Beethoven, Bach, und Billionen Bytes

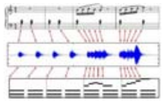
Musik trifft Informatik

Meinard Müller
International Audio Laboratories Erlangen

Informatik-Kolloquium
Universität Bamberg
23. Oktober 2014

Meinard Müller

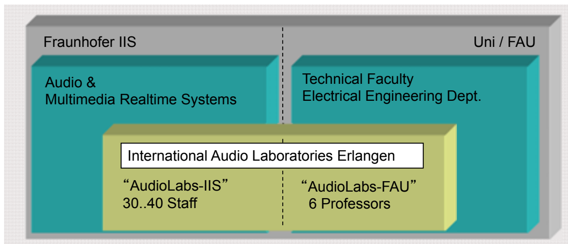
- 2007 Habilitation, Bonn
- 2007 MPI Informatik, Saarbrücken
  Senior Researcher
  Music Processing & Motion Processing



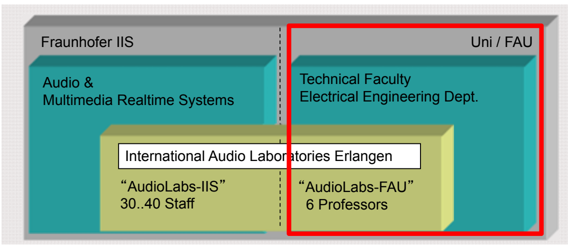
- 2012 W3-Professur, AudioLabs Erlangen
  Semantic Audio Processing

International Audio Laboratories Erlangen

International Audio Laboratories Erlangen



International Audio Laboratories Erlangen



International Audio Laboratories Erlangen

International Audio Laboratories Erlangen

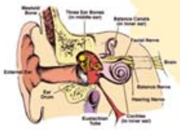
Audio

International Audio Laboratories Erlangen

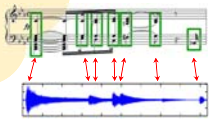
Audio Coding

3D Audio

Audio



Psychoacoustics

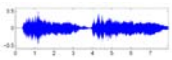


Music Processing

Music Representations

Sheet Music (Image)

CD / MP3 (Audio)

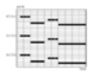


MusicXML (Text)

Dance / Motion (Mocap)

Music

MIDI



Singing / Voice (Audio)

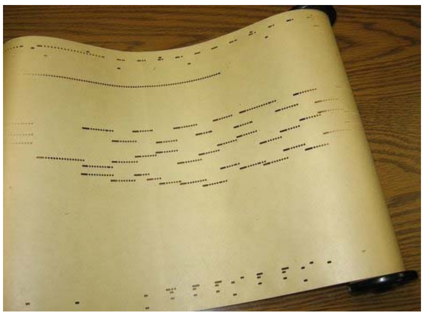
Music Film (Video)

Music Literature (Text)

Research Goals

- Music Information Retrieval (MIR) → ISMIR
- Analysis of music signals (harmonic, melodic, rhythmic, motivic aspects)
- Design of musically relevant audio features
- Tools for multimodal search and interaction

Piano Roll Representation

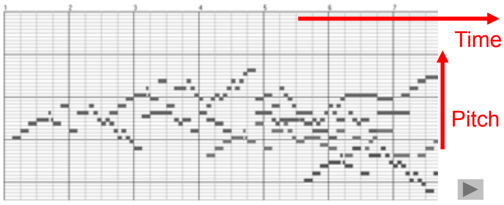


Player Piano (1900)

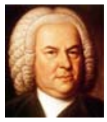
Piano Roll Representation (MIDI)

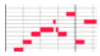
J.S. Bach, C-Major Fuge

(Well Tempered Piano, BWV 846)

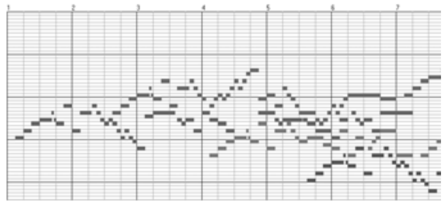


Piano Roll Representation (MIDI)

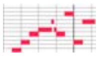


Query: 

Goal: Find all occurrences of the query

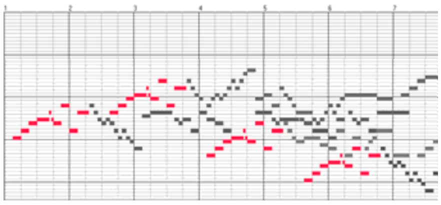


Piano Roll Representation (MIDI)

Query: 

Goal: Find all occurrences of the query

Matches:



Audio Data

Various interpretations – Beethoven's Fifth

Bernstein
Karajan
Scherbakov (piano)
MIDI (piano)

Audio Data (Memory Requirements)

1 Bit	=	1: on, 0: off
1 Byte	=	8 Bits
1 Kilobyte (KB)	=	1 Thousand Bytes
1 Megabyte (MB)	=	1 Million Bytes
1 Gigabyte (GB)	=	1 **Billion Bytes**
1 Terabyte (TB)	=	1000 Billion Bytes

Two audio CDs	>	1 **Billion Bytes**
1000 audio CDs	≈	**Billions of Bytes**
12.000 MIDI files	<	350 MB

Music Synchronization: Audio-Audio

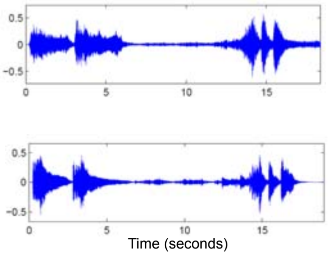
Beethoven's Fifth

Music Synchronization: Audio-Audio

Beethoven's Fifth

Orchester (Karajan)

Piano (Scherbakov)

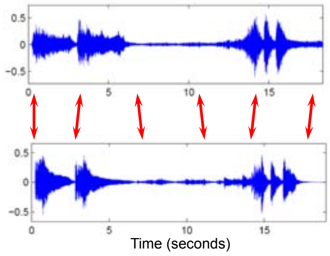


Music Synchronization: Audio-Audio

Beethoven's Fifth

Orchester (Karajan)

Piano (Scherbakov)

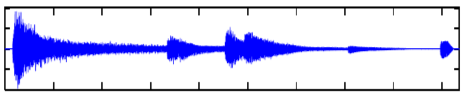


Application: Interpretation Switcher

Music Synchronization: Image-Audio

Image

Audio

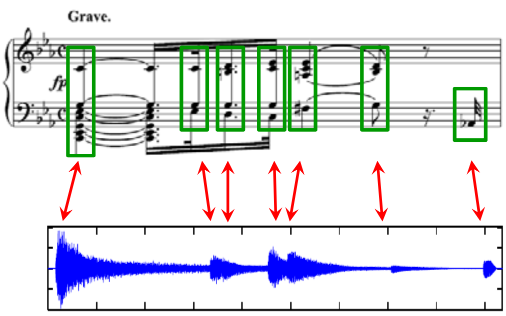


Time (seconds)

Music Synchronization: Image-Audio

Image

Audio

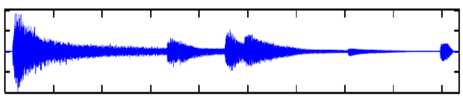


Time (seconds)

How to make the data comparable?

Image

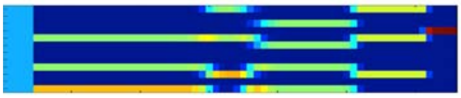
Audio



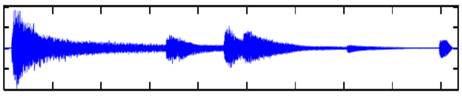
How to make the data comparable?

Image Processing: Optical Music Recognition

Image



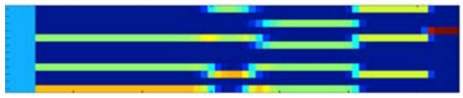
Audio



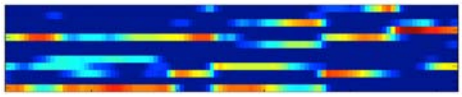
How to make the data comparable?

Image Processing: Optical Music Recognition

Image



Audio



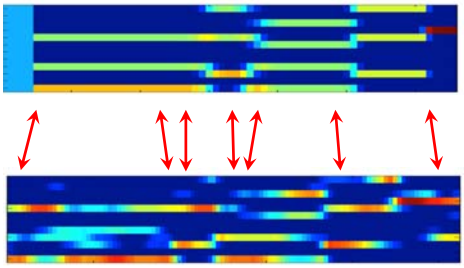
Audio Processing: Fourier Analyse

How to make the data comparable?

Image Processing: Optical Music Recognition

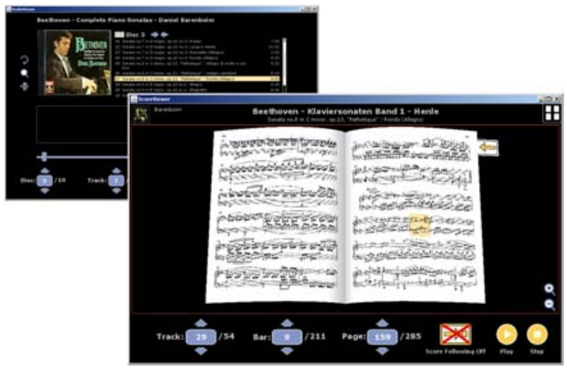
Image

Audio



Audio Processing: Fourier Analyse

Application: Score Viewer



Feature Representation

General goal: Convert an audio recording into a mid-level representation that captures certain musical properties while supressing other properties.

- Timbre / Instrumentation
- Tempo / Rhythm
- Pitch / Harmony

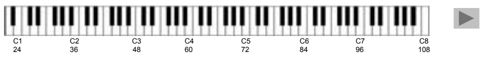
Feature Representation

General goal: Convert an audio recording into a mid-level representation that captures certain musical properties while supressing other properties.

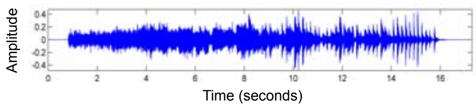
- Timbre / Instrumentation
- Tempo / Rhythm
- Pitch / Harmony

Feature Representation

Example: Chromatic scale

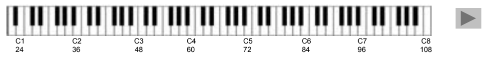


Waveform

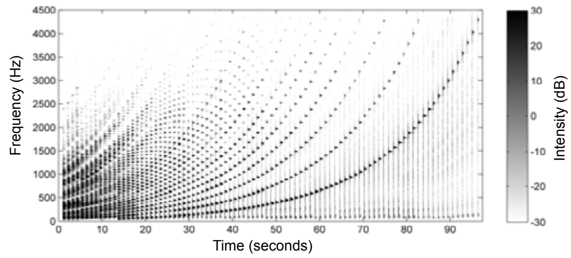


Feature Representation

Example: Chromatic scale

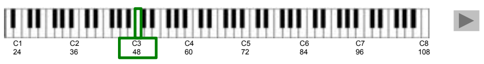


Spectrogram

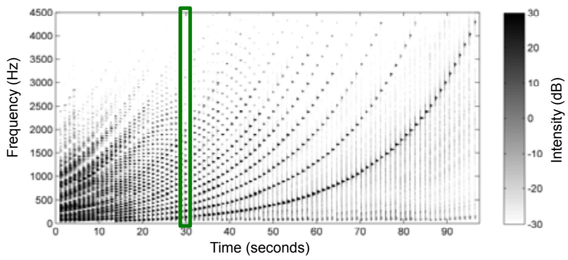


Feature Representation

Example: Chromatic scale

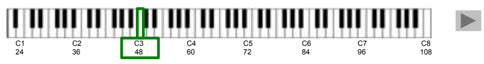


Spectrogram

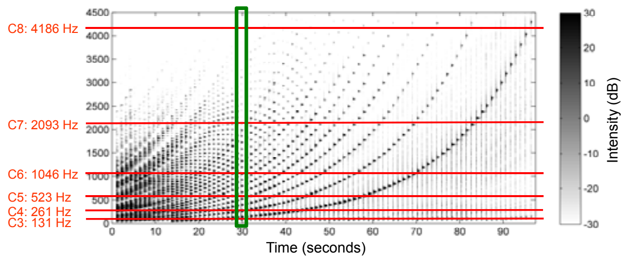


Feature Representation

Example: Chromatic scale

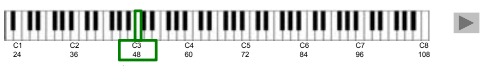


Spectrogram

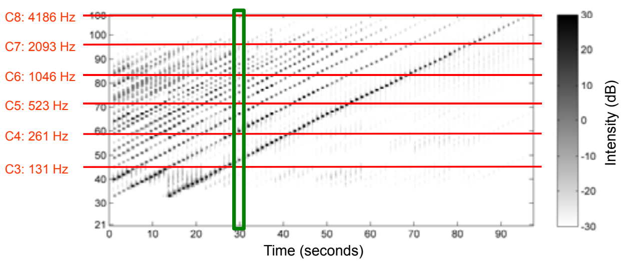


Feature Representation

Example: Chromatic scale

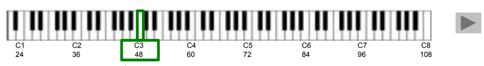


Log-frequency spectrogram

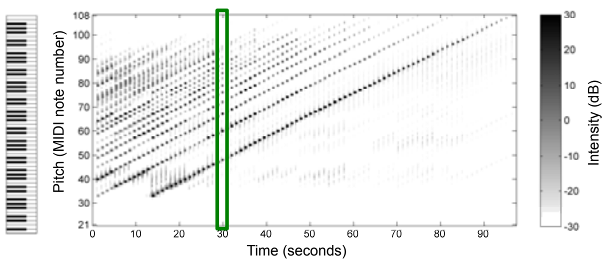


Feature Representation

Example: Chromatic scale

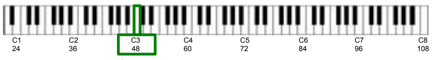


Log-frequency spectrogram

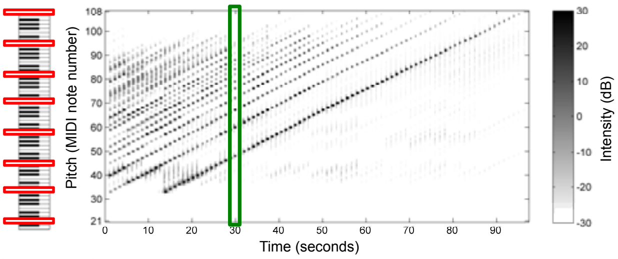


Feature Representation

Example: Chromatic scale



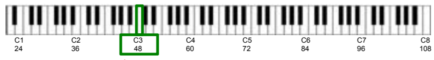
Log-frequency spectrogram



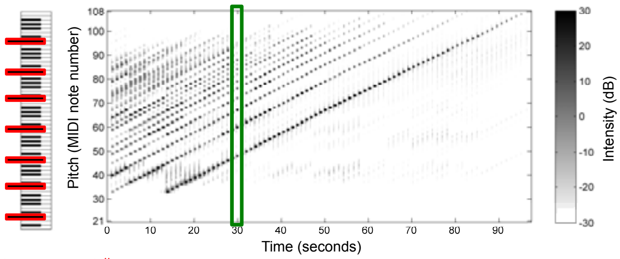
Chroma C

Feature Representation

Example: Chromatic scale



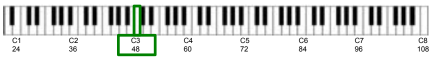
Log-frequency spectrogram



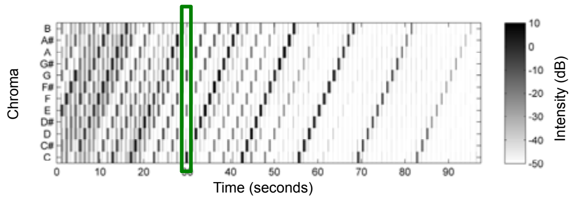
Chroma C#

Feature Representation

Example: Chromatic scale



Chroma representation



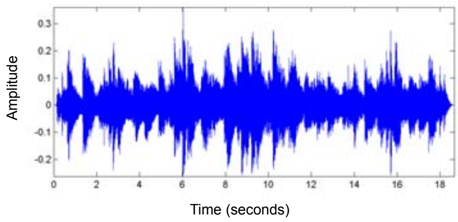
Why is Music Processing Challenging?

Example: Chopin, Mazurka Op. 63 No. 3

Why is Music Processing Challenging?

Example: Chopin, Mazurka Op. 63 No. 3

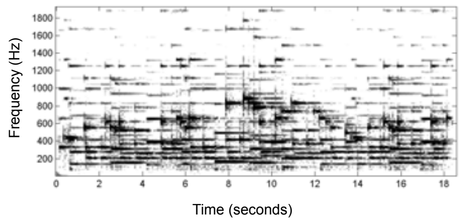
- Waveform



Why is Music Processing Challenging?

Example: Chopin, Mazurka Op. 63 No. 3

- Waveform / Spectrogram



Why is Music Processing Challenging?

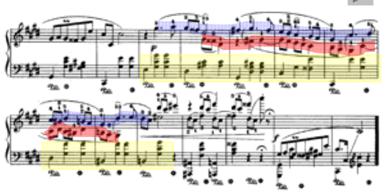
Example: Chopin, Mazurka Op. 63 No. 3

- 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

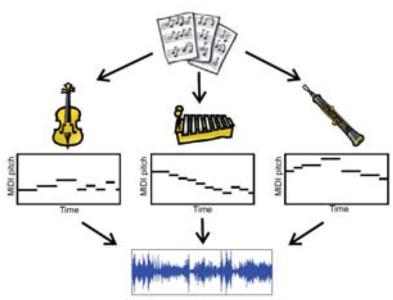
Source Separation

- Decomposition of audio stream into different sound sources
- Central task in digital signal processing
- "Cocktail party effect"
- Sources are often assumed to be statistically independent
- This is often not the case in music

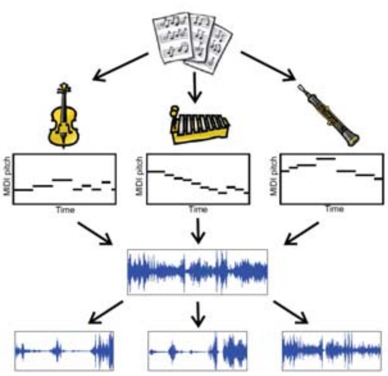
Strategy: Exploit additional information (e.g. musical score) to support the seperation process

Score-Informed Source Separation

Score-Informed Source Separation

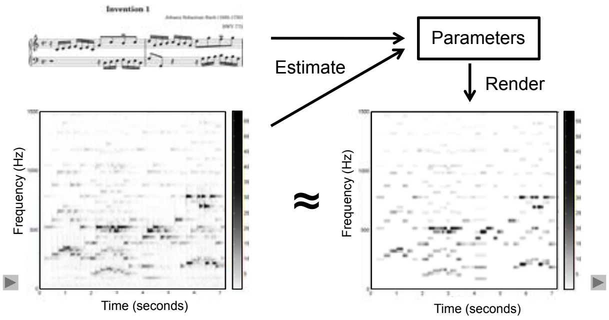


Score-Informed Source Separation

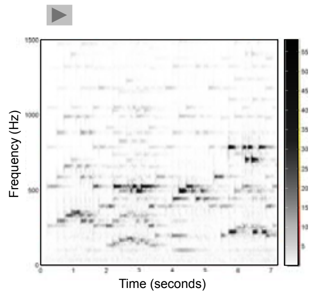


Score-Informed Source Separation

Goal: Approximate spectrogram using a parametric model exploiting availablity of score information

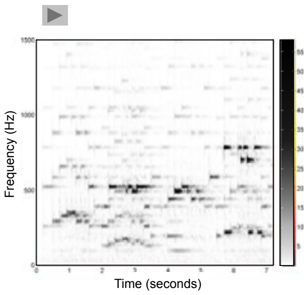


Score-Informed Source Separation

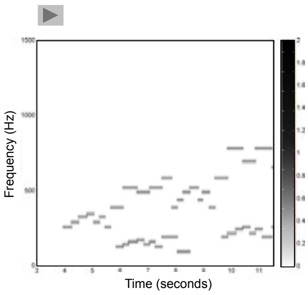


Original audio recording

Score-Informed Source Separation

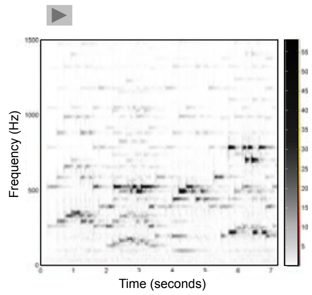


Original audio recording

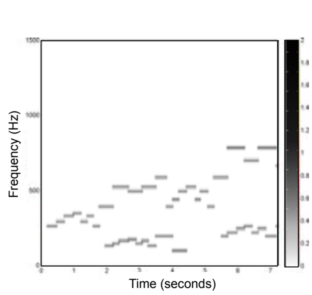


Model initialized with MIDI information

Score-Informed Source Separation

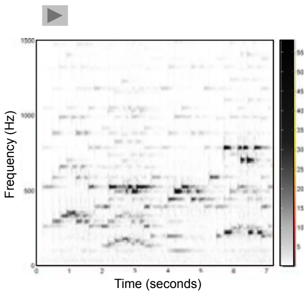


Original audio recording

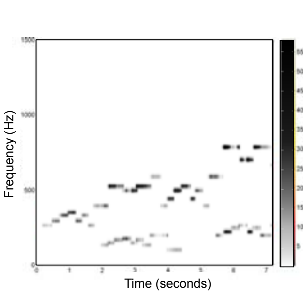


- Temporal synchronization
- Tuning estimation

Score-Informed Source Separation

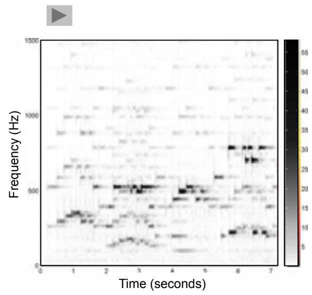


Original audio recording

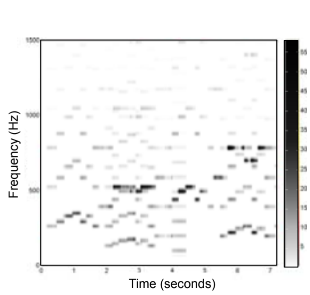


Activity parameters updated

Score-Informed Source Separation

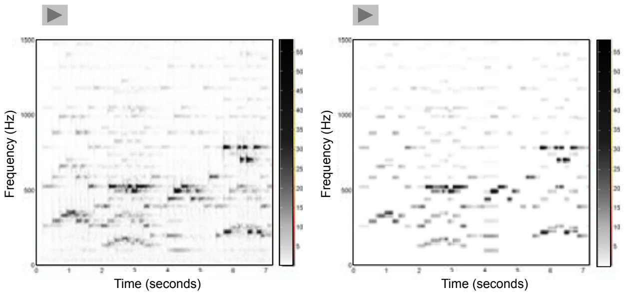


Original audio recording



- Partial energy distribution
- Resonance body properties

Score-Informed Source Separation



Original audio recording　　Model after three iterations

Note: Each note specified by the score parameterizes a portion of the spectrogram

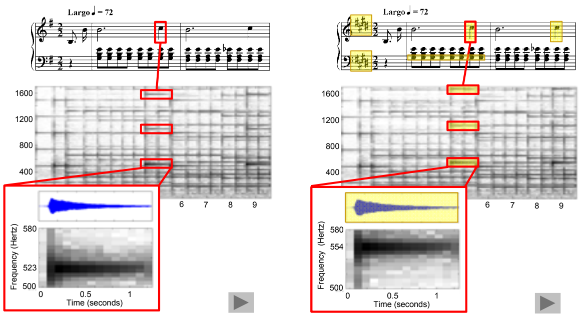
Score-Informed Source Separation

Experimental results for separating left and right hands for piano recordings:

Composer	Piece	Database	Results			
			L	R	Eq	Org
Bach	BWV 875, Prelude	SMD				
Chopin	Op. 28, No. 15	SMD				
Chopin	Op. 64, No. 1	European Archive				

Score-Informed Source Separation

Audio editing



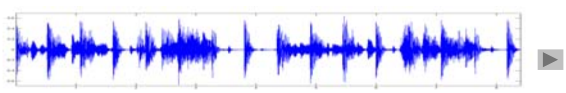
Tempo Estimation and Beat Tracking

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

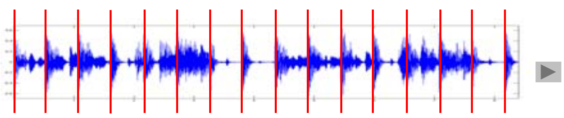


Time (seconds)

Tempo Estimation and Beat Tracking

Basic task: "Tapping the foot when listening to music"

Example: Queen – Another One Bites The Dust



Time (seconds)

Tempo Estimation and Beat Tracking

Example: Happy Birthday to you

Pulse level: **Measure**

Tempo Estimation and Beat Tracking

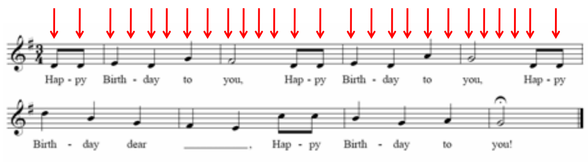
Example: Happy Birthday to you

Pulse level: **Tactus (beat)**

Tempo Estimation and Beat Tracking

Example: Happy Birthday to you

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



Tempo Estimation and Beat Tracking

Example: Chopin – Mazurka Op. 68-3

Pulse level: Quarter note

Tempo: ???

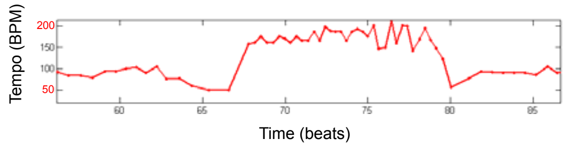
Tempo Estimation and Beat Tracking

Example: Chopin – Mazurka Op. 68-3

Pulse level: Quarter note

Tempo: 50-200 BPM

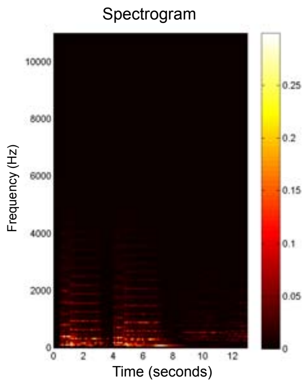
Tempo curve



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)

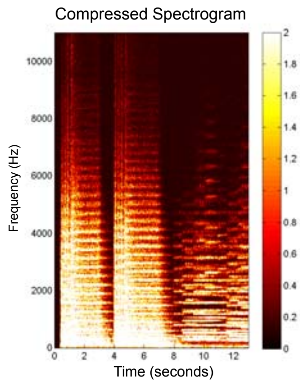
Tempo Estimation and Beat Tracking



Steps:

1. Spectrogram

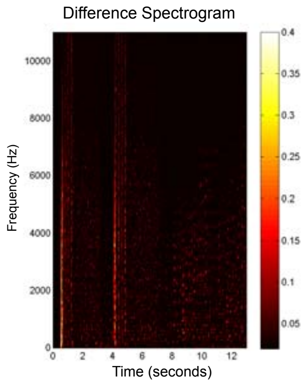
Tempo Estimation and Beat Tracking



Steps:

1. Spectrogram
2. Log Compression

Tempo Estimation and Beat Tracking



Steps:

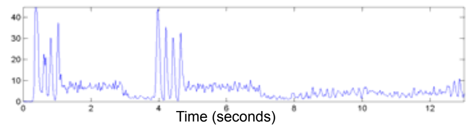
1. Spectrogram
2. Log Compression
3. Differentiation

Tempo Estimation and Beat Tracking

Steps:

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

Novelty Curve

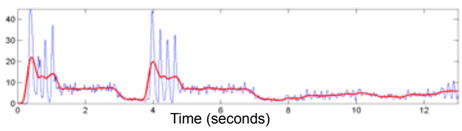


Tempo Estimation and Beat Tracking

Steps:

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

Novelty Curve
Local Average

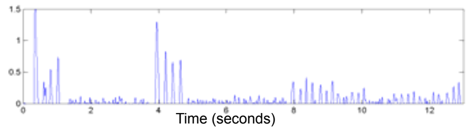


Tempo Estimation and Beat Tracking

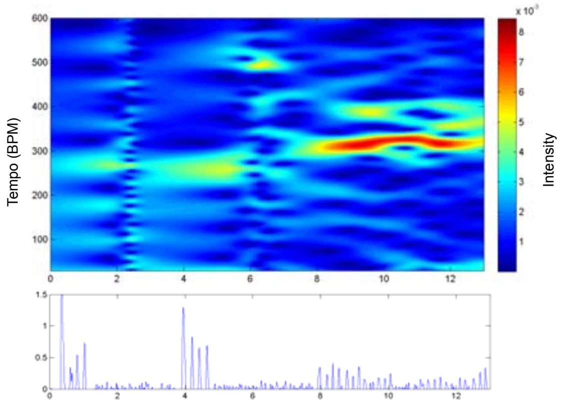
Steps:

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

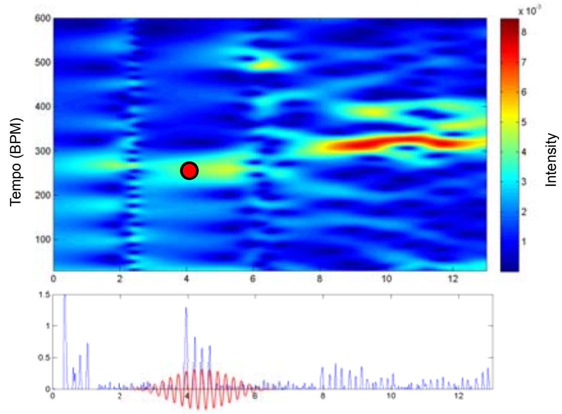
Novelty Curve



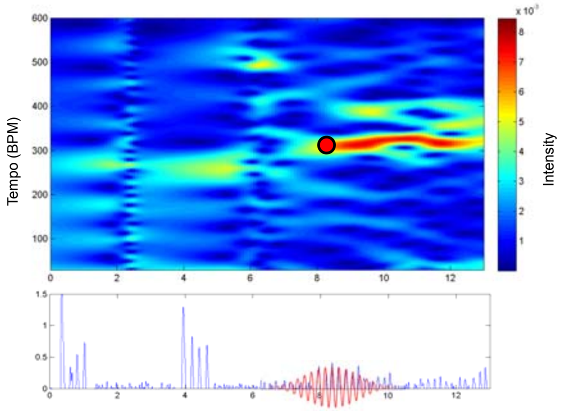
Tempo Estimation and Beat Tracking



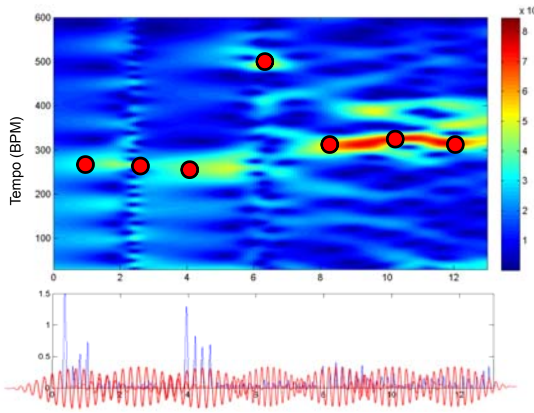
Tempo Estimation and Beat Tracking



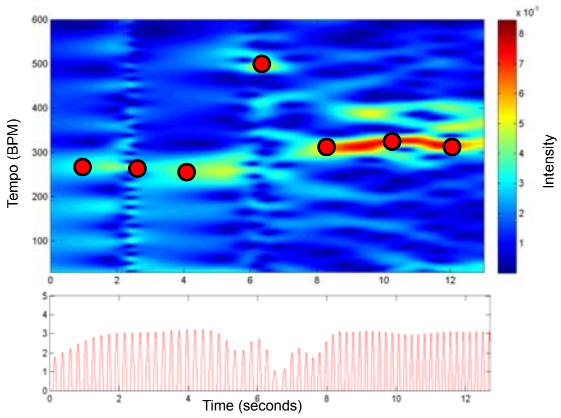
Tempo Estimation and Beat Tracking



Tempo Estimation and Beat Tracking

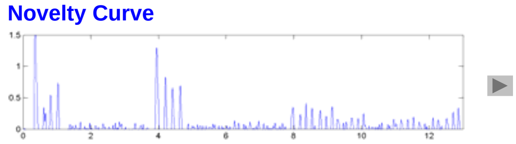


Tempo Estimation and Beat Tracking

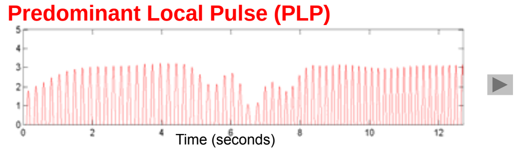


Tempo Estimation and Beat Tracking

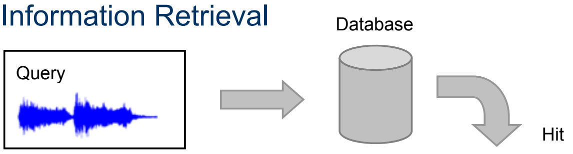
Novelty Curve



Predominant Local Pulse (PLP)



Information Retrieval



Audio-ID

Bernstein (1962)
Beethoven, Symphony No. 5

Version-ID

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

Category-ID

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

Information Retrieval

Musiksynchronisation

Audio-ID

Version-ID

Category-ID

Information Retrieval

Musiksynchronisation

Audio-ID

Version-ID

Category-ID

Similarity?

Relevance?

User!

Motivic Similarity

Beethoven's Fifth (1st Mov.)

Motivic Similarity

Beethoven's Fifth (1st Mov.)

Beethoven's Fifth (3rd Mov.)

Motivic Similarity

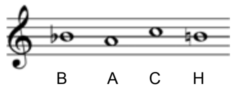
Beethoven's Fifth (1st Mov.)

Beethoven's Fifth (3rd Mov.)

Beethoven's Appassionata

Motivic Similarity

Motivic Similarity



Music Processing

Computer Sciene

Information Retrieval
Pattern Matching
Multimedia
User Interfaces

EEI

Signal Processing
Audio Processing
Computational Acoustics
Sensors

Humanities

Music Analysis
Performance Analysis
Music Education

Thanks

- Michael Clausen (Bonn University)
- David Damm (Bonn University)
- Jonathan Driedger (Universität Erlangen-Nürnberg)
- Sebastian Ewert (Bonn University)
- Christian Fremerey (Bonn University)
- Peter Grosche (Saarland University)
- Nanzhu Jiang (Universität Erlangen-Nürnberg)
- Verena Konz (Saarland University)
- Frank Kurth (Fraunhofer-FKIE, Wachtberg)
- Thomas Prätzlich (Universität Erlangen-Nürnberg)
- Verena Thomas (Bonn University)

Projekte & Kooperationen

- DFG-Projekt: METRUM
  Mehrschichtige Analyse und Strukturierung von Musiksignalen
  Kooperation: Michael Clausen
  Laufzeit: 2011-2015
- BMBF-Projekt: Freischütz Digital
  Freischütz Digital – Paradigmatische Umsetzung eines genuin digitalen Editionskonzepts
  Kooperation: Joachim Veit, Thomas Betzwieser, Gerd Szwillus
  Laufzeit: 2012-2015
- DFG-Projekt: SIAMUS
  Notentext-Informierte Audioparametrisierung von Musiksignalen
  Laufzeit: 2014-2017
- Projekt Musikwissenschaften
  Computergestützte Analyse harmonischer Strukturen
  Kooperation: Rainer Kleinertz
  Laufzeit: 2015-2018

Selected Publications (Music Processing)

- M. Müller, N. Jiang, P. Grosche (2013):
  A robust fitness measure for capturing repetitions in music recordings with applications to audio thumbnailing.
  IEEE Trans. on Audio, Speech & Language Processing, Vol. 21, No. 3, pp. 531–543.
- M. Müller, P.W. Ellis, A. Klapuri, G. Richard (2011):
  Signal Processing for Music Analysis.
  IEEE Journal of Selected Topics in Signal Processing, Vol. 5, No. 6, pp. 1088-1110.
- P. Grosche and M. Müller (2011):
  Extracting Predominant Local Pulse Information from Music Recordings.
  IEEE Trans. on Audio, Speech & Language Processing, Vol. 19, No. 6, pp. 1688-1701.
- M. Müller and S. Ewert (2010):
  Towards Timbre-Invariant Audio Features for Harmony-Based Music.
  IEEE Trans. on Audio, Speech & Language Processing, Vol. 18, No. 3, pp. 649-662.
- F. Kurth, M. Müller (2008):
  Efficient Index-Based Audio Matching.
  IEEE Trans. Audio, Speech & Language Processing, Vol. 16, No. 2, 382-395.
- M. Müller (2007):
  Information Retrieval for Music and Motion.
  Monograph, Springer, 318 pages