

Music Information Retrieval

When Music Meets Computer Science

Meinard Müller

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

Berlin MIR Meetup
20.03.2017

Meinard Müller



- 2001 PhD, Bonn University
- 2002/2003 Postdoc, Keio University, Japan
- 2007 Habilitation, Bonn University
“Information Retrieval for Music and Motion”
- 2007-2012 Senior Researcher
Max-Planck Institut für Informatik, Saarland
- 2012: Professor
Semantic Audio Processing
Universität Erlangen-Nürnberg

Group Members

- Stefan Balke
- Christian Dittmar
- Patricio López-Serrano
- Christof Weiß
- Frank Zalkow
- Thomas Prätzlich

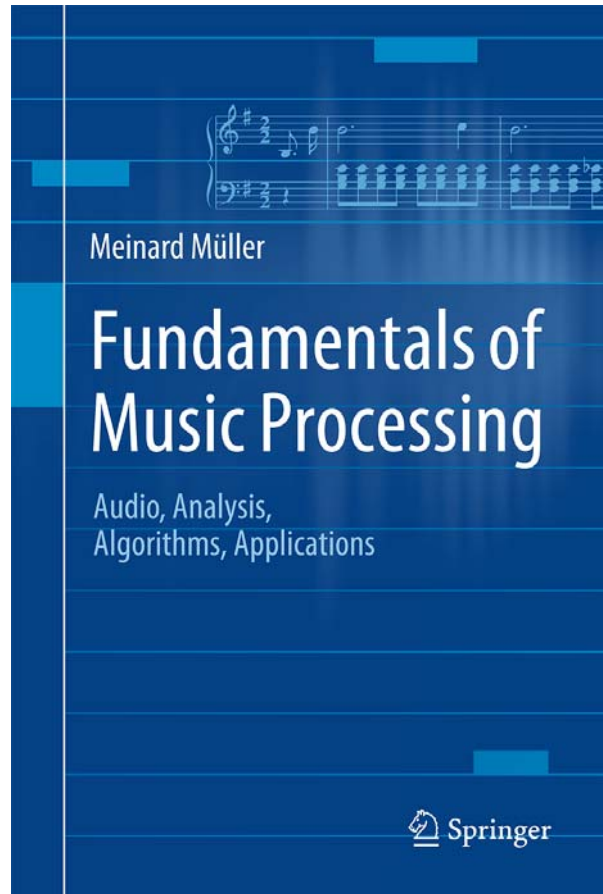


Group Members

- Stefan Balke
- Christian Dittmar
- Patricio López-Serrano
- Christof Weiß
- Frank Zalkow
- Thomas Prätzlich



Book: Fundamentals of Music Processing



Meinard Müller

Fundamentals of Music Processing

Audio, Analysis, Algorithms, Applications

483 p., 249 illus., 30 illus. in color, hardcover

ISBN: 978-3-319-21944-8

Springer, 2015

Accompanying website:

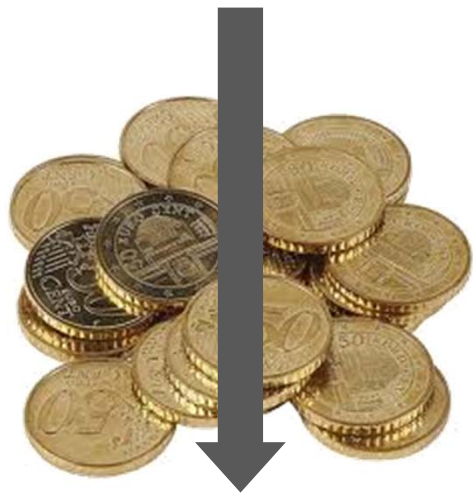
www.music-processing.de

International Audio Laboratories Erlangen

 **Fraunhofer**
IIS




FAU FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG



AUDIO
LABS

International Audio Laboratories Erlangen



Audio

International Audio Laboratories Erlangen

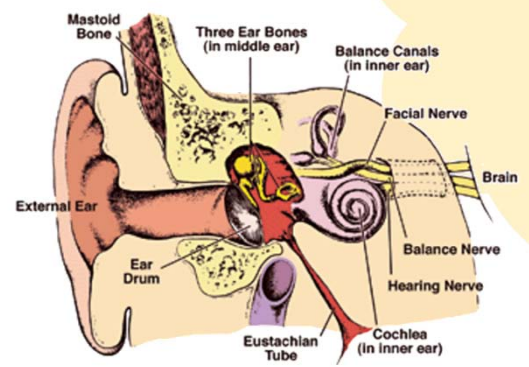
Audio Coding



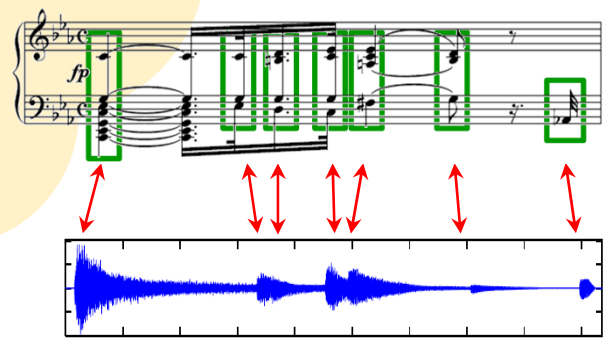
3D Audio



Audio



Psychoacoustics



Music Processing

Music

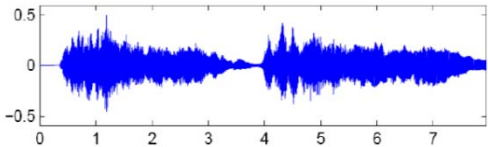


Music Information Retrieval

Sheet Music (Image)



CD / MP3 (Audio)



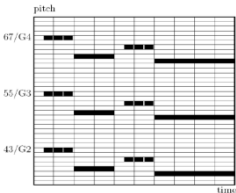
MusicXML (Text)

```
<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)



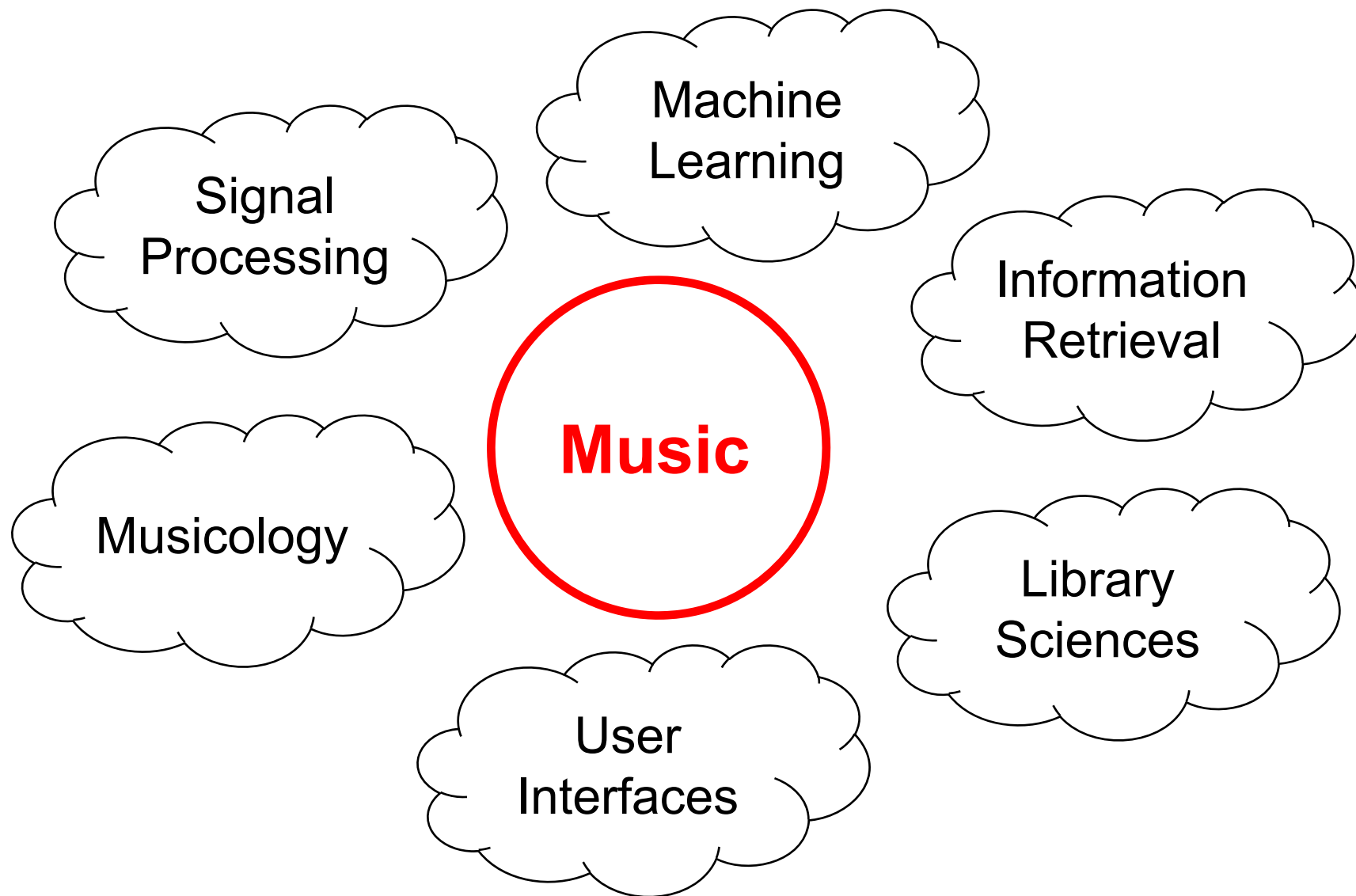
Music Film (Video)



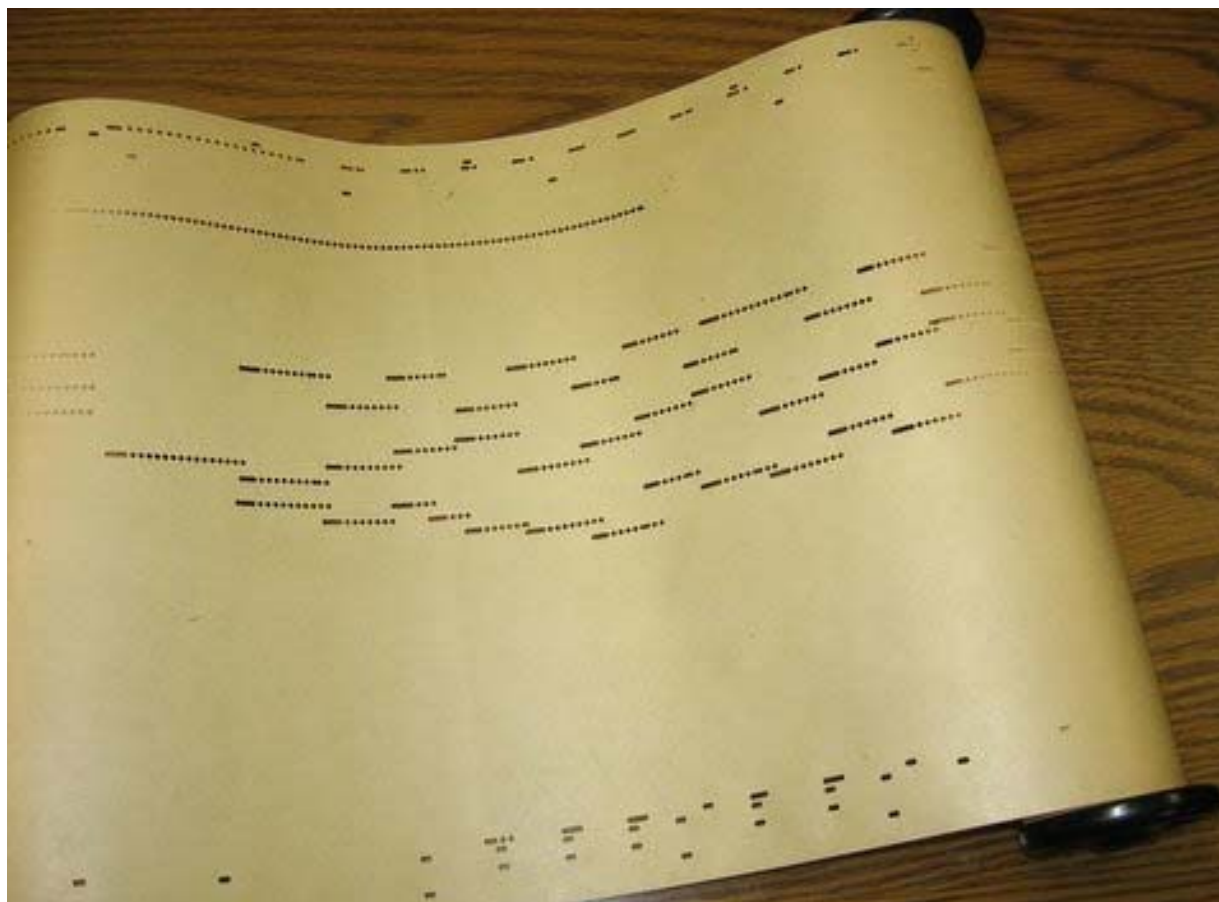
Music Literature (Text)



Music Information Retrieval



Piano Roll Representation



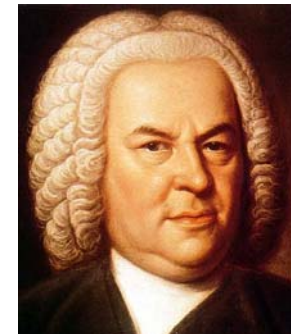
Player Piano (1900)



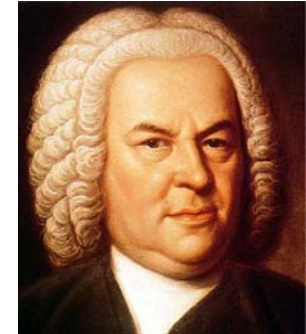
Piano Roll Representation (MIDI)

J.S. Bach, C-Major Fuge

(Well Tempered Piano, BWV 846)



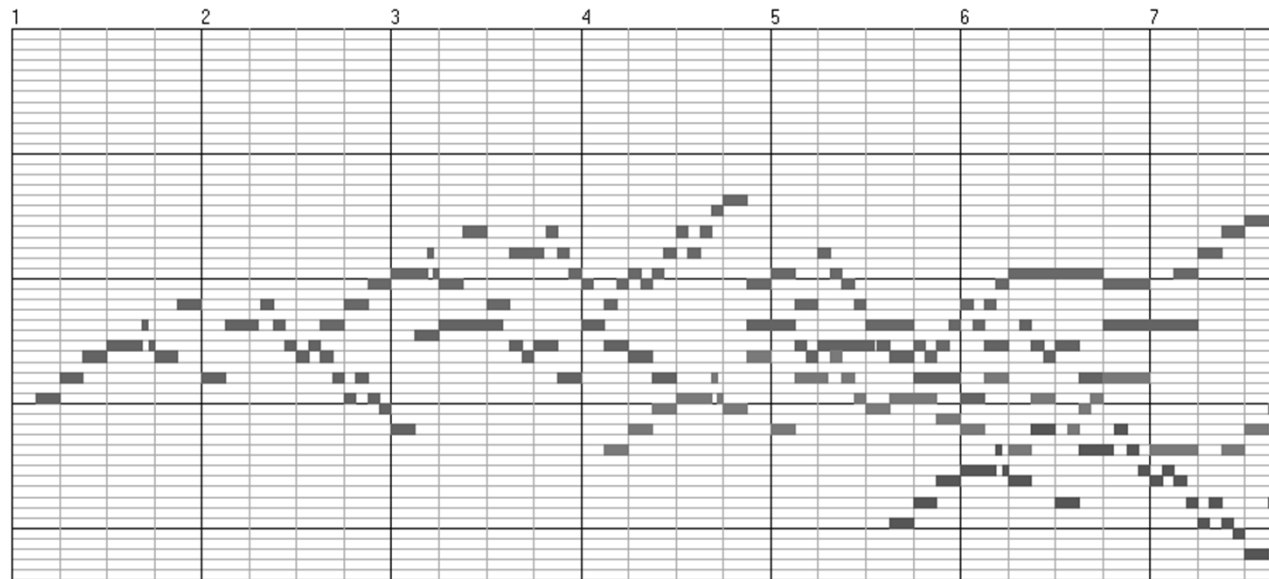
Piano Roll Representation (MIDI)



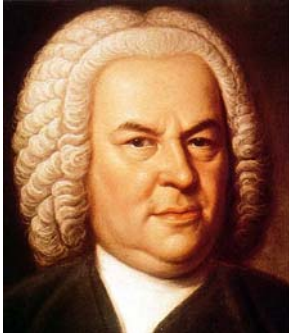
Query:



Goal: Find all occurrences of the query



Piano Roll Representation (MIDI)

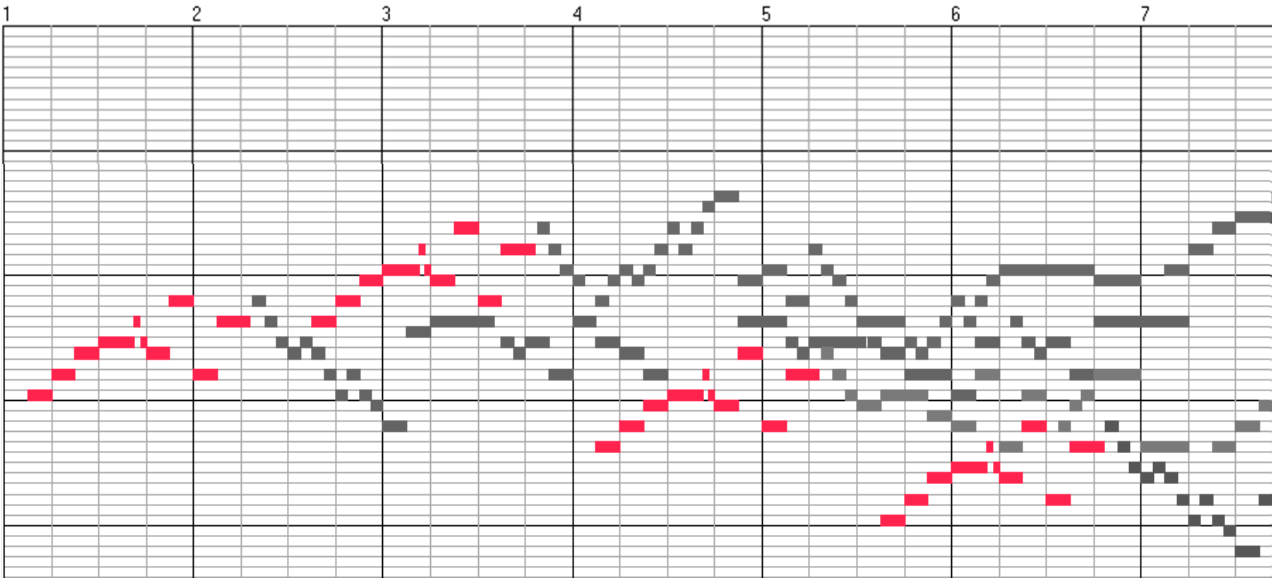


Query:

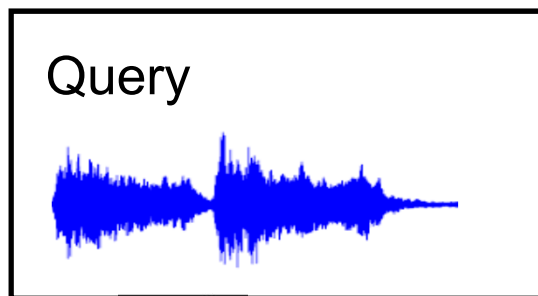


Goal: Find all occurrences of the query

Matches:



Music Retrieval

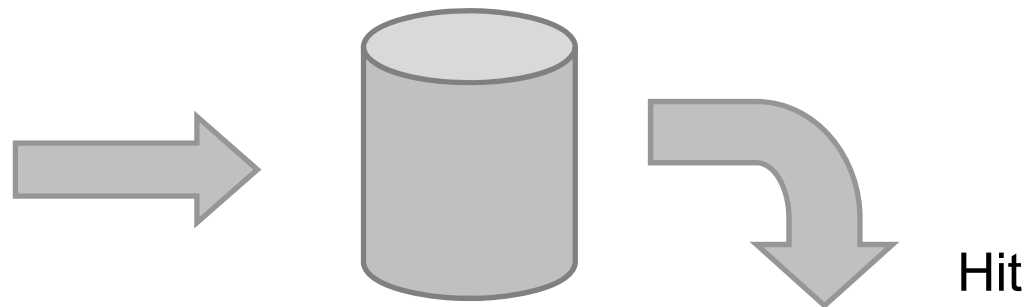


Audio-ID

Version-ID

Category-ID

Database



Bernstein (1962)
Beethoven, Symphony No. 5

Beethoven, Symphony No. 5:

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

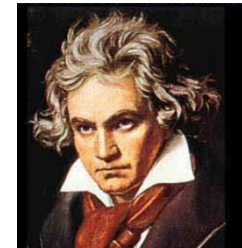


- Beethoven, Symphony No. 9
- Beethoven, Symphony No. 3
- Haydn Symphony No. 94



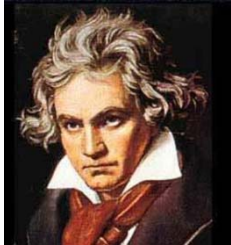
Music Synchronization: Audio-Audio

Beethoven's Fifth

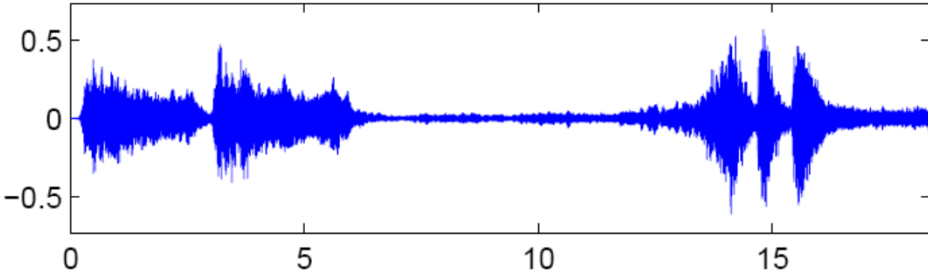


Music Synchronization: Audio-Audio

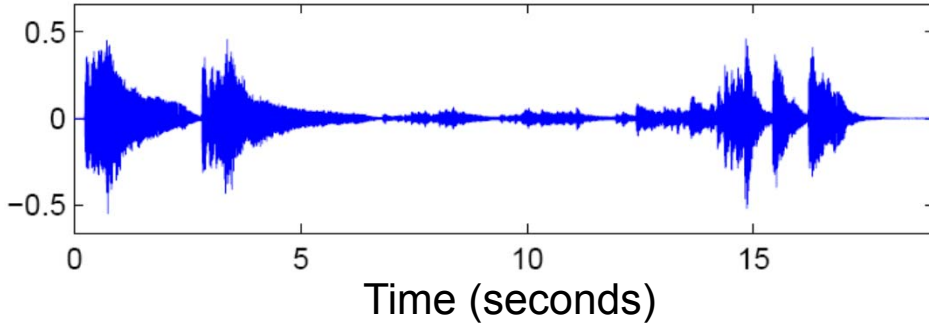
Beethoven's Fifth



Orchester
(Karajan)

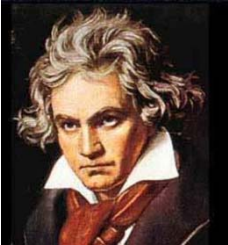


Piano
(Scherbakov)

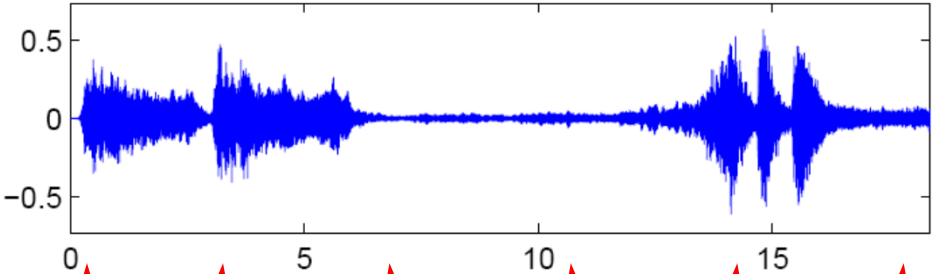


Music Synchronization: Audio-Audio

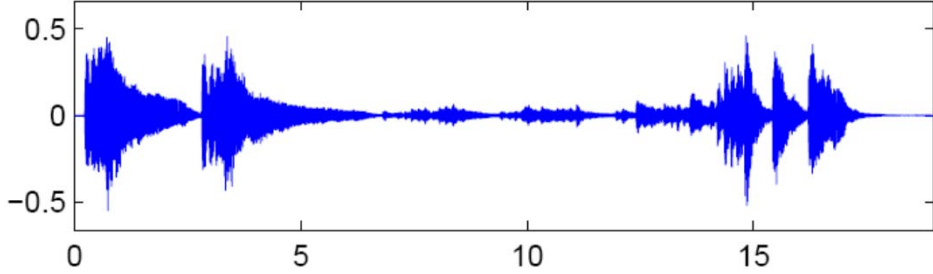
Beethoven's Fifth



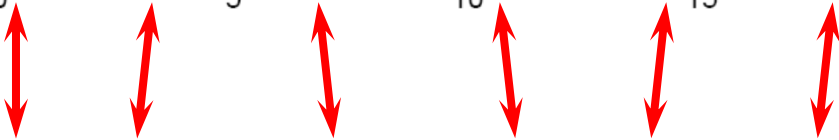
Orchester
(Karajan)



Piano
(Scherbakov)



Time (seconds)



Application: Interpretation Switcher

Interpretation Switcher
Beethoven, Op067-1_Symphony5

Interpretation	Duration
midi	00:44.18
Bernstein	01:00.64
Sawallisch	00:58.35
Scherbakov	00:52.45

Control Panel:

- midi
- Bernstein
- Sawallisch
- Scherbakov

Buttons: Deselect all

Status Bar:

- Absolute ✓
- Relative
- Reference
- ⏸
- ⏹
- ⏮ Movement selection
- Interval Repeat
- ⓘ Info



Music Synchronization: Image-Audio

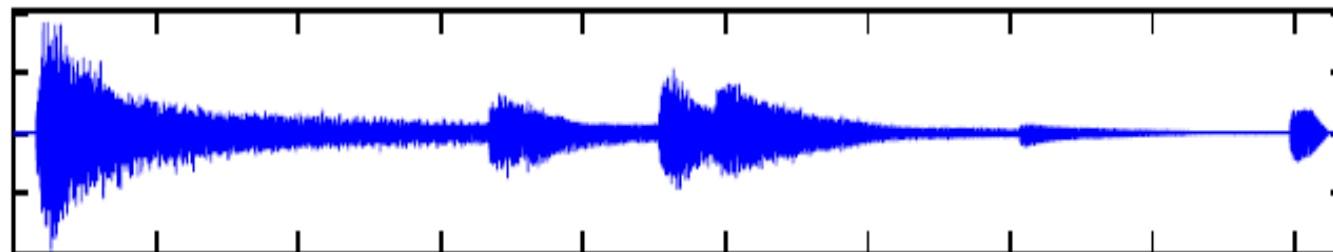
Image

Grave.

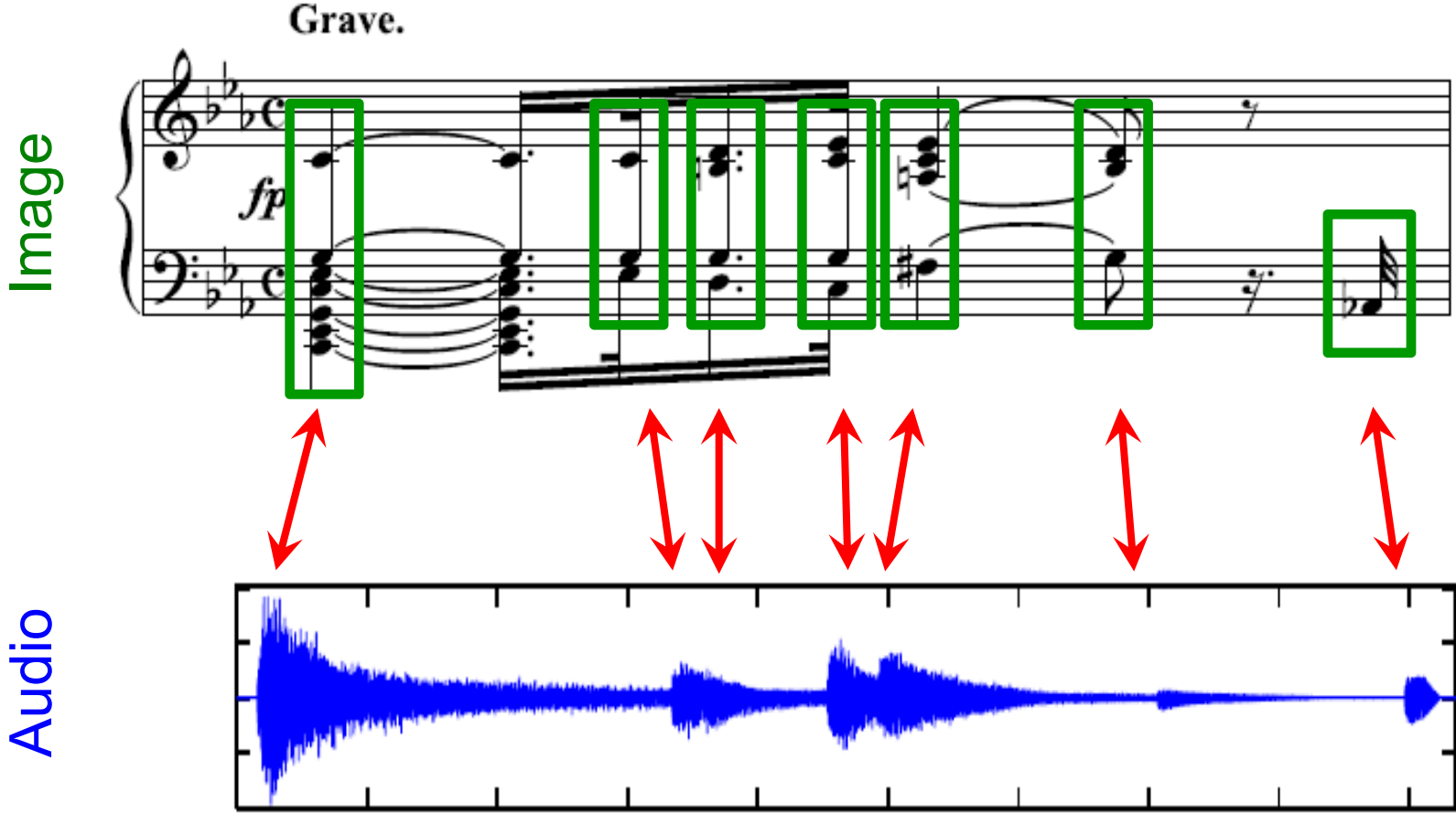


The image shows a musical score for piano, marked "Grave." and "fp". The score is written in G major (one sharp) and common time (C). It consists of two staves: a treble clef staff and a bass clef staff. The music is characterized by slow, sustained notes and chords, with a focus on the lower register of the piano.

Audio



Music Synchronization: Image-Audio

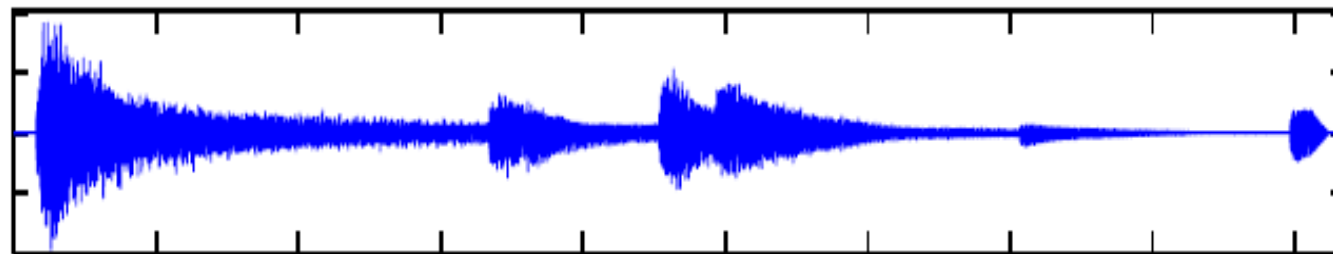


How to make the data comparable?

Image



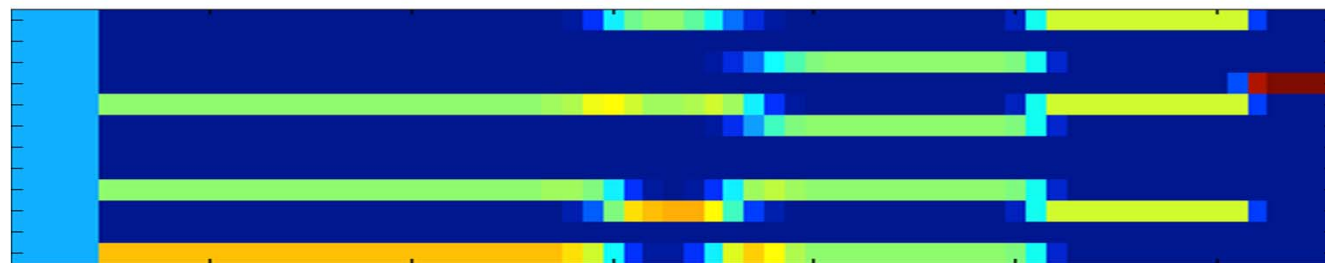
Audio



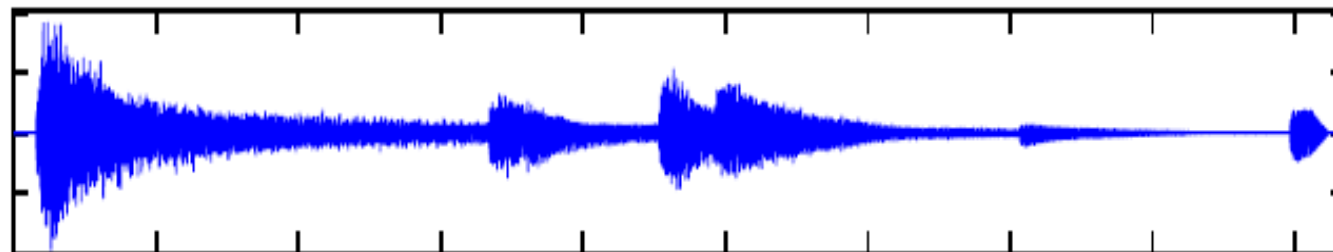
How to make the data comparable?

Image Processing: Optical Music Recognition

Image



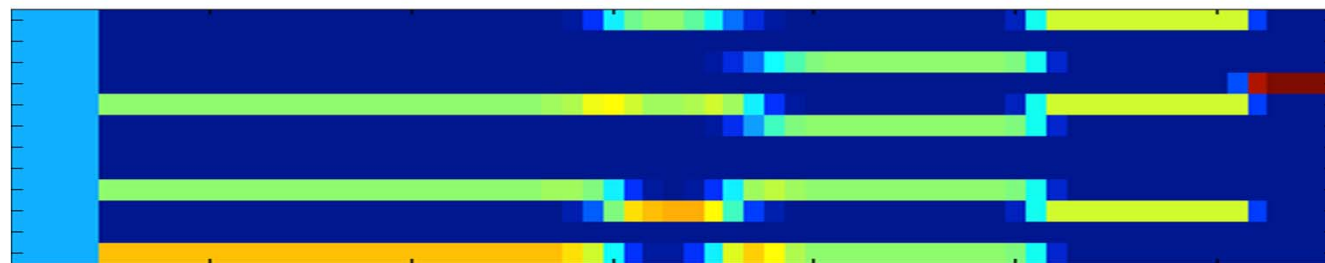
Audio



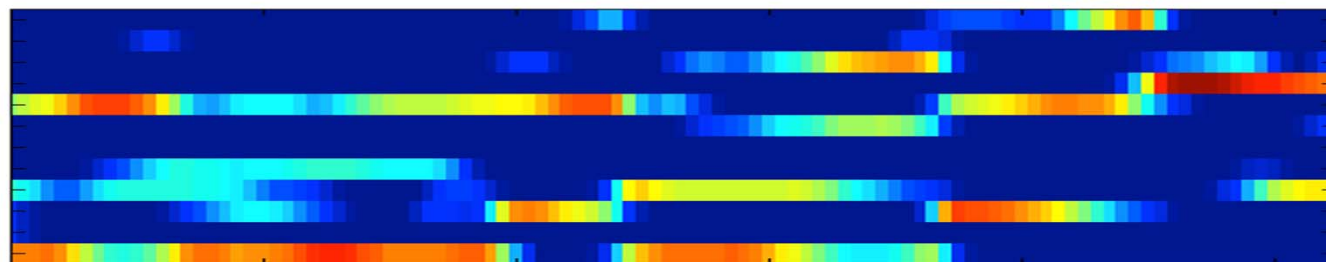
How to make the data comparable?

Image Processing: Optical Music Recognition

Image



Audio

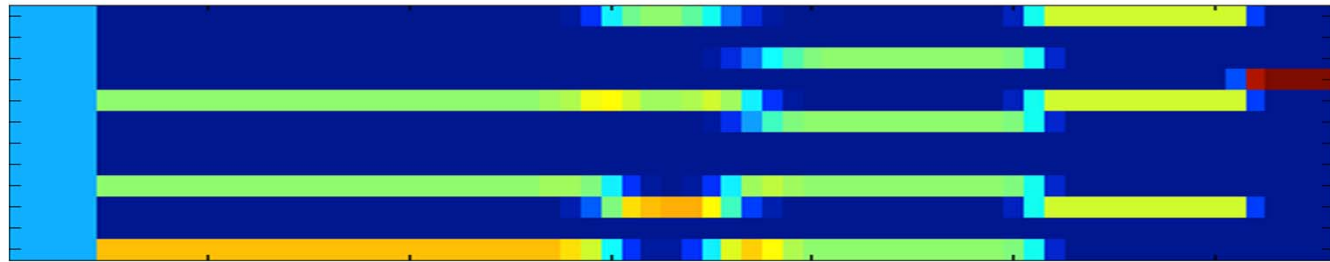


Audio Processing: Fourier Analysis

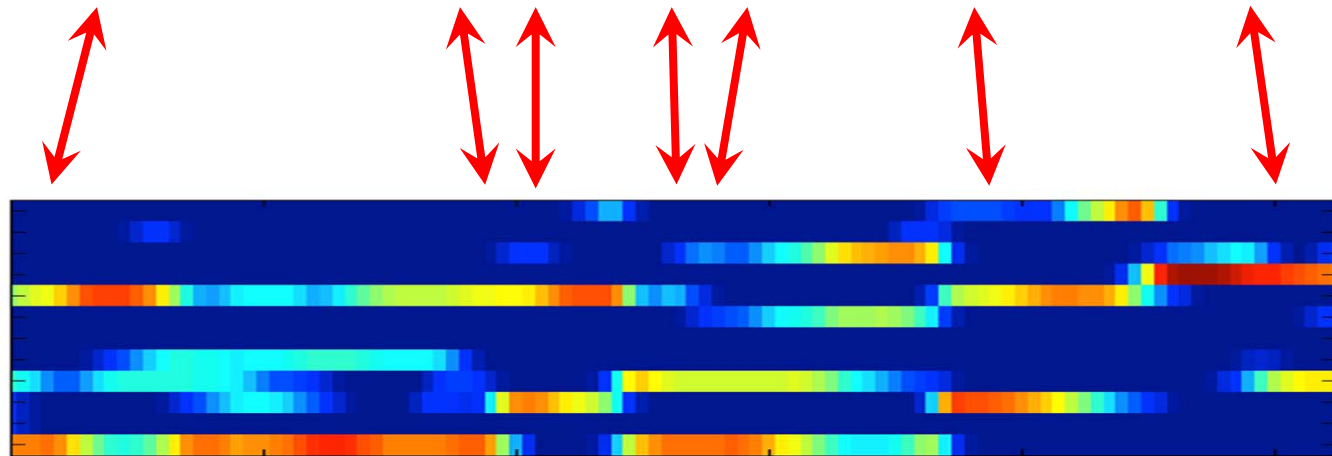
How to make the data comparable?

Image Processing: Optical Music Recognition

Image



Audio




Audio Processing: Fourier Analysis



Application: Score Viewer

AudioViewer

Beethoven - Complete Piano Sonatas - Daniel Barenboim



Disc 3

01 Sonata no.7 in D major, op.10 no.3: Presto	7:08
02 Sonata no.7 in D major, op.10 no.3: Largo e mesto	10:02
03 Sonata no.7 in D major, op.10 no.3: Menuetto (Allegro)	2:53
04 Sonata no.7 in D major, op.10 no.3: Rondo (Allegro)	4:05
05 Sonata no.8 in C minor, op.13, "Pathetique" / Allegro di molto e con brio	9:32
06 Sonata no.8 in C minor, op.13, "Pathetique" / Adagio cantabile	5:19
07 Sonata no.8 in C minor, op.13, "Pathetique" / Rondo (Allegro)	4:53
08 Sonata no.9 in E major, op.14 no.1: Allegro	6:48
09 Sonata no.9 in E major, op.14 no.1: Allegretto	4:16
10 Sonata no.9 in E major, op.14 no.1: Adagio	


Disc: 3 / 10 Track: 7

ScoreViewer

Barenboim

Beethoven - Klaviersonaten Band 1 - Henle

Sonata no.8 in C minor, op.13, "Pathetique" / Rondo (Allegro)



Track: 29 / 54 Bar: 9 / 211 Page: 159 / 285

Score Following Off Play Stop



Music Processing

Coarse Level	Fine Level
What do different versions have in common?	What are the characteristics of a specific version?

Music Processing

Coarse Level	Fine Level
What do different versions have in common?	What are the characteristics of a specific version?
What makes up a piece of music?	What makes music come alive?

Music Processing

Coarse Level	Fine Level
What do different versions have in common?	What are the characteristics of a specific version?
What makes up a piece of music?	What makes music come alive?
Identify despite of differences	Identify the differences

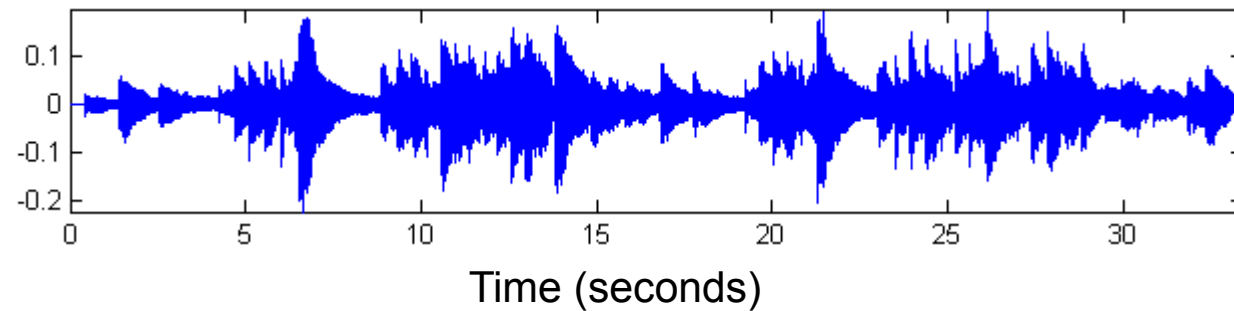
Music Processing

Coarse Level	Fine Level
What do different versions have in common?	What are the characteristics of a specific version?
What makes up a piece of music?	What makes music come alive?
Identify despite of differences	Identify the differences
Example tasks: Audio Matching Cover Song Identification	Example tasks: Tempo Estimation Performance Analysis

Performance Analysis

Schumann: Träumerei

Performance:



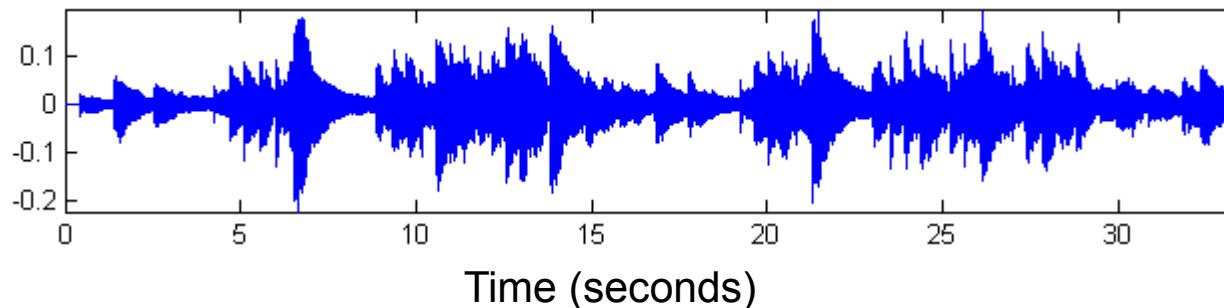
Performance Analysis

Schumann: Träumerei

Score (reference):



Performance:



Performance Analysis

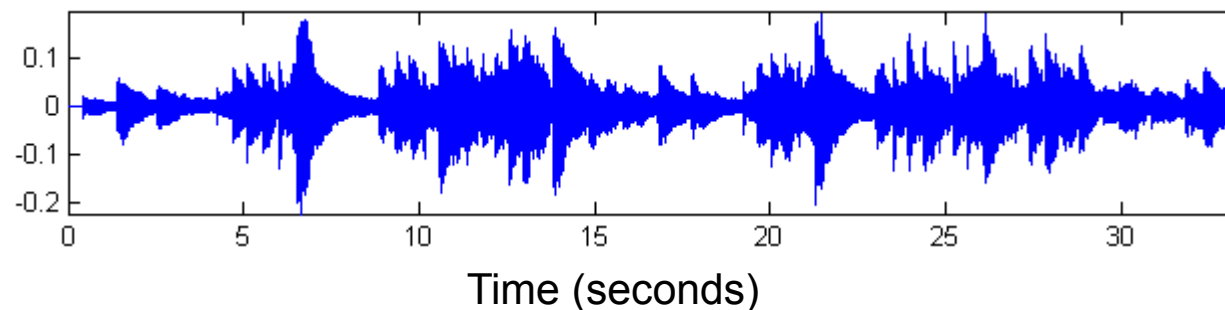
Schumann: Träumerei

Score (reference):



Strategy: Compute score-audio synchronization and derive tempo curve

Performance:



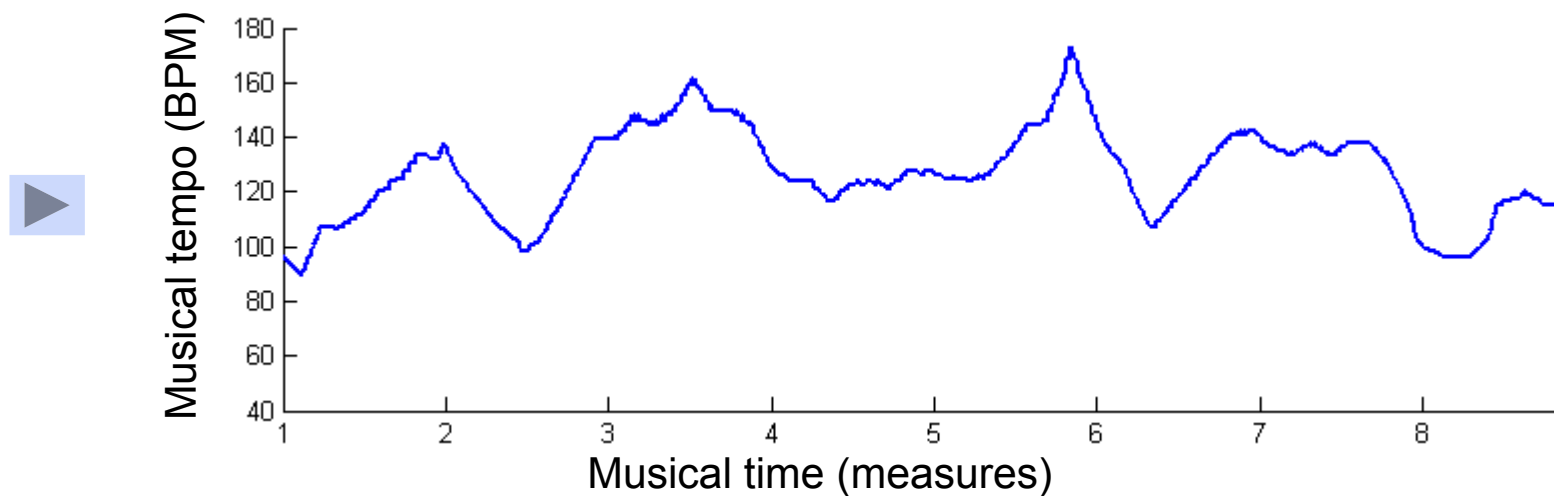
Performance Analysis

Schumann: Träumerei

Score (reference):



Tempo Curve:



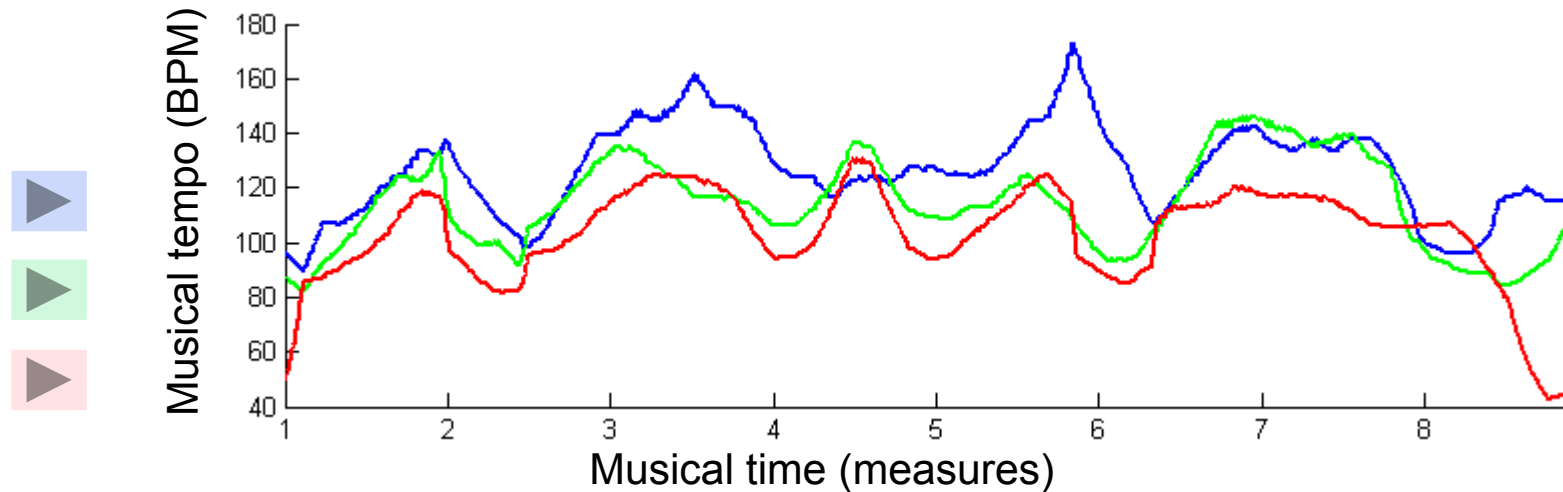
Performance Analysis

Schumann: Träumerei

Score (reference):



Tempo Curves:



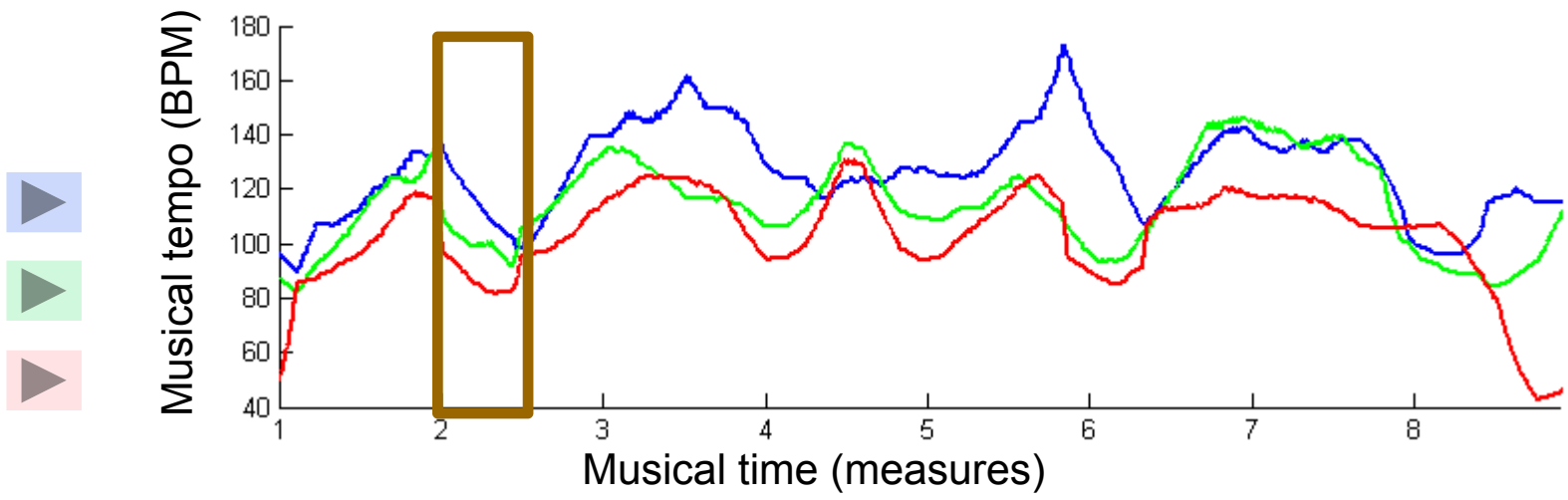
Performance Analysis

Schumann: Träumerei

Score (reference):



Tempo Curves:



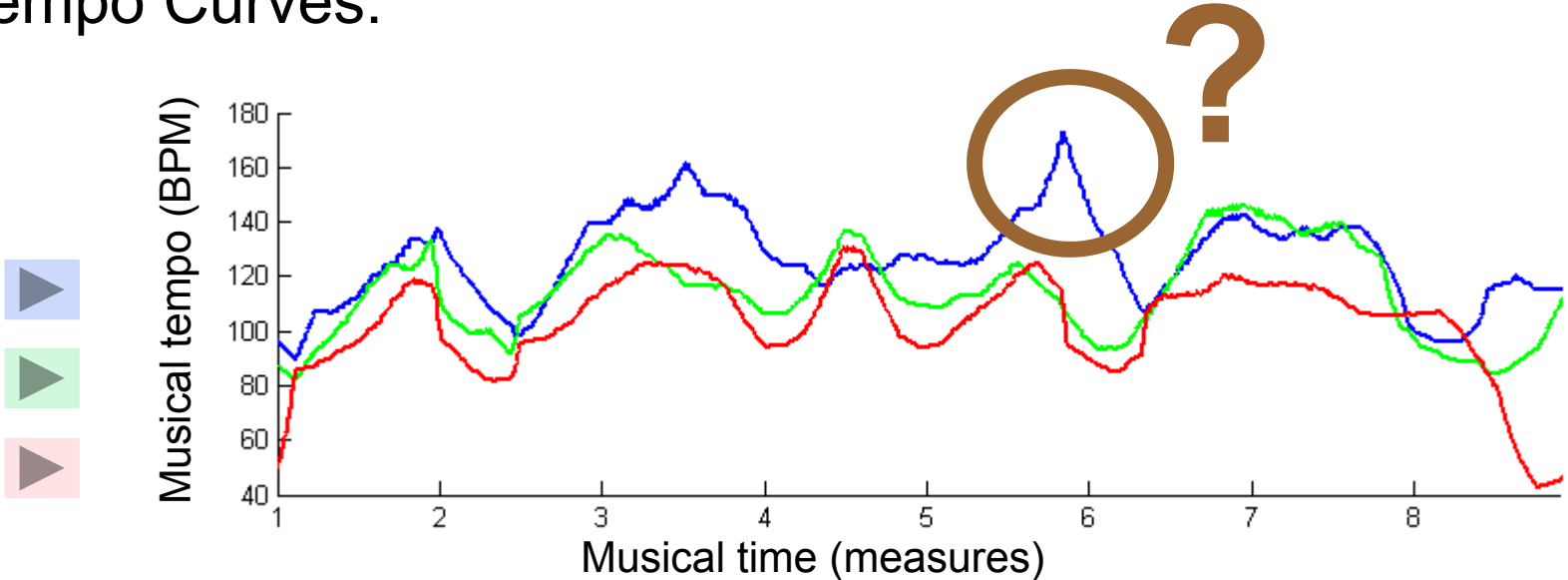
Performance Analysis

Schumann: Träumerei

Score (reference):



Tempo Curves:

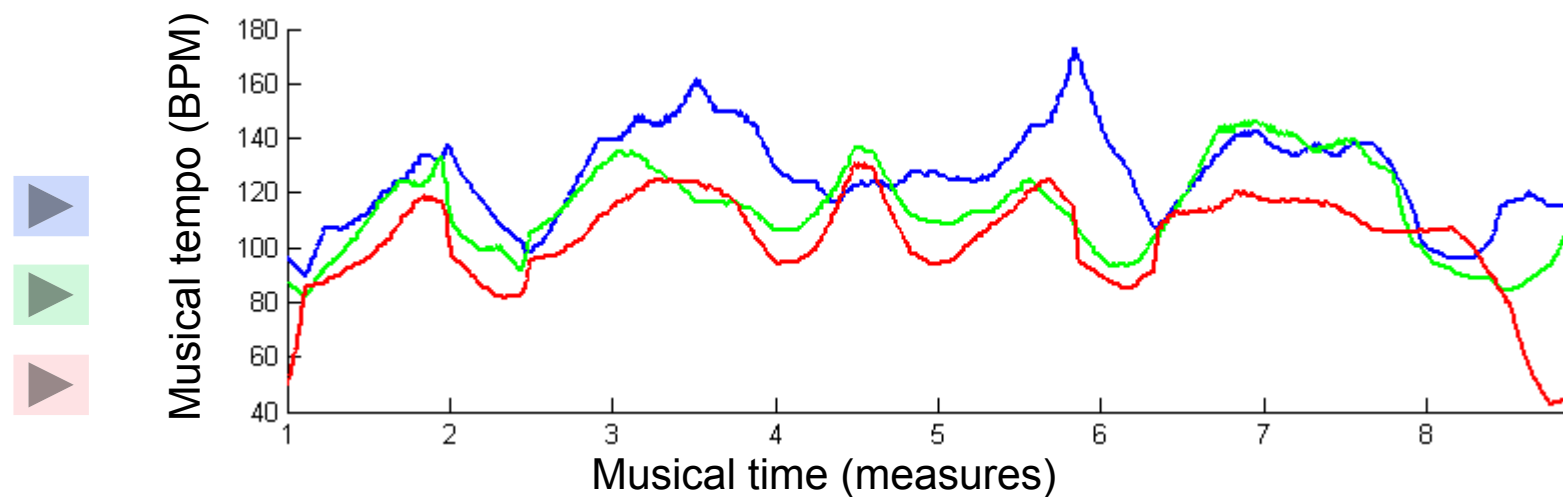


Performance Analysis

Schumann: Träumerei

What can be done if no reference is available?

Tempo Curves:



Music Processing

Relative	Absolute
Given: Several versions	Given: One version

Music Processing

Relative	Absolute
Given: Several versions	Given: One version
Comparison of extracted parameters	Direct interpretation of extracted parameters

Music Processing

Relative	Absolute
Given: Several versions	Given: One version
Comparison of extracted parameters	Direct interpretation of extracted parameters
Extraction errors have often no consequence on final result	Extraction errors immediately become evident

Music Processing

Relative	Absolute
Given: Several versions	Given: One version
Comparison of extracted parameters	Direct interpretation of extracted parameters
Extraction errors have often no consequence on final result	Extraction errors immediately become evident
Example tasks: Music Synchronization Genre Classification	Example tasks: Music Transcription Tempo Estimation

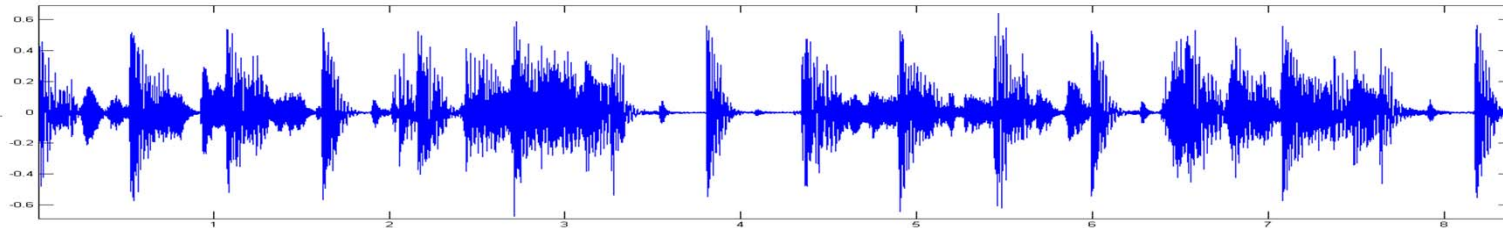
Tempo Estimation and Beat Tracking

Basic task: “Tapping the foot when listening to music”

Tempo Estimation and Beat Tracking

Basic task: “Tapping the foot when listening to music”

Example: Queen – Another One Bites The Dust

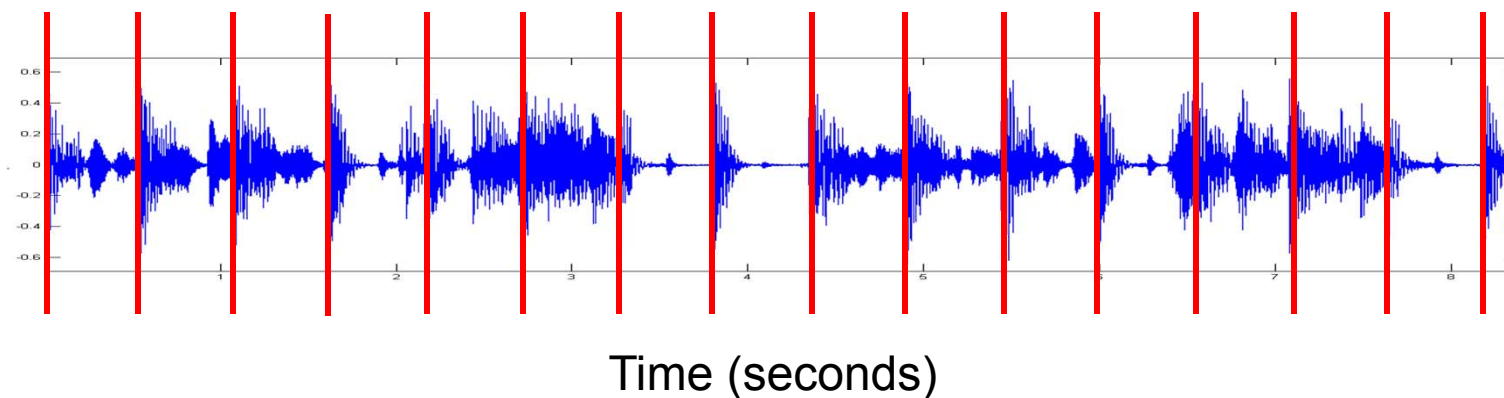


Time (seconds)

Tempo Estimation and Beat Tracking

Basic task: “Tapping the foot when listening to music”

Example: Queen – Another One Bites The Dust



Tempo Estimation and Beat Tracking

Example: Happy Birthday to you

Pulse level: **Measure**

The image shows a musical score for the song "Happy Birthday to you" in 3/4 time. The score is written on two staves. The first staff contains the first two phrases of the song: "Hap - py Birth - day to you," and "Hap - py Birth - day to you," followed by the start of "Hap - py". The second staff contains the continuation: "Birth - day dear _____," followed by "Hap - py Birth - day to you!". Four red arrows point downwards to the first note of each of the four measures in the first staff, indicating the pulse level at the measure level.

Tempo Estimation and Beat Tracking

Example: Happy Birthday to you

Pulse level: **Tactus (beat)**

The image shows a musical score for the song "Happy Birthday to you" in 3/4 time. The score is written on two staves. The first staff contains the first two phrases of the song: "Hap - py Birth - day to you, Hap - py Birth - day to you, Hap - py". The second staff contains the final phrase: "Birth - day dear _____, Hap - py Birth - day to you!". Above the first staff, there are 13 red arrows pointing down to the first note of each measure, indicating the pulse level (Tactus) at the start of each measure. The time signature is 3/4, and the key signature is one sharp (F#).

Tempo Estimation and Beat Tracking

Example: Happy Birthday to you

Pulse level: **Tatum (temporal atom)**

The image shows a musical score for the song "Happy Birthday to you" in 3/4 time. The score is written on two staves. The first staff contains the melody for the first two phrases: "Hap - py Birth - day to you, Hap - py Birth - day to you, Hap - py". The second staff contains the melody for the final phrase: "Birth - day dear _____, Hap - py Birth - day to you!". Above the first staff, there are 24 red arrows pointing downwards, indicating the pulse level (Tatum) for each note. The arrows are placed above the notes, showing a regular interval corresponding to the 3/4 time signature.

Tempo Estimation and Beat Tracking

Example: Chopin – Mazurka Op. 68-3

Pulse level: Quarter note

Tempo: ???



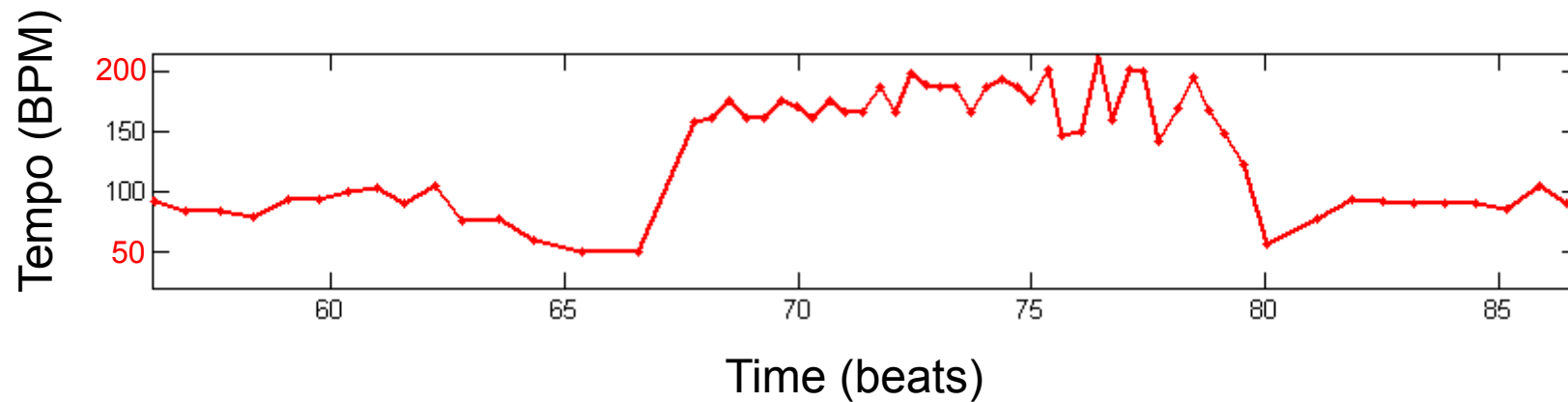
Tempo Estimation and Beat Tracking

Example: Chopin – Mazurka Op. 68-3

Pulse level: Quarter note

Tempo: **50-200 BPM** 

Tempo curve

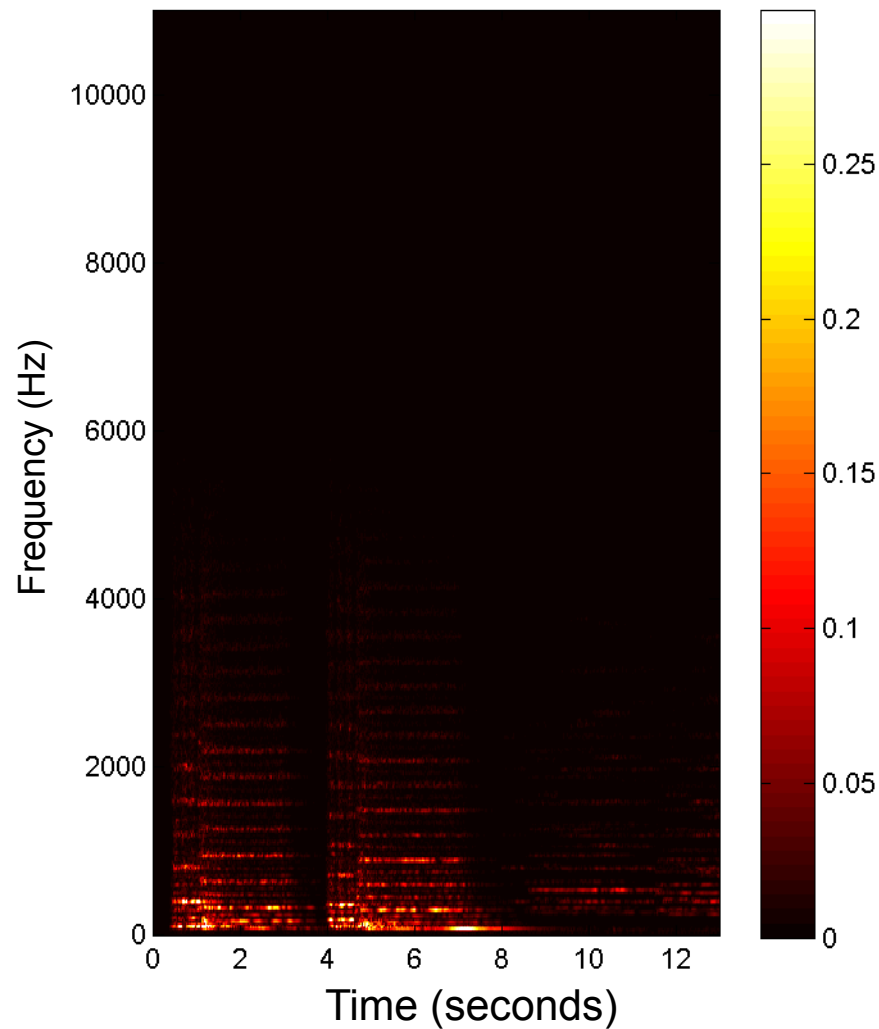


Tempo Estimation and Beat Tracking

- Which temporal level?
- Local tempo deviations
- Sparse information
(e.g., only note onsets available)
- Vague information
(e.g., extracted note onsets corrupt)

Tempo Estimation and Beat Tracking

Spectrogram

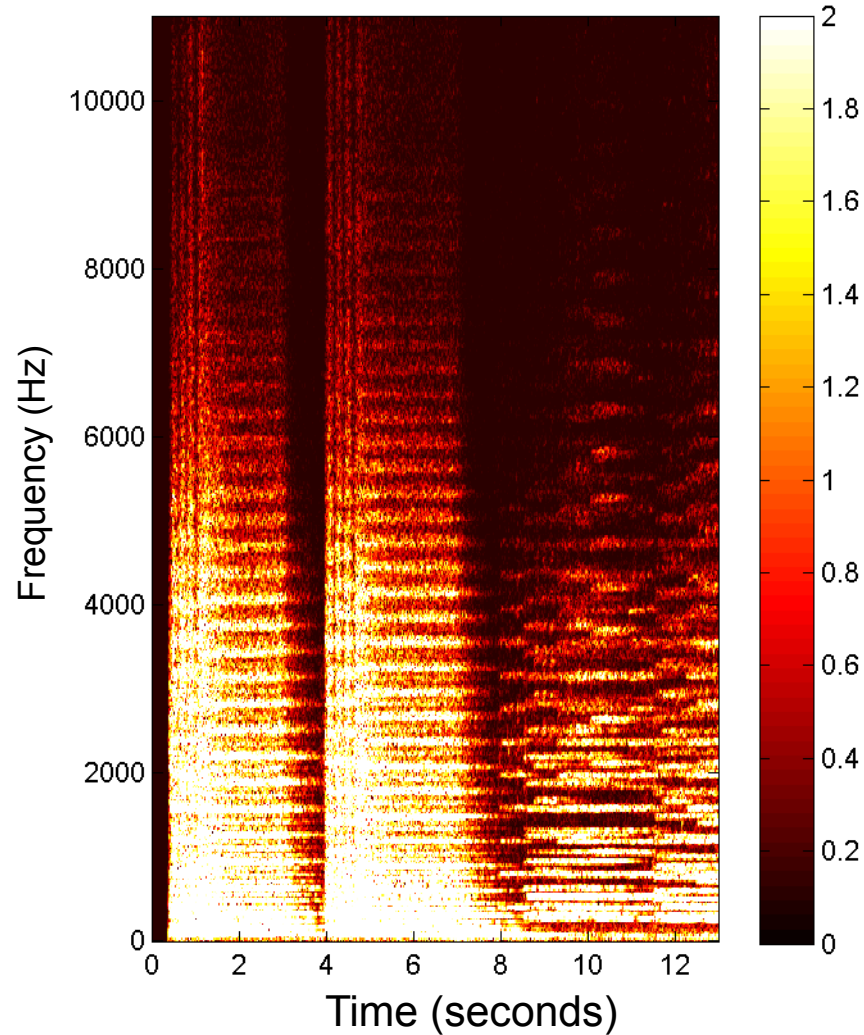


Steps:

1. Spectrogram

Tempo Estimation and Beat Tracking

Compressed Spectrogram

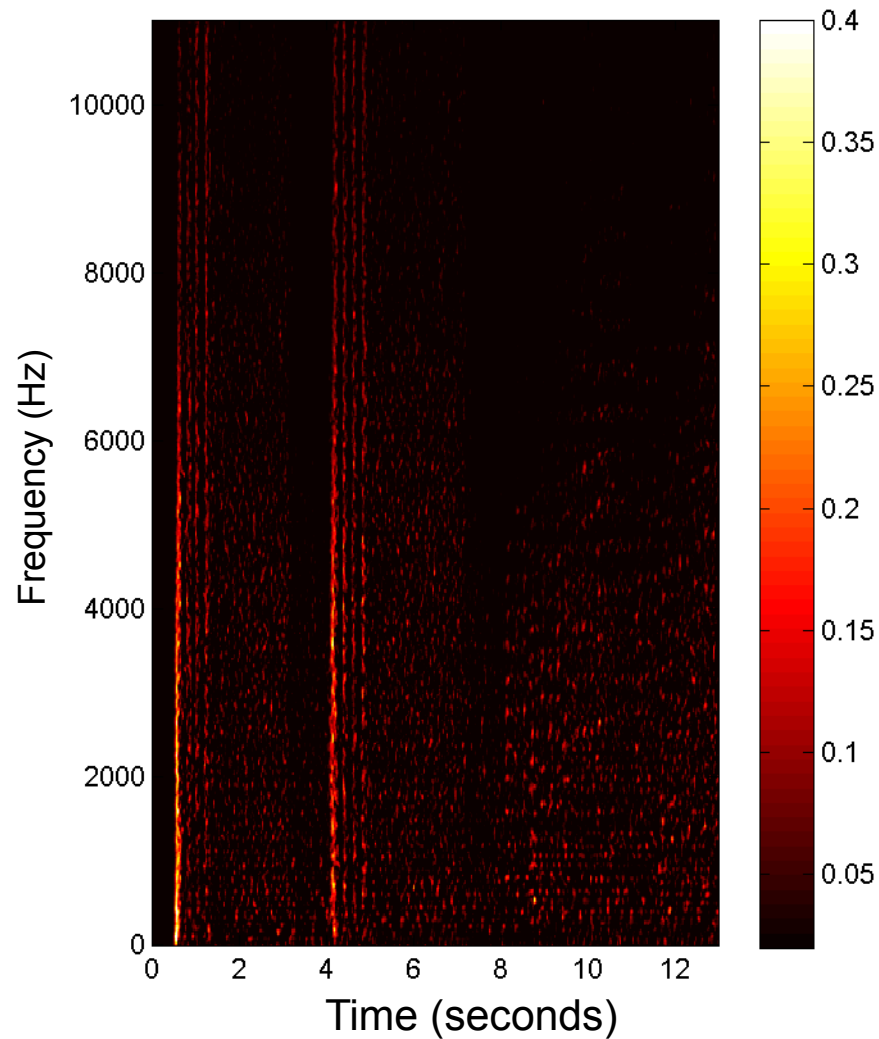


Steps:

1. Spectrogram
2. Log Compression

Tempo Estimation and Beat Tracking

Difference Spectrogram



Steps:

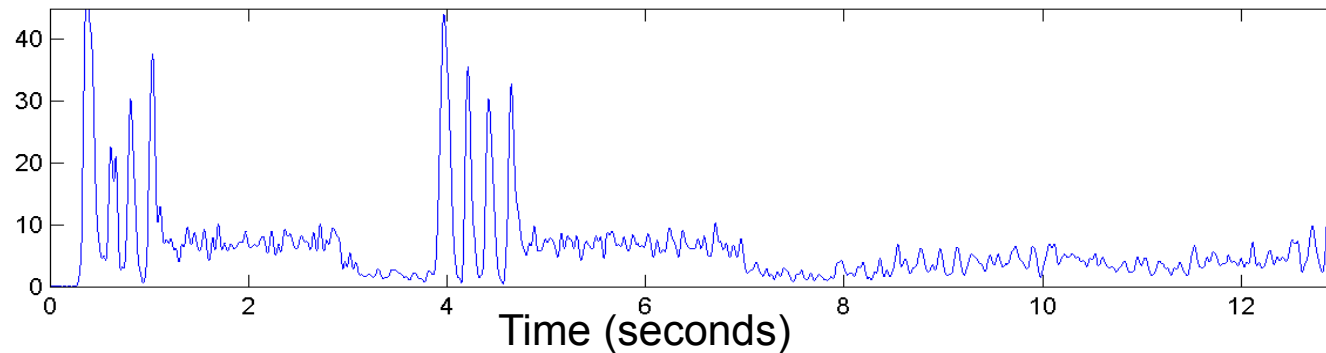
1. Spectrogram
2. Log Compression
3. Differentiation

Tempo Estimation and Beat Tracking

Steps:

1. Spectrogram
2. Log Compression
3. Differentiation
4. Accumulation

Novelty Curve

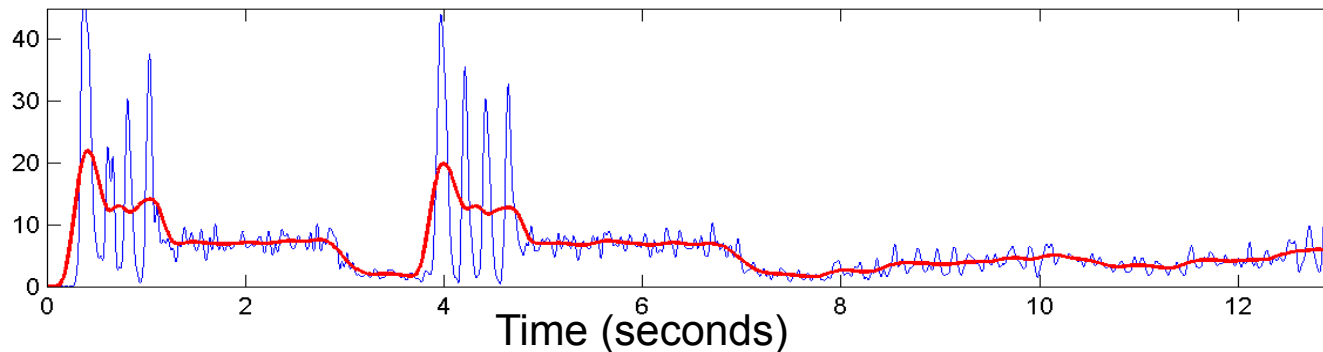


Tempo Estimation and Beat Tracking

Steps:

1. Spectrogram
2. Log Compression
3. Differentiation
4. Accumulation

Novelty Curve Local Average

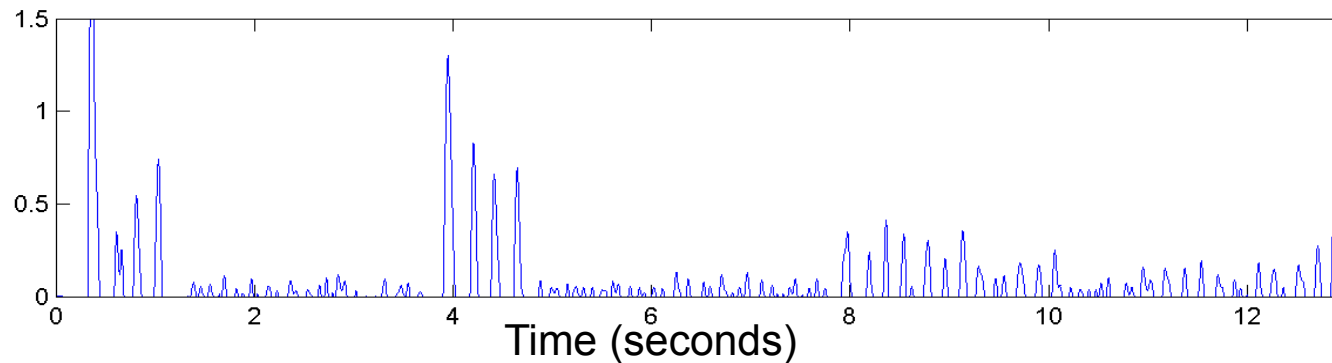


Tempo Estimation and Beat Tracking

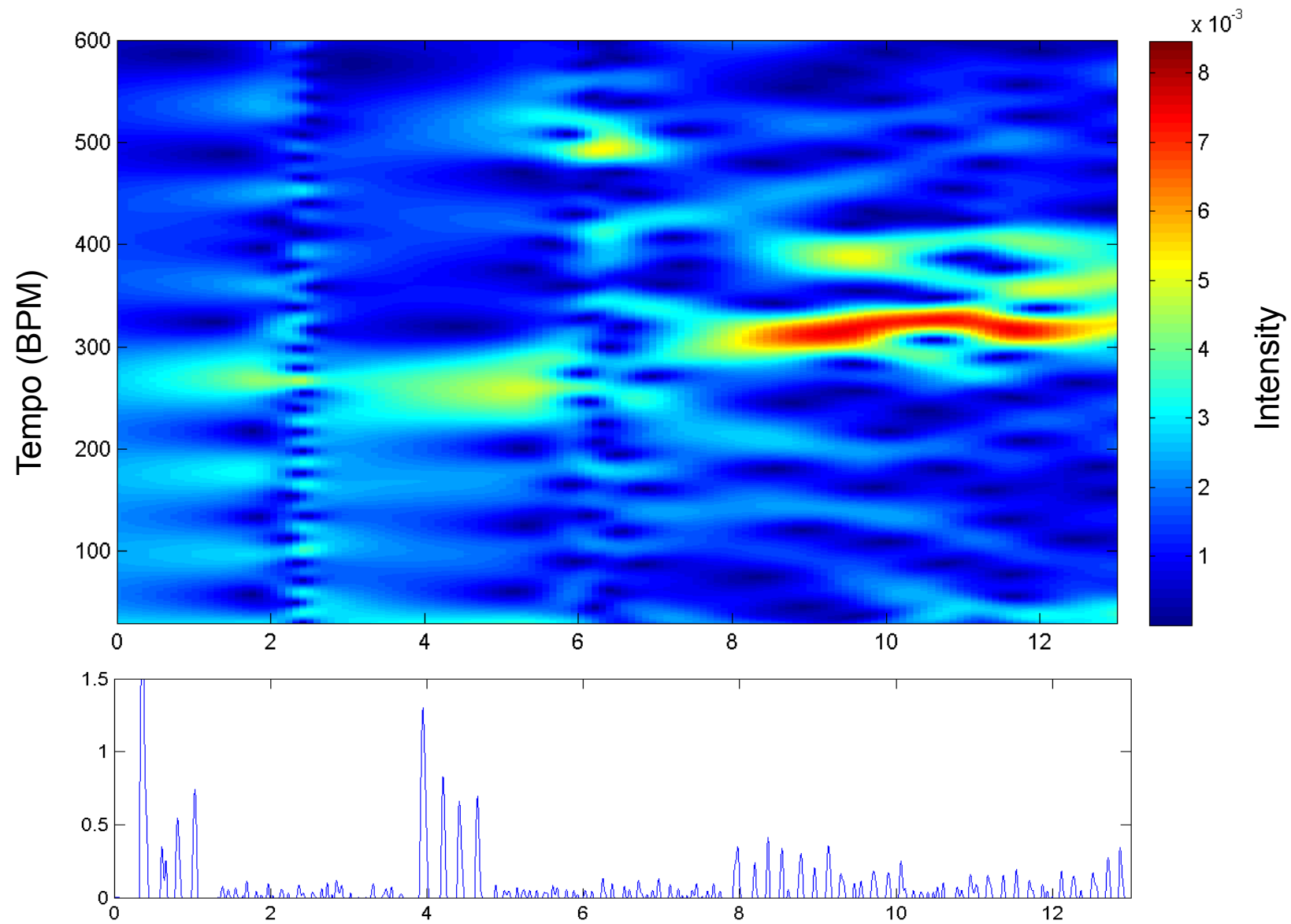
Steps:

1. Spectrogram
2. Log Compression
3. Differentiation
4. Accumulation
5. Normalization

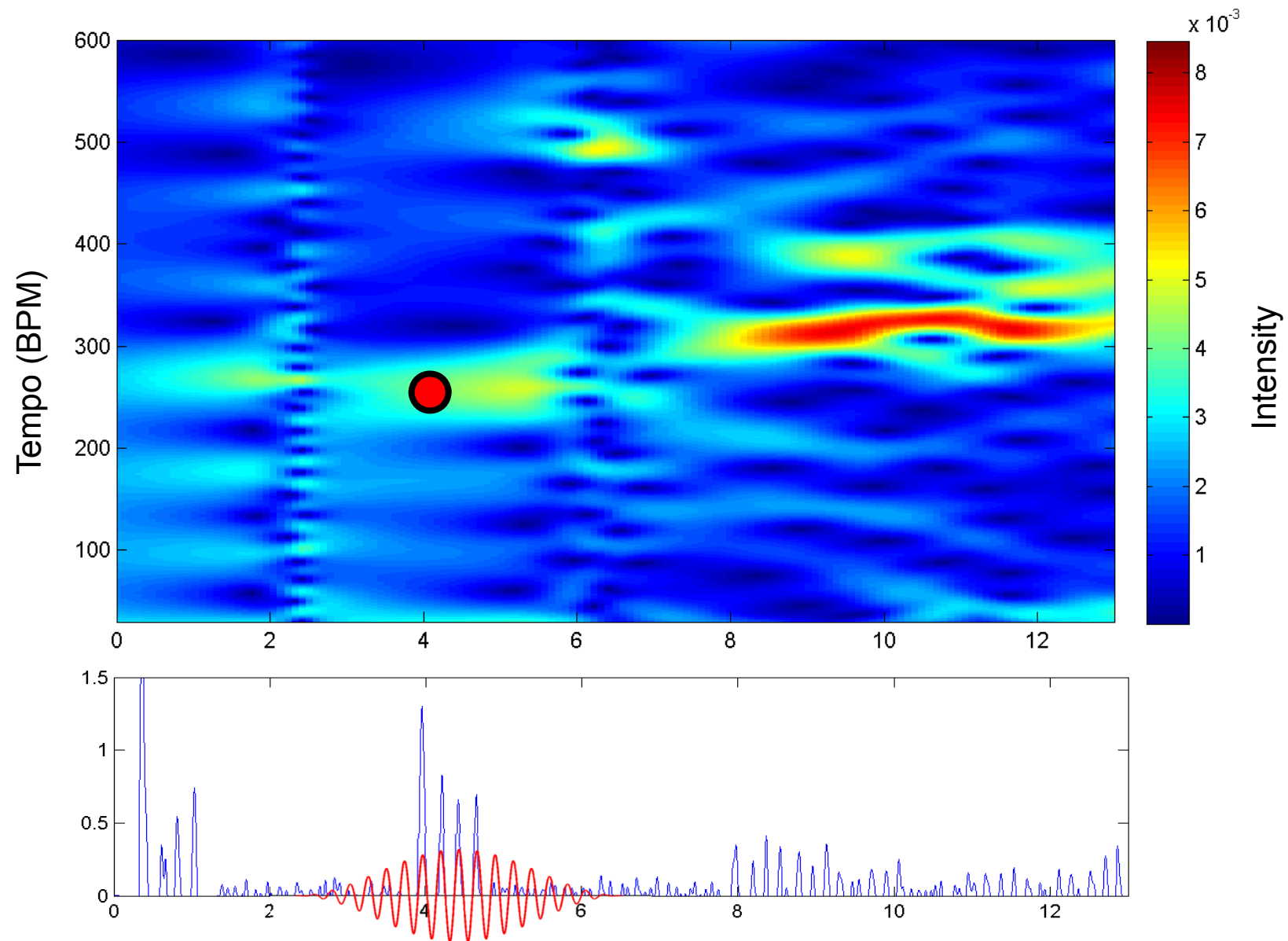
Novelty Curve



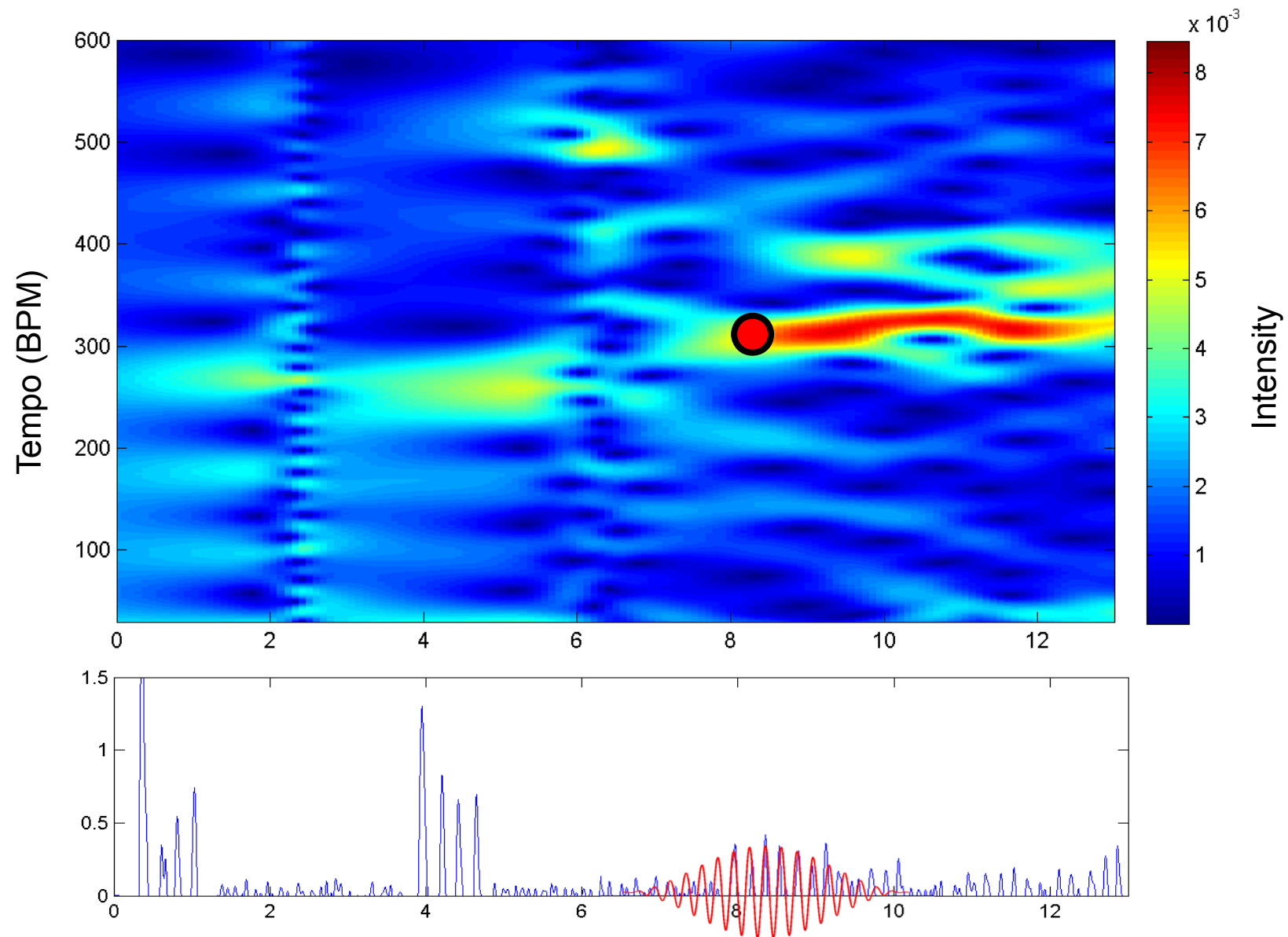
Tempo Estimation and Beat Tracking



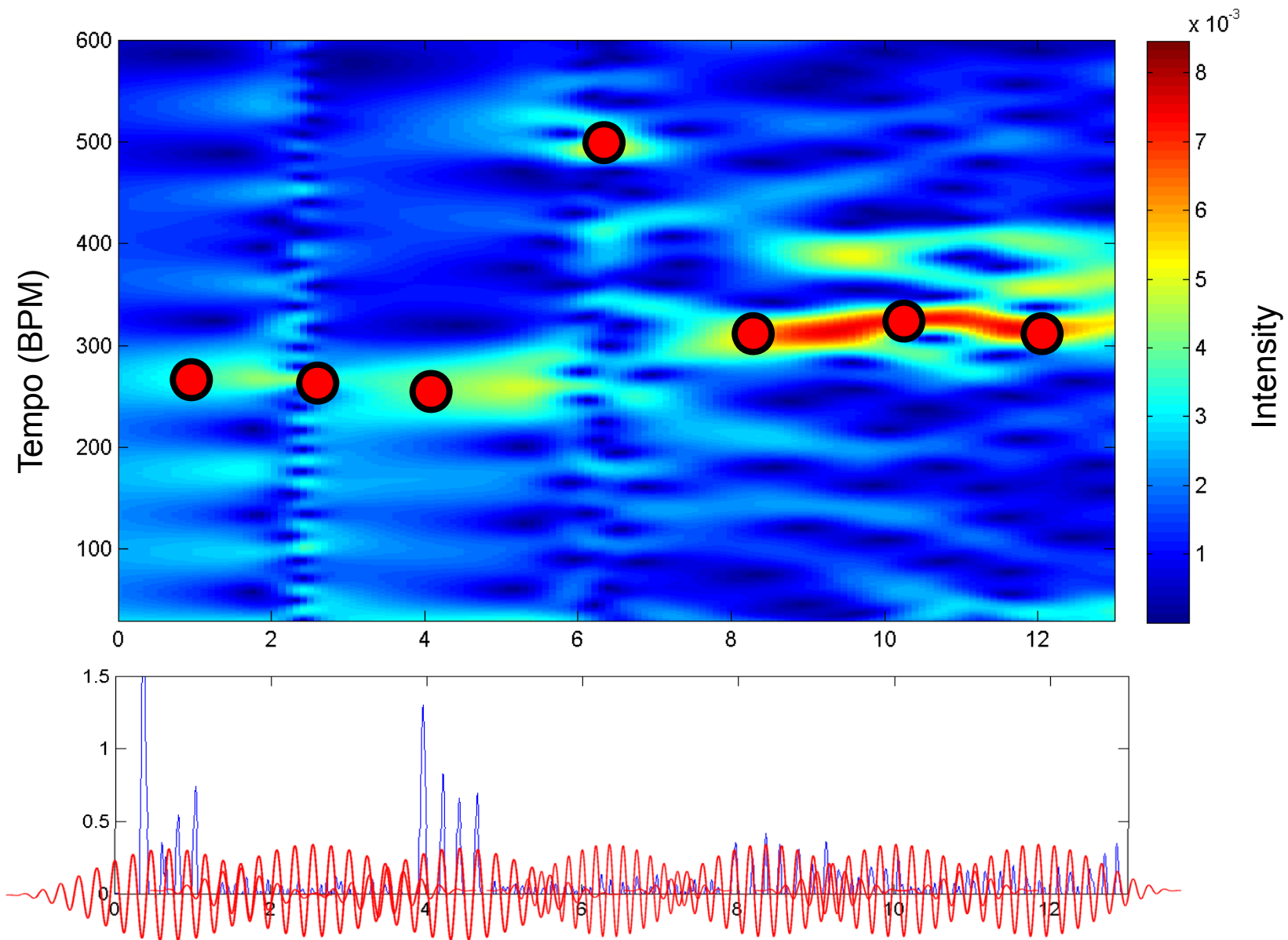
Tempo Estimation and Beat Tracking



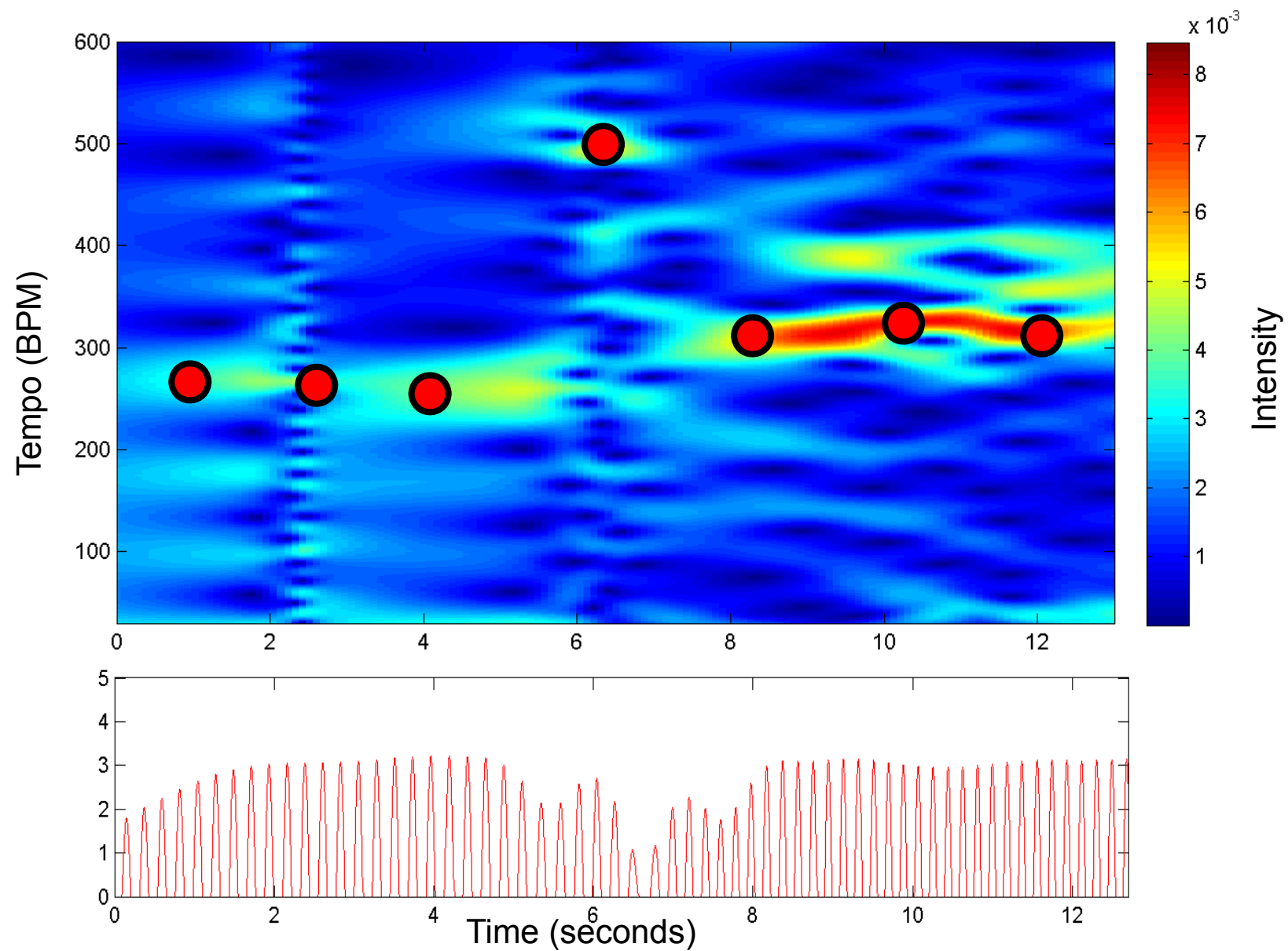
Tempo Estimation and Beat Tracking



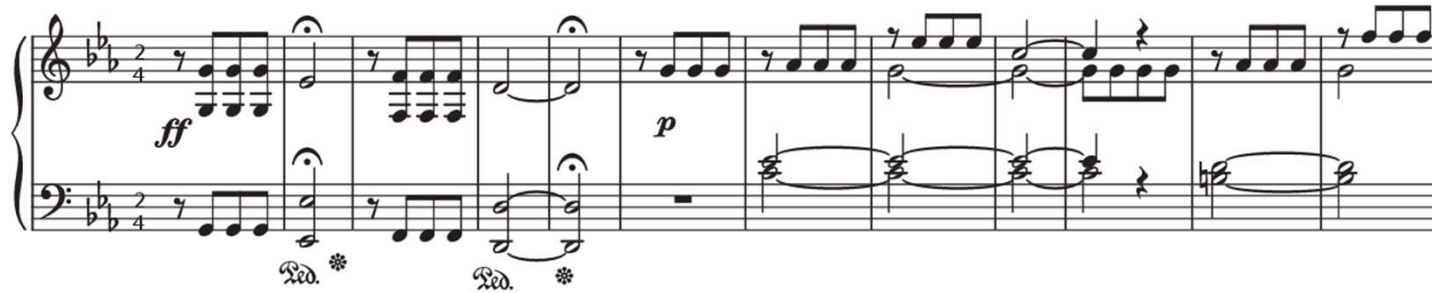
Tempo Estimation and Beat Tracking



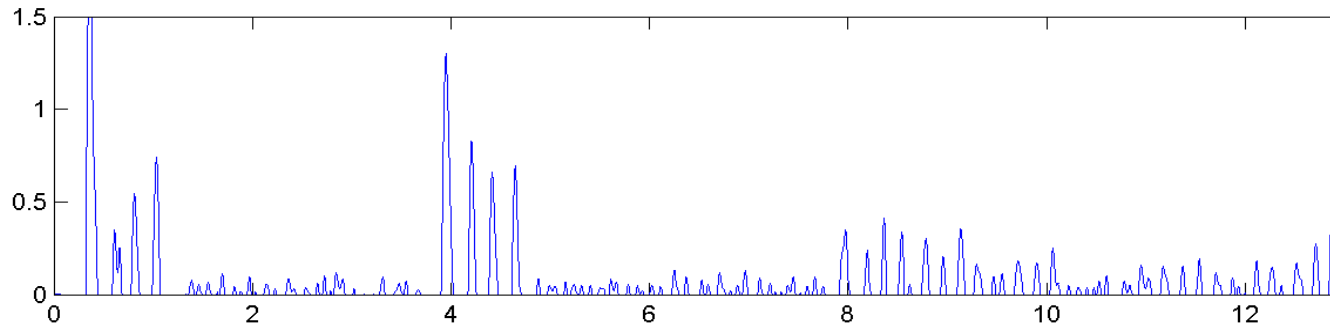
Tempo Estimation and Beat Tracking



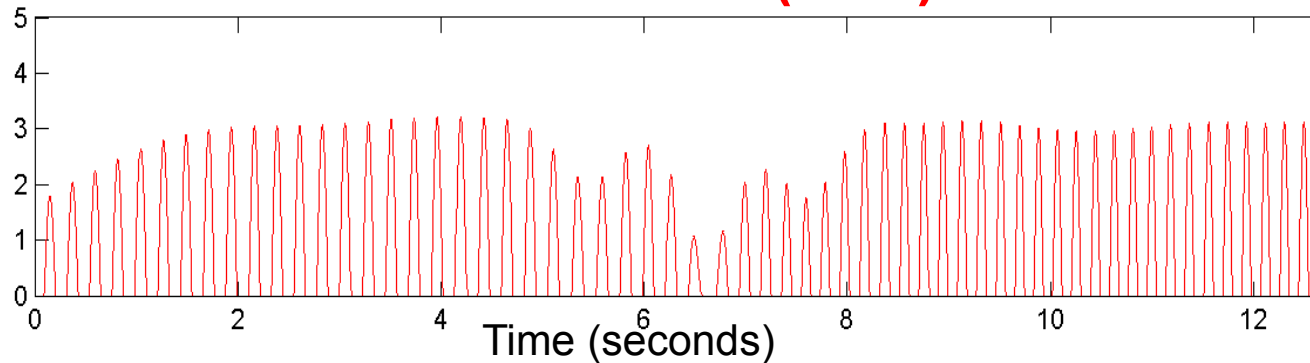
Tempo Estimation and Beat Tracking



Novelty Curve



Predominant Local Pulse (PLP)



Tempo Estimation and Beat Tracking

Light effects

Music recommendation

DJ

Audio editing



Why is Music Processing Challenging?

Example: Chopin, Mazurka Op. 63 No. 3



Mazurka.

F. CHOPIN. Op. 63, № 3.

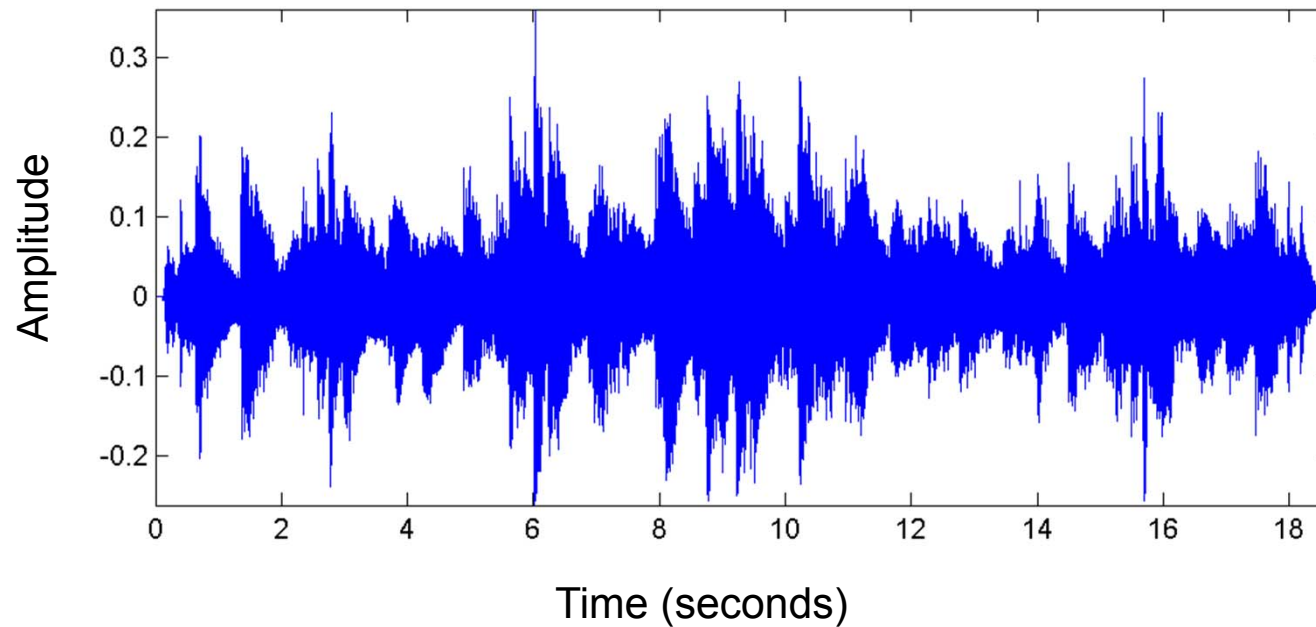
41. Allegretto. *p*

The image shows two systems of musical notation for Chopin's Mazurka Op. 63 No. 3, measures 41-50. The first system (measures 41-46) features a treble clef with a key signature of three sharps (F#, C#, G#) and a 3/4 time signature. The melody is marked 'Allegretto' and 'p' (piano). It includes triplets and a fourth-note figure. The bass clef accompaniment consists of chords, with some marked 'Ped.' and asterisks. The second system (measures 47-50) continues the melody and accompaniment with similar markings.

Why is Music Processing Challenging?

Example: Chopin, Mazurka Op. 63 No. 3

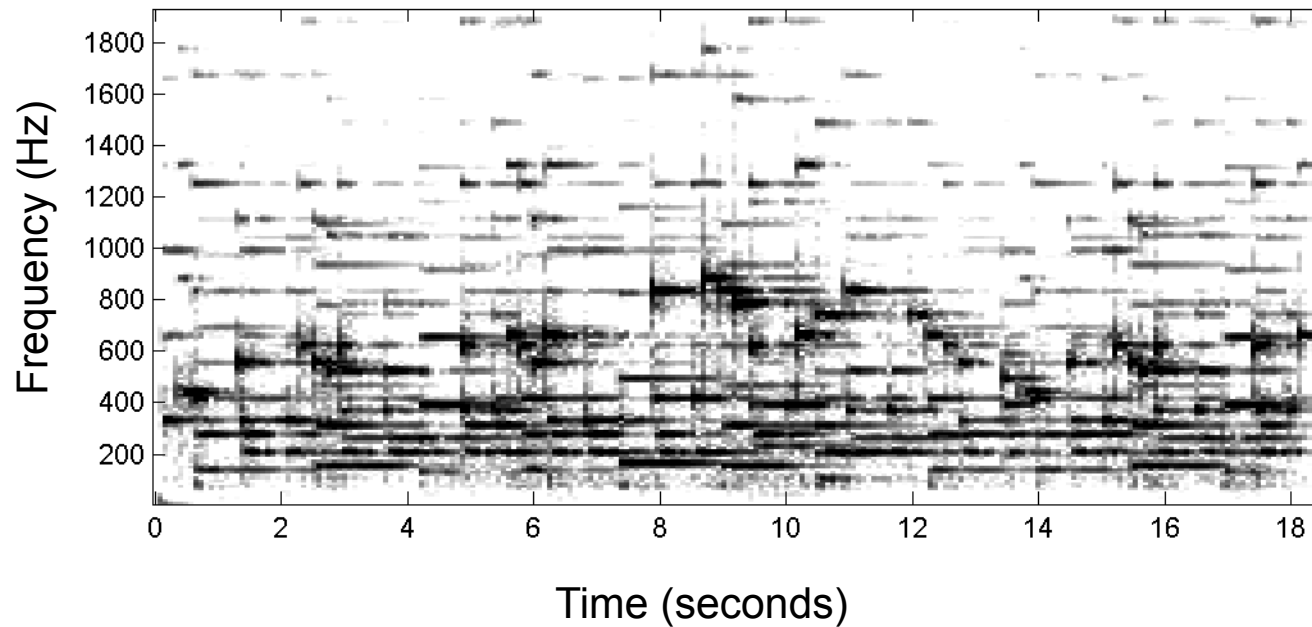
- Waveform



Why is Music Processing Challenging?

Example: Chopin, Mazurka Op. 63 No. 3

- Waveform / Spectrogram



Why is Music Processing Challenging?

Example: Chopin, Mazurka Op. 63 No. 3

- Waveform / Spectrogram
- Performance
 - Tempo
 - Dynamics
 - Note deviations
 - Sustain pedal

Why is Music Processing Challenging?

Example: Chopin, Mazurka Op. 63 No. 3

- Waveform / Spectrogram



- Performance

- Tempo
- Dynamics
- Note deviations
- Sustain pedal

A musical score for Chopin's Mazurka Op. 63 No. 3, showing two systems of music. The score is annotated with performance markings: blue highlights for the main melody, red highlights for an additional melody line, and yellow highlights for the accompaniment. The score includes various musical notations such as notes, rests, and dynamic markings like 'p' and 'f'. There are also asterisks and 'Ped' markings indicating pedal use.

- Polyphony



Main Melody



Additional melody line



Accompaniment

Source Separation

- Decomposition of audio stream into different sound sources
- Central task in digital signal processing
- “Cocktail party effect”

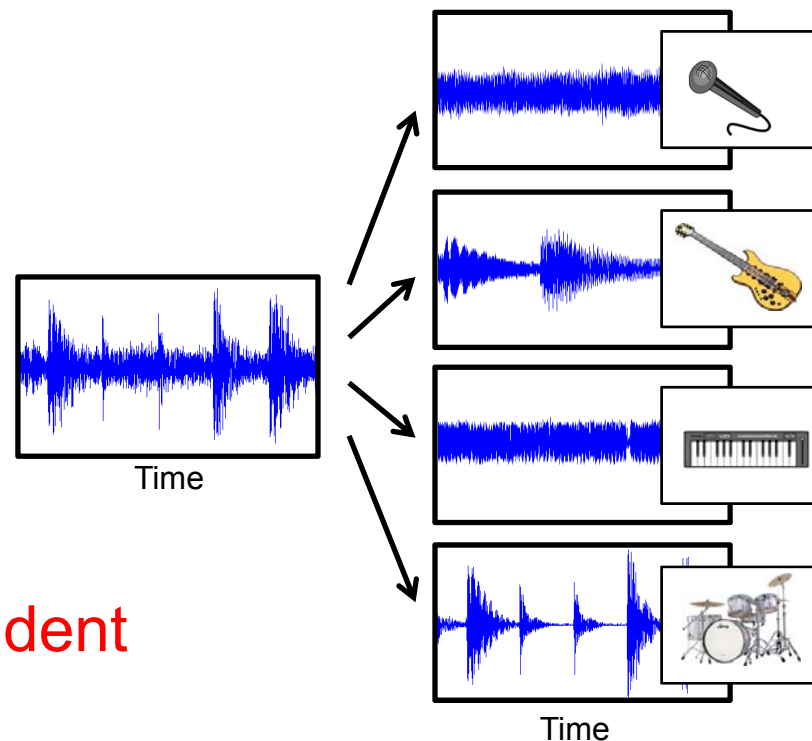


Source Separation

- Decomposition of audio stream into different sound sources
- Central task in digital signal processing
- “Cocktail party effect”
- Several input signals
- Sources are assumed to be statistically independent

Source Separation (Music)

- Main melody, accompaniment, drum track
- Instrumental voices
- Individual note events
- Only mono or stereo
- Sources are often highly dependent



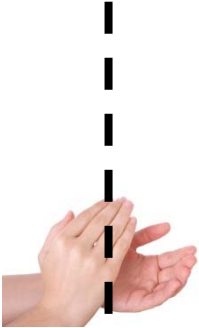
Harmonic-Percussive Decomposition



Harmonic-Percussive Decomposition



Clearly harmonic sounds



Clearly percussive sounds

Harmonic component

Percussive component

Harmonic-Percussive Decomposition



Clearly harmonic sounds



Harmonic component



Residual component

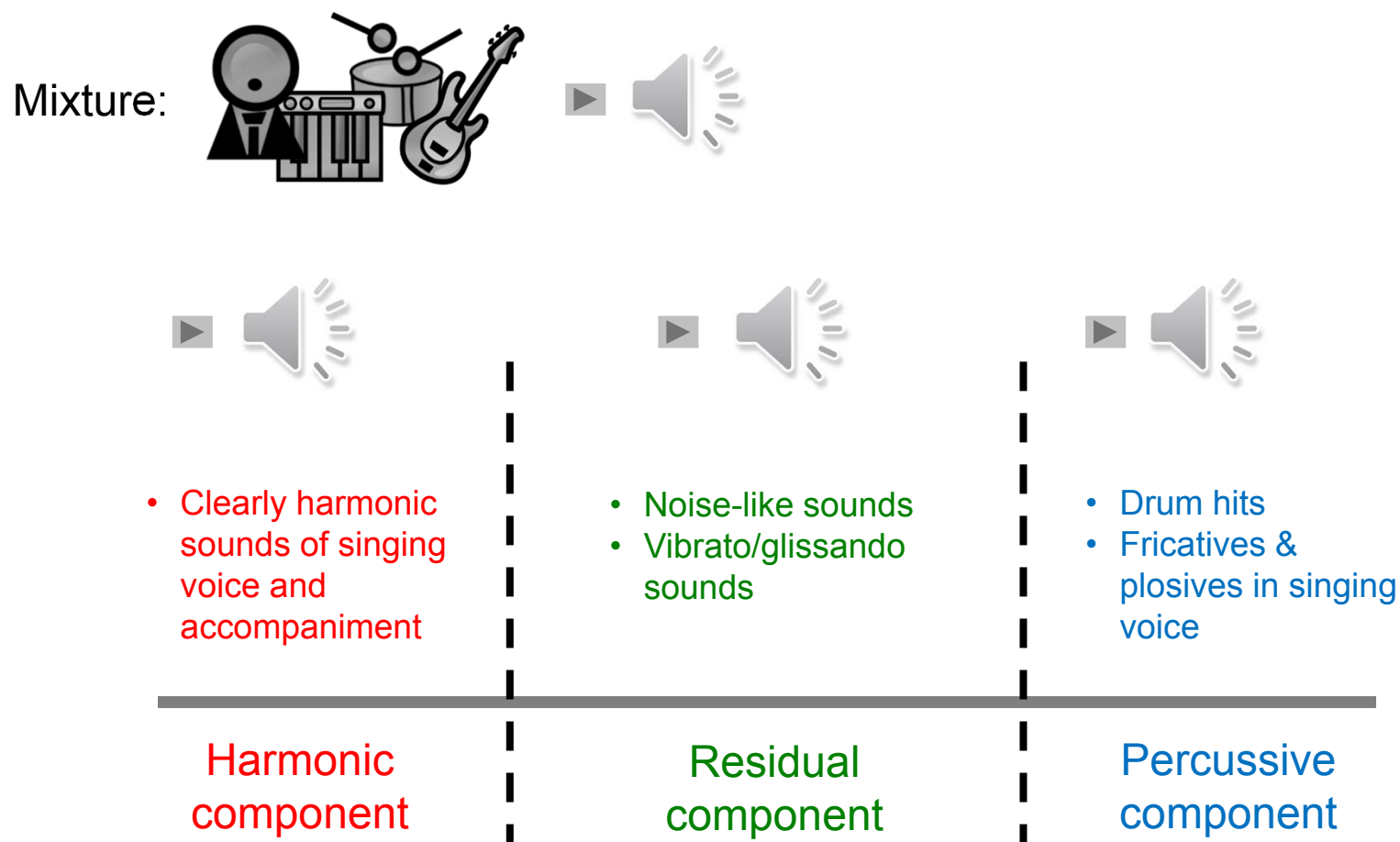


Clearly percussive sounds



Percussive component

Harmonic-Percussive Decomposition

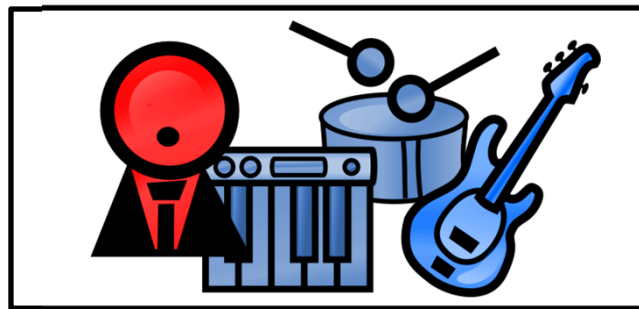


Literature: [Driedger/Müller/Disch, ISMIR 2014]

Demo: <https://www.audiolabs-erlangen.de/resources/2014-ISMIR-ExtHPSep/>

Singing Voice Extraction

Original Recording



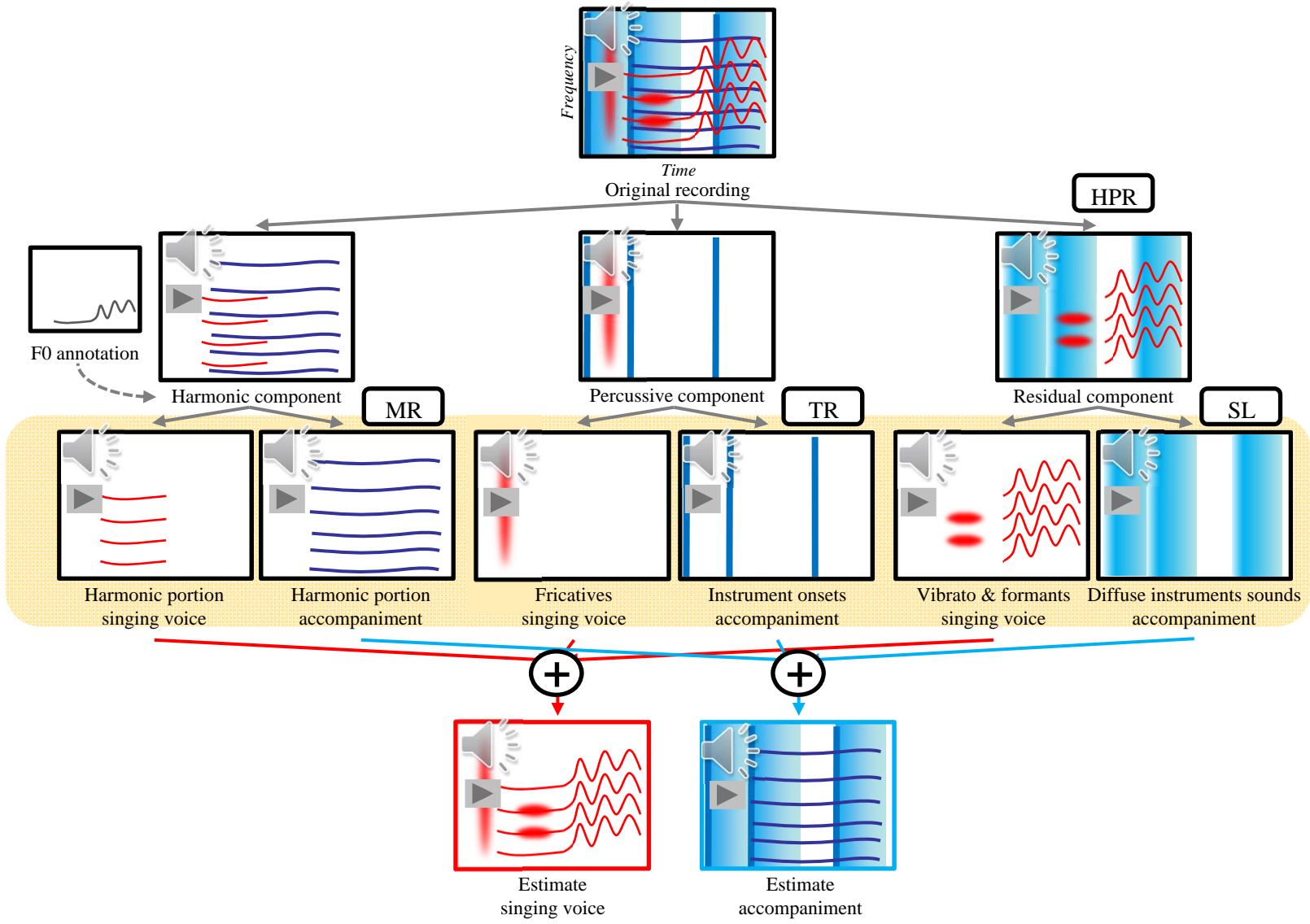
Singing voice



Accompaniment

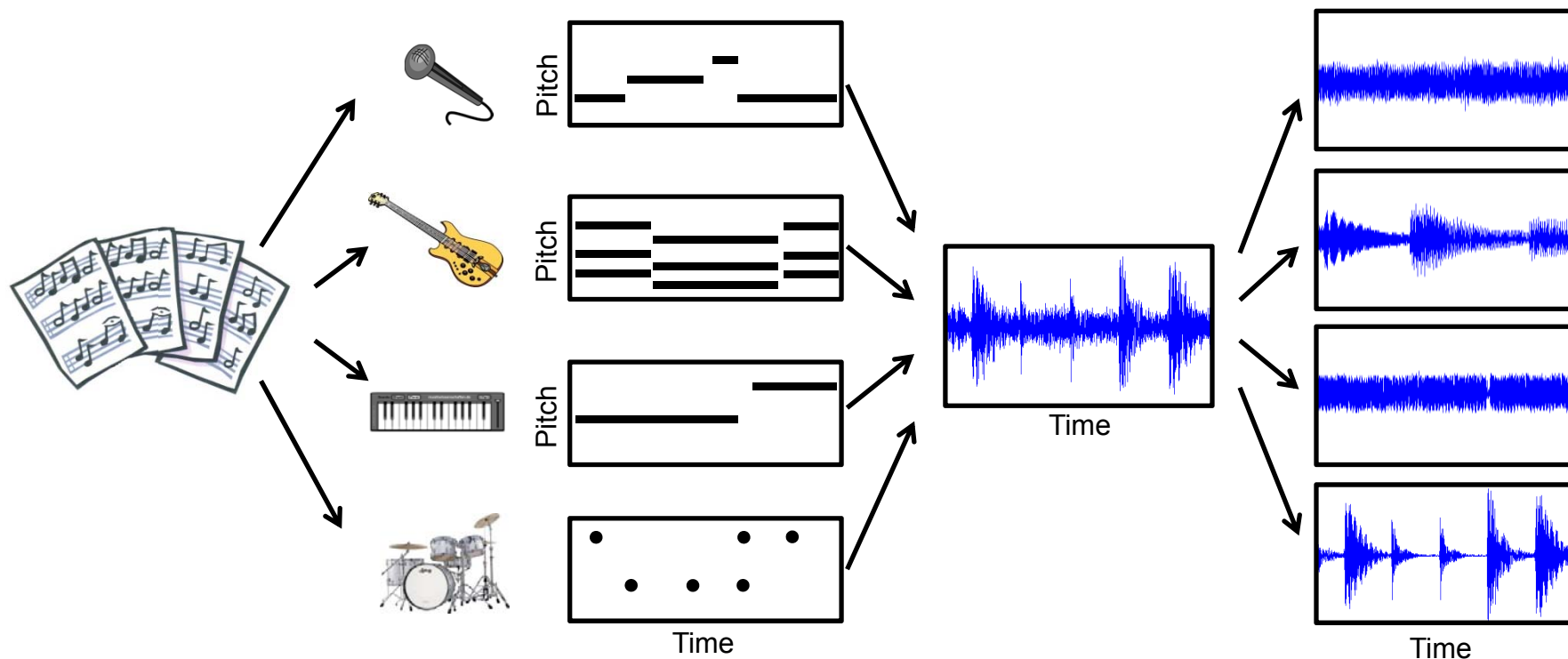


Singing Voice Extraction



Score-Informed Source Separation

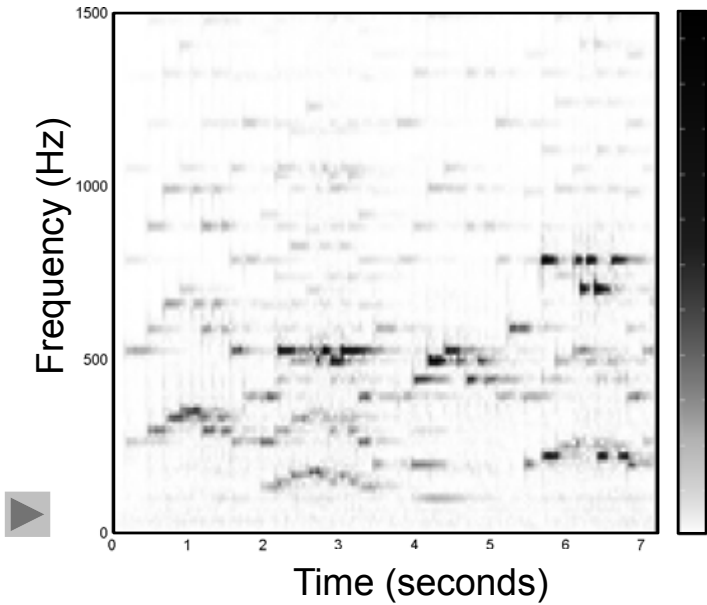
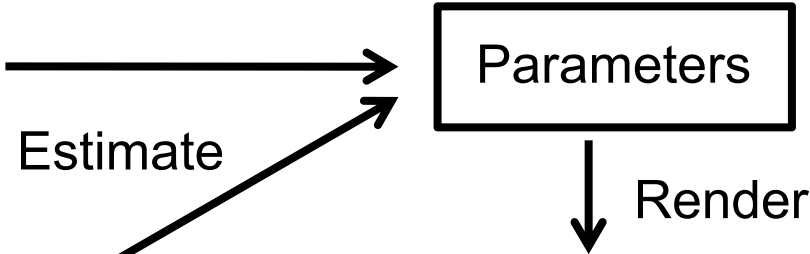
Exploit musical score to support separation process



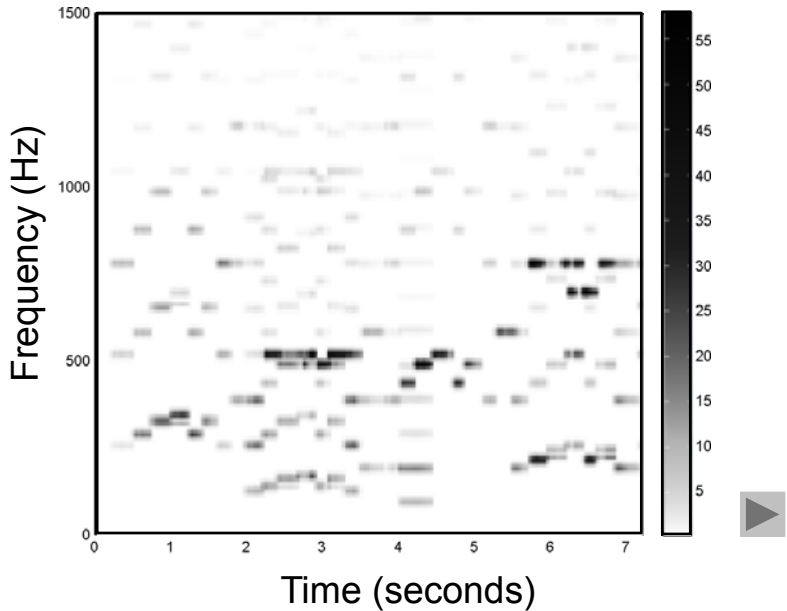
Parametric Model Approach

Rebuild spectrogram information

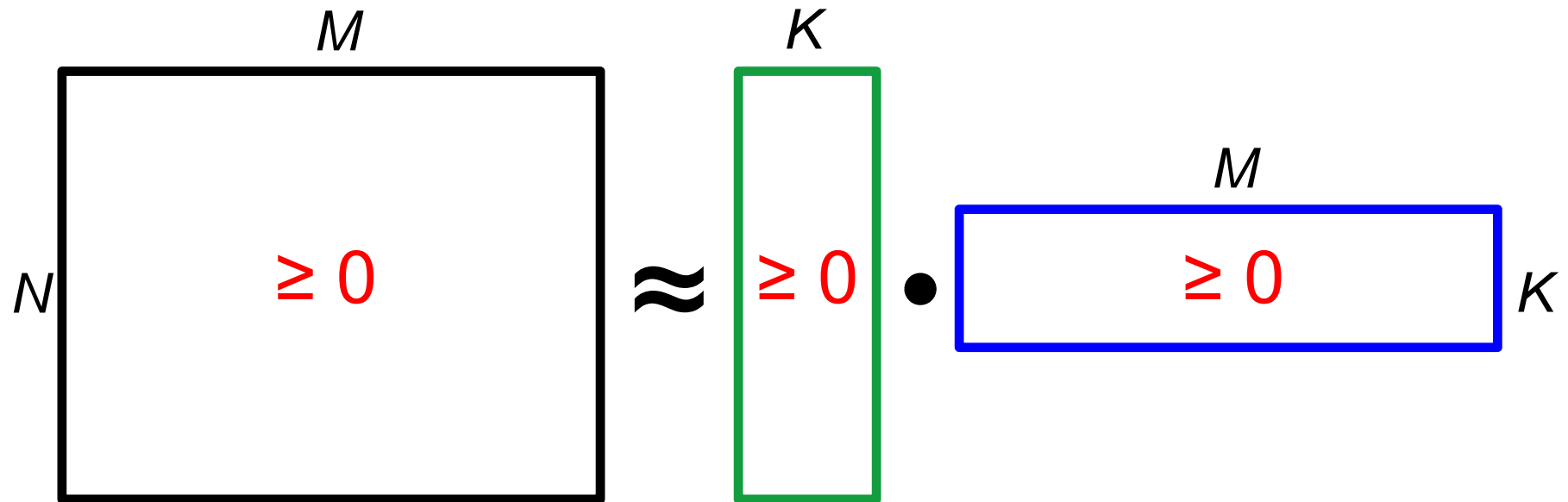
Invention 1
Johann Sebastian Bach (1685-1750)
BWV 772



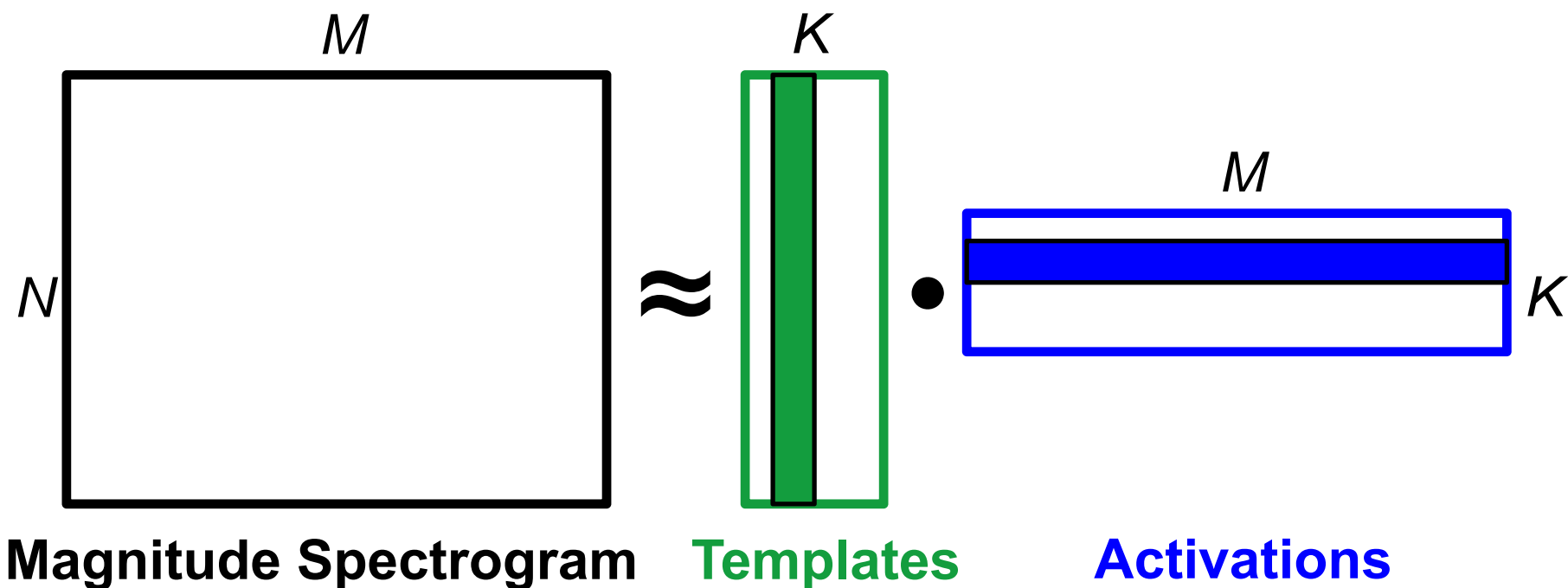
≈



NMF (Nonnegative Matrix Factorization)



NMF (Nonnegative Matrix Factorization)



Templates: Pitch + Timbre

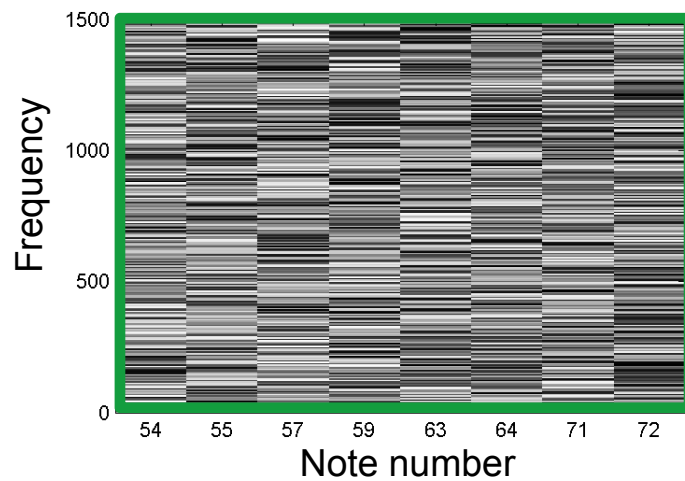
Activations: Onset time + Duration

“How does it sound”

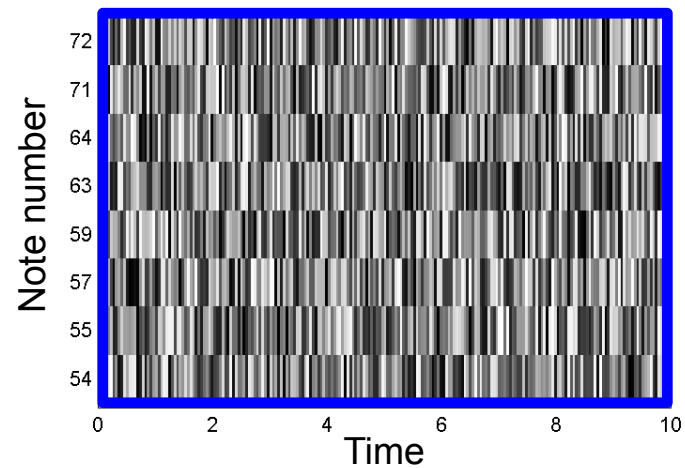
“When does it sound”

NMF-Decomposition

Initialized template



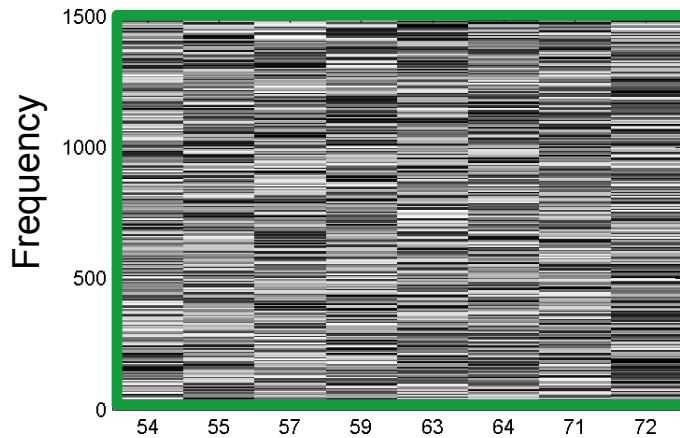
Initialized activations



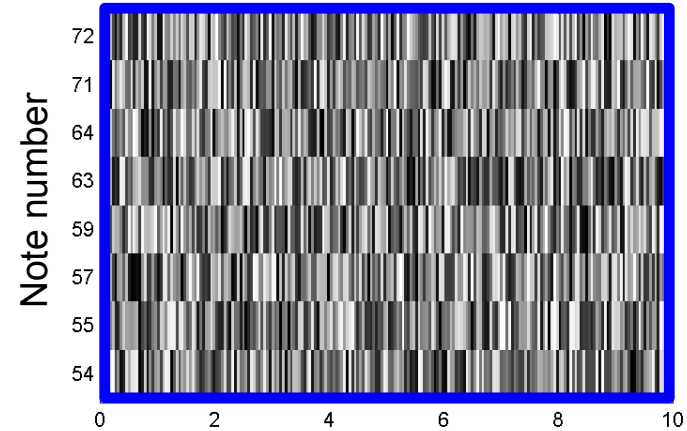
Random initialization

NMF-Decomposition

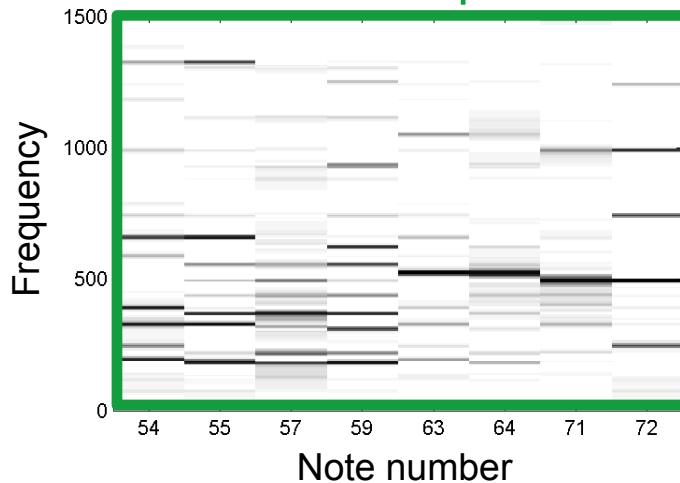
Initialized template



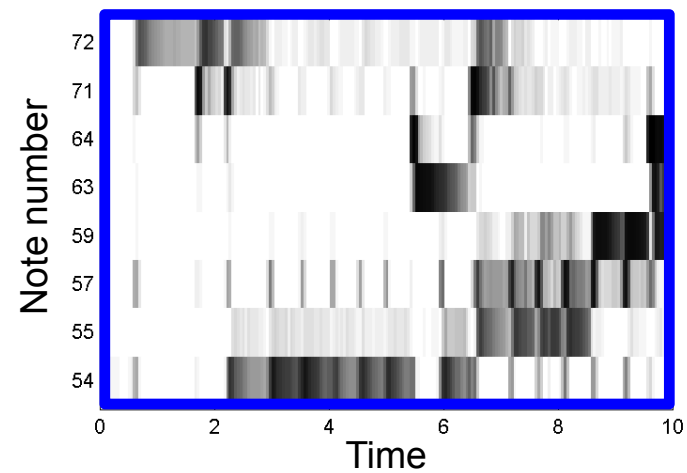
Initialized activations



Learnt templates



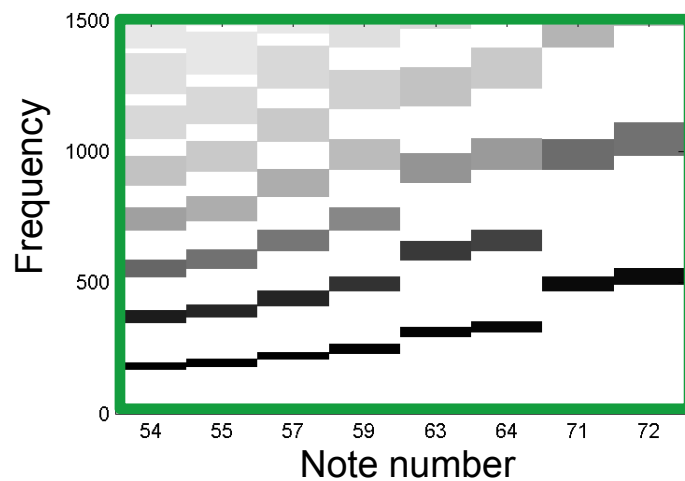
Learnt activations



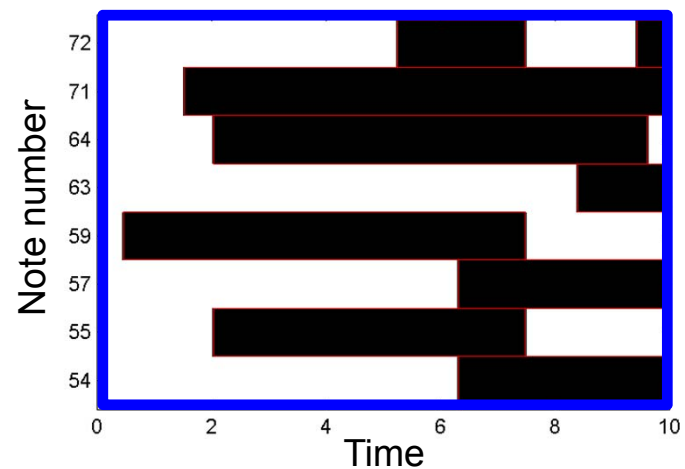
Random initialization → No semantic meaning

NMF-Decomposition

Initialized template



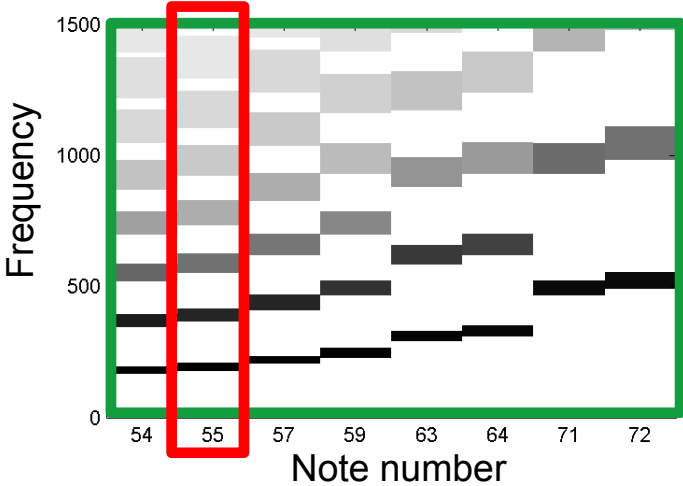
Initialized activations



Constrained initialization

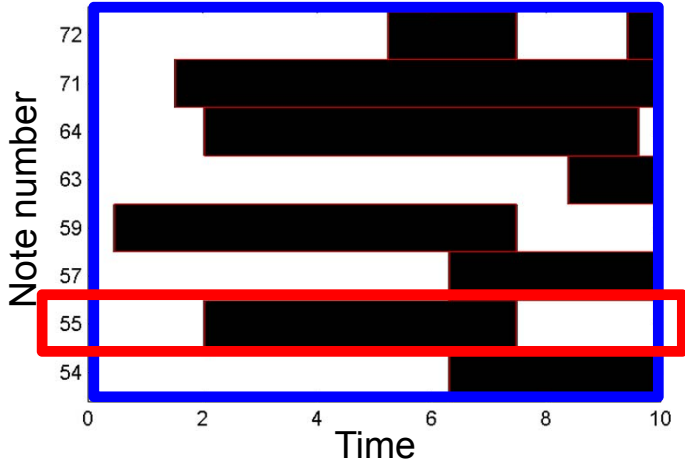
NMF-Decomposition

Initialized template



Template constraint for $p=55$

Initialized activations

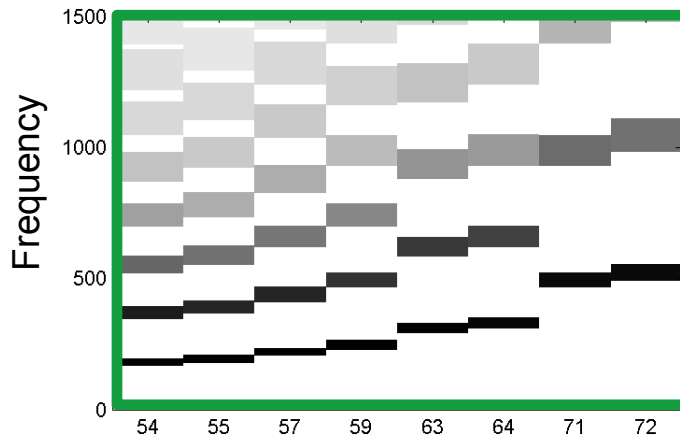


Activation constraints for $p=55$

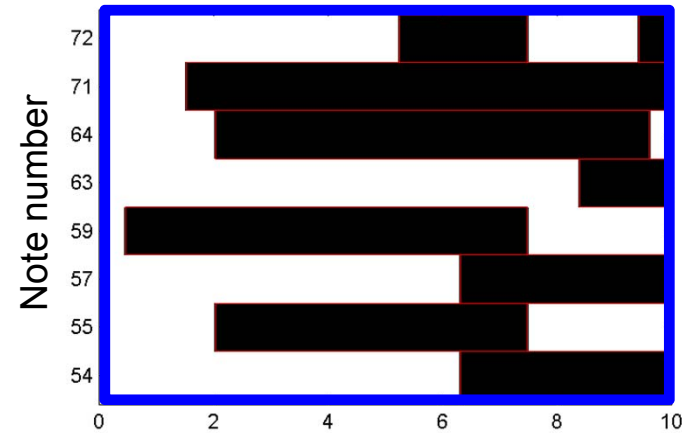
Constrained initialization

NMF-Decomposition

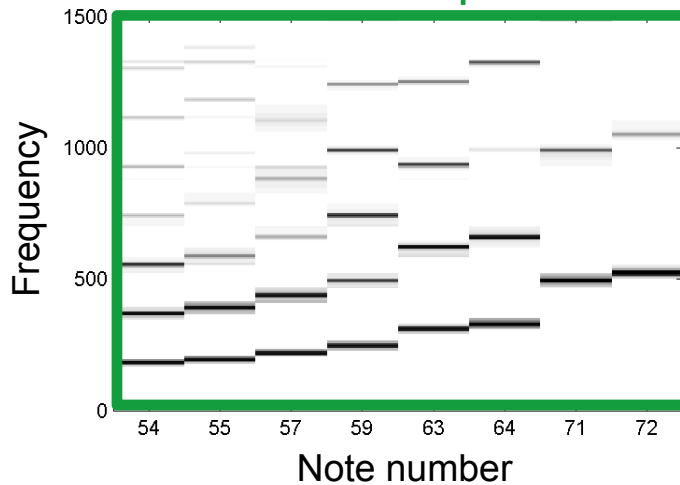
Initialized template



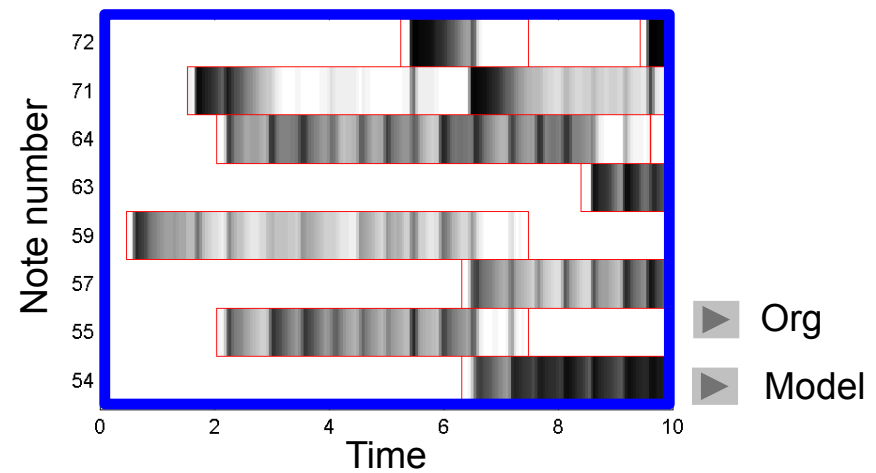
Initialized activations



Learnt templates



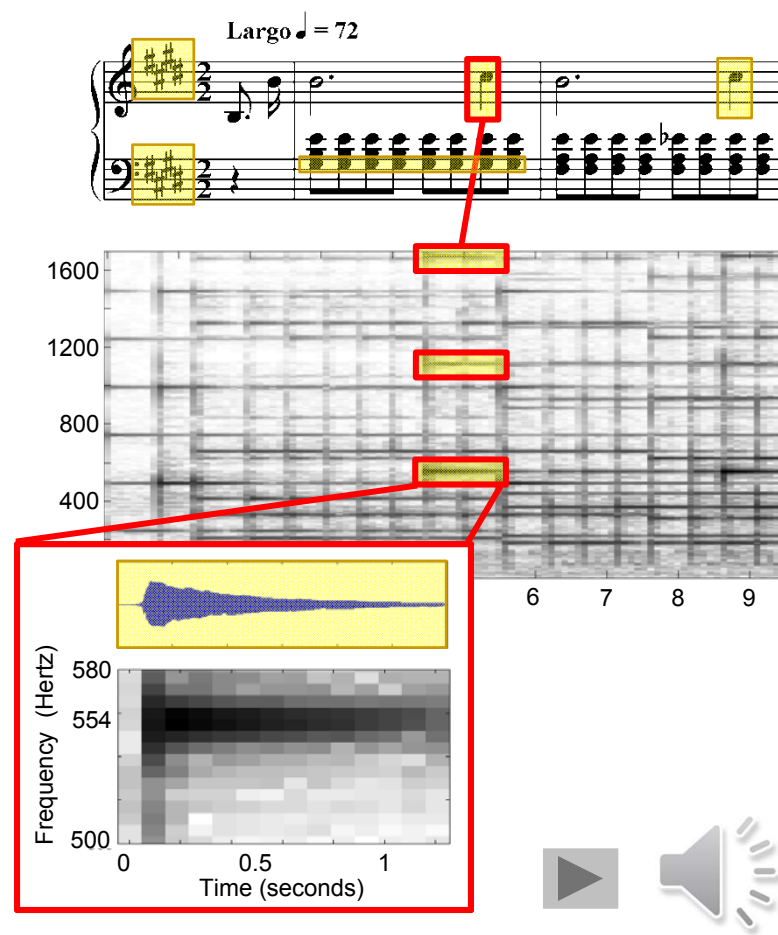
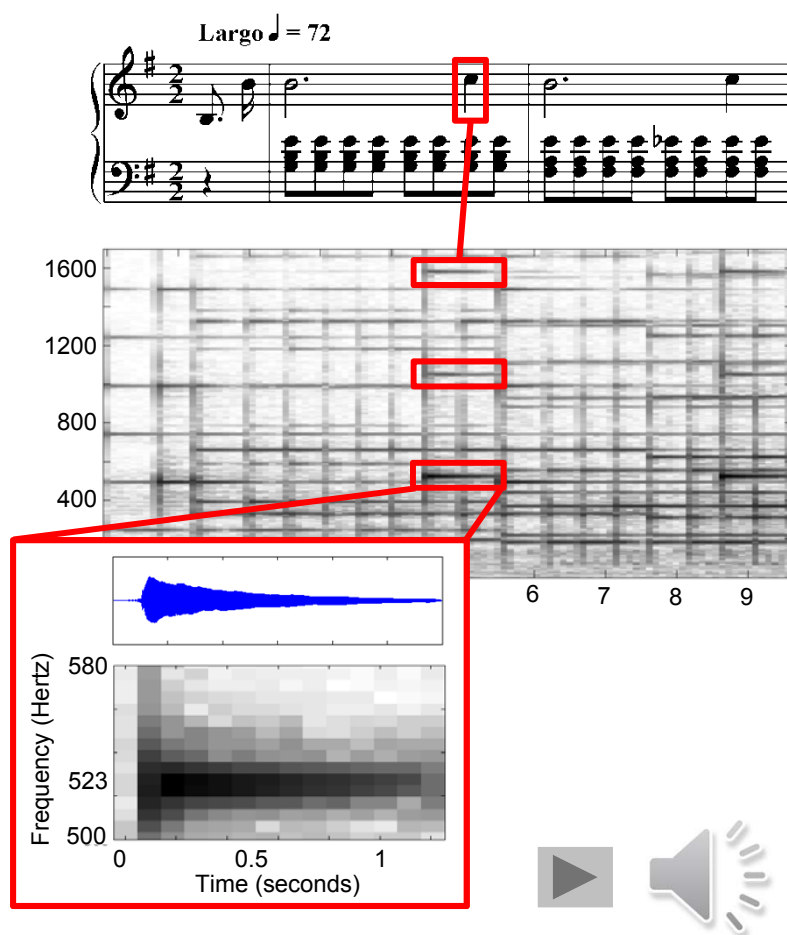
Learnt activations



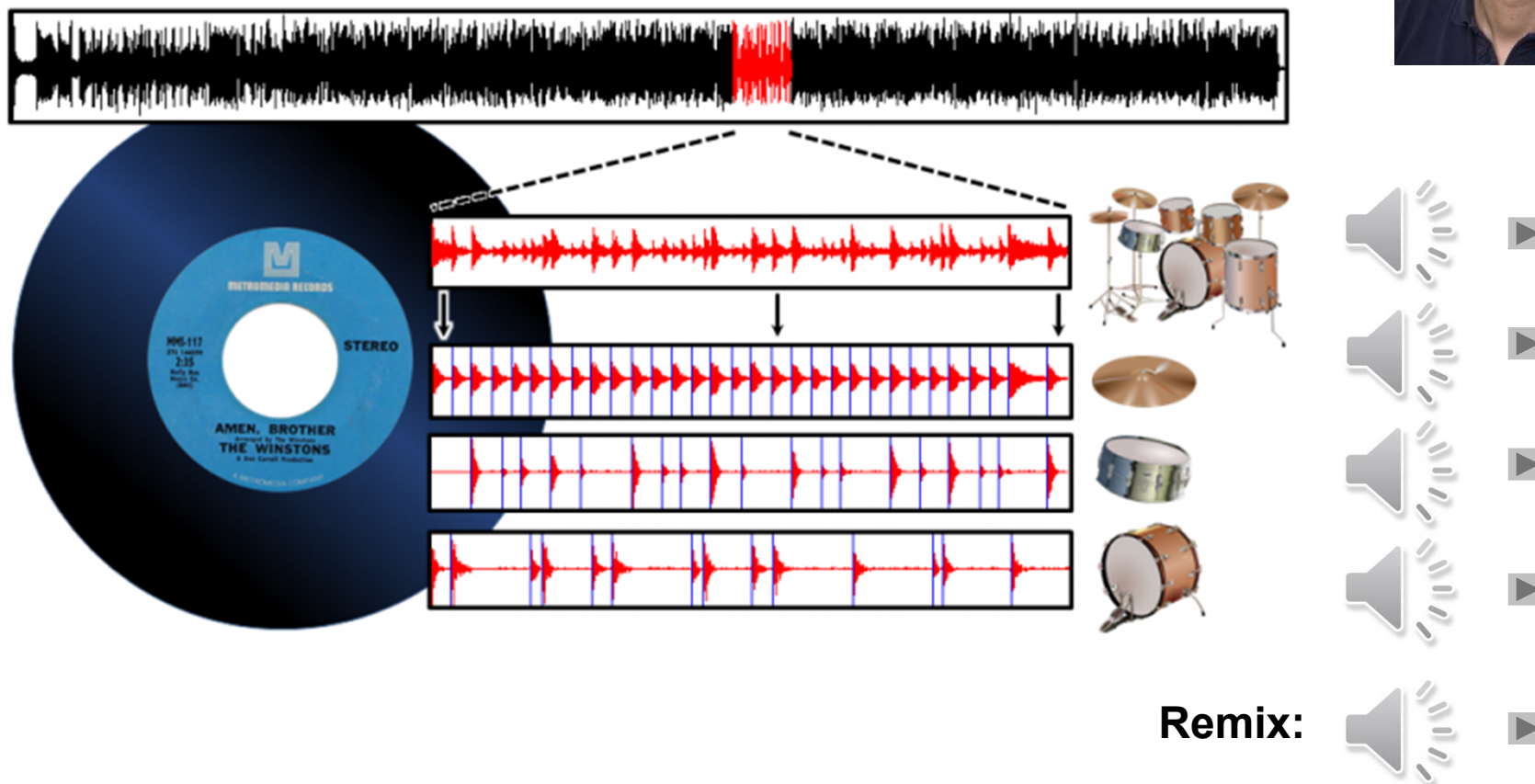
Constrained initialization → NMF as refinement

Score-Informed Audio Decomposition

Application: Audio editing



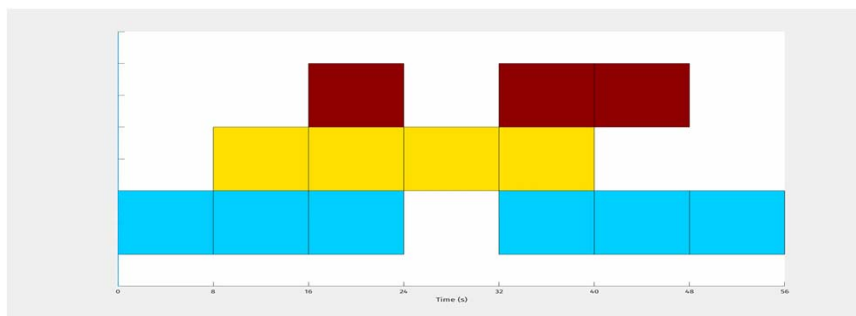
Informed Drum-Sound Decomposition



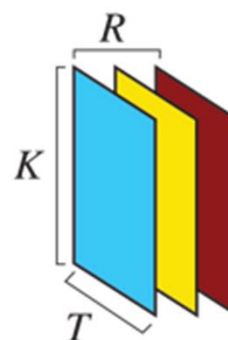
Literature: [Dittmar/Müller, IEEE/ACM-TASLP 2016]

Demo: <https://www.audiolabs-erlangen.de/resources/MIR/2016-IEEE-TASLP-DrumSeparation>

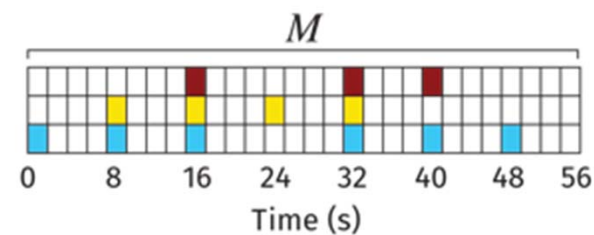
Loop Decomposition of EDM



Decomposition



Patterns



Activations

Literature: [López-Serrano/Dittmar/Müller, ISMIR 2016]

Demo: <https://www.audiolabs-erlangen.de/resources/MIR/2016-ISMIR-EMLoop>

Audio Mosaicing

Target signal: Beatles–Let it be



Source signal: Bees



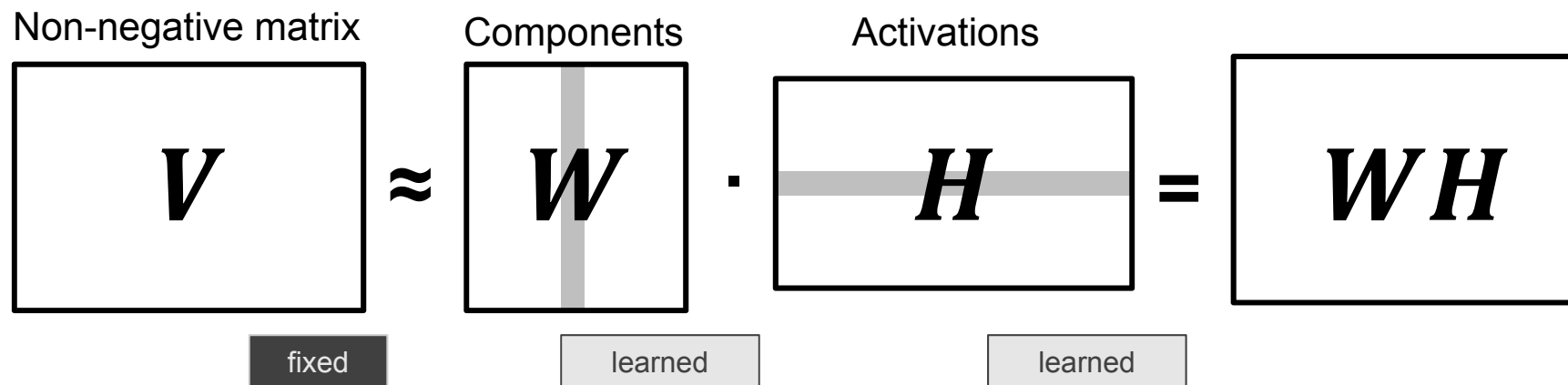
Mosaic signal: Let it Bee

Literature: [Driedger/Müller, ISMIR 2015]

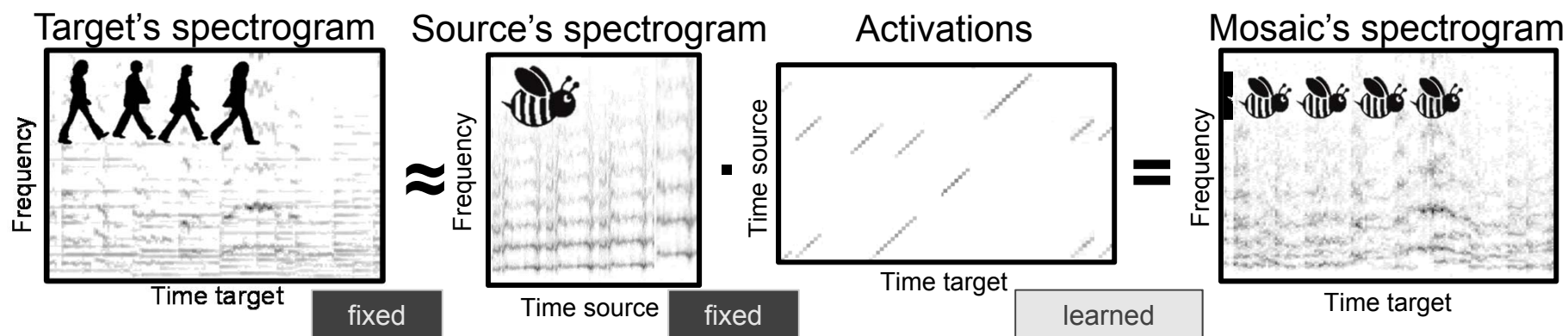
Demo: <https://www.audiolabs-erlangen.de/resources/MIR/2015-ISMIR-LetItBee>

NMF-Inspired Audio Mosaicing

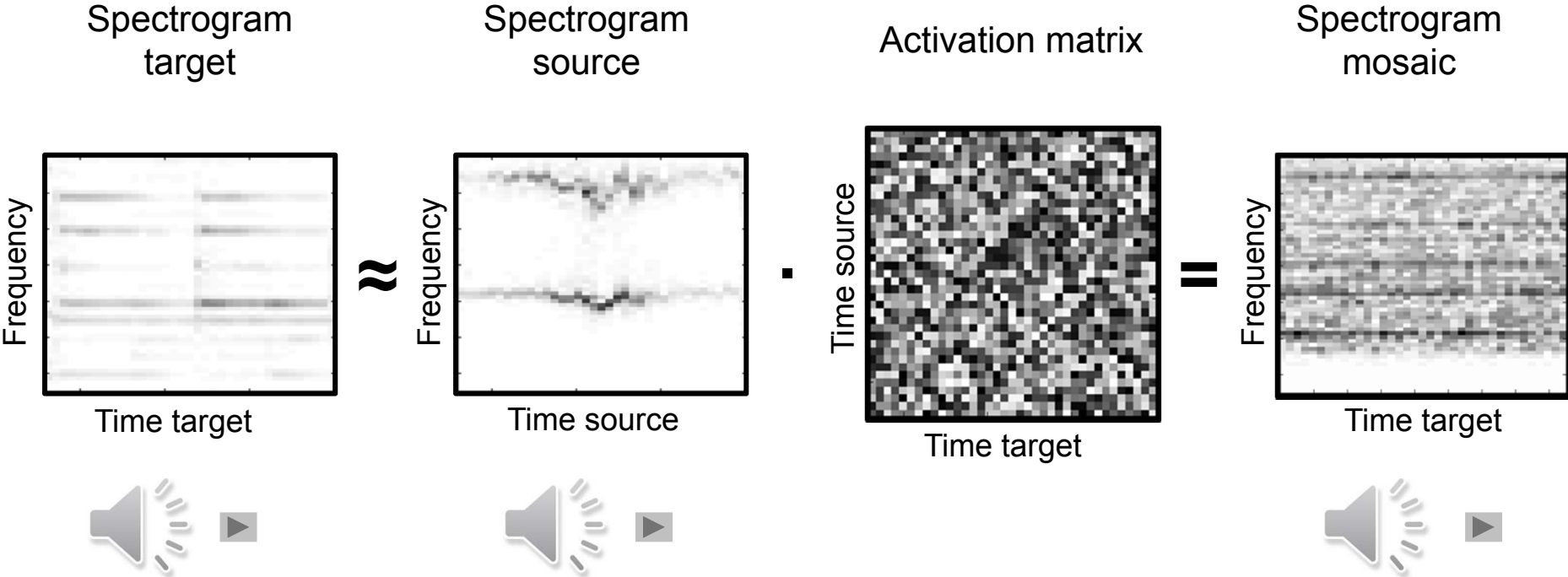
Non-negative matrix factorization (NMF)



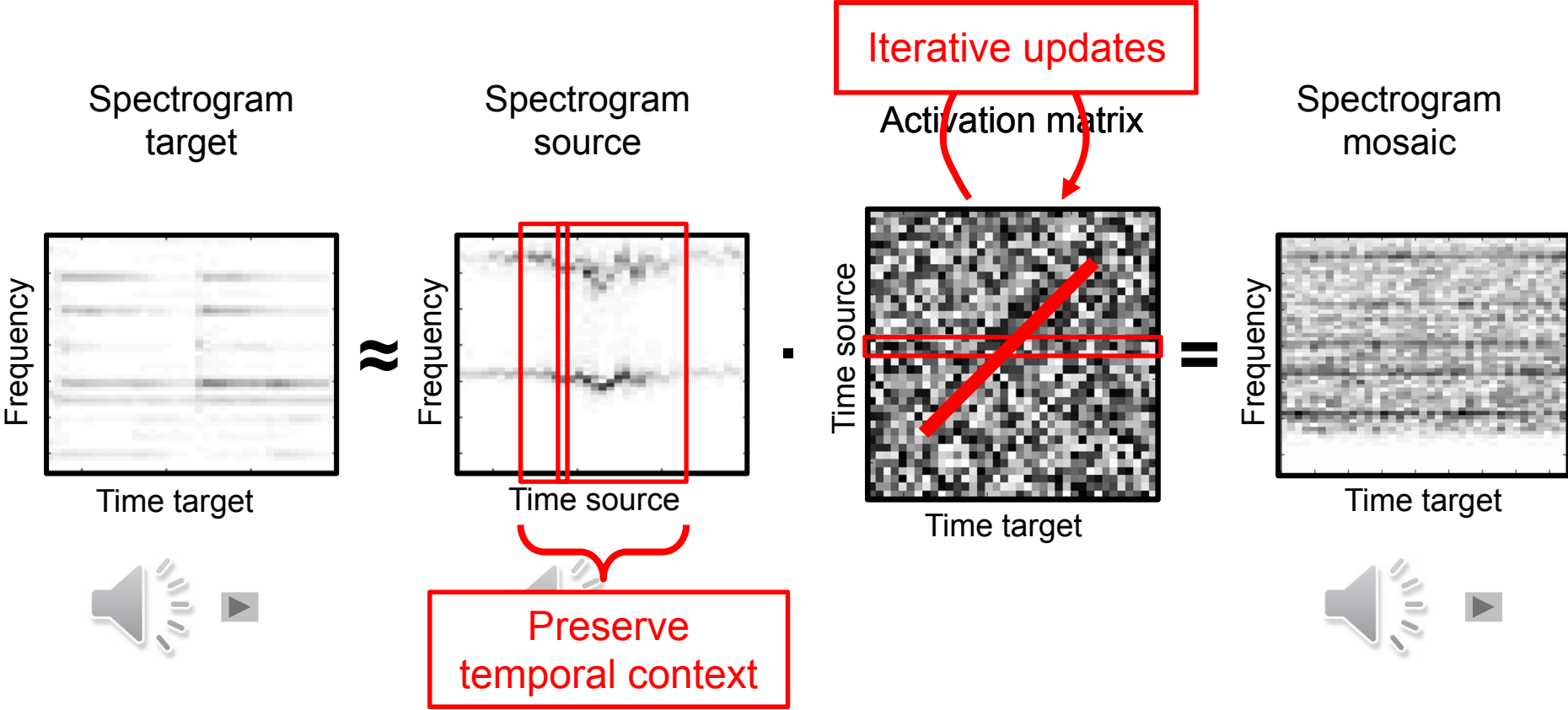
Proposed audio mosaicing approach



NMF-Inspired Audio Mosaicing

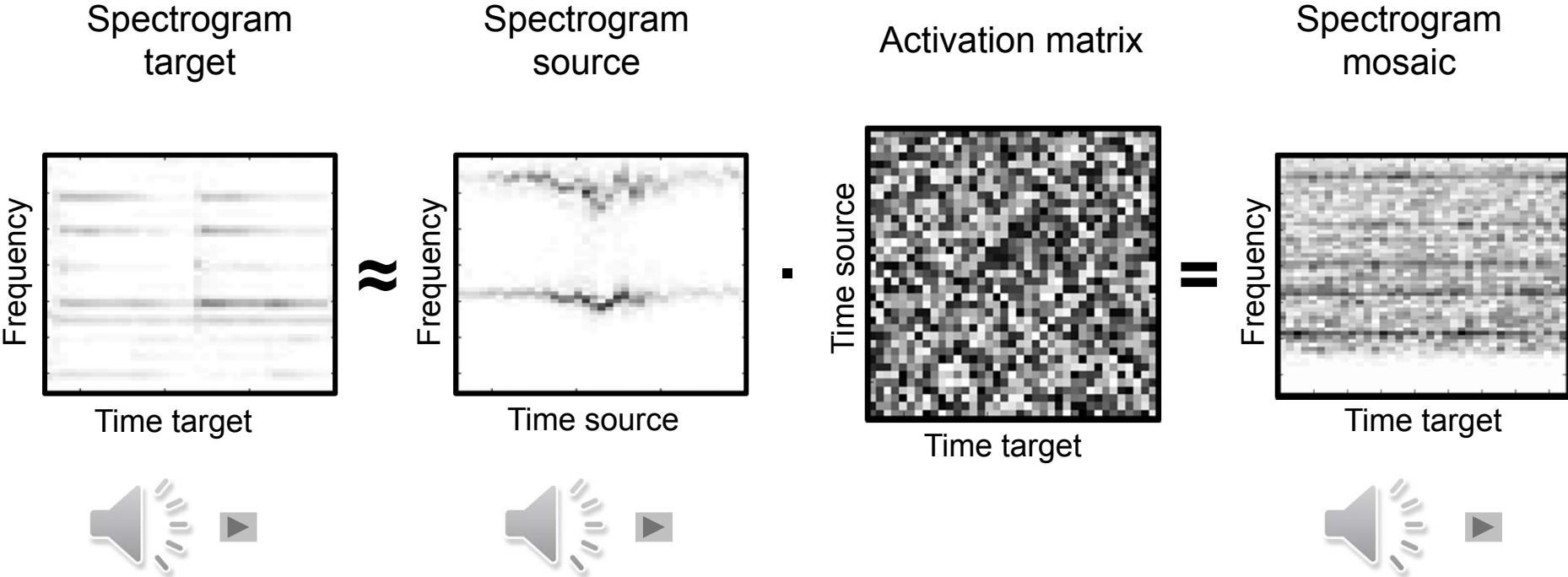


NMF-Inspired Audio Mosaicing

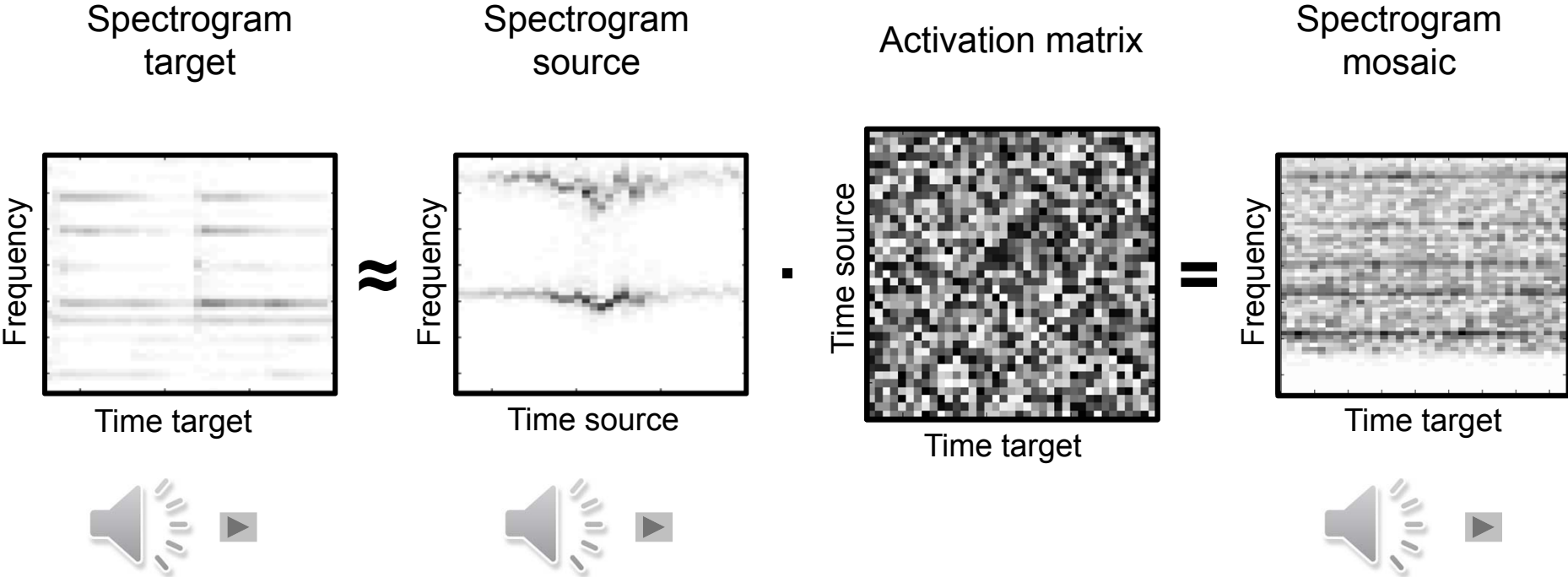


Core idea: support the development of sparse diagonal activation structures

NMF-Inspired Audio Mosaicing



NMF-Inspired Audio Mosaicing

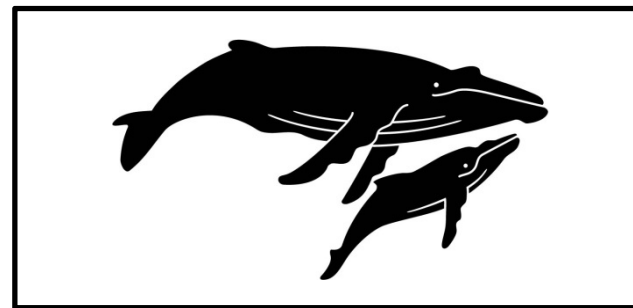


Audio Mosaicing

Target signal: Chic–Good times



Source signal: Whales



Mosaic signal

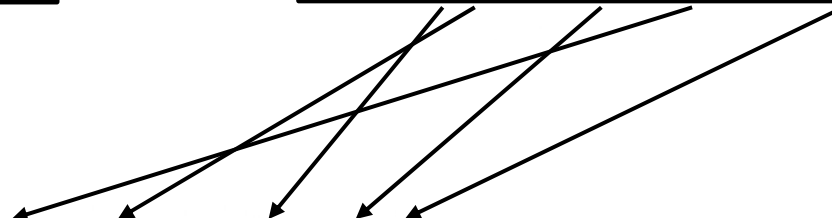
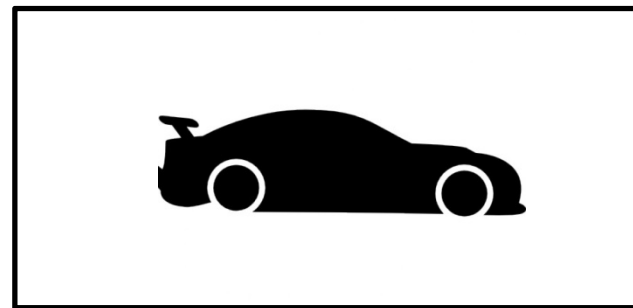


Audio Mosaicing

Target signal: Adele—Rolling in the Deep



Source signal: Race car



Mosaic signal

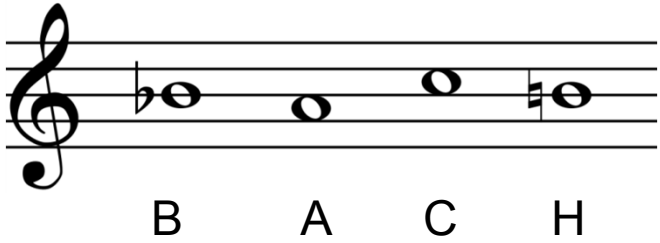
Motivic Similarity

Var. 4: Vivace

The musical score consists of four staves of music in bass clef, 2/4 time, and a key signature of two flats. The first staff begins with a forte (*f*) dynamic and contains six measures of eighth-note patterns. The second staff continues with similar eighth-note patterns and includes a triplet of eighth notes. The third staff features more complex eighth-note patterns with some slurs. The fourth staff shows a dynamic contrast, alternating between piano (*p*) and forte (*f*) dynamics, and includes slurs over groups of notes. The piece concludes with a double bar line.



Motivic Similarity



A musical staff in treble clef showing four notes: B (B-flat), A, C, and H (B-natural). Below the staff are the letters B, A, C, and H.



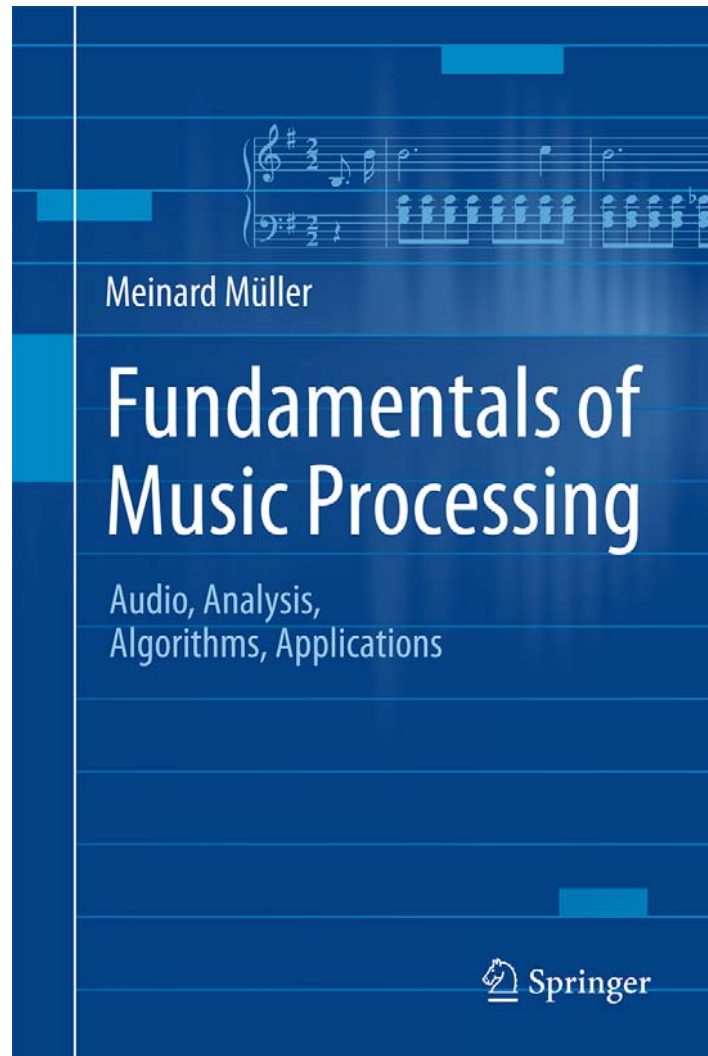
A musical score for four voices: Soprano (S), Alto (A), Tenor (T), and Bass (B). The score is in 4/4 time and G major. The lyrics are: "auf - - ge - rafft, und nie - mand ach - - tet und nie - mand ach - - tet drauf". A red box highlights a four-note melodic phrase in the Alto part, with the notes labeled 'b', 'a', 'c', and 'h' in red.



Summary

- Music information retrieval
- Audio decomposition techniques
- Machine learning
- Music applications & musicology
- Multimedia scenarios
- Web-based interfaces
- Teaching
- Academic training of students
- Fundamental research

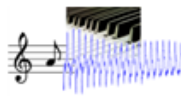

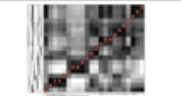


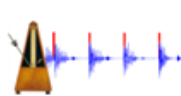
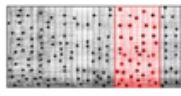
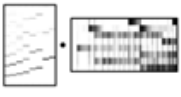
Book: Fundamentals of Music Processing



Meinard Müller
Fundamentals of Music Processing
Audio, Analysis, Algorithms, Applications
483 p., 249 illus., hardcover
ISBN: 978-3-319-21944-8
Springer, 2015

Accompanying website:
www.music-processing.de

Book: Fundamentals of Music Processing

Chapter		Music Processing Scenario
1		Music Representations
2		Fourier Analysis of Signals
3		Music Synchronization
4		Music Structure Analysis
5		Chord Recognition
6		Tempo and Beat Tracking
7		Content-Based Audio Retrieval
8		Musically Informed Audio Decomposition

Meinard Müller

Fundamentals of Music Processing

Audio, Analysis, Algorithms, Applications

483 p., 249 illus., hardcover

ISBN: 978-3-319-21944-8

Springer, 2015

Accompanying website:

www.music-processing.de

MIR-Related Events in Germany

AES Conference on
Semantic Audio
22 – 24 June 2017
Erlangen



GI Jahrestagung
25 – 29 September 2017
Chemnitz

- Workshop: Musik trifft Informatik
26 September 2017
- Tutorial: Musikverarbeitung
25 September 2017

Gesellschaft
für Informatik

