Audio Decomposition

Meinard Müller, Christof Weiß

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

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

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
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
Harmonic-Percussive Decomposition

Mixture
Harmonic-Percussive Decomposition

Mixture:
- Clearly harmonic sounds
- Clearly percussive sounds

Harmonic component
Percussive component
Harmonic-Percussive Decomposition

Mixture

Clearly harmonic sounds

Harmonic component

Residual component

Clearly percussive sounds

Percussive component
Harmonic-Percussive Decomposition

Mixture:
- Clearly harmonic sounds of singing voice and accompaniment
- Drum hits
- Fricatives & plosives in singing voice
- Noise-like sounds
- Vibrato/glissando sounds

Harmonic component

Residual component

Percussive component

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

Original recording

Harmonic component

MR

Harmonic portion

singing voice

HR

Harmonic portion

accompaniment

TR

Percussive component

Vibrato & formants

singing voice

Estimate

singing voice

Estimate

accompaniment

HPR

Residual component

SL

Instrument onsets

accompaniment

Diffuse instruments sounds

component
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

Estimate → Parameters

≈

Time (seconds)
Frequency (Hz)

≈

Time (seconds)
Frequency (Hz)
NMF (Nonnegative Matrix Factorization)
NMF (Nonnegative Matrix Factorization)

**Magnitude Spectrogram**  \( \approx \)  **Templates**  \( \approx \)  **Activations**

**Templates:** Pitch + Timbre

**Activations:** Onset time + Duration

“How does it sound”

“How when does it sound”
NMF-Decomposition

Initialized template

Initialized activations

Random initialization
NMF-Decomposition

Initialized template

初始化模板

Learnt templates

学习模板

Initialized activations

初始化激活

Learnt activations

学习激活

Random initialization  →  No semantic meaning

随机初始化  →  无语义意义
NMF-Decomposition

Initialized template

Initialized activations

Constrained initialization
NMF-Decomposition

Template constraint for $p=55$

Activation constraints for $p=55$

Constrained initialization
NMF-Decomposition

Initialized template

Learnt templates

Initialized activations

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
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)

Non-negative matrix

\[ V \approx W \cdot H = WH \]

Components

Activations

Target’s spectrogram

Source’s spectrogram

Activations

Mosaic’s spectrogram

Proposed audio mosaicing approach

Target’s spectrogram

Source’s spectrogram

Activations

Mosaic’s spectrogram
NMF-Inspired Audio Mosaicing

Spectrogram target

≈

Spectrogram source

Activation matrix

Spectrogram mosaic

Frequency

Time target

Time source

Time target

Time source

Time target
NMF-Inspired Audio Mosaicing

Core idea: support the development of sparse diagonal activation structures
NMF-Inspired Audio Mosaicing

Spectrogram target

Spectrogram source

Activation matrix

Spectrogram mosaic

≈

Time target

Time source

Time target
NMF-Inspired Audio Mosaicing

\[ \text{Spectrogram target} \approx \text{Spectrogram source} \times \text{Activation matrix} \Rightarrow \text{Spectrogram mosaic} \]
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/