INTERNATIONAL AUDIO LABORATORIES ERLANGEN

AUDIO LABS

An Introduction to Music Information Retrieval

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

Deep Learning Indaba𝕏

Nigeria, 24 Sep – 25 Sep 2021

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Fraunhofer IIS

Meinard Müller

- Mathematics (Diplom/Master)
  Computer Science (PhD)
  Information Retrieval (Habilitation)
- Since 2012: Full Professor
  Semantic Audio Processing
- President of the International Society for Music Information Retrieval (MIR)
- Member of the Senior Editorial Board of the IEEE Signal Processing Magazine
- IEEE Fellow for contributions to Music Signal Processing

universitätbonn

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

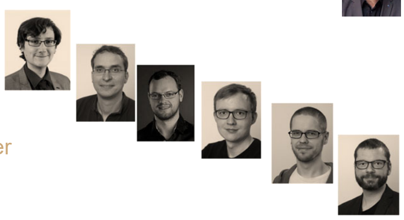
ISMIR

IEEE

Meinard Müller: Research Group

- Sebastian Rosenzweig
- Michael Krause
- Yigitcan Özer
- Peter Meier (external)

- Frank Zalkow
- Christian Dittmar
- Christof Weiß
- Stefan Balke
- Jonathan Driedger
- Thomas Prätzlich
- …



International Audio Laboratories Erlangen

Fraunhofer IIS

- Fraunhofer Institute for Integrated Circuits IIS
- Largest Fraunhofer institute with ≈ 1000 members
- Applied research for sensor, audio, and media technology

mp3

AUDIO LABS

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

- Friedrich-Alexander Universität Erlangen-Nürnberg (FAU)
- One of Germany's largest universities with ≈ 40,000 students
- Strong Technical Faculty

International Audio Laboratories Erlangen

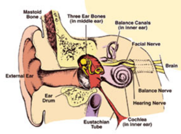
Audio

International Audio Laboratories Erlangen

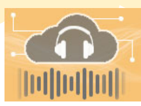
Audio

Audio Coding

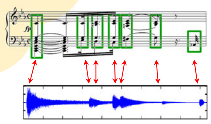
3D Audio



Psychoacoustics



Internet of Things



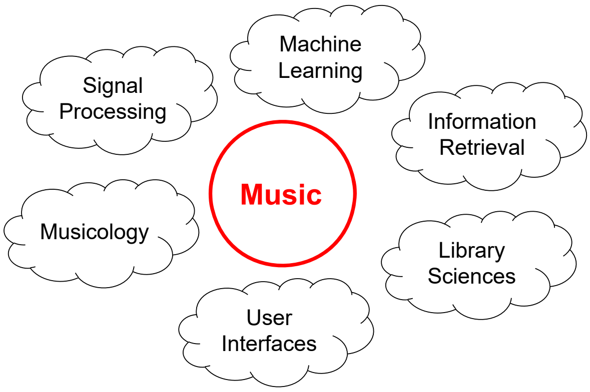
Music Processing

International Audio Laboratories Erlangen

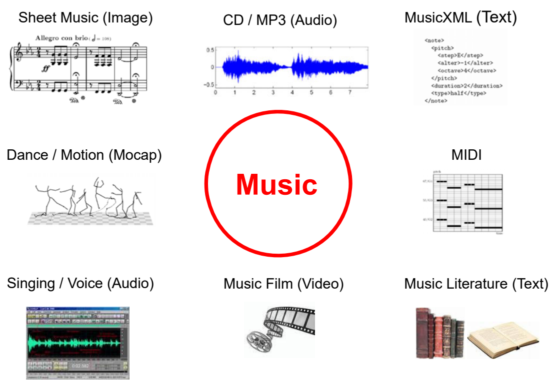
- Prof. Dr. Jürgen Herre
  Audio Coding
- Prof. Dr. Bernd Edler
  Audio Signal Analysis
- Prof. Dr. Meinard Müller
  Semantic Audio Processing
- Prof. Dr. Emanuël Habets
  Spatial Audio Signal Processing
- Prof. Dr. Nils Peters
  Audio Signal Processing
- Dr. Stefan Turowski
  Coordinator AudioLabs-FAU

Music

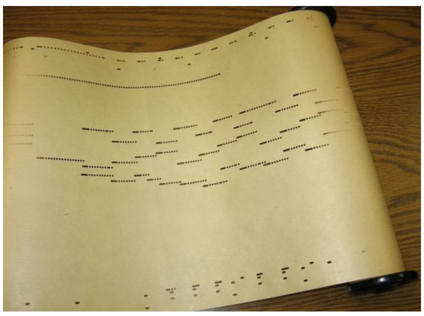
Music Information Retrieval (MIR)



Music Information Retrieval (MIR)

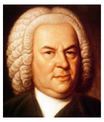


Piano Roll Representation



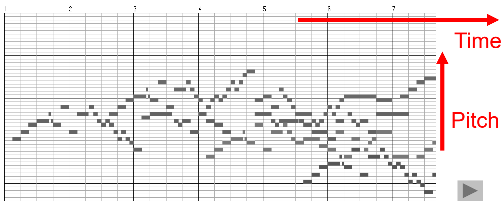
Player Piano (1900)

Piano Roll Representation (MIDI)

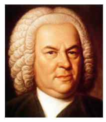


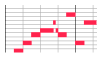
J.S. Bach, C-Major Fuge

(Well Tempered Piano, BWV 846)

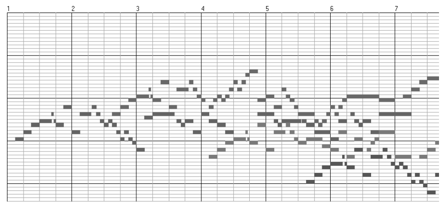


Piano Roll Representation (MIDI)

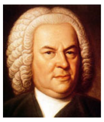


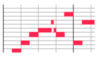
Query: 

Goal: Find all occurrences of the query



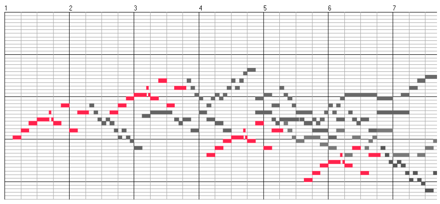
Piano Roll Representation (MIDI)



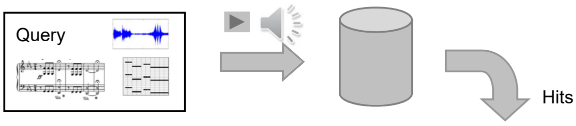
Query: 

Goal: Find all occurrences of the query

Matches:



Music Retrieval



Retrieval tasks:

Audio identification

Audio matching

Version identification

Category-based music retrieval

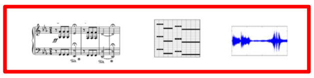
Bernstein (1962)
Beethoven, Symphony No. 5

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

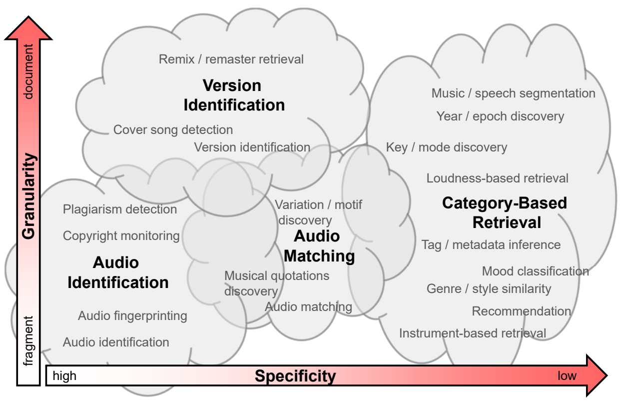
Low specificity

Granularity

Fragment-based retrieval

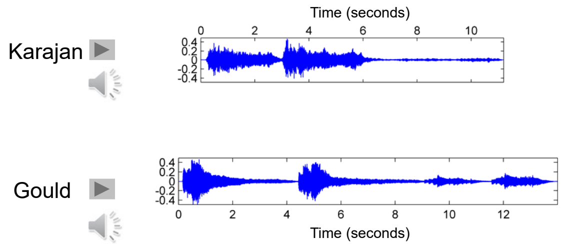
Document-based retrieval

Music Retrieval



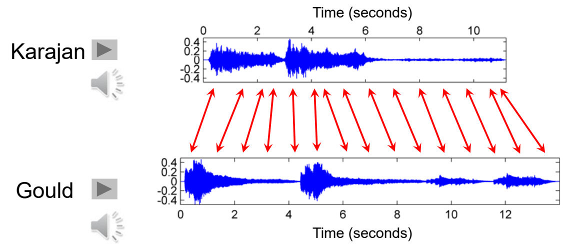
Music Synchronization: Audio-Audio

Beethoven's Fifth

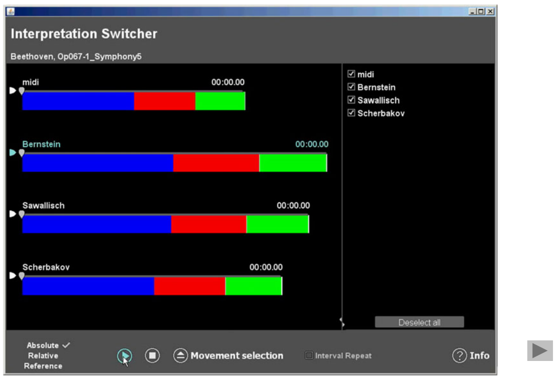


Music Synchronization: Audio-Audio

Beethoven's Fifth



Application: Interpretation Switcher



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

Traditional Engineering Approach:

1.) Feature extraction

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

➡ **Chroma features**

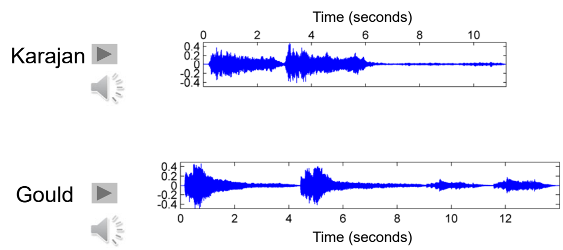
2.) Temporal alignment

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

➡ **Dynamic time warping (DTW)**

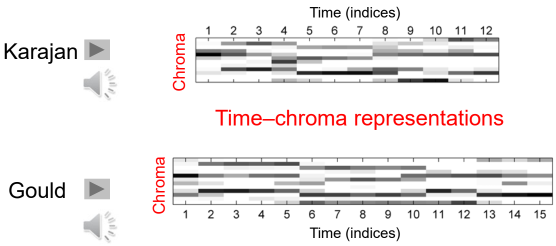
Music Synchronization: Audio-Audio

Beethoven's Fifth



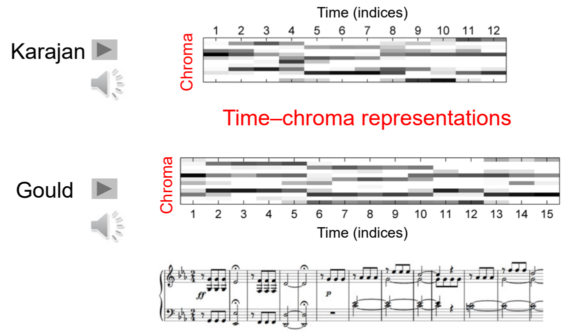
Music Synchronization: Audio-Audio

Beethoven's Fifth



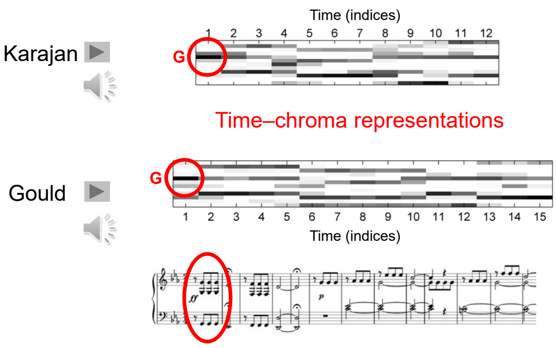
Music Synchronization: Audio-Audio

Beethoven's Fifth



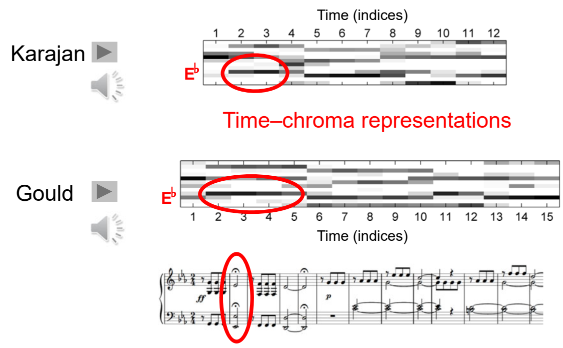
Music Synchronization: Audio-Audio

Beethoven's Fifth

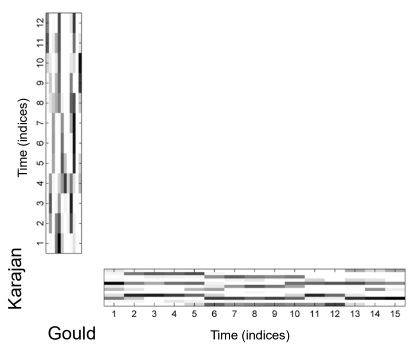


Music Synchronization: Audio-Audio

Beethoven's Fifth

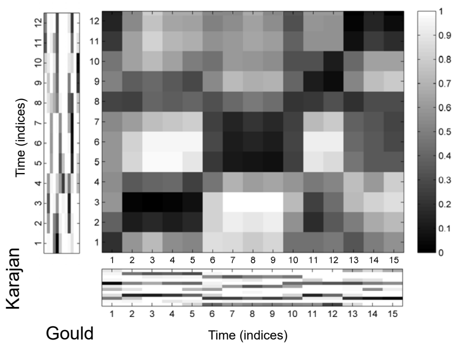


Music Synchronization: Audio-Audio



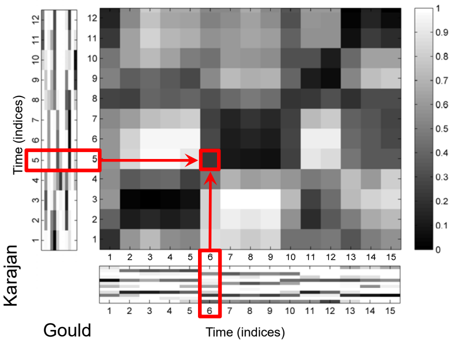
Music Synchronization: Audio-Audio

Cost matrix



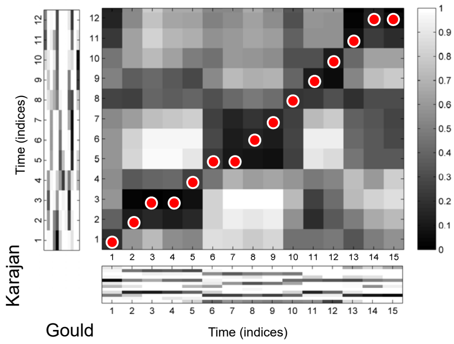
Music Synchronization: Audio-Audio

Cost matrix



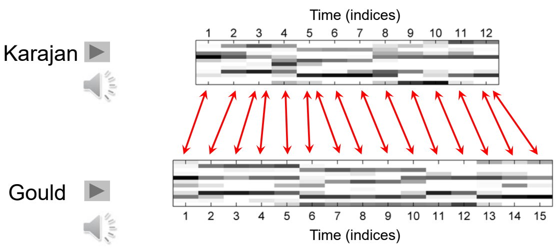
Music Synchronization: Audio-Audio

Cost-minimizing warping path



Music Synchronization: Audio-Audio

Optimal alignment (cost-minimizing warping path)



Music Synchronization: Audio-Audio

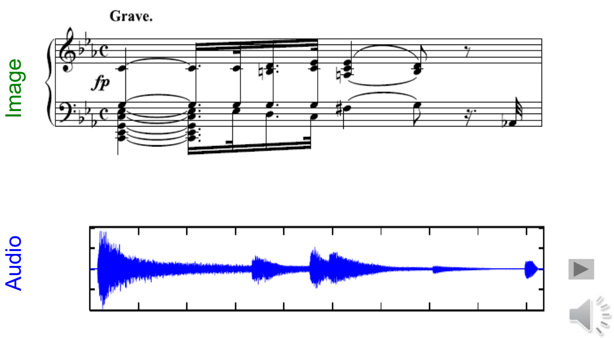
Deep Learning Approaches

- Learn audio features from data
  - Should be able to achieve high alignment accuracy
  - Should be robust to performance variations
  - Musical relevance?
- Alignment problem
  - Pre-aligned data for training
  - Part of loss function → differentiability?

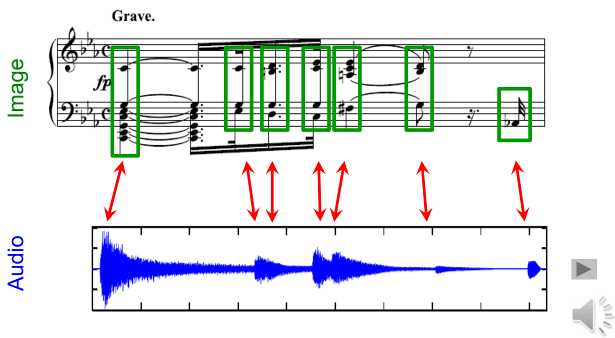
CTC-Loss
Graves et al.: Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks. ICML 2006

Soft-DTW
Cuturi, Blondel: Soft-DTW: A Differentiable Loss Function for Time-Series. ICML 2017

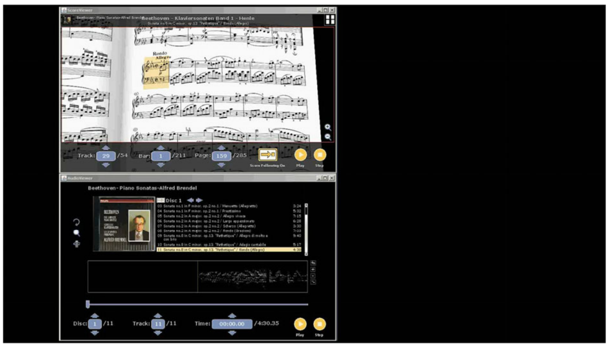
Music Synchronization: Image-Audio



Music Synchronization: Image-Audio



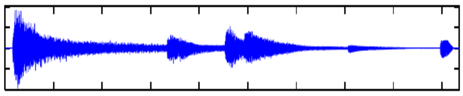
Application: Score Viewer



How to make the data comparable?

Image

Audio

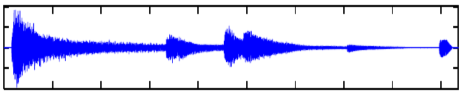


How to make the data comparable?

Image Processing: Optical Music Recognition

Image

Audio



How to make the data comparable?

Image Processing: Optical Music Recognition

Image

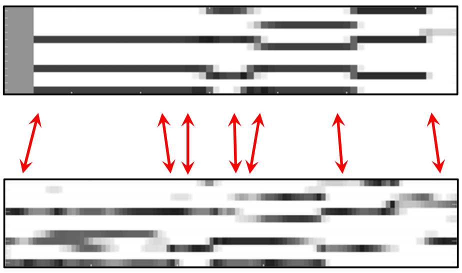
Audio

Audio Processing: Fourier Analysis

How to make the data comparable?

Image Processing: Optical Music Recognition

Image

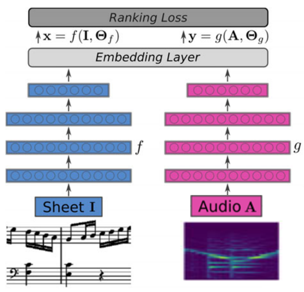


Audio

Audio Processing: Fourier Analysis

Music Synchronization: Image-Audio

Deep Learning Approach

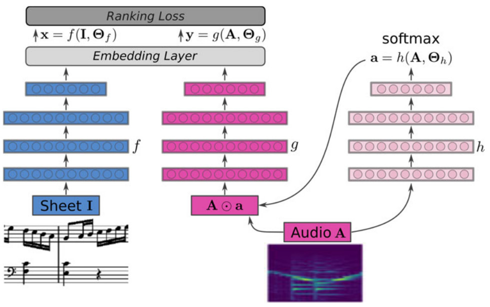


- Cross-modal embedding
- Requires corresponding snippets of audio and sheet music for training
- Triplet Loss function $\max(0, d(x^a, y^p) - d(x^a, y^n) + \alpha)$
- Problem very hard
  - Performance variations
  - Layout variations

Cross-Modal Retrieval
Dorfer et al.: End-to-End Cross-Modality Retrieval with CCA Projections and Pairwise Ranking Loss. International Journal of Multimedia Information Retrieval, 2018.

Music Synchronization: Image-Audio

Deep Learning Approach: Soft Attention Mechanism

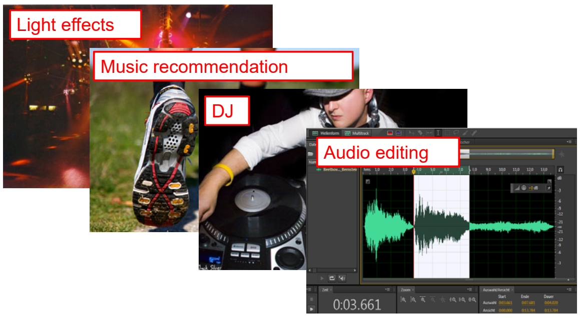


Music Processing

Coarse/Relative Level	**Fine/Absolute Level**
What do different versions or instances have in common?	What are the characteristics of a specific version or instance?
Provide coarse description: What makes up a piece of music?	Capture nuances and subtleties: What makes music come alive?
Identify despite of differences	Identify the differences
Example tasks: **Music Retrieval** **Genre Classification** **Global Tempo Estimation**	Example tasks: **Music Transcription** **Performance Analysis** **Local Tempo Estimation**

Tempo Estimation and Beat Tracking

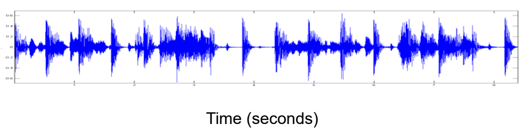
Basic task: “Tapping the foot when listening to music”



Tempo Estimation and Beat Tracking

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

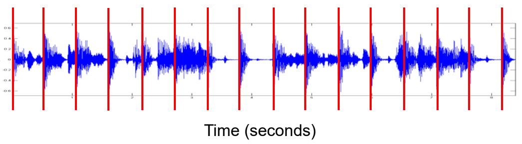
Example: Queen – Another One Bites The Dust



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: Chopin – Mazurka Op. 68-3

Pulse level: Quarter note

Tempo: ???

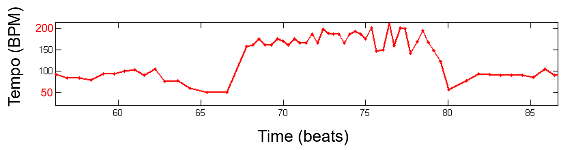
Tempo Estimation and Beat Tracking

Example: Chopin – Mazurka Op. 68-3

Pulse level: Quarter note

Tempo: 50-200 BPM

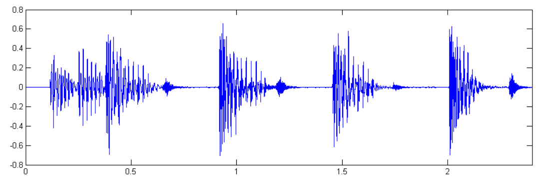
Tempo curve



Tempo Estimation and Beat Tracking

Tasks

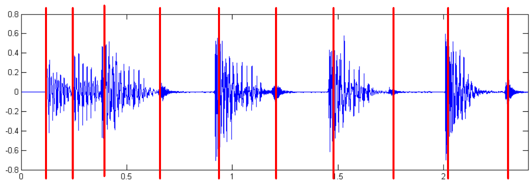
- Onset detection
- Beat tracking
- Tempo estimation



Tempo Estimation and Beat Tracking

Tasks

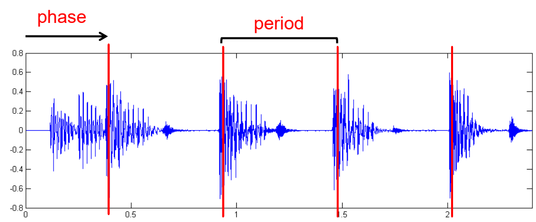
- Onset detection
- Beat tracking
- Tempo estimation



Tempo Estimation and Beat Tracking

Tasks

- Onset detection
- Beat tracking
- Tempo estimation



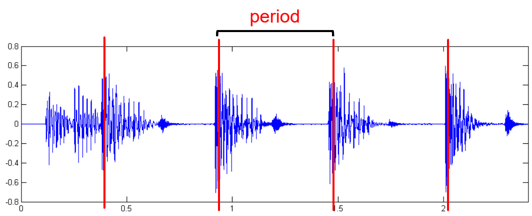
Tempo Estimation and Beat Tracking

Tasks

- Onset detection
- Beat tracking
- Tempo estimation

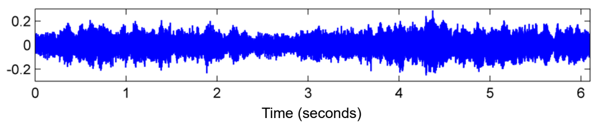
Tempo := 60 / period

Beats per minute (BPM)



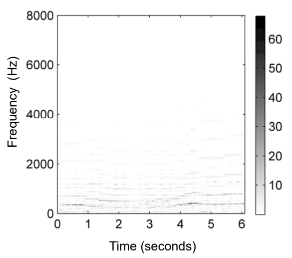
Onset Detection (Spectral Flux)

Audio recording



Onset Detection (Spectral Flux)

Magnitude spectrogram $|X|$

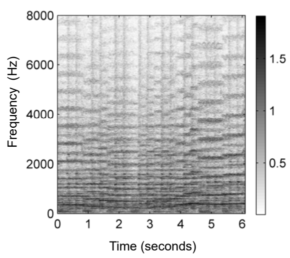


Steps:

1. Spectrogram

Onset Detection (Spectral Flux)

Compressed spectrogram Y

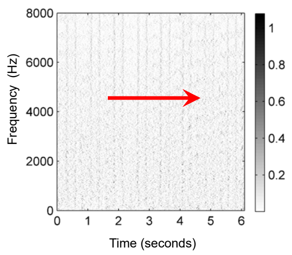


Steps:

1. Spectrogram
2. Logarithmic compression

Onset Detection (Spectral Flux)

Spectral difference

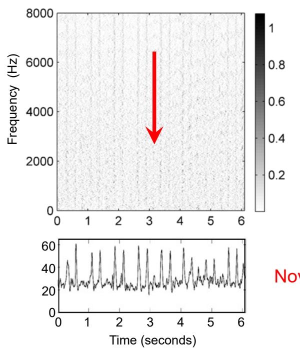


Steps:

1. Spectrogram
2. Logarithmic compression
3. Differentiation & half wave rectification

Onset Detection (Spectral Flux)

Spectral difference



Novelty curve

Steps:

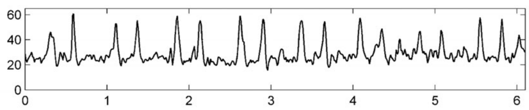
1. Spectrogram
2. Logarithmic compression
3. Differentiation & half wave rectification
4. Accumulation

Onset Detection (Spectral Flux)

Steps:

1. Spectrogram
2. Logarithmic compression
3. Differentiation & half wave rectification
4. Accumulation

Novelty function



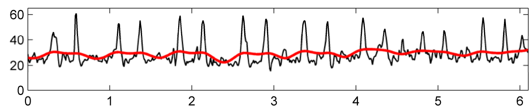
Onset Detection (Spectral Flux)

Steps:

1. Spectrogram
2. Logarithmic compression
3. Differentiation & half wave rectification
4. Accumulation
5. Normalization

Novelty function

Substraction of local average

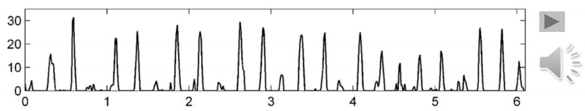


Onset Detection (Spectral Flux)

Steps:

1. Spectrogram
2. Logarithmic compression
3. Differentiation & half wave rectification
4. Accumulation
5. Normalization

Normalized novelty function



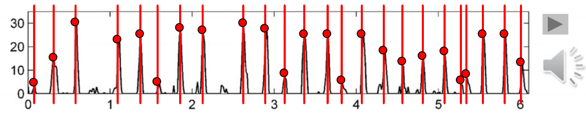
Onset Detection (Spectral Flux)

Steps:

1. Spectrogram
2. Logarithmic compression
3. Differentiation & half wave rectification
4. Accumulation
5. Normalization

Normalized novelty function

Peak positions indicate beat candidates



Onset Detection (Spectral Flux)

Deep Learning Approach

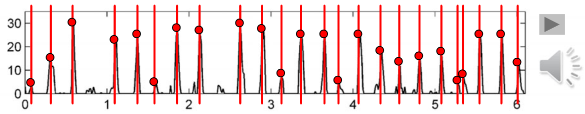
1. Input representation
2. Sigmoid activation
3. Convolution & rectified linear unit (ReLU)
4. Pooling
5. Convolution & ReLU

Steps:

1. Spectrogram
2. Logarithmic compression
3. Differentiation & half wave rectification
4. Accumulation
5. Normalization

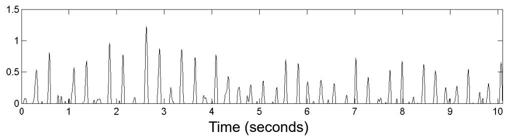
Normalized novelty function

Peak positions indicate beat candidates



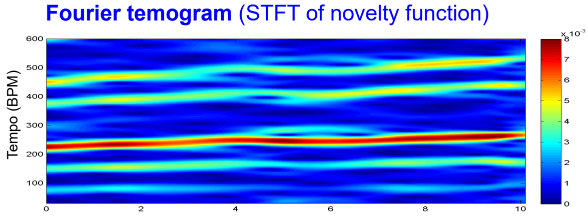
Local Pulse and Tempo Tracking

Normalized novelty function

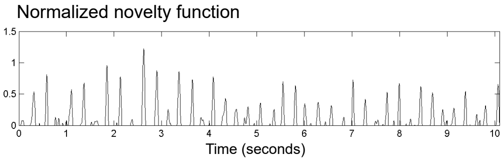


Local Pulse and Tempo Tracking

Fourier temogram (STFT of novelty function)

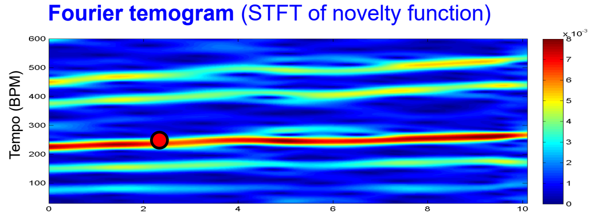


Normalized novelty function

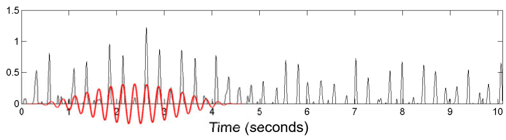


Local Pulse and Tempo Tracking

Fourier temogram (STFT of novelty function)

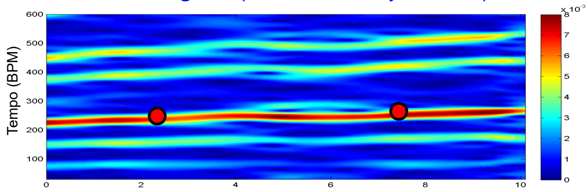


Optimizing local periodicity kernel

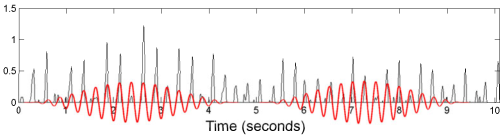


Local Pulse and Tempo Tracking

Fourier temogram (STFT of novelty function)

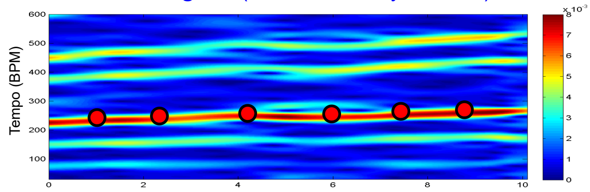


Optimizing local periodicity kernel

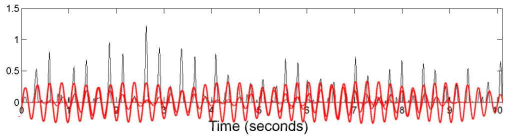


Local Pulse and Tempo Tracking

Fourier temogram (STFT of novelty function)

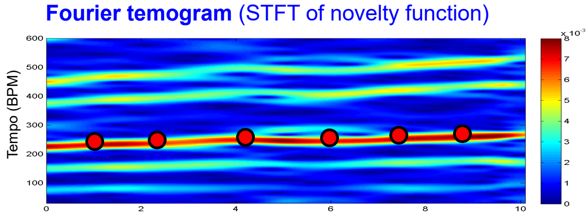


Optimizing local periodicity kernel

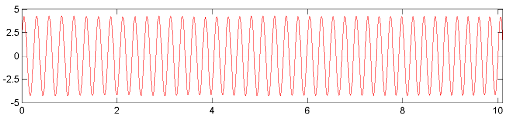


Local Pulse and Tempo Tracking

Fourier temogram (STFT of novelty function)

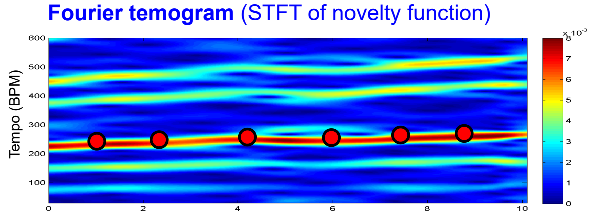


Accumulation of kernels

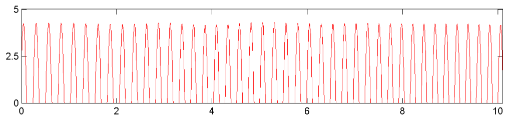


Local Pulse and Tempo Tracking

Fourier temogram (STFT of novelty function)

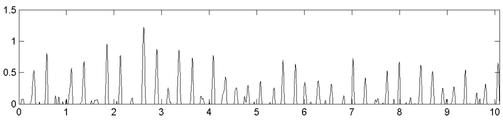


Halfwave rectification

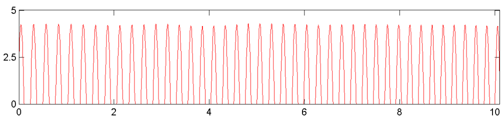


Local Pulse and Tempo Tracking

Novelty Curve



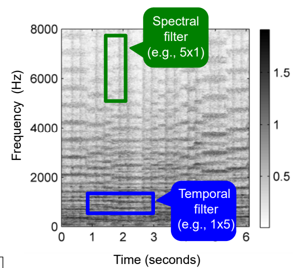
Predominant Local Pulse (PLP)



Local Pulse and Tempo Tracking

Deep Learning Approach

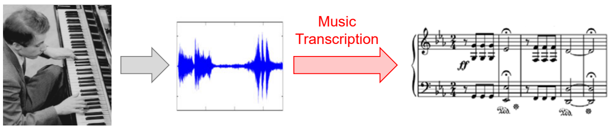
- End-to-end approach
  - Input: Short audio snippets
  - Output: Tempo value
- DL architecture inspired by traditional engineering
  - Layers and activation functions
  - Shape of convolutional kernels



Tempo Estimation
Schreiber, Müller: A Single-Step Approach to Musical Tempo Estimation Using a Convolutional Neural Network, ISMIR 2018.

Automatic Music Transcription

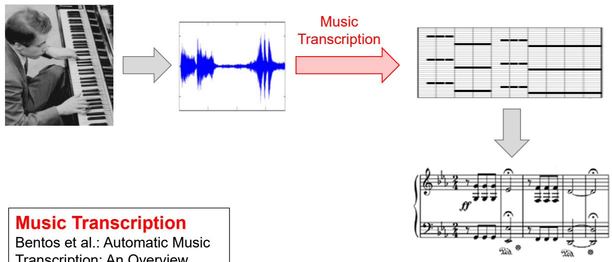
Task: Convert a music recording into sheet music



Music Transcription
Bentos et al.: Automatic Music Transcription: An Overview. IEEE Signal Processing Magazine 36(1), 2019.

Automatic Music Transcription

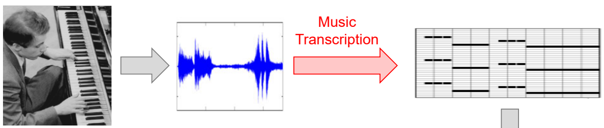
Task: Convert a music recording into sheet music
(or another symbolic music representation)



Music Transcription
Bentos et al.: Automatic Music Transcription: An Overview. IEEE Signal Processing Magazine 36(1), 2019.

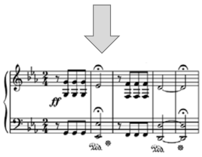
Automatic Music Transcription

Task: Convert a music recording into sheet music
(or another symbolic music representation)



Multitask Learning for estimating

- pitches,
- note onsets & offsets,
- beat & measure positions,
- musical voices & instrumentation,
- pedalling, dynamics, …



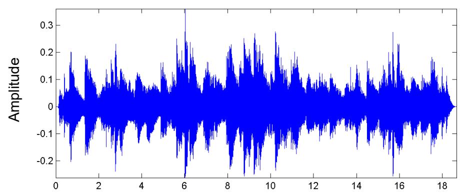
Why is Music Processing Challenging?

Example: Chopin, Mazurka Op. 63 No. 3

Why is Music Processing Challenging?

Example: Chopin, Mazurka Op. 63 No. 3

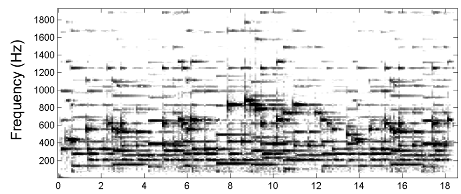
- Waveform



Why is Music Processing Challenging?

Example: Chopin, Mazurka Op. 63 No. 3

- Waveform / Spectrogram



Why is Music Processing Challenging?

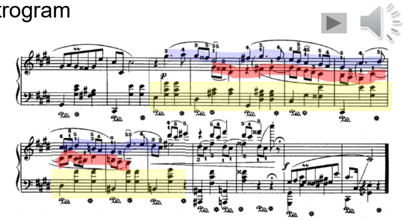
Example: Chopin, Mazurka Op. 63 No. 3

- 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
- Polyphony



Main Melody
Additional melody line
Accompaniment

Source Separation

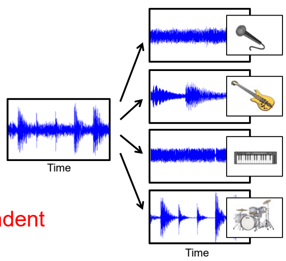
- Decomposition of audio stream into different sound sources
- Central task in digital signal processing
- "Cocktail party effect"

Source Separation

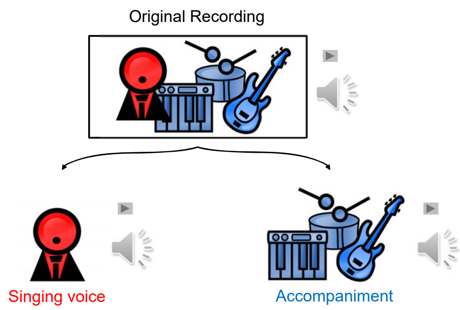
- Decomposition of audio stream into different sound sources
- Central task in digital signal processing
- "Cocktail party effect"
- Several input signals
- Sources are assumed to be statistically independent

Source Separation (Music)

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

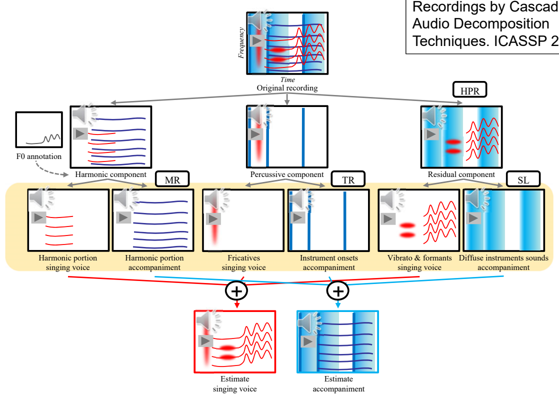


Singing Voice Extraction



Singing Voice Extraction

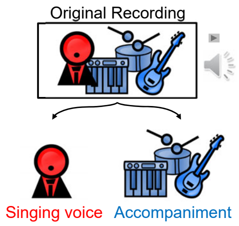
Traditional Approach
Driedger, Müller: Extracting Singing Voice from Music Recordings by Cascading Audio Decomposition Techniques. ICASSP 2015.



Singing Voice Extraction

DL-Based Approach
Stöter, Uhlich Luitkus, Mitsufuji: Open-Unmix – A Reference Implementation for Music Source Separation. JOSS 2019.

Deep learning has lead to breakthrough



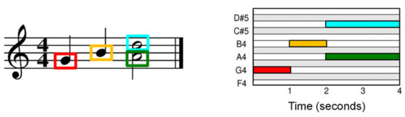
Reference voices:

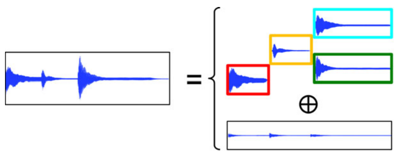
Engineering approach:

Deep learning approach:

Score-Informed Audio Decomposition

Exploit musical score to support decomposition process



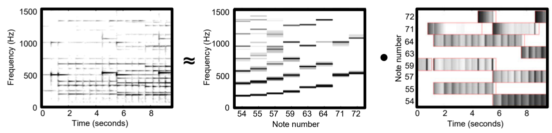


Prior Knowledge
Ewert, Pardo, Müller, Plumbley: Score-Informed Source Separation for Musical Audio Recordings. IEEE SPM, 2014.

Score-Informed Audio Decomposition

Exploit musical score to support decomposition process

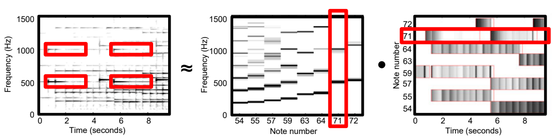
NMF-based spectrogram decomposition



Score-Informed Audio Decomposition

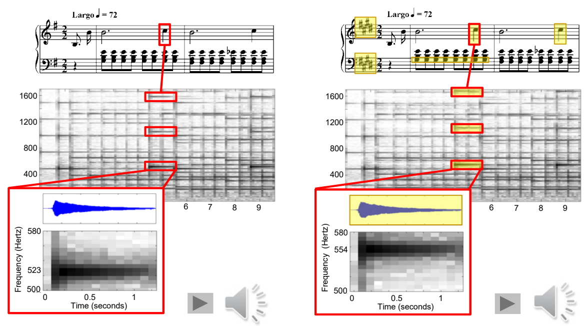
Exploit musical score to support decomposition process

NMF-based spectrogram decomposition

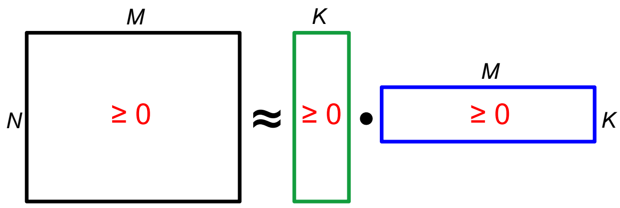


Score-Informed Audio Decomposition

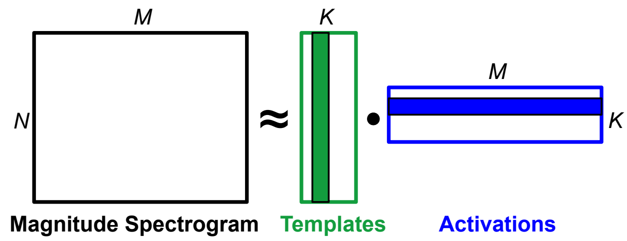
Application: Audio editing



NMF (Nonnegative Matrix Factorization)



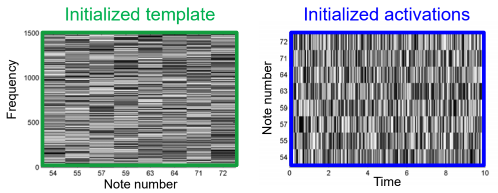
NMF (Nonnegative Matrix Factorization)



Templates: **Pitch + Timbre** **"How does it sound"**

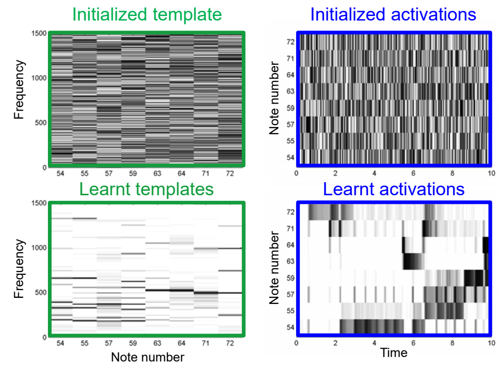
Activations: **Onset time + Duration** **"When does it sound"**

NMF-Decomposition



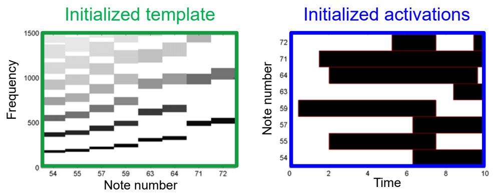
Random initialization

NMF-Decomposition



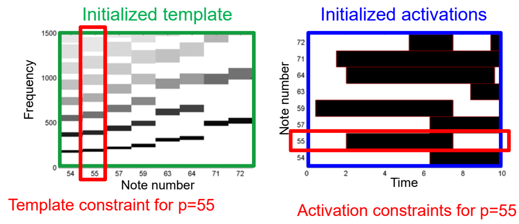
Random initialization → No semantic meaning

NMF-Decomposition



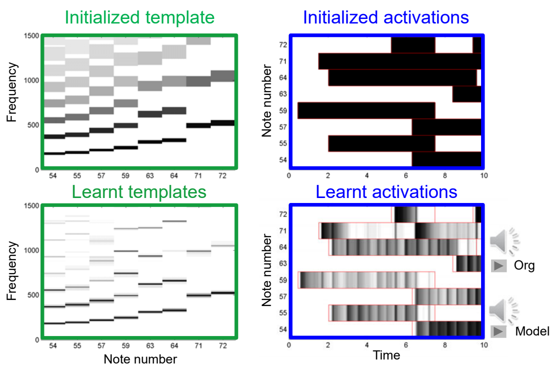
Constrained initialization

NMF-Decomposition



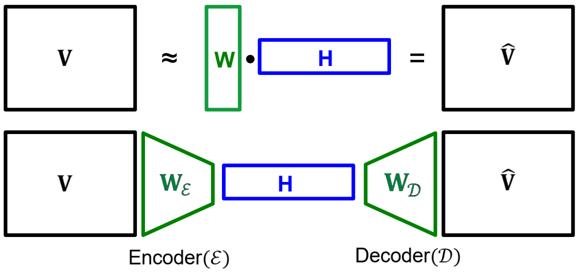
Constrained initialization

NMF-Decomposition



Constrained initialization → NMF as refinement

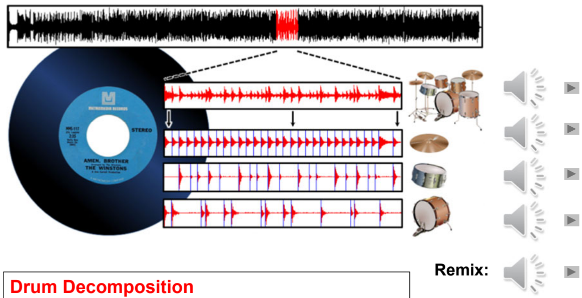
NMF-Decomposition



NMF as Autoencoder
Smaragdis, Venkataramani: A Neural Network Alternative to Non-Negative Audio Models. ICASSP 2017.

Constraint Autoencoders
Ewert, Sandler: Structured dropout for weak label and multi-instance learning and its application to score-informed source separation. ICASSP 2017

Informed Drum-Sound Decomposition



Drum Decomposition
Dittmar, Müller: Reverse Engineering the Amen Break – Score-Informed Separation and Restoration Applied to Drum Recordings. IEEE/ACM TASLP, 2016.

Informed Drum-Sound Decomposition

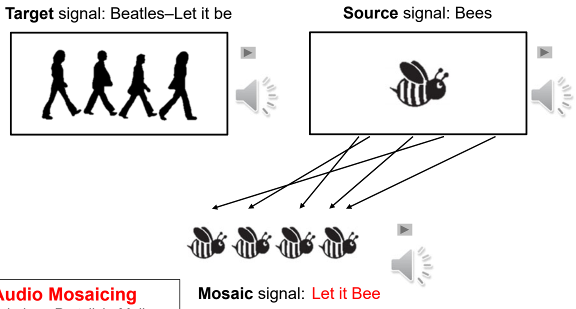
Major challenge: Reconstructed sound events often have artifacts

Approaches:

- Resynthesize certain sound components
- Differentiable Digital Signal Processing (DDSP) combines classical DSP and deep learning
- Generative adversarial networks may help to reduce the artifacts

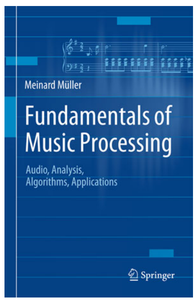
DDSP
Engel et al.: DDSP: Differentiable Digital Signal Processing. ICLR 2020.

Audio Mosaicing



Audio Mosaicing
Driedger, Prätzlich, Müller: Let It Bee – Towards NMF-Inspired Audio Mosaicing. ISMIR 2015.

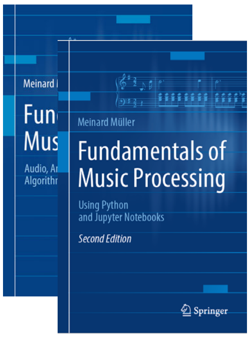
Fundamentals of Music Processing (FMP)



Meinard Müller
Fundamentals of Music Processing
Audio, Analysis, Algorithms, Applications
Springer, 2015

Accompanying website:
www.music-processing.de

Fundamentals of Music Processing (FMP)



Meinard Müller
Fundamentals of Music Processing
Audio, Analysis, Algorithms, Applications
Springer, 2015

Accompanying website:
www.music-processing.de

2nd edition
Meinard Müller
Fundamentals of Music Processing
Using Python and Jupyter Notebooks
Springer, 2021

Fundamentals of Music Processing (FMP)

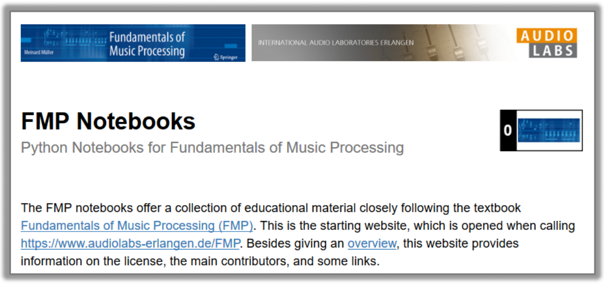
Chapter	Music Processing Scenario
1	Music Representations
2	Fourier Analysis of Signals
3	Music Synchronization
4	Music Structure Analysis
5	Chord Recognition
6	Tempo and Beat Tracking
7	Content-Based Audio Retrieval
8	Musically Informed Audio Decomposition

Meinard Müller
Fundamentals of Music Processing
Audio, Analysis, Algorithms, Applications
Springer, 2015

Accompanying website:
www.music-processing.de

2nd edition
Meinard Müller
Fundamentals of Music Processing
Using Python and Jupyter Notebooks
Springer, 2021

FMP Notebooks: Education & Research



https://www.audiolabs-erlangen.de/FMP