



ISMIR
2017, SUZHOU, CHINA

Tutorial T3
**A Basic Introduction to Audio-Related
Music Information Retrieval**

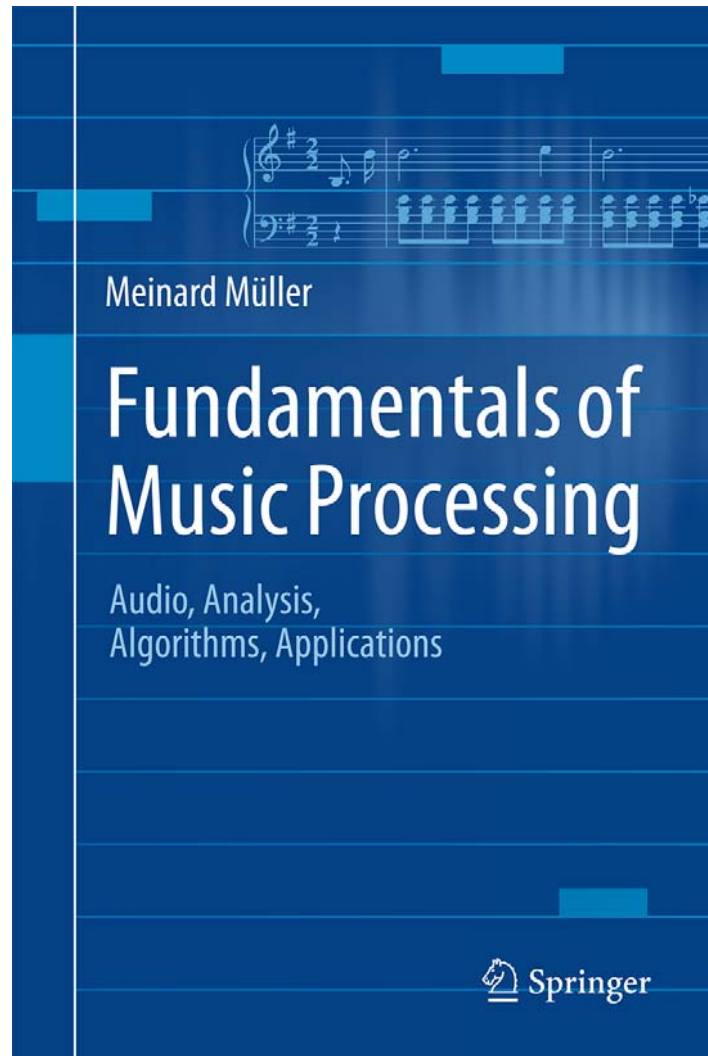
Audio Decomposition

Meinard Müller, Christof Weiß

International Audio Laboratories Erlangen

meinard.mueller@audiolabs-erlangen.de, christof.weiss@audiolabs-erlangen.de

Book: Fundamentals of Music Processing



Meinard Müller

Fundamentals of Music Processing

Audio, Analysis, Algorithms, Applications

483 p., 249 illus., hardcover

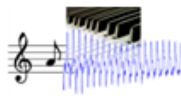

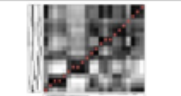


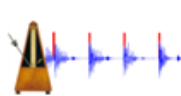
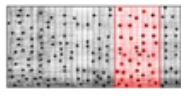
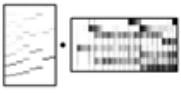
ISBN: 978-3-319-21944-8

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

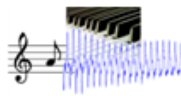

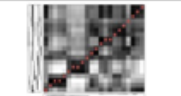
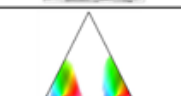

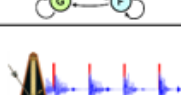


ISBN: 978-3-319-21944-8

Springer, 2015

Accompanying website:

www.music-processing.de

Book: Fundamentals of Music Processing

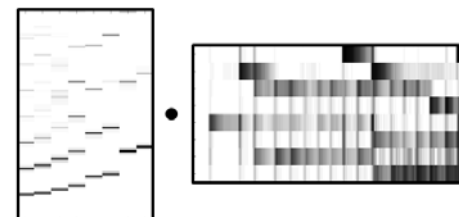
Chapter		Music Processing Scenario
1		Music Representations
2		Fourier Analysis of Signals
3		Music Synchronization
4		Music Structure Analysis
5		Chord Recognition
6		Tempo and Beat Tracking
7		Content-Based Audio Retrieval
8		Musically Informed Audio Decomposition

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

Chapter 8: Audio Decomposition

- 8.1 Harmonic-Percussive Separation
- 8.2 Melody Extraction
- 8.3 NMF-Based Audio Decomposition
- 8.4 Further Notes



In the final Chapter 8 on audio decomposition, we present a challenging research direction that is closely related to source separation. Within this wide research area, we consider three subproblems: harmonic–percussive separation, main melody extraction, and score-informed audio decomposition. Within these scenarios, we discuss a number of key techniques including instantaneous frequency estimation, fundamental frequency (F0) estimation, spectrogram inversion, and nonnegative matrix factorization (NMF). Furthermore, we encounter a number of acoustic and musical properties of audio recordings that have been introduced and discussed in previous chapters, which rounds off the book.

Why is Music Processing Challenging?

Example: Chopin, Mazurka Op. 63 No. 3



Mazurka.

F. CHOPIN. Op. 63, № 3.

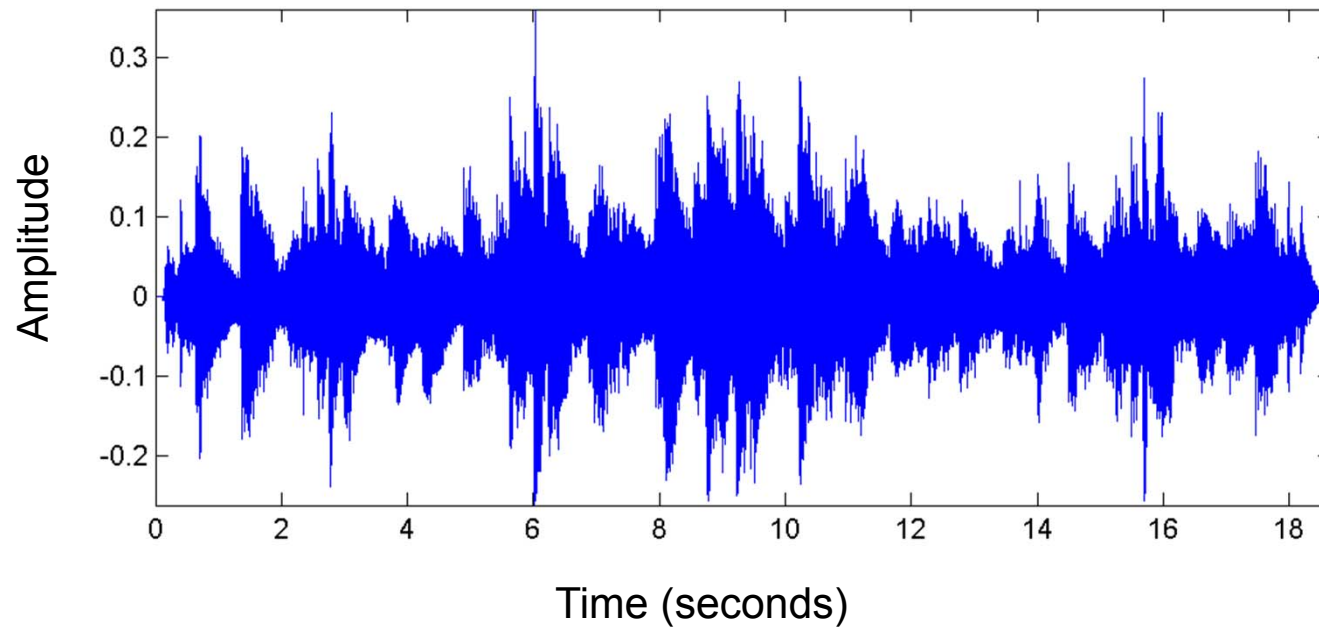
41. Allegretto. *p*

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 melody starting on G4, marked 'Allegretto' and 'p'. The bass clef accompaniment consists of chords and single notes. The second system (measures 47-50) continues the melody and accompaniment. Fingerings are indicated by numbers 1-4. The key signature is three sharps (F#, C#, G#) and the time signature is 3/4. The notation includes slurs, accents, and dynamic markings.

Why is Music Processing Challenging?

Example: Chopin, Mazurka Op. 63 No. 3

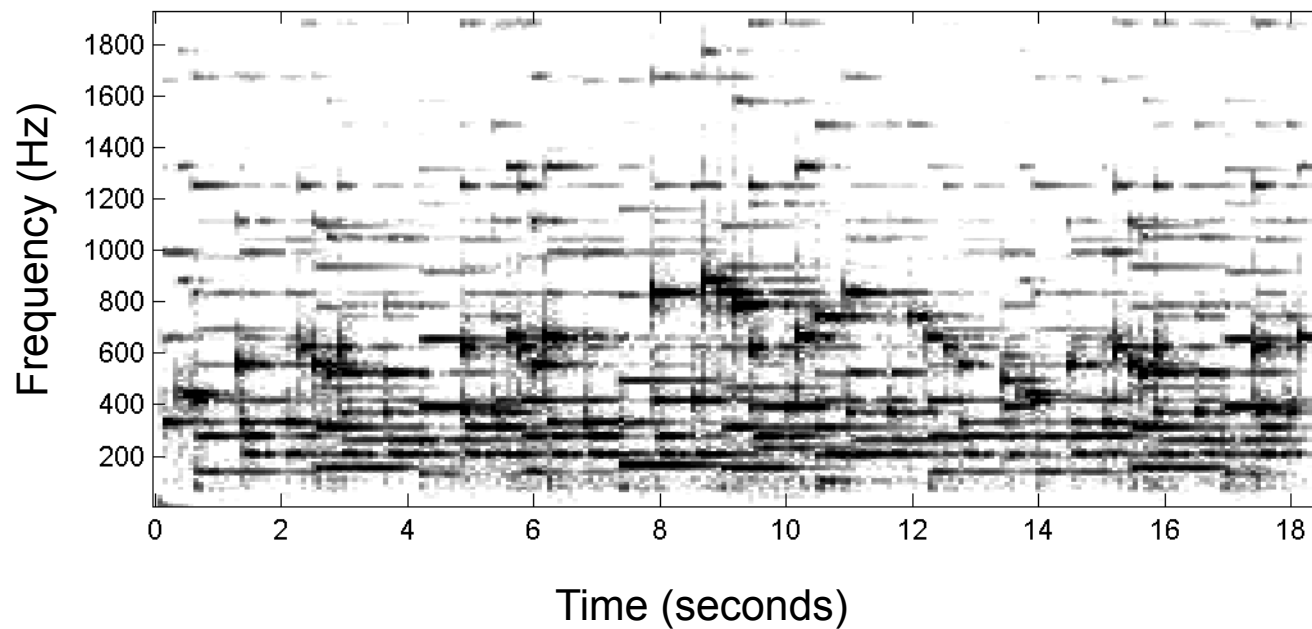
- Waveform



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. Fingerings (1-5) and dynamics (p, f) are also indicated. The score includes a treble and bass clef, a key signature of one sharp (F#), and a 3/4 time signature. The first system ends with a repeat sign and a fermata. The second system begins with a forte (f) dynamic and ends with a fermata.

- 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”

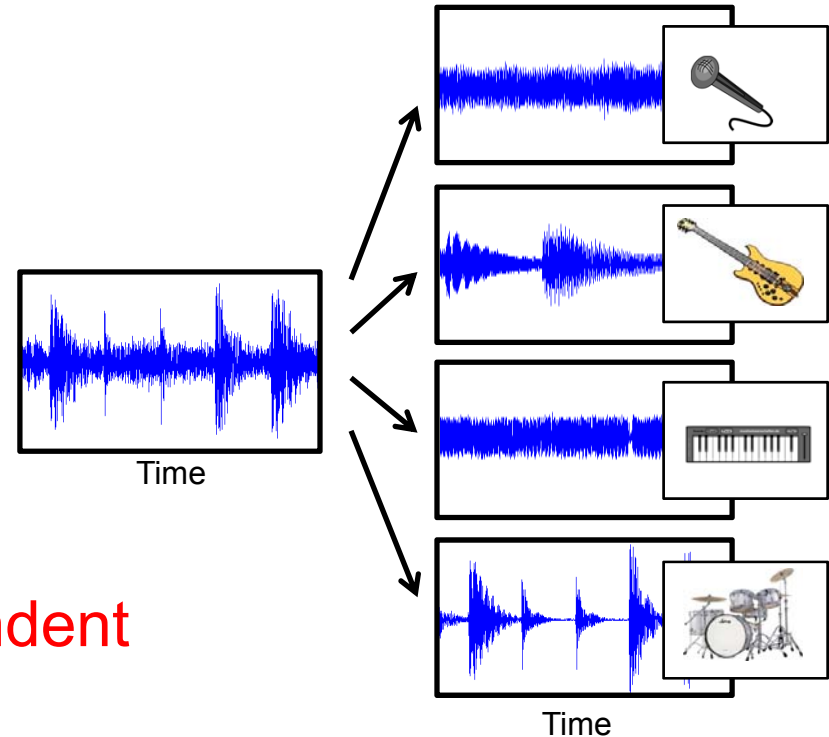


Source Separation

- 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

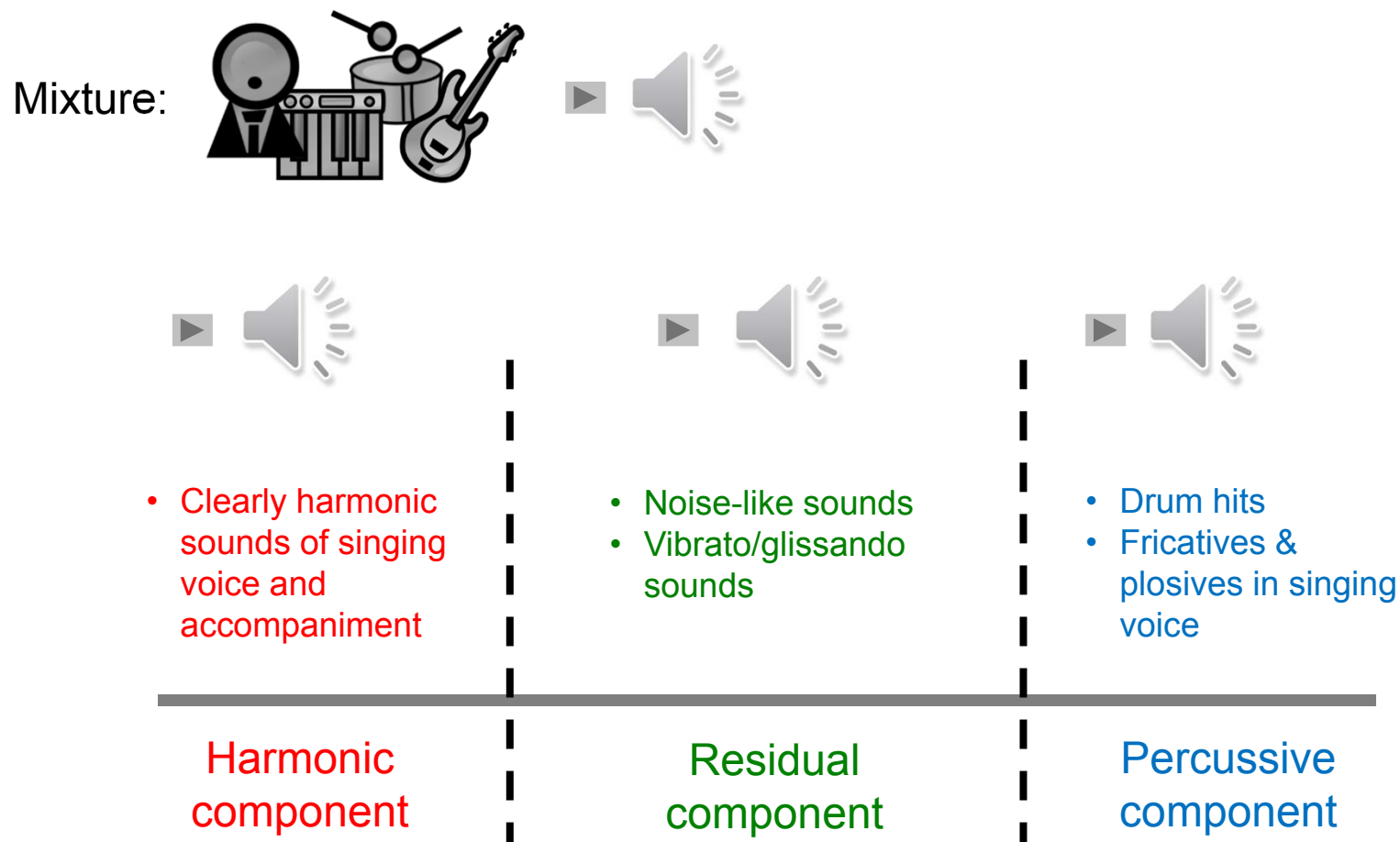


Clearly percussive sounds



Percussive component

Harmonic-Percussive Decomposition

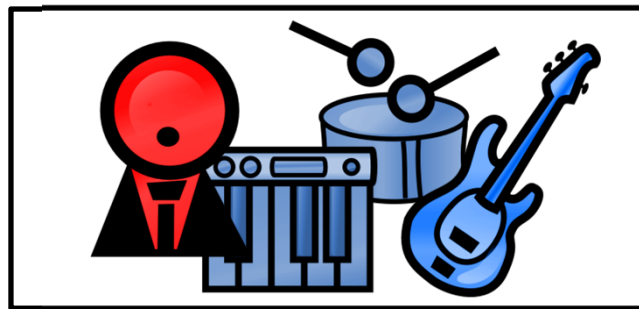


Literature: [Driedger/Müller/Disch, ISMIR 2014]

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

Singing Voice Extraction

Original Recording

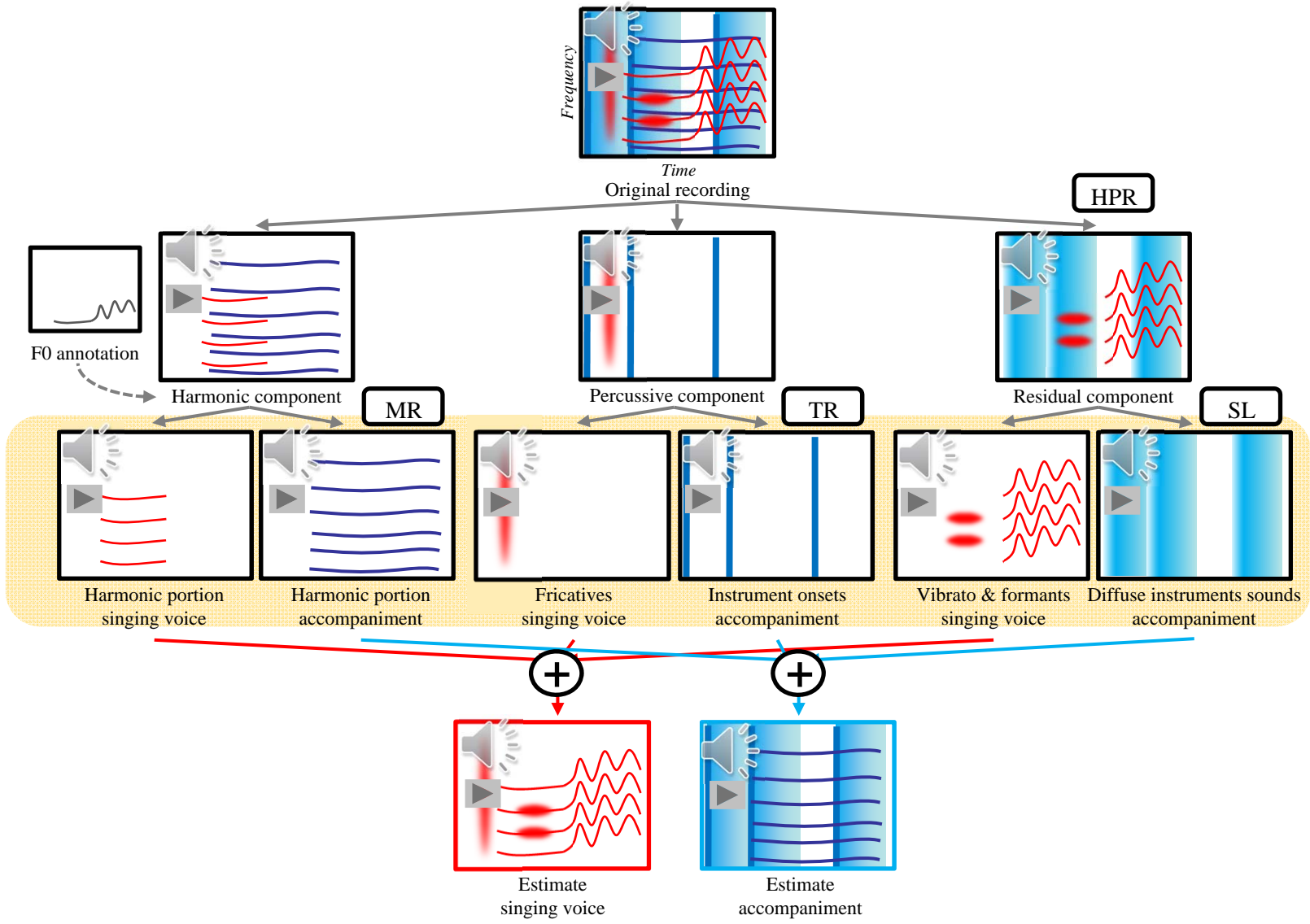


Singing voice



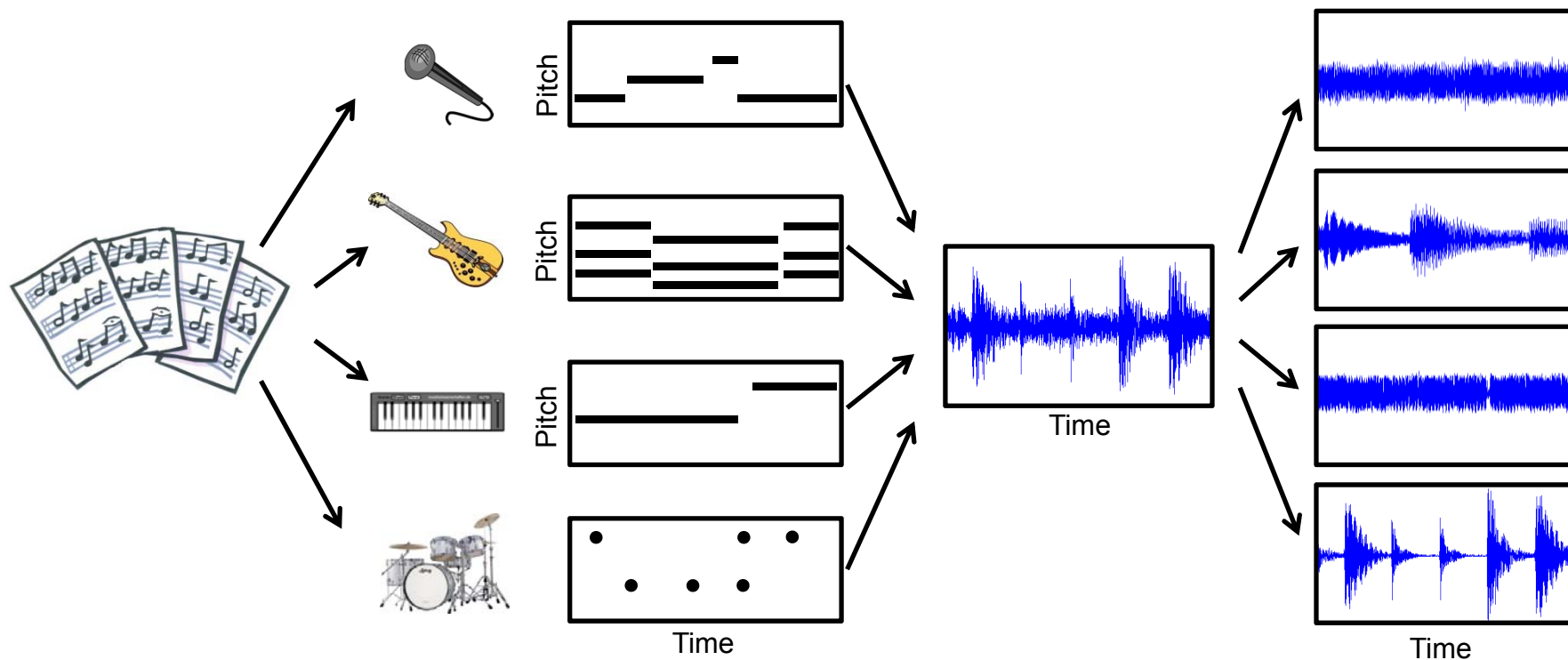
Accompaniment

Singing Voice Extraction



Score-Informed Source Separation

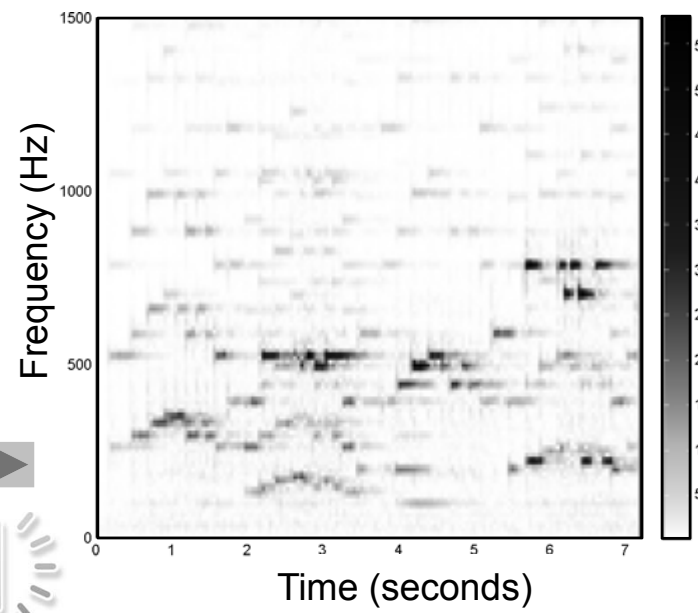
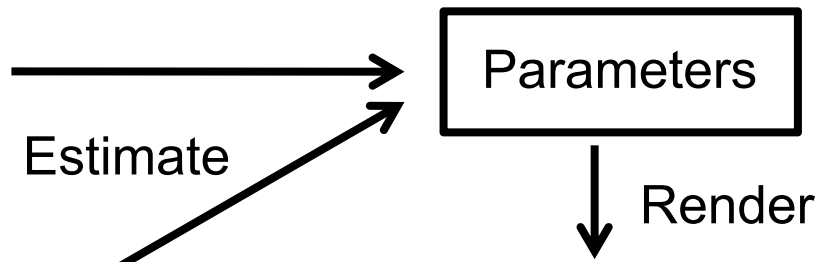

Exploit musical score to support separation process



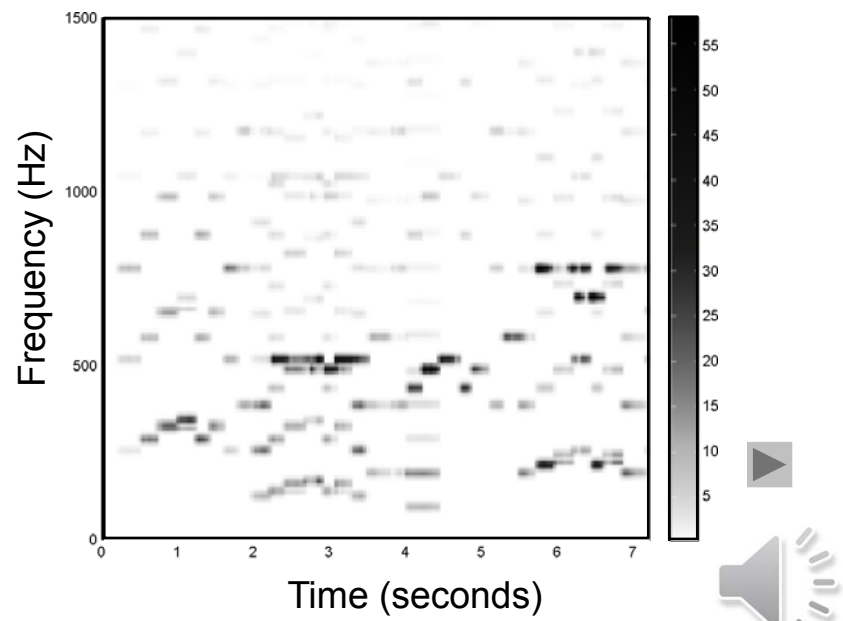
Parametric Model Approach

Rebuild spectrogram information

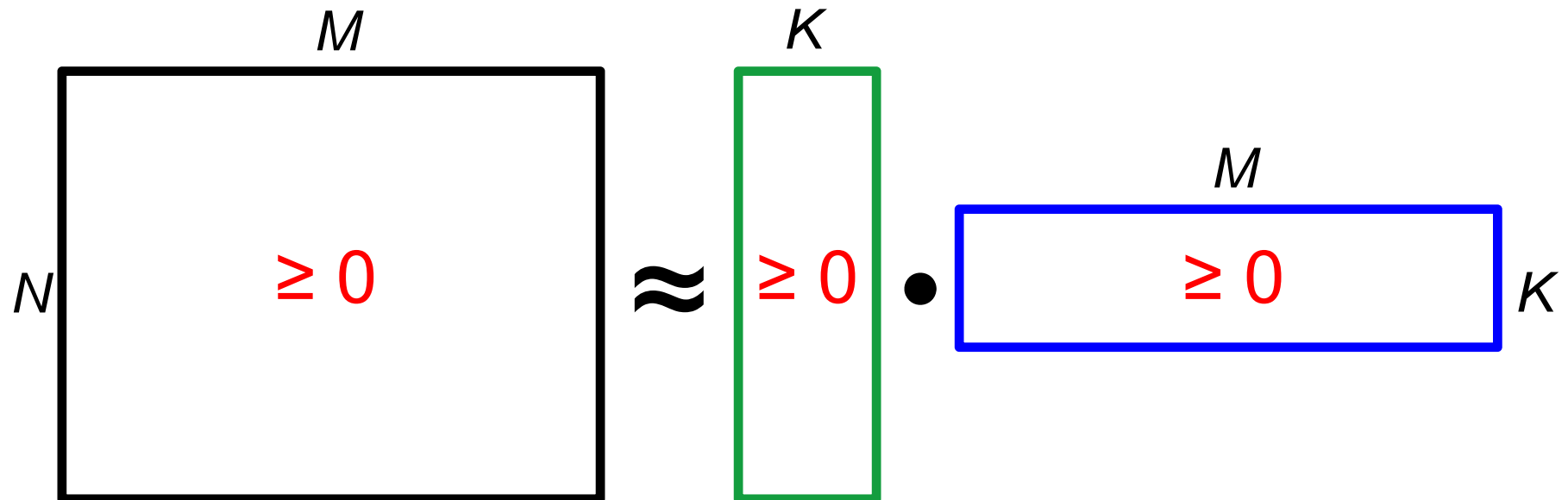
Invention 1
Johann Sebastian Bach (1685-1750)
BWV 772



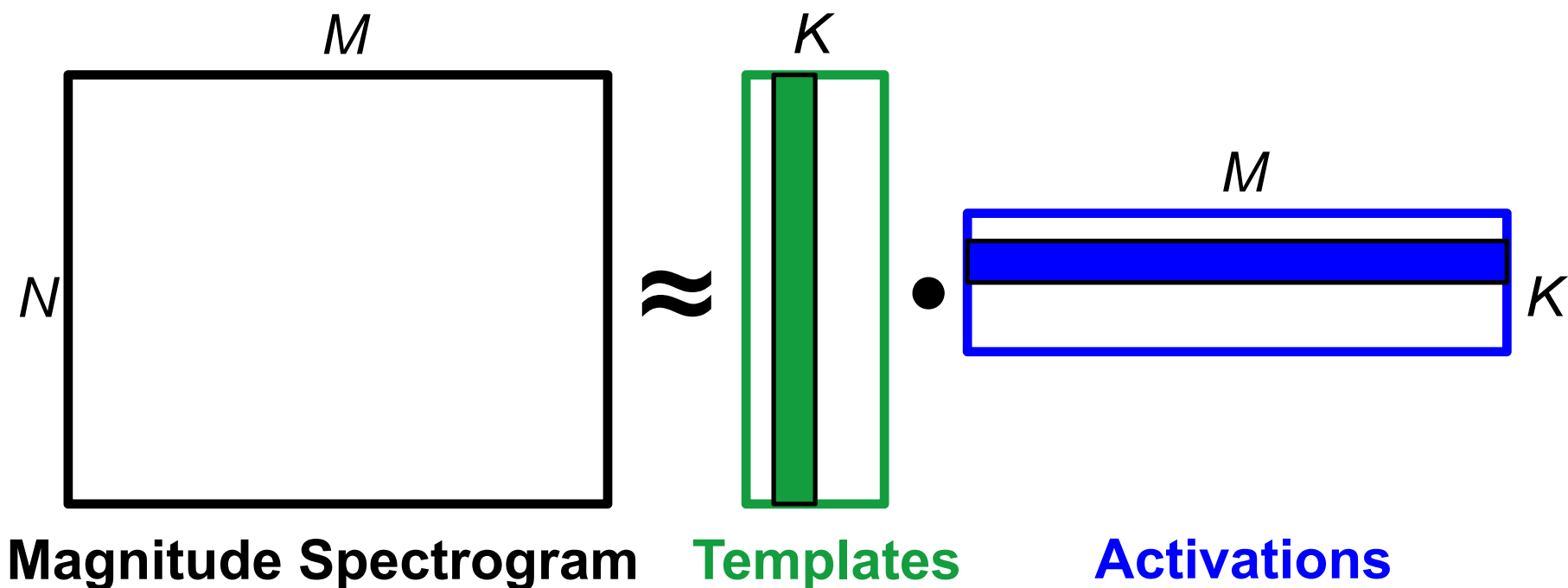
≈



NMF (Nonnegative Matrix Factorization)



NMF (Nonnegative Matrix Factorization)



Templates: Pitch + Timbre

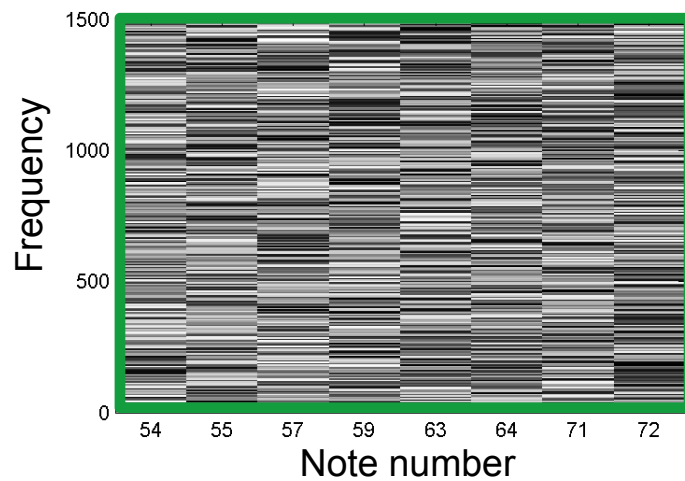
Activations: Onset time + Duration

“How does it sound”

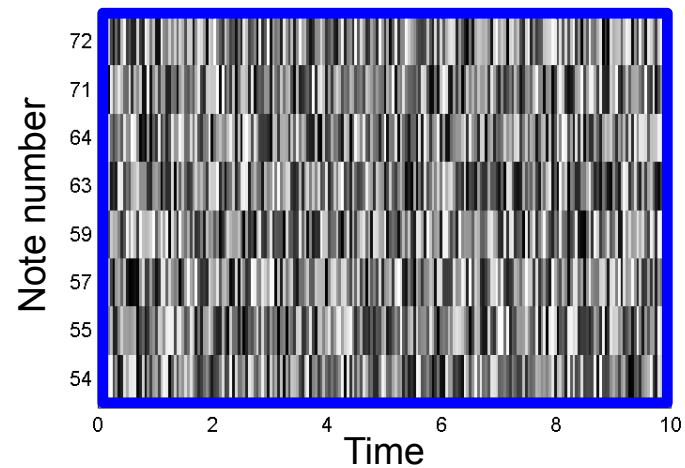
“When does it sound”

NMF-Decomposition

Initialized template



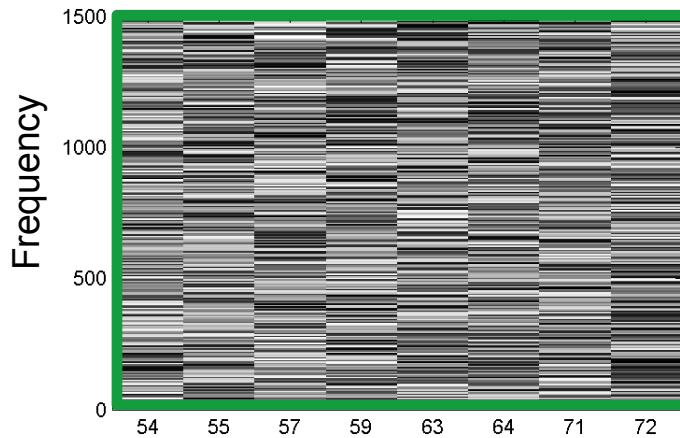
Initialized activations



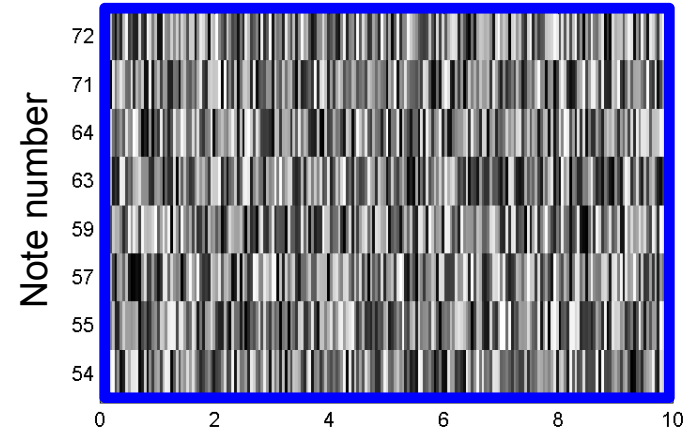
Random initialization

NMF-Decomposition

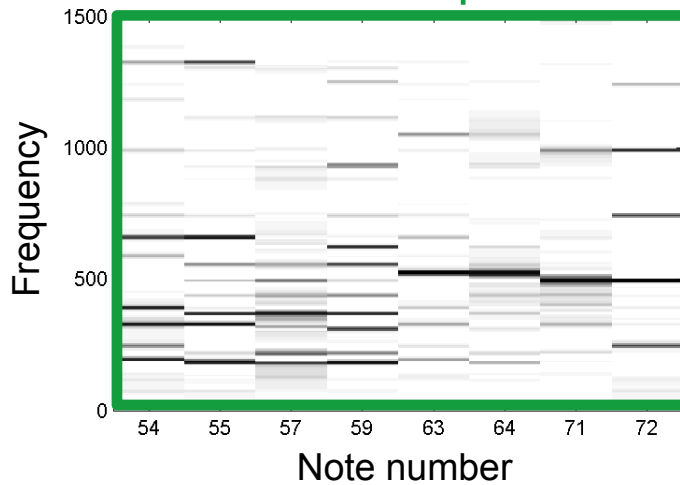
Initialized template



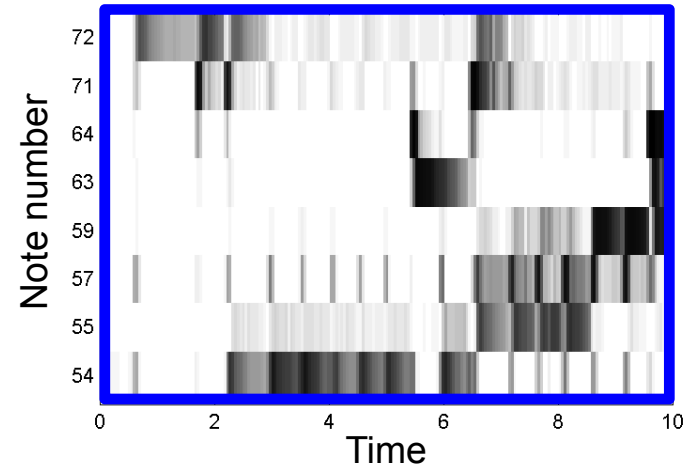
Initialized activations



Learnt templates



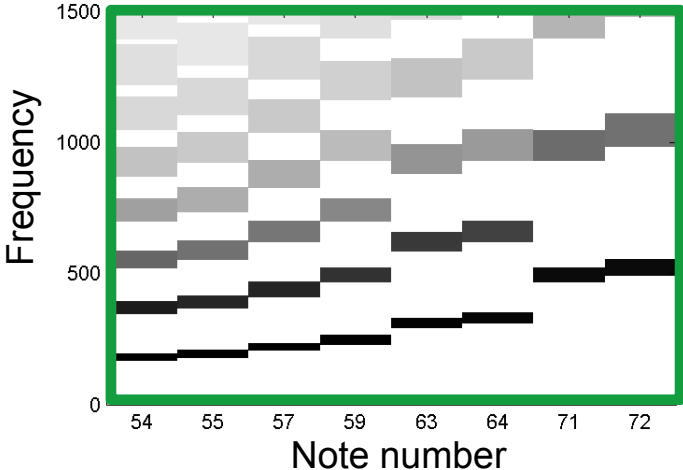
Learnt activations



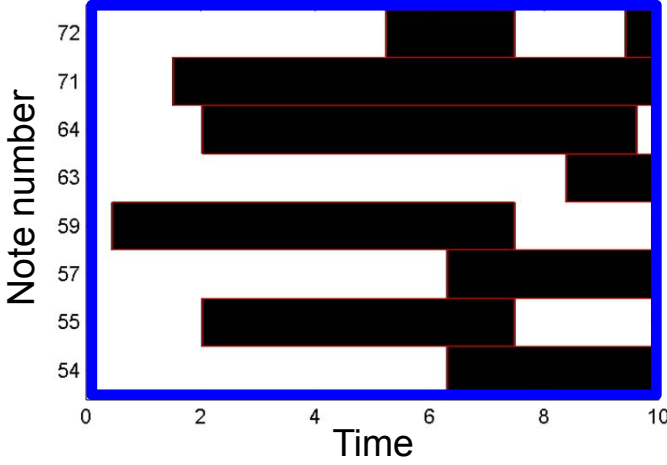
Random initialization → No semantic meaning

NMF-Decomposition

Initialized template



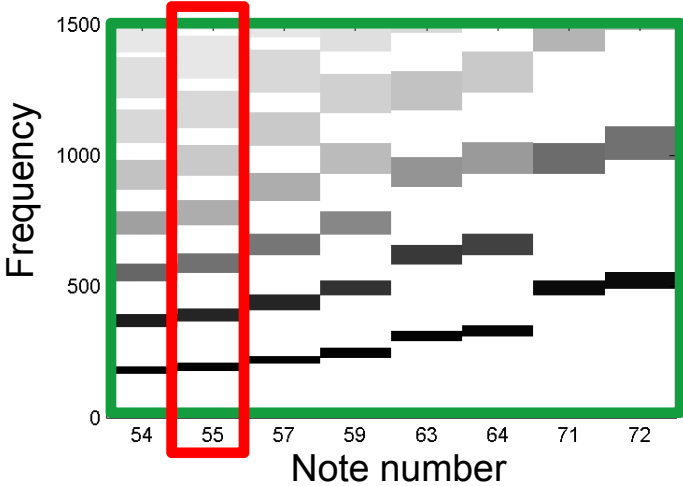
Initialized activations



Constrained initialization

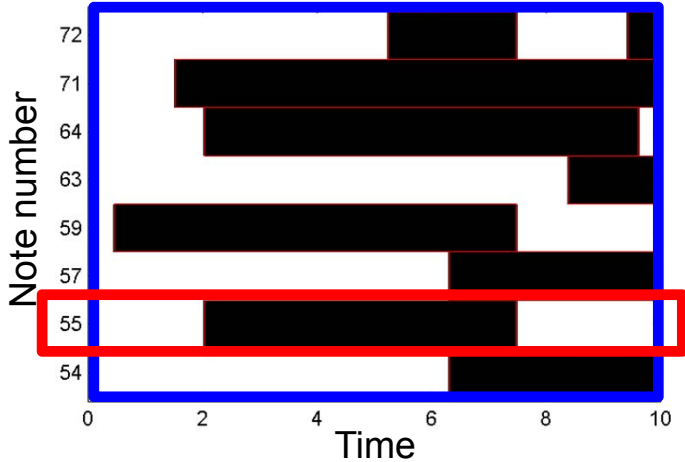
NMF-Decomposition

Initialized template



Template constraint for $p=55$

Initialized activations

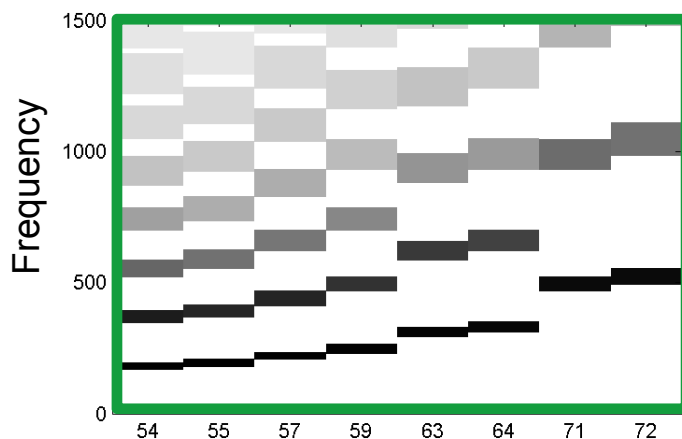


Activation constraints for $p=55$

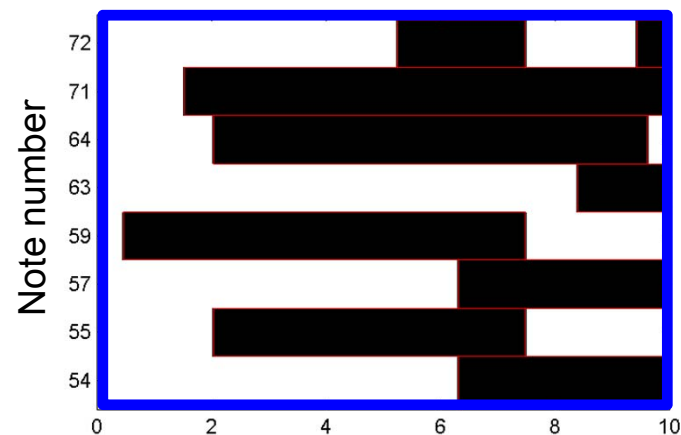
Constrained initialization

NMF-Decomposition

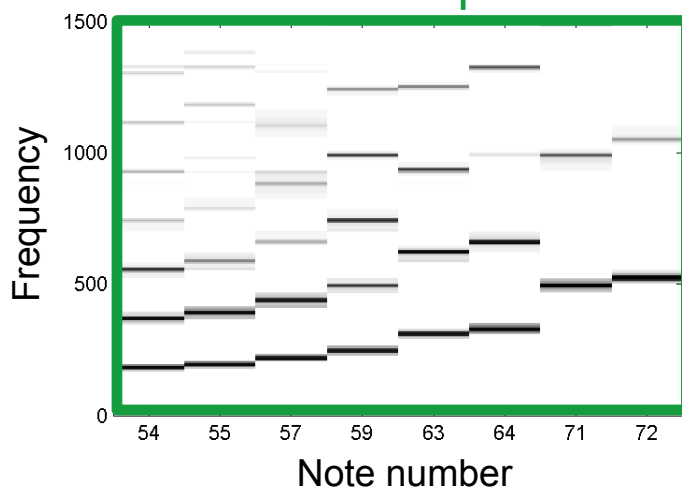
Initialized template



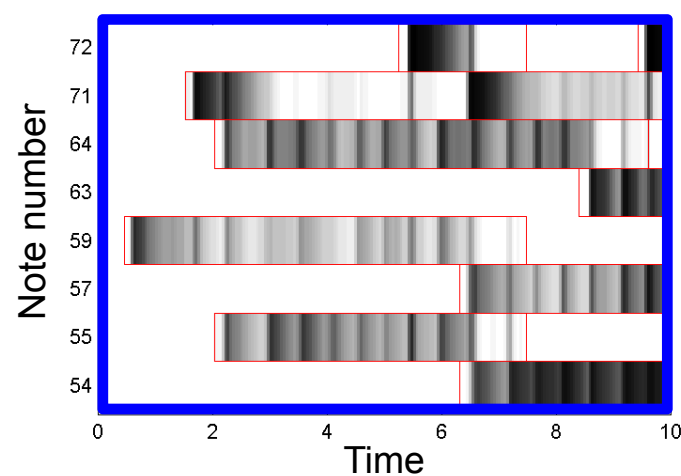
Initialized activations



Learnt templates



Learnt activations



Constrained initialization → NMF as refinement

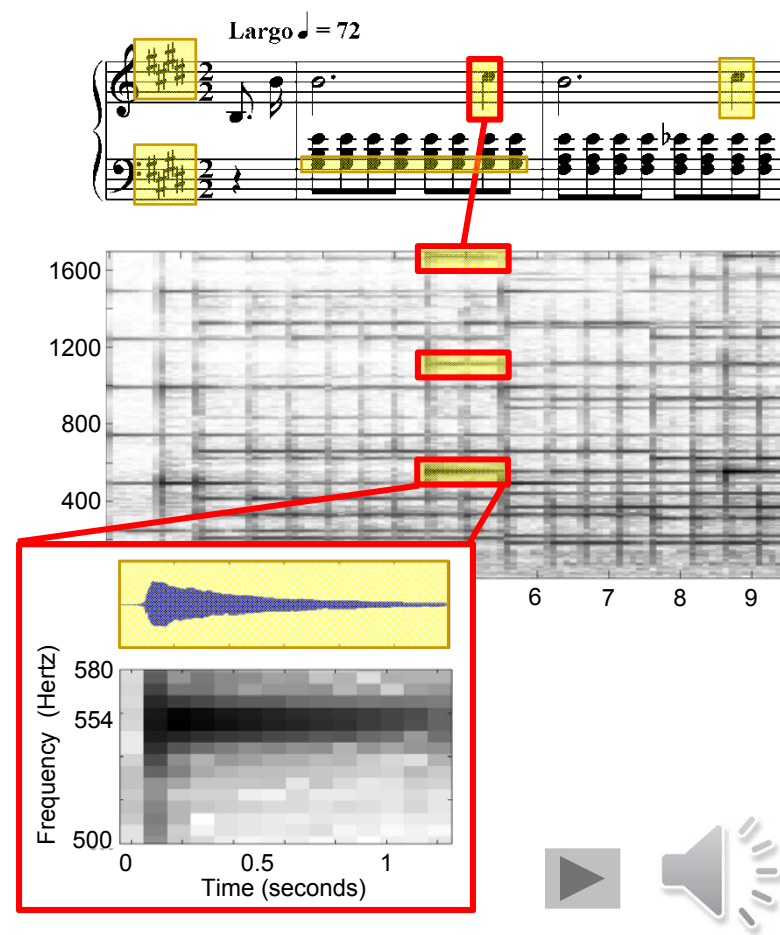
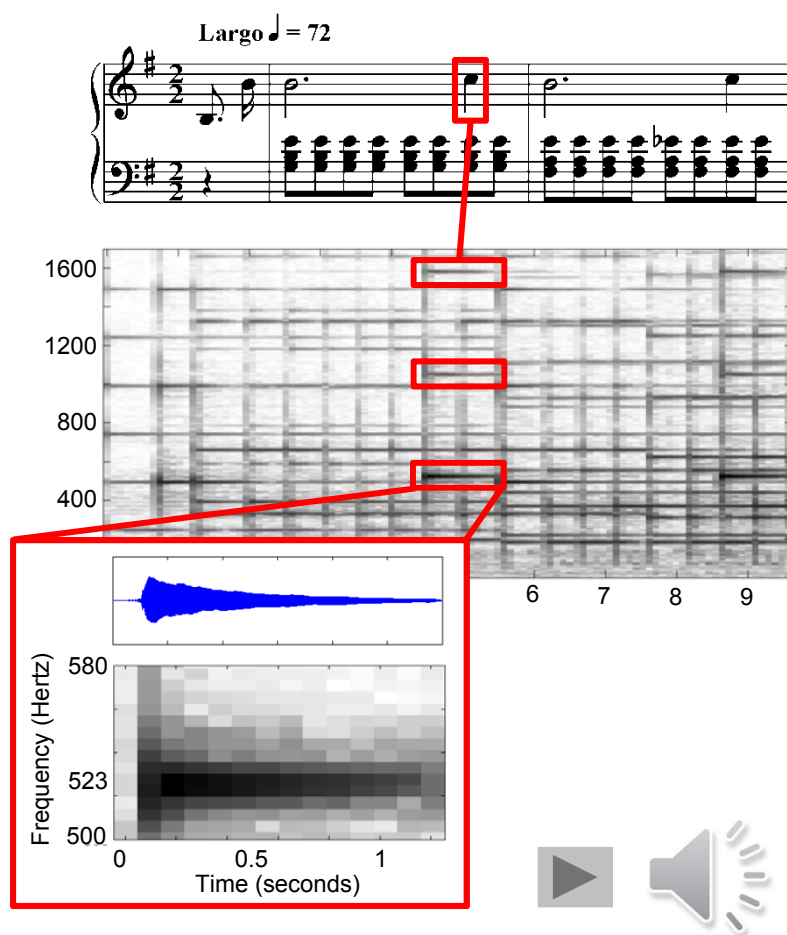


▶ Org
▶ Model

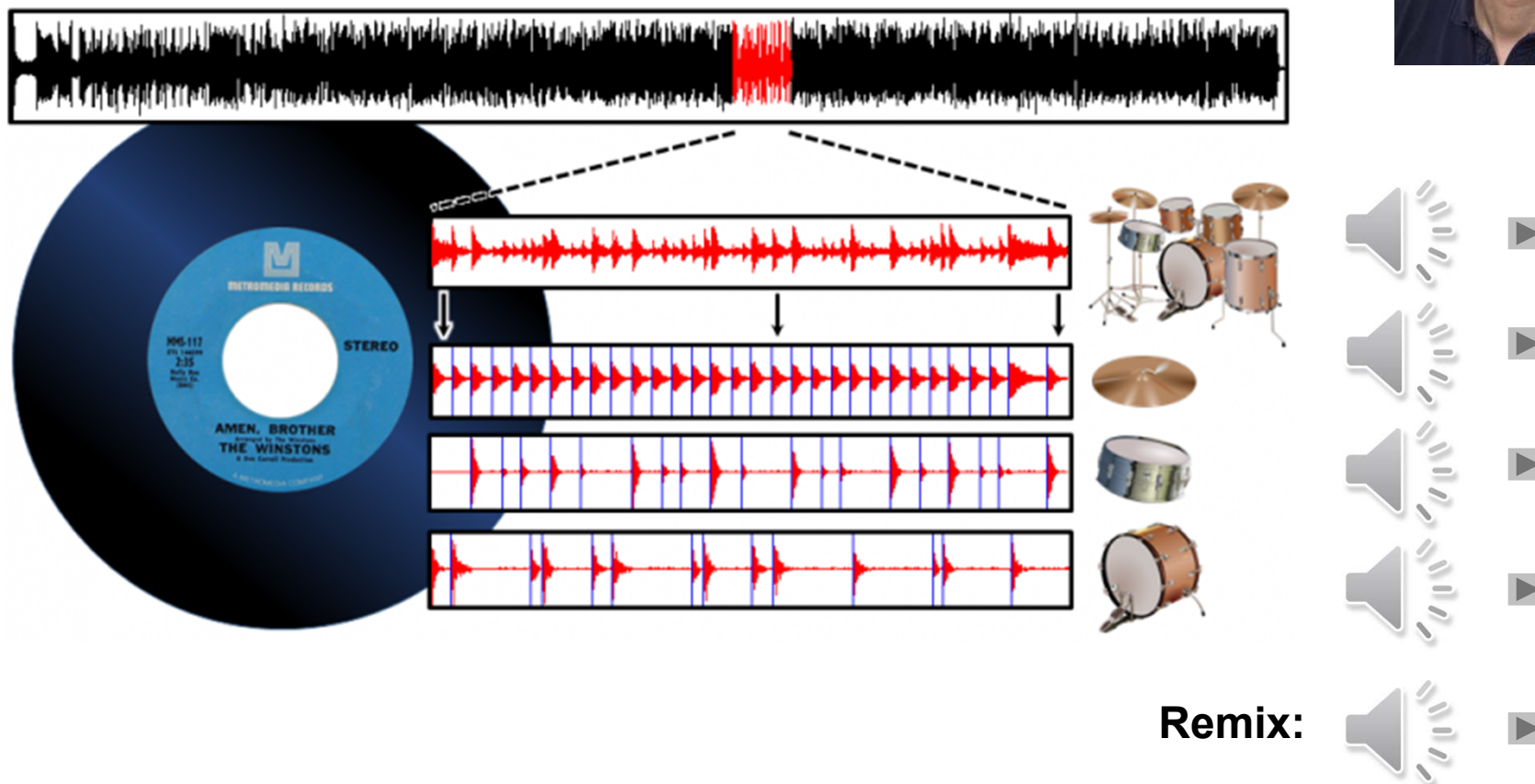


Score-Informed Audio Decomposition

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>

Audio Mosaicing

Target signal: Beatles–Let it be



Source signal: Bees



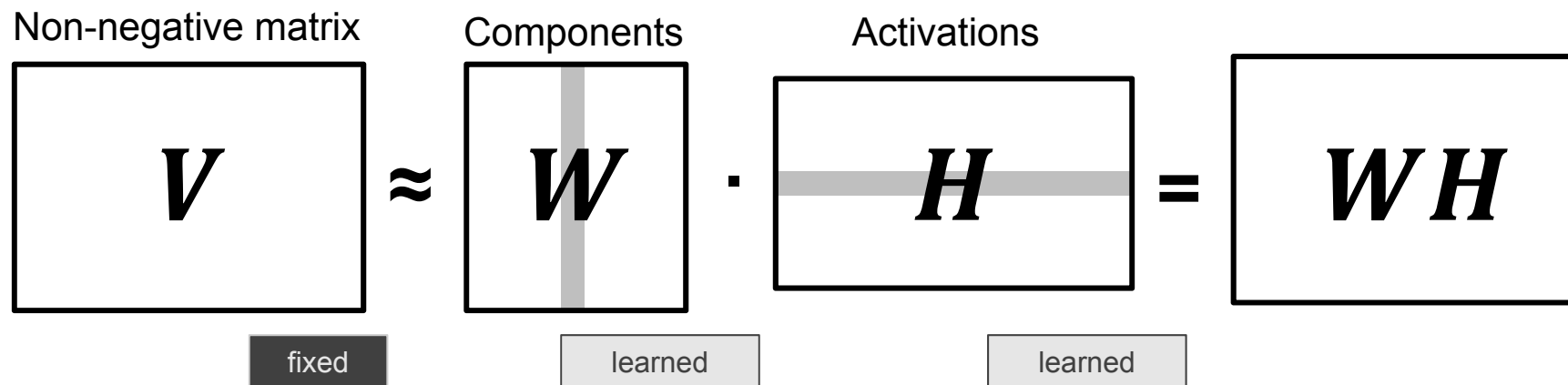
Mosaic signal: Let it Bee

Literature: [Driedger/Müller, ISMIR 2015]

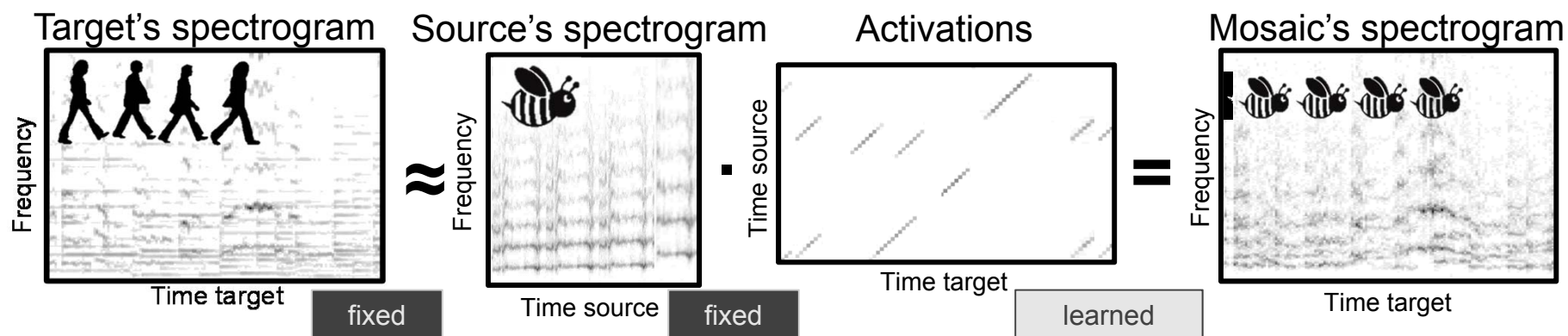
Demo: <https://www.audiolabs-erlangen.de/resources/MIR/2015-ISMIR-LetItBee>

NMF-Inspired Audio Mosaicing

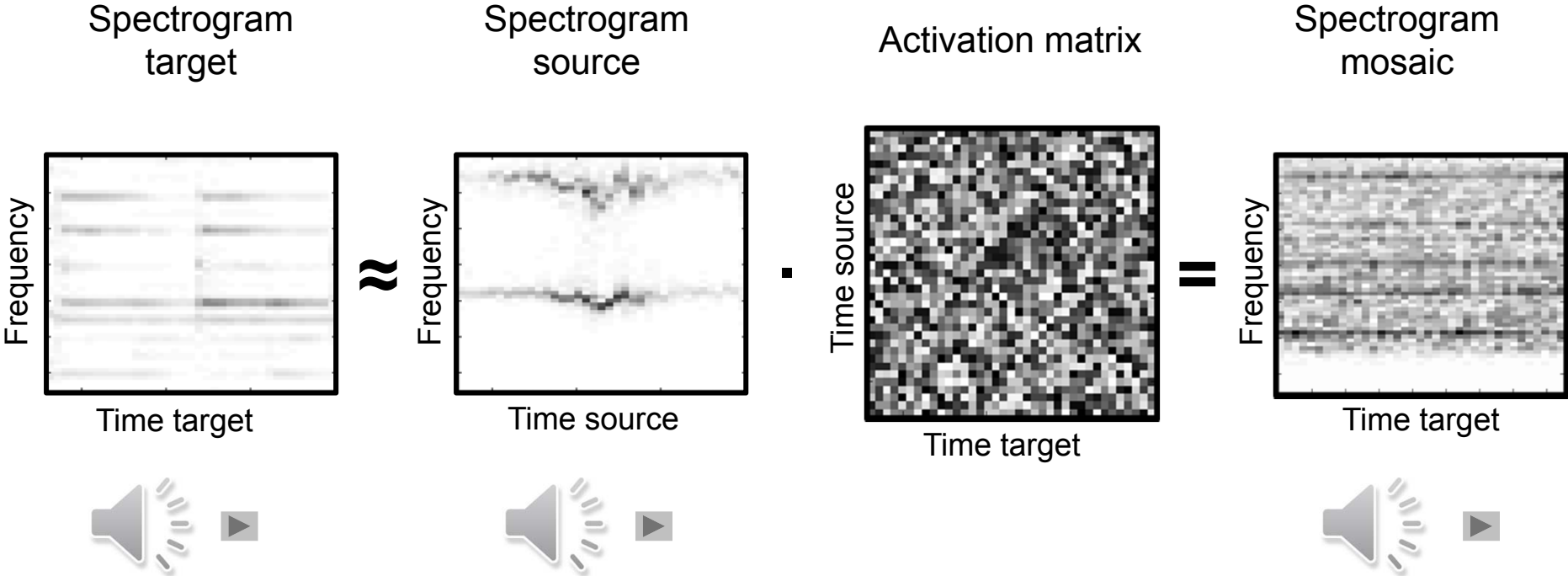
Non-negative matrix factorization (NMF)



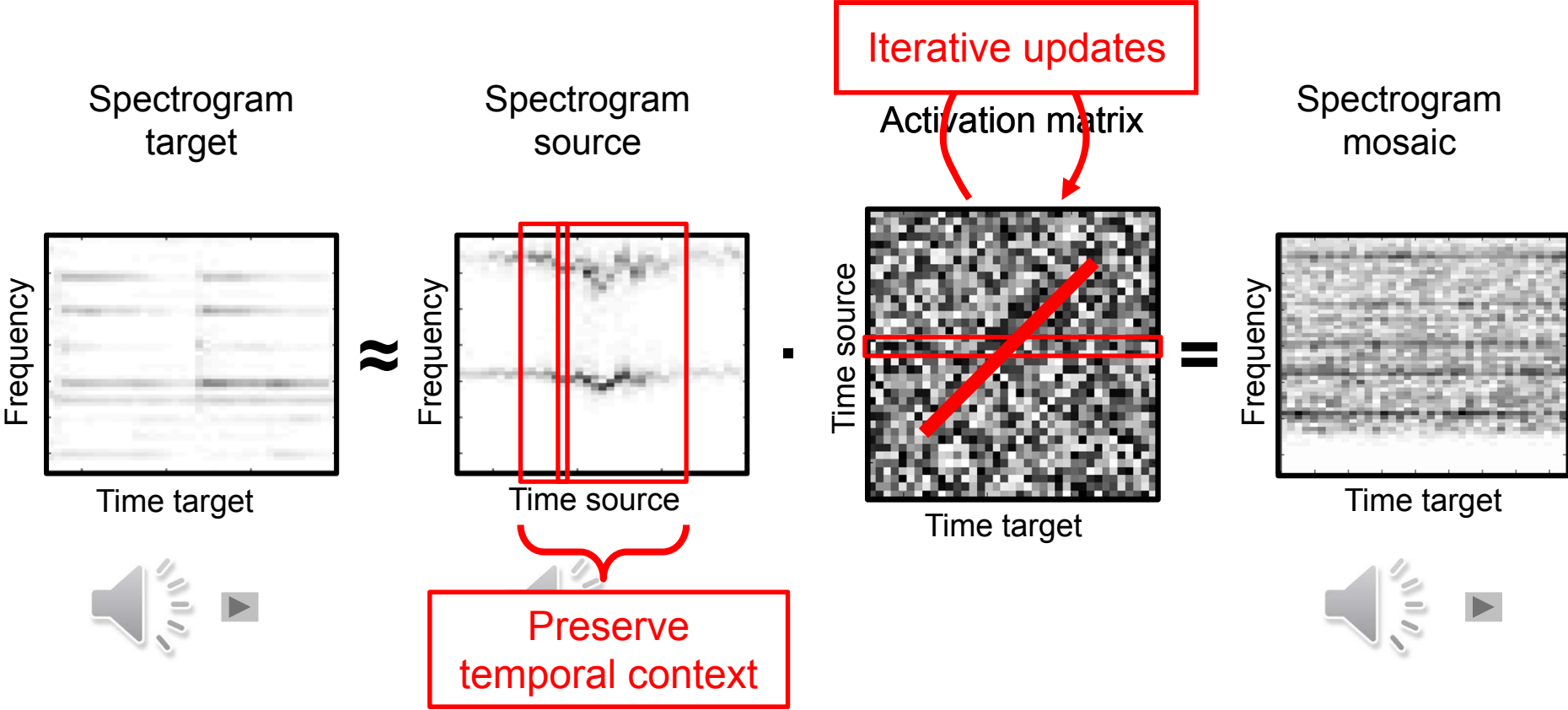
Proposed audio mosaicing approach



NMF-Inspired Audio Mosaicing

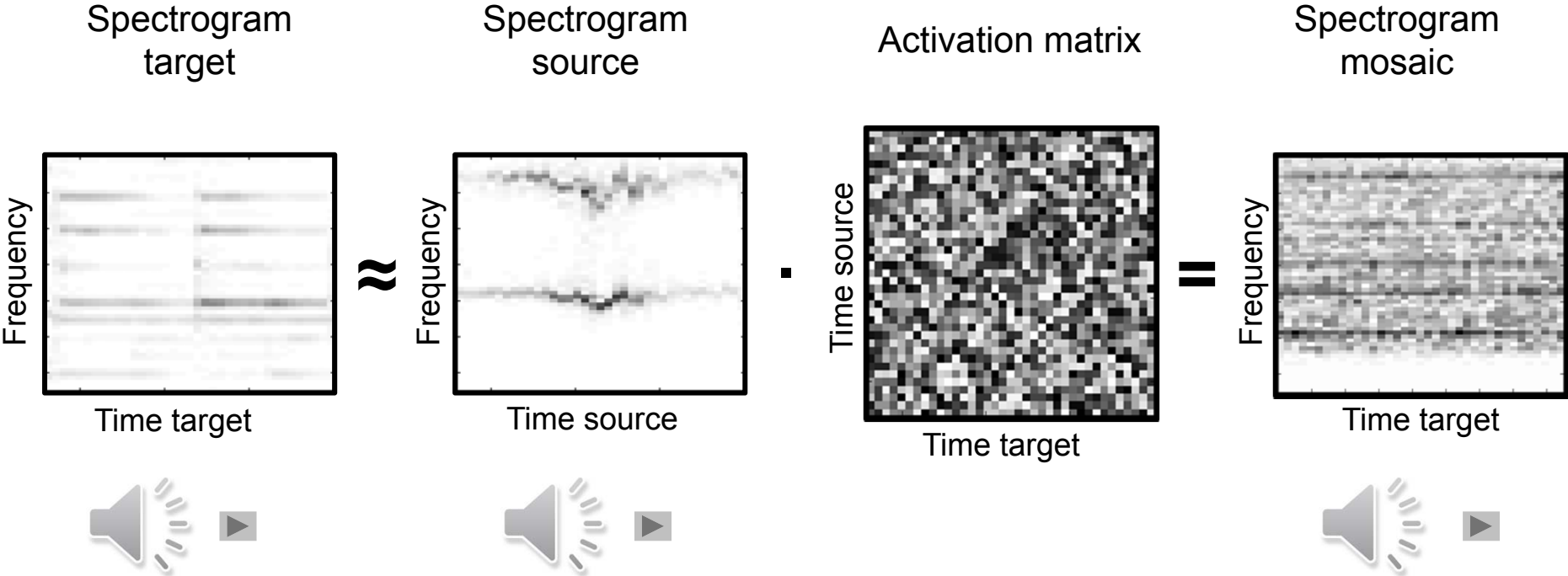


NMF-Inspired Audio Mosaicing

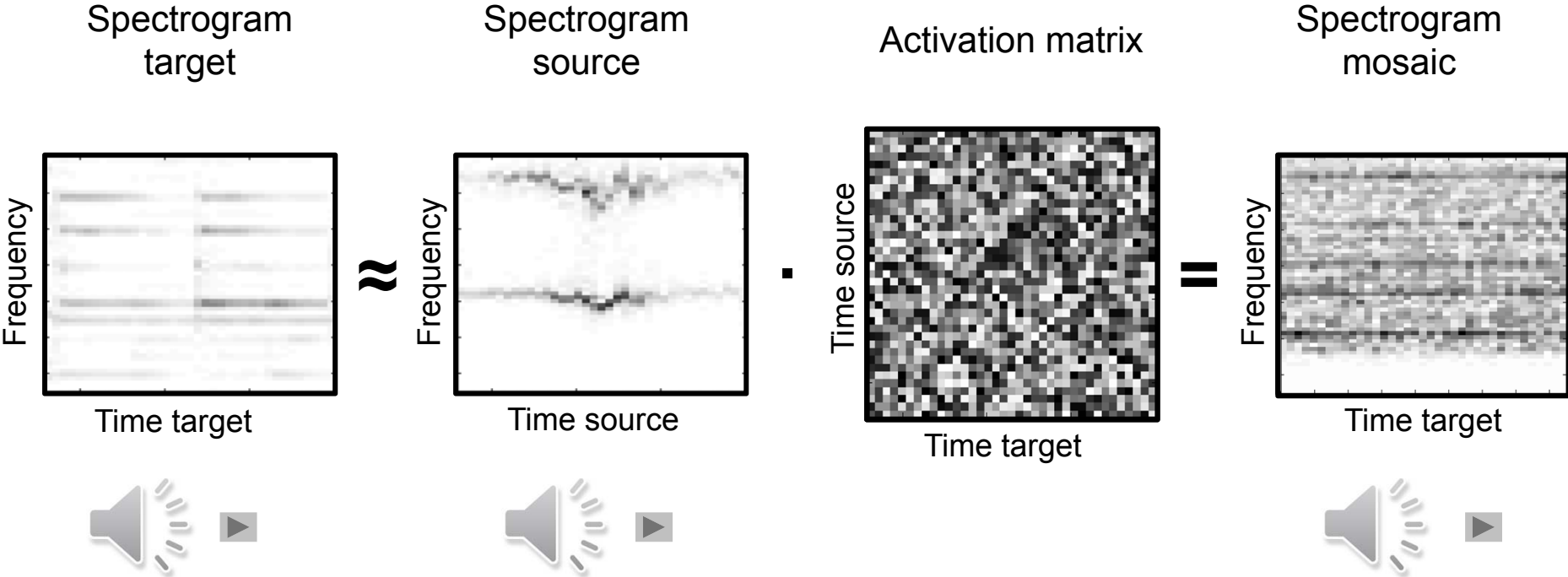


Core idea: support the development of sparse diagonal activation structures

NMF-Inspired Audio Mosaicing



NMF-Inspired Audio Mosaicing

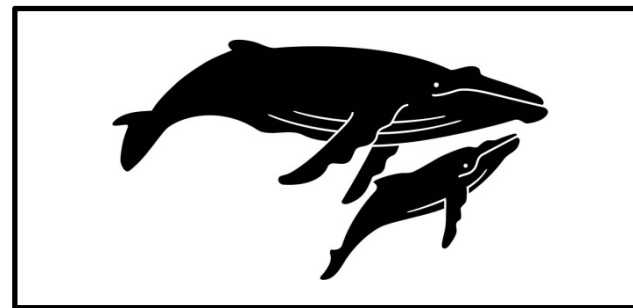


Audio Mosaicing

Target signal: Chic–Good times



Source signal: Whales



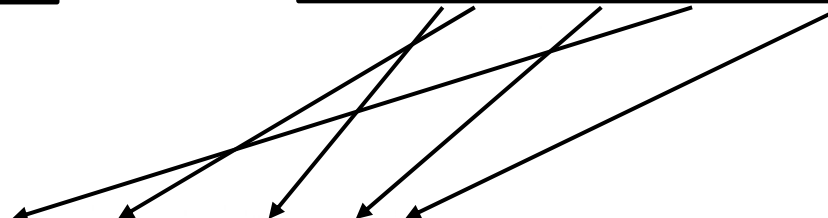
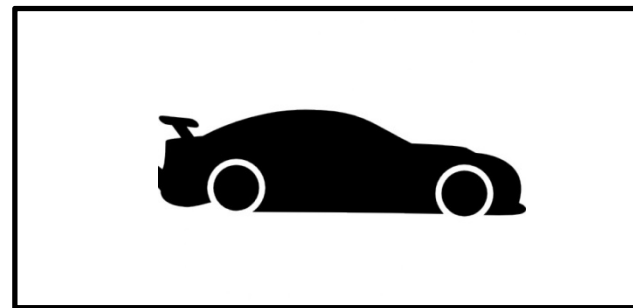
Mosaic signal

Audio Mosaicing

Target signal: Adele—Rolling in the Deep



Source signal: Race car



Mosaic signal

Links

- SiSEC: Signal Separation Evaluation Campaign
<https://www.sisec17.audiolabs-erlangen.de/>
- MedleyDB: A Dataset of Multitrack Audio
<http://steinhardt.nyu.edu/marl/research/medleydb>
- LibROSA (Python)
<https://librosa.github.io/librosa/>