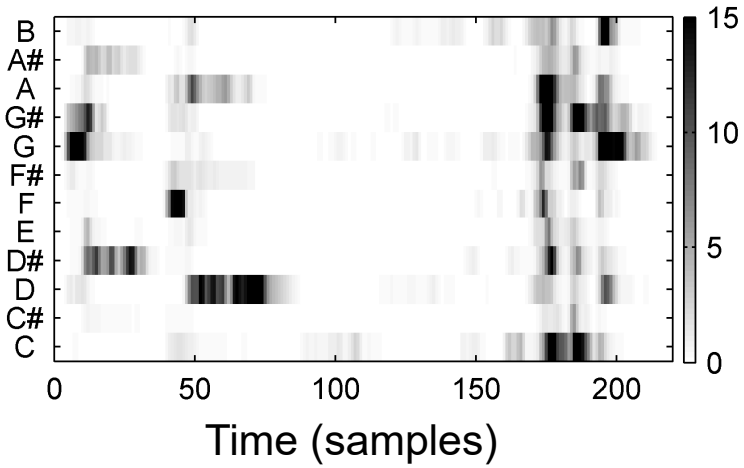


Subsequence matching

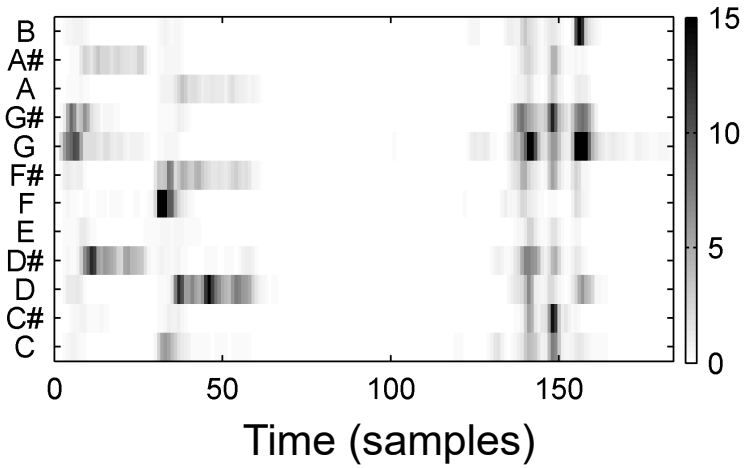
Audio Features

Example: Beethoven's Fifth

Bernstein



Karajan

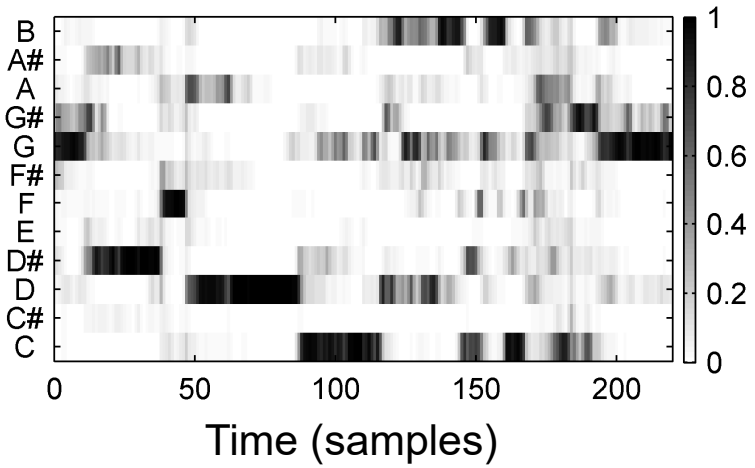


Chroma representation (10 Hz)

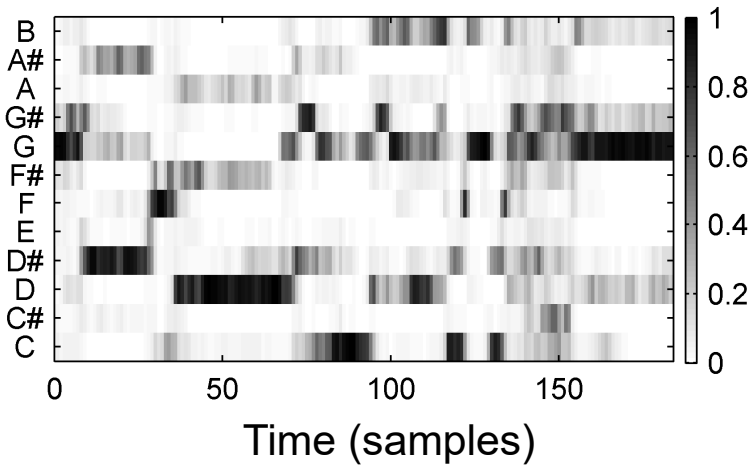
Audio Features

Example: Beethoven's Fifth

Bernstein



Karajan



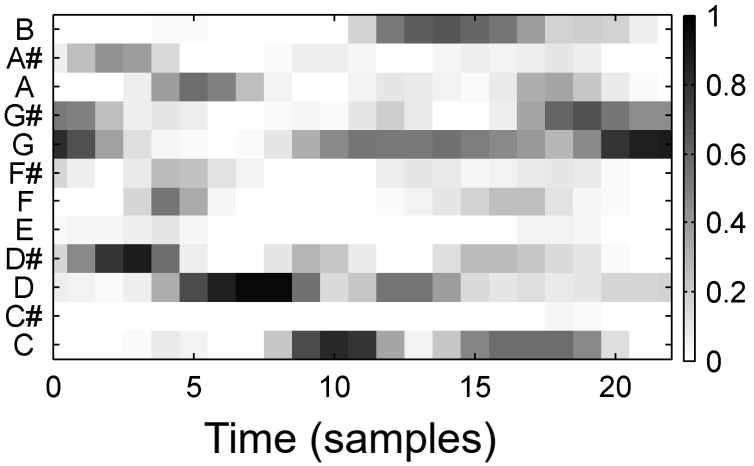
Chroma representation (10 Hz)

- Normalization

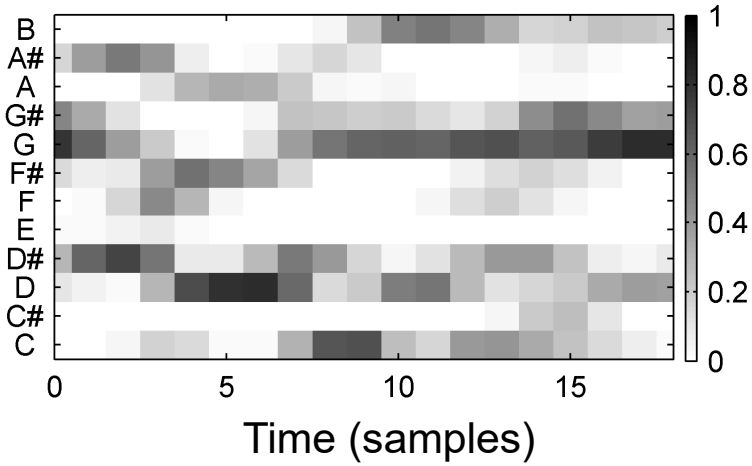
Audio Features

Example: Beethoven's Fifth

Bernstein



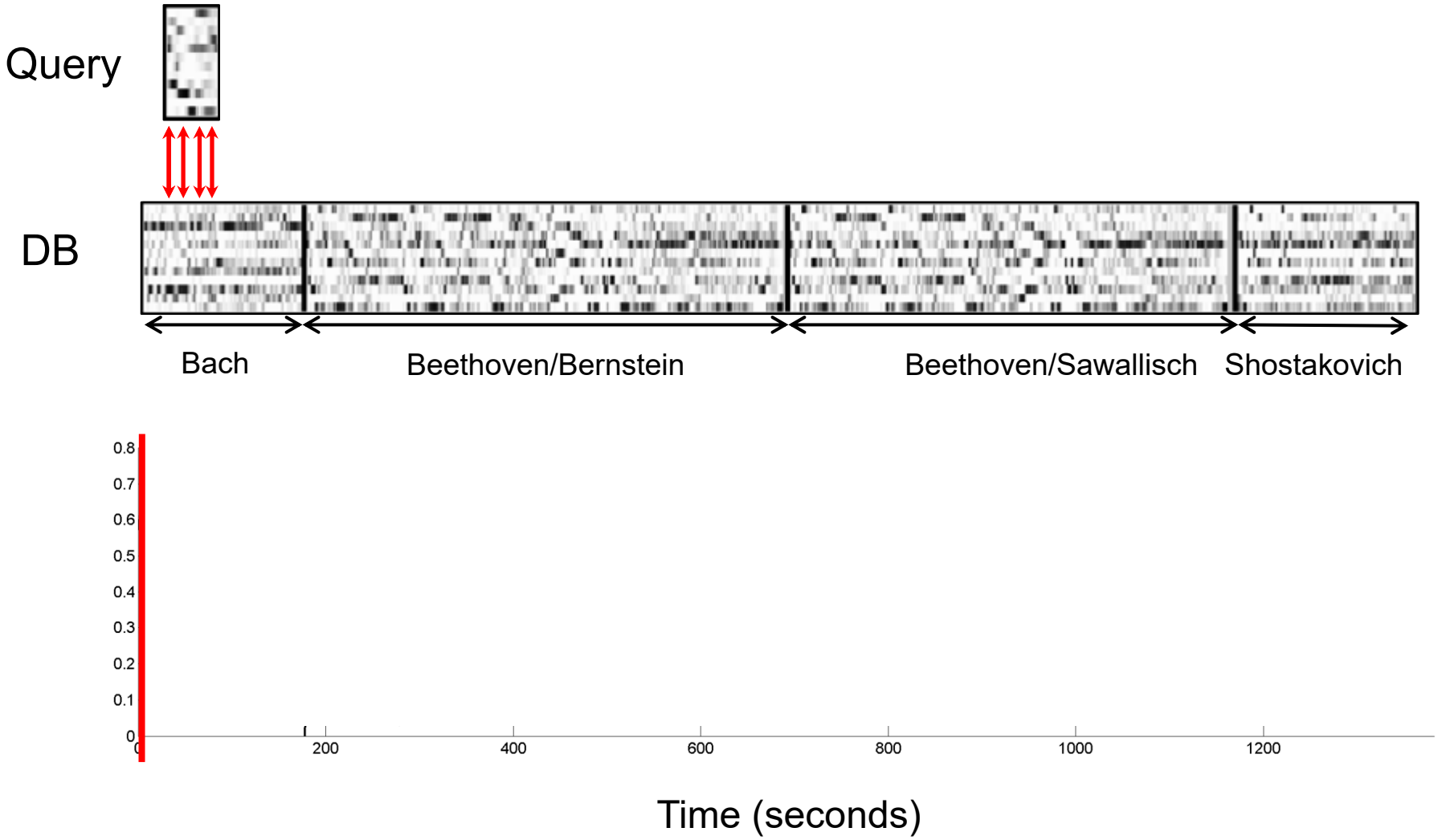
Karajan



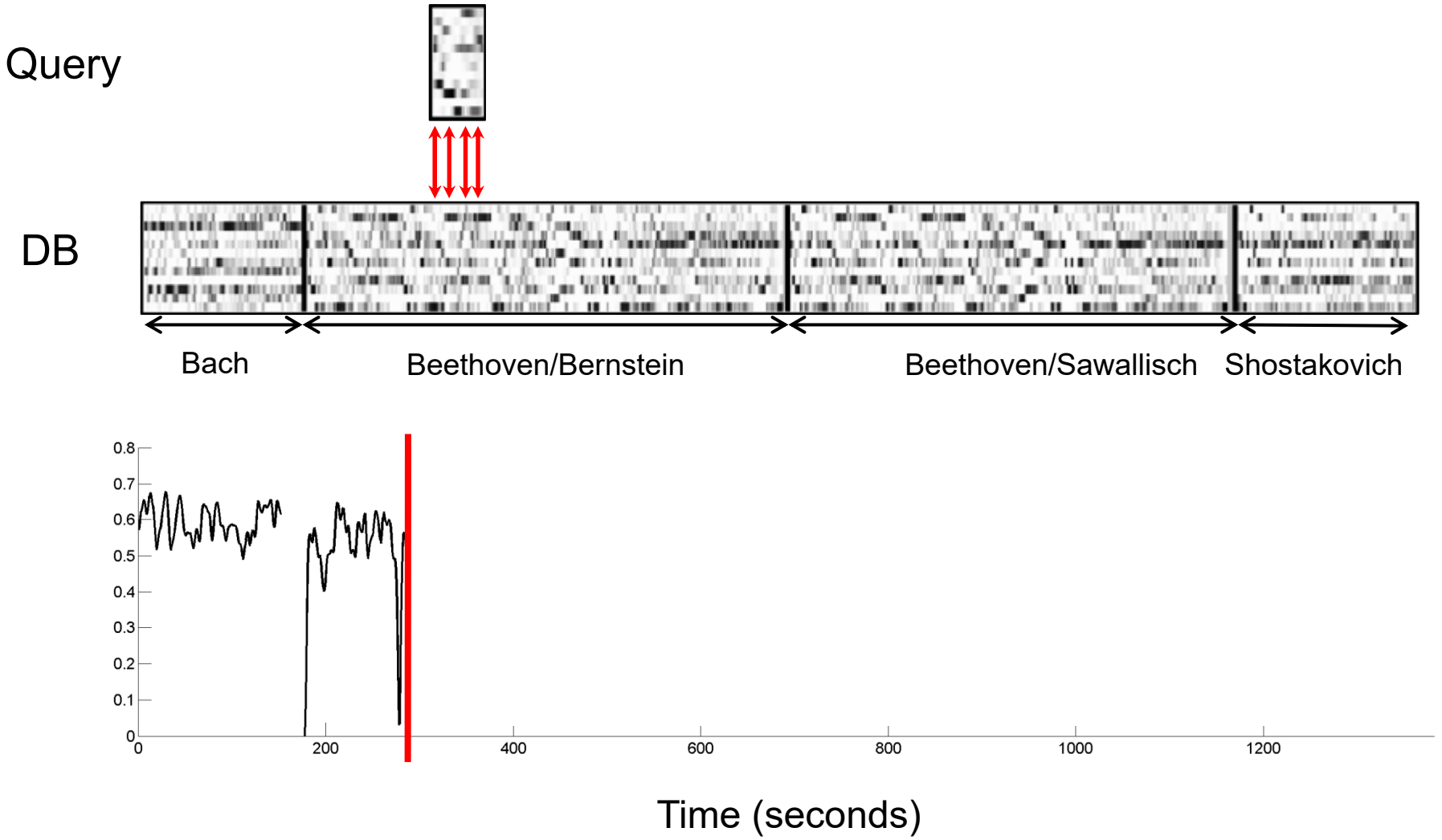
Chroma representation (1 Hz)

- Normalization
- Smoothing & downsampling

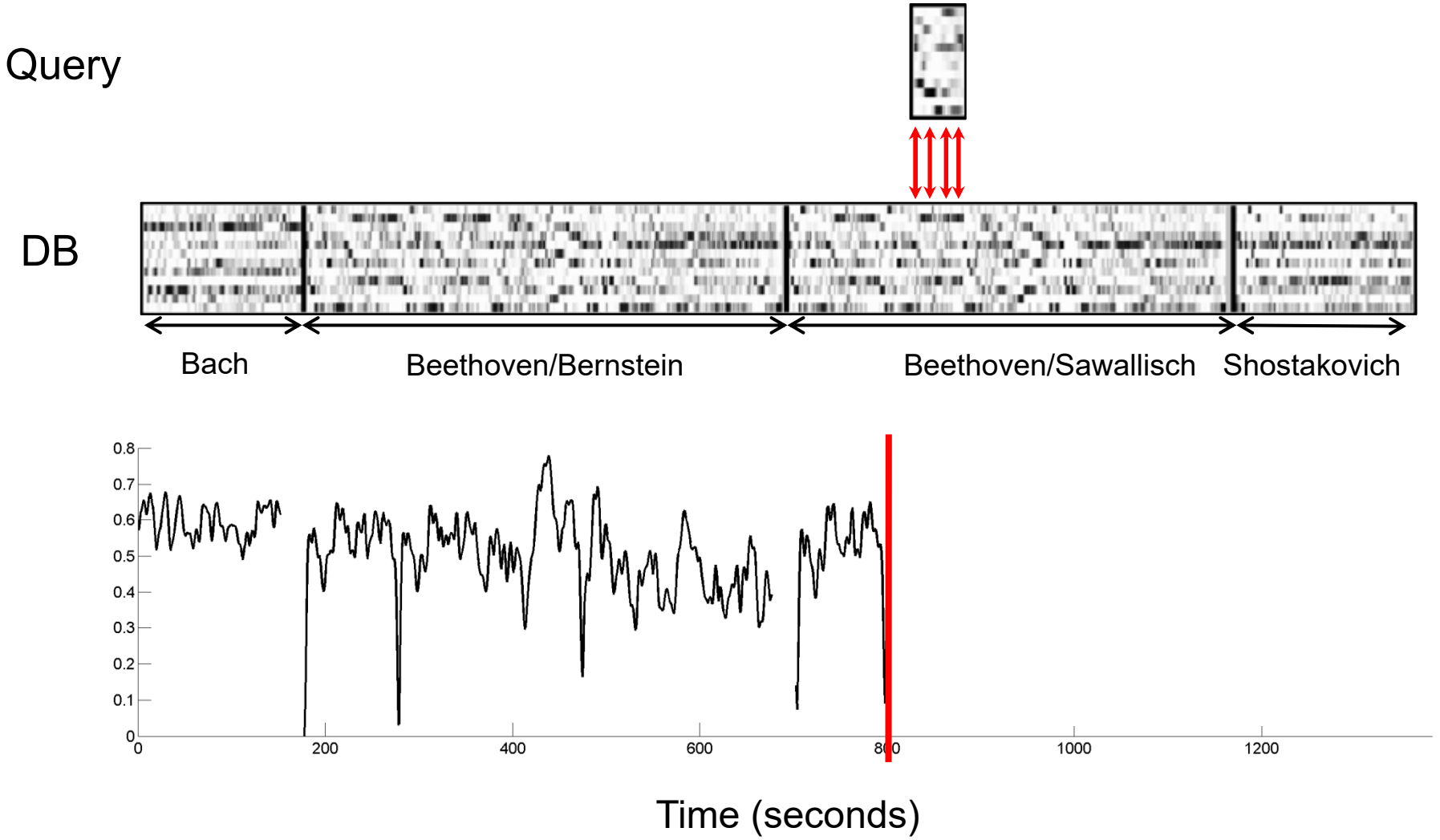
Matching Procedure



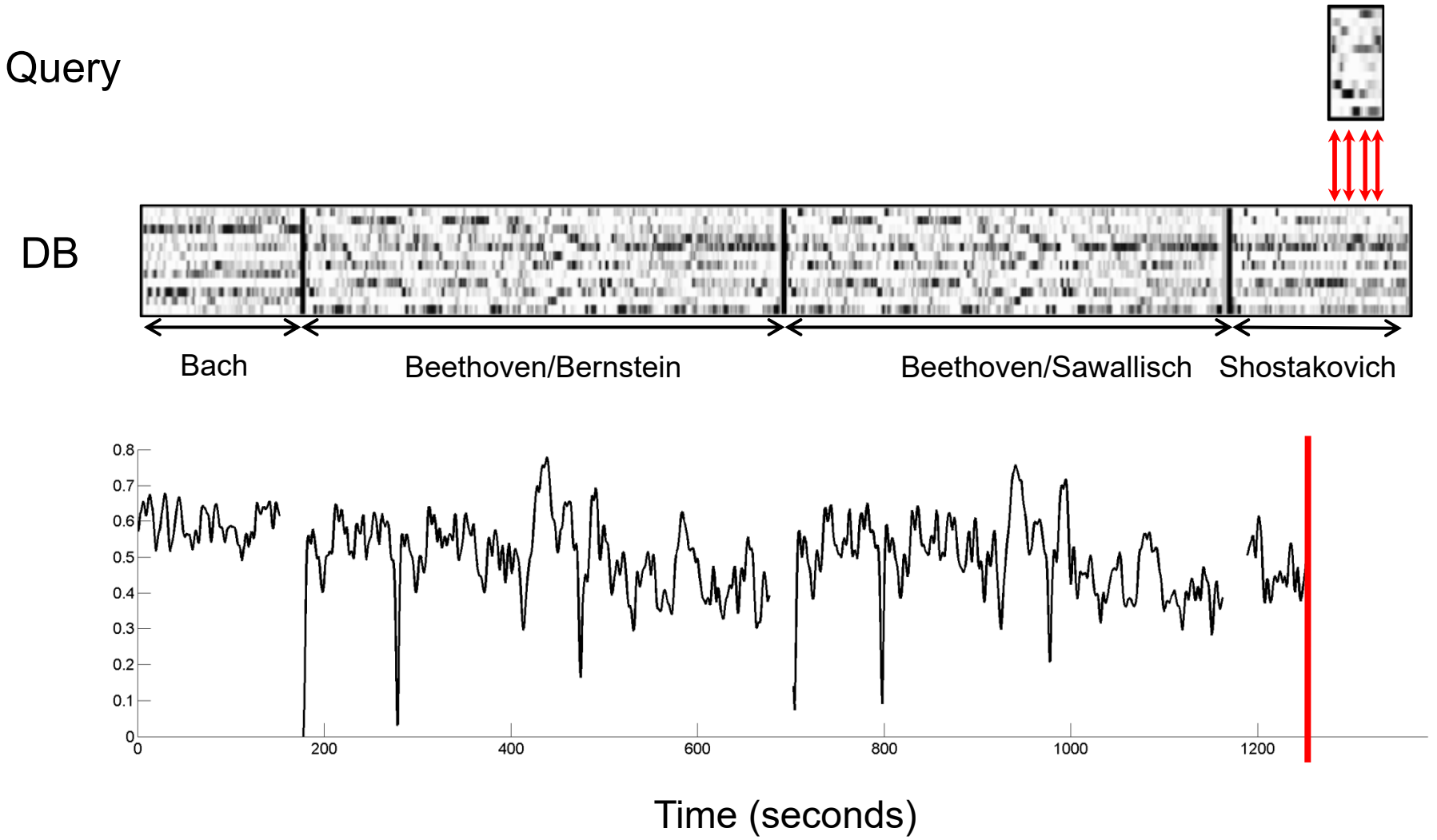
Matching Procedure



Matching Procedure



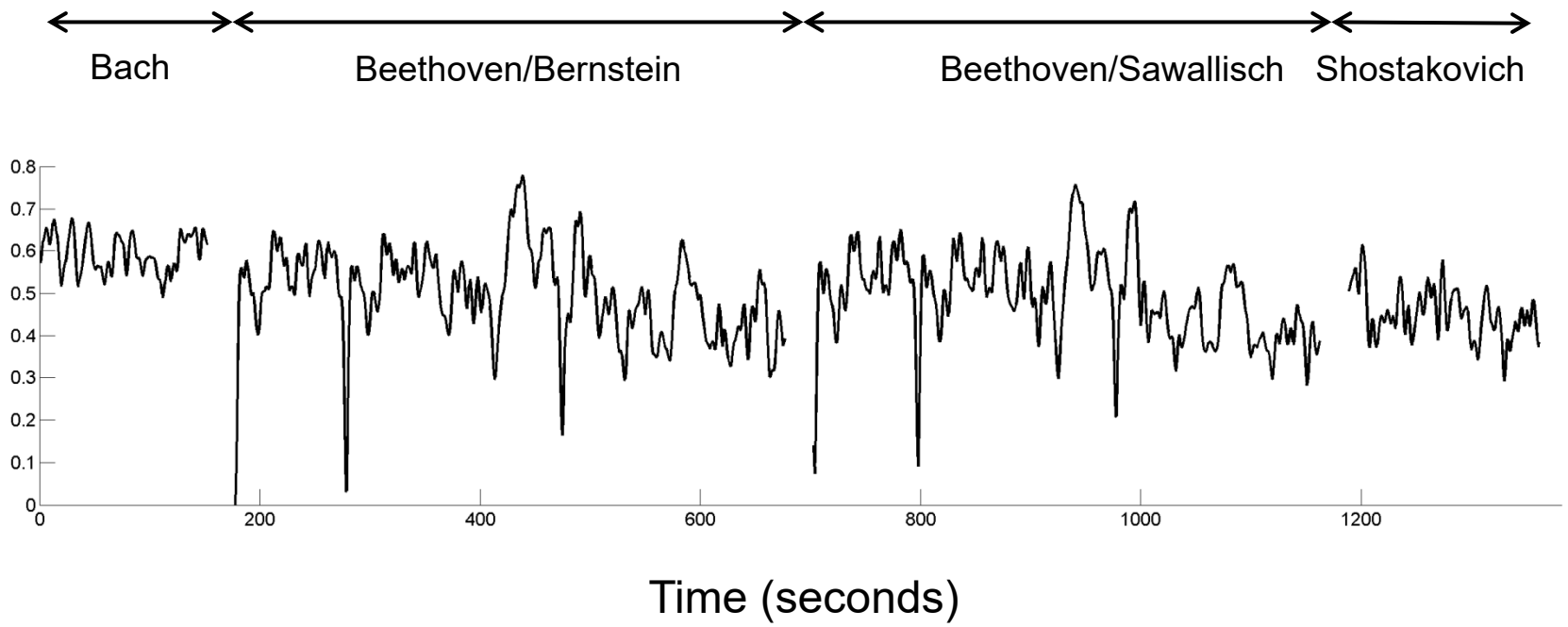
Matching Procedure



Matching Procedure

Matching curve

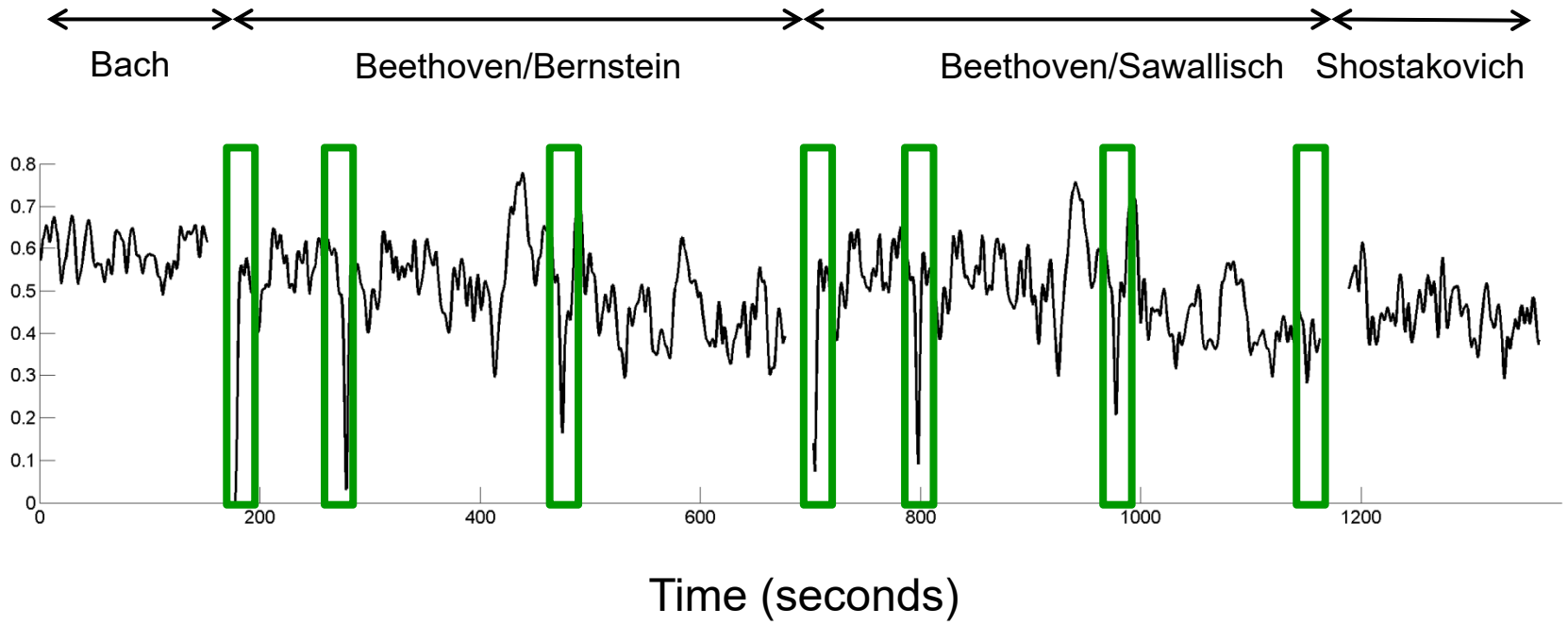
Query: Beethoven's Fifth / Bernstein (first 20 seconds)



Matching Procedure

Matching curve

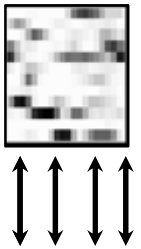
Query: Beethoven's Fifth / Bernstein (first 20 seconds)

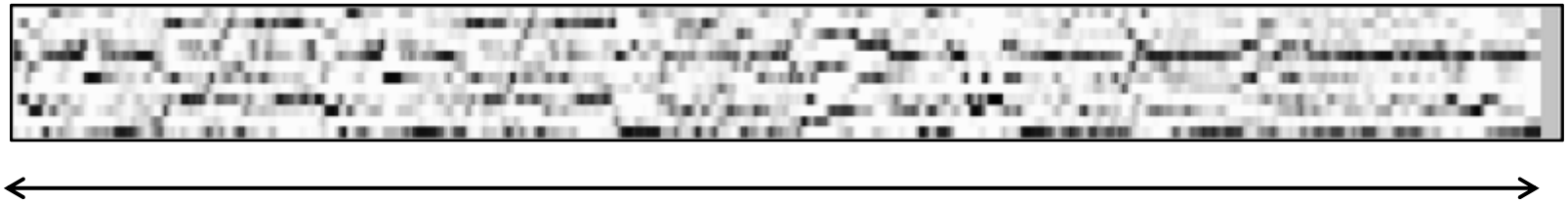


Matching Procedure

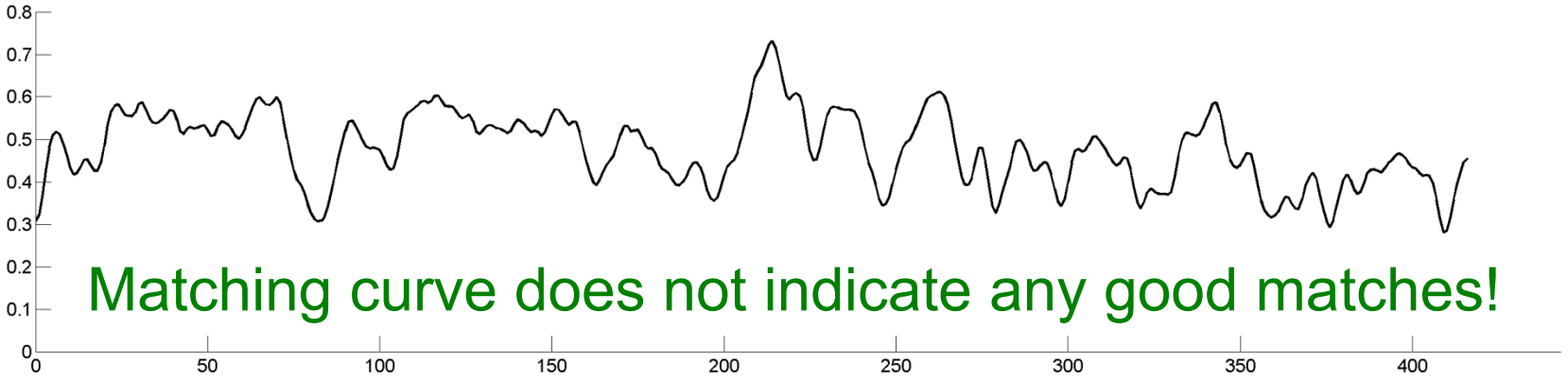
Problem: How to deal with tempo differences?

Karajan is much
faster than Bernstein!





Beethoven/Karajan



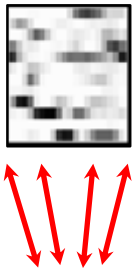
Matching curve does not indicate any good matches!

Time (seconds)

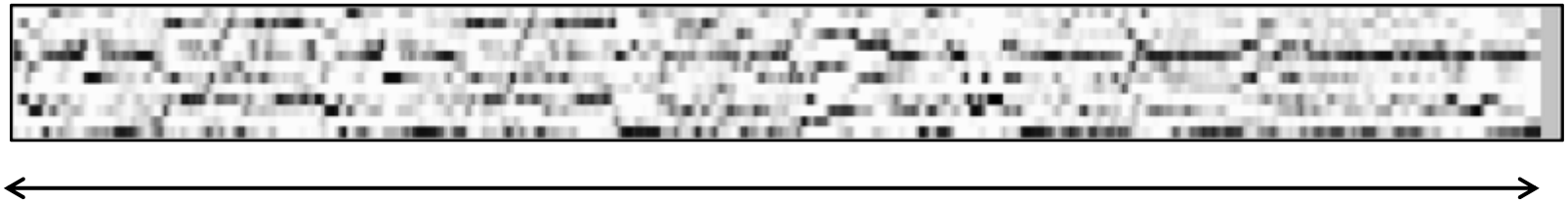
Matching Procedure

1. Strategy: Usage of local warping

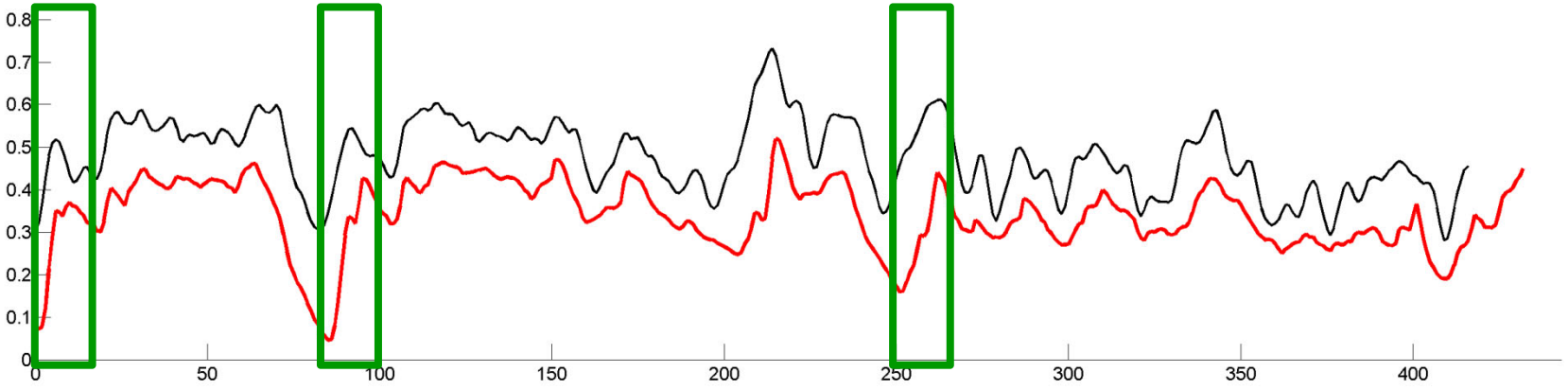
Karajan is much
faster than Bernstein!



Warping strategies are computationally expensive and hard for indexing.



Beethoven/Karajan

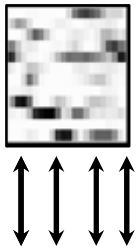


Time (seconds)

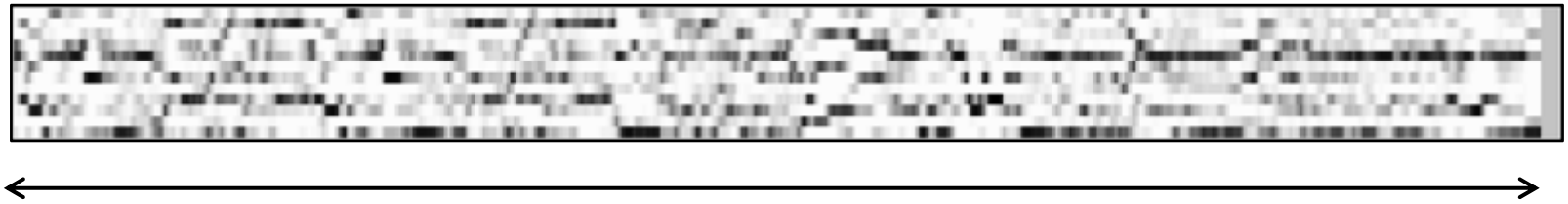
Matching Procedure

2. Strategy: Usage of multiple scaling

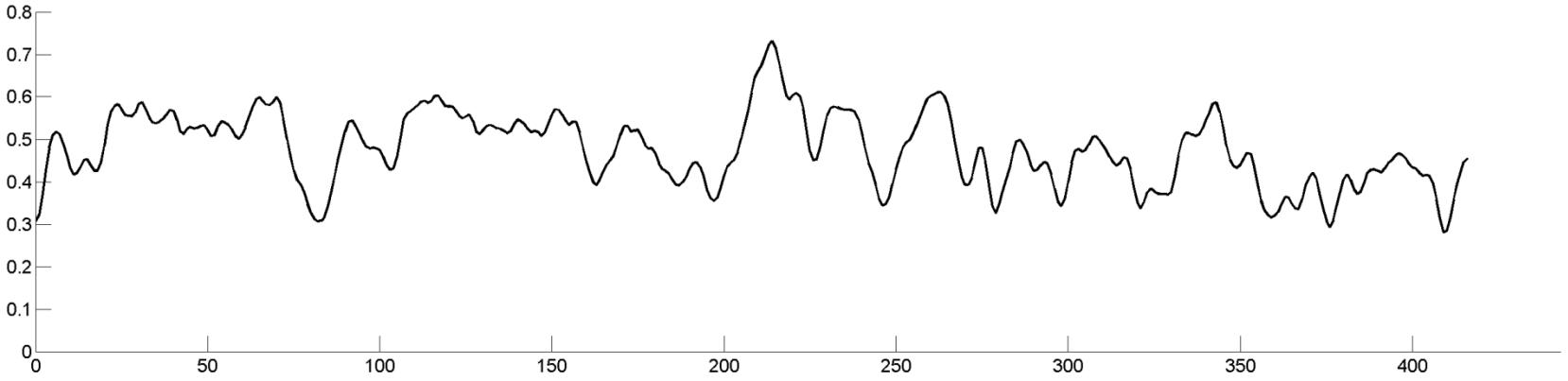
Query



DB



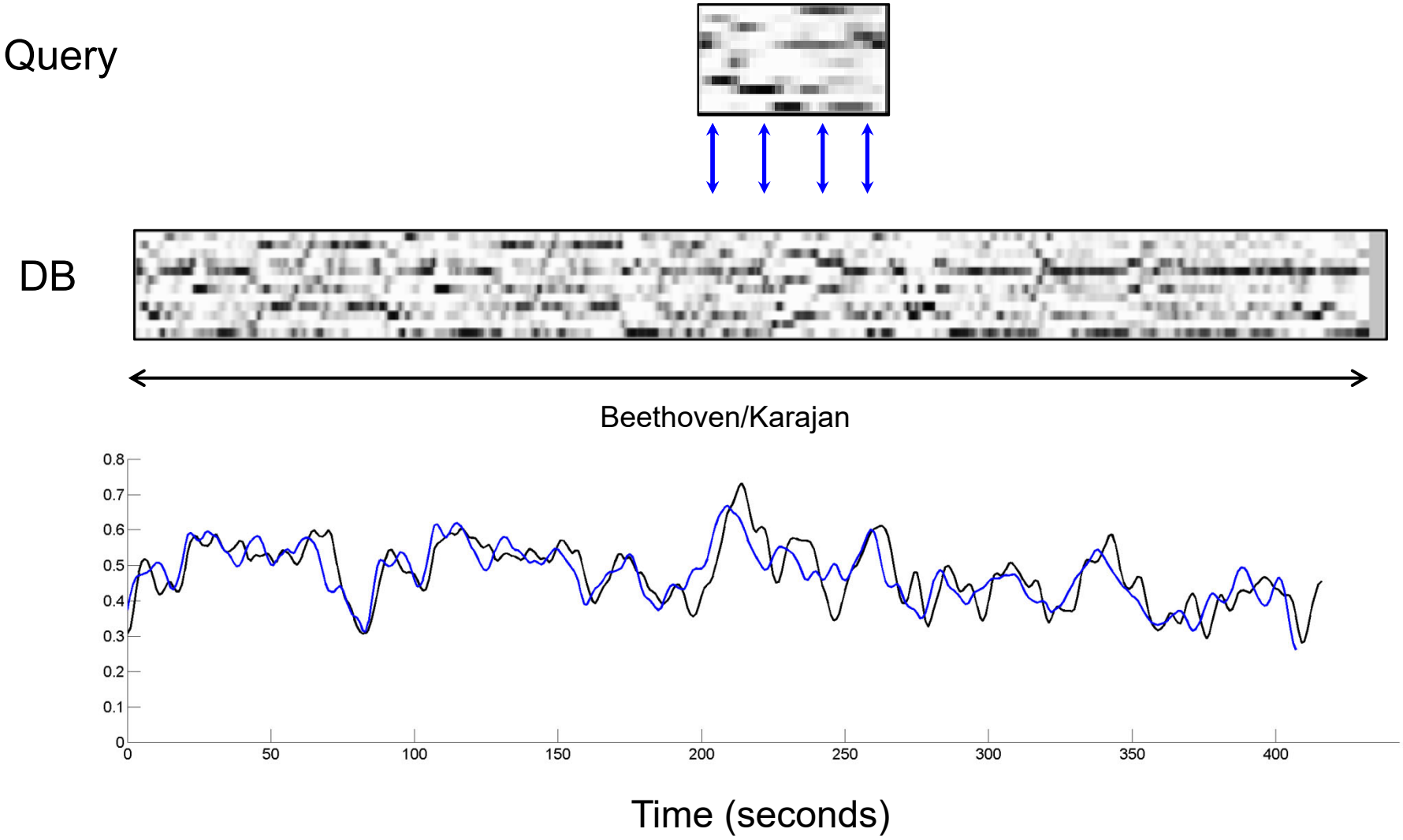
Beethoven/Karajan



Time (seconds)

Matching Procedure

2. Strategy: Usage of multiple scaling

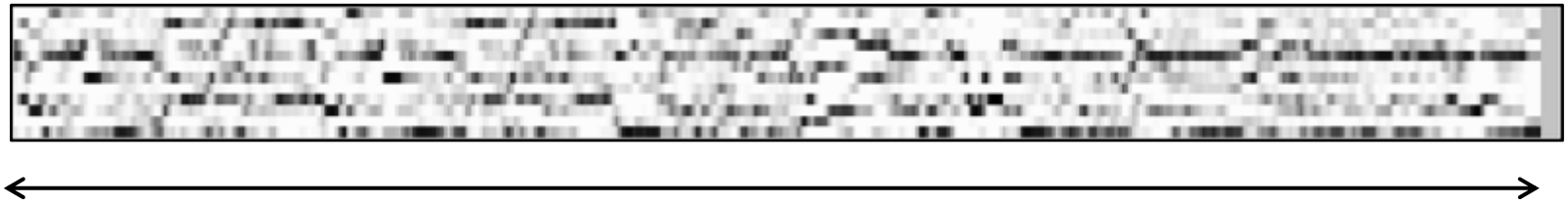


Matching Procedure

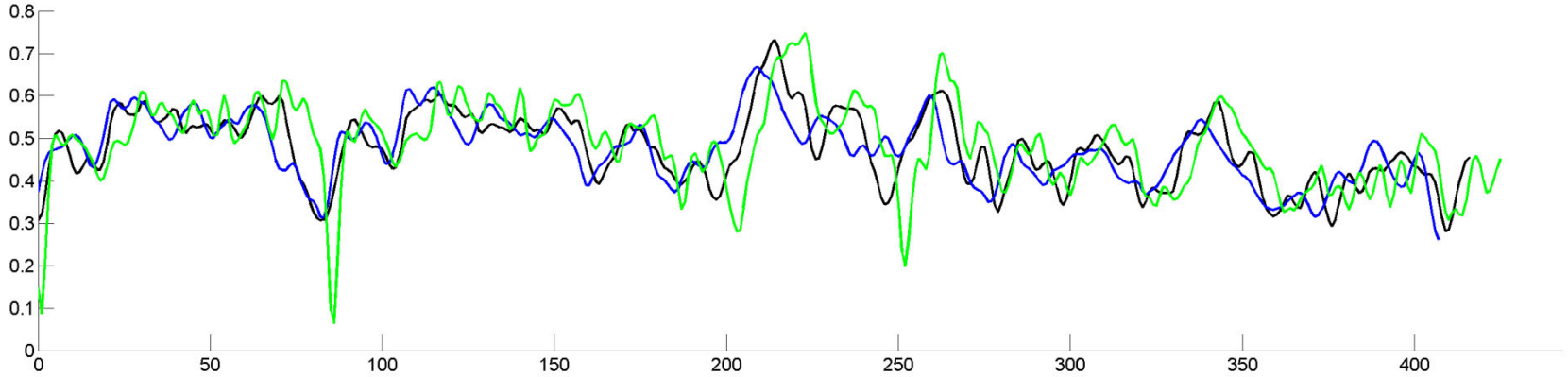
2. Strategy: Usage of multiple scaling

Query

DB



Beethoven/Karajan

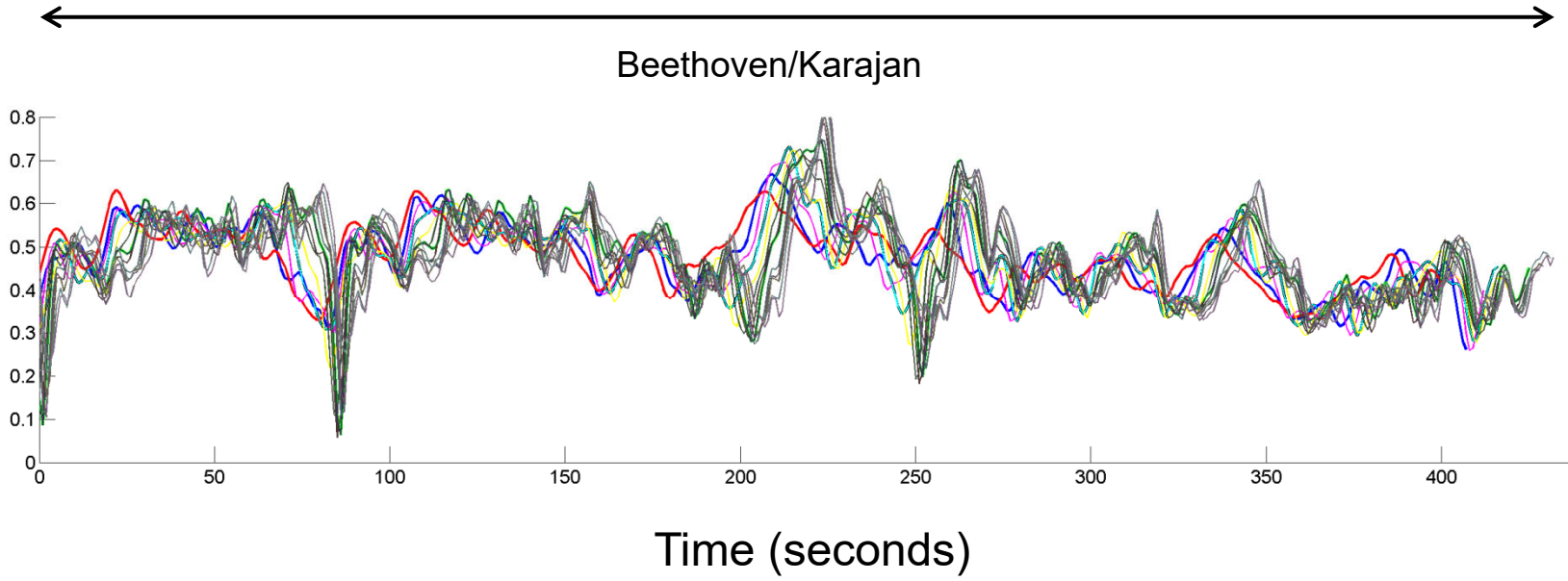


Time (seconds)

Matching Procedure

2. Strategy: Usage of multiple scaling

Query resampling simulates tempo changes

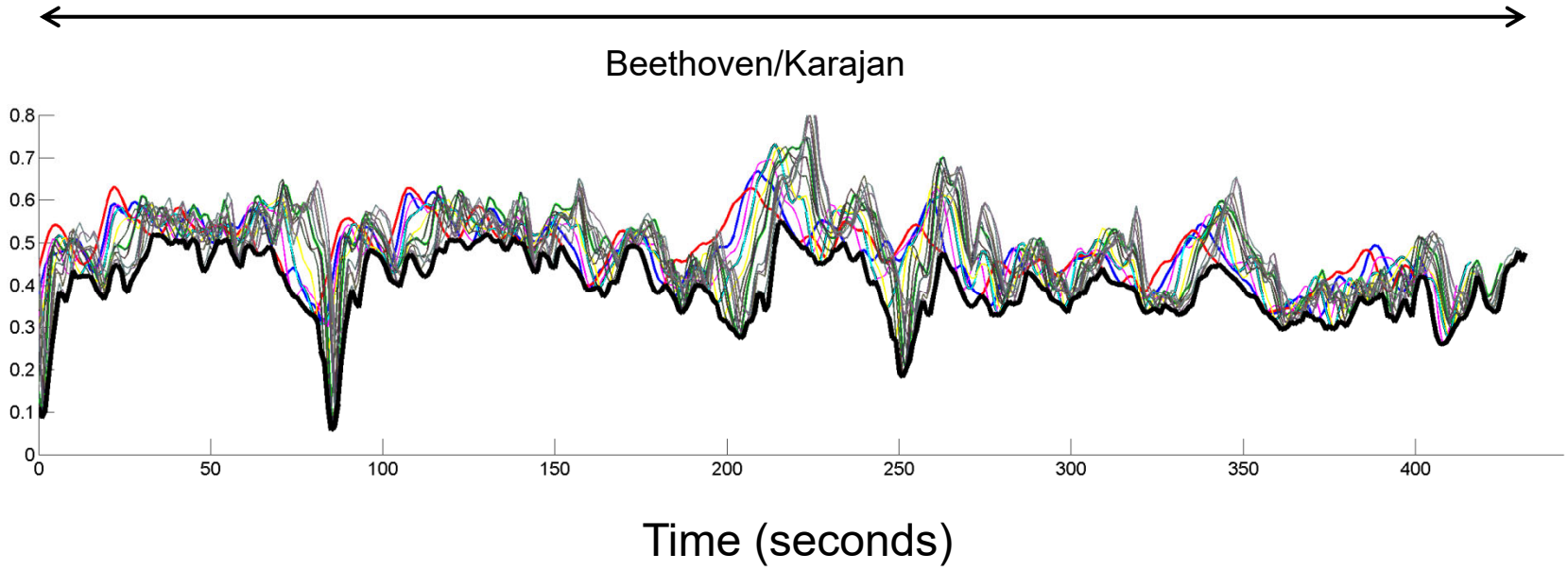


Matching Procedure

2. Strategy: Usage of multiple scaling

Query resampling simulates tempo changes

Minimize over all curves



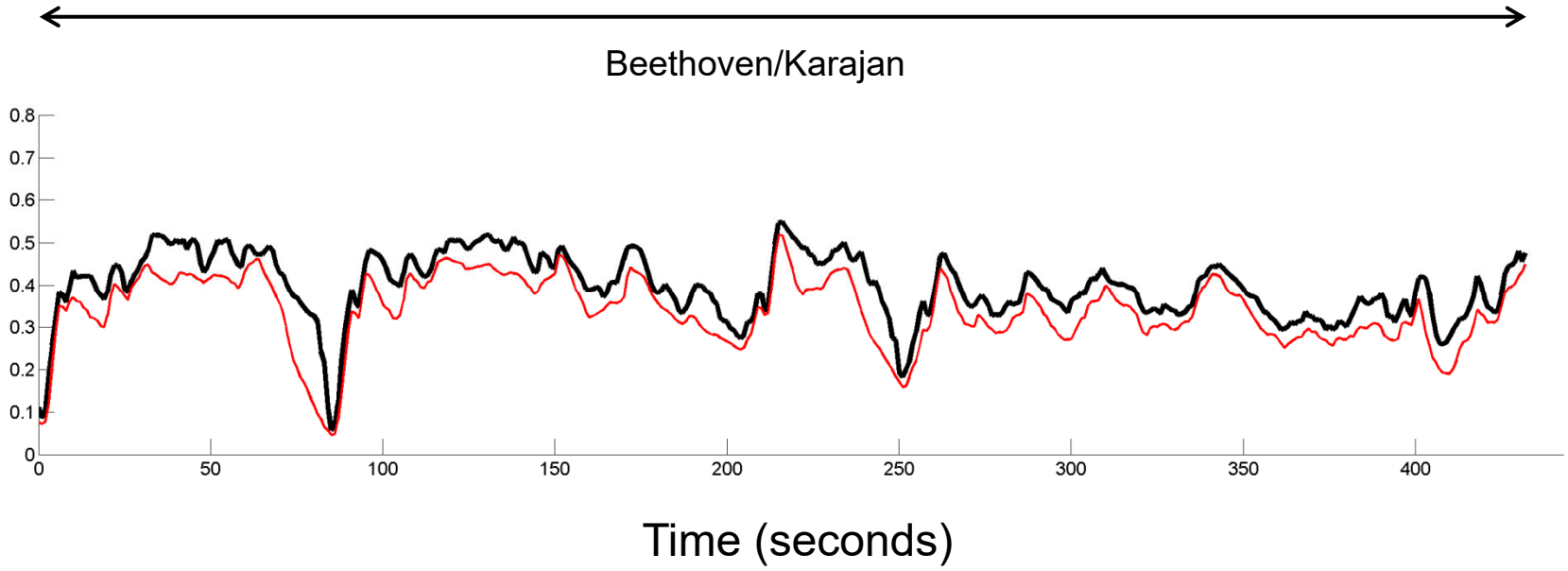
Matching Procedure

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