INTERNATIONAL AUDIO LABORATORIES ERLANGEN

AUDIO LABS

Informed Feature Representations

Meinard Müller

FAST IMPACt All-hands Meeting

18-20 March, Oxford

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Fraunhofer IIS

Meinard Müller

- 2007 Habilitation, Bonn
- 2007 MPI Informatik, Saarbrücken
  Senior Researcher
  Music Processing & Motion Processing
- 2012 W3-Professur, AudioLabs Erlangen
  Semantic Audio Processing

Thanks

- Jonathan Driedger
- Thomas Prätzlich
- Harald Grohganz
- Tido Röder
- Sebastian Ewert

International Audio Laboratories Erlangen

Fraunhofer IIS

mp3

AUDIO LABS

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

International Audio Laboratories Erlangen

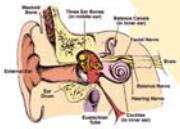
Audio

International Audio Laboratories Erlangen

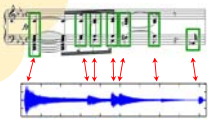
Audio Coding

3D Audio

Audio

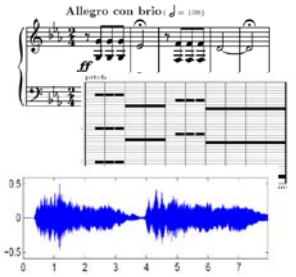


Psychoacoustics



Music Processing

Music and Motion



Overview

- Audio Features based on Chroma Information
  Application: Audio Matching
- Motion Features based on Geometric Relations
  Application: Motion Retrieval
- Musically Informed Audio Decomposition
  Application: Audio Editing

Overview

- Audio Features based on Chroma Information
  Application: Audio Matching
- Motion Features based on Geometric Relations
  Application: Motion Retrieval
- Musically Informed Audio Decomposition
  Application: Audio Editing

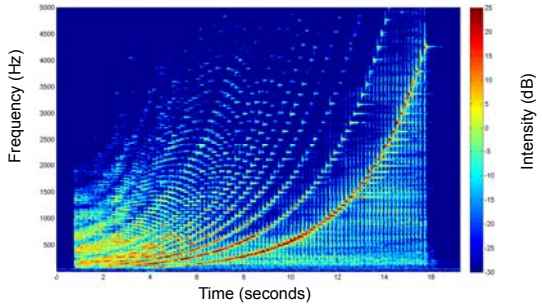
Chroma-based Audio Features

- Very popular in music signal processing
- Based equal-tempered scale of Western music
- Captures information related to harmony
- Robust to variations in instrumentation or timbre

Chroma-based Audio Features

Example: Chromatic scale

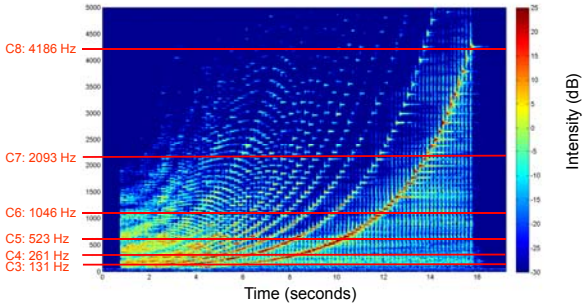
Spectrogram



Chroma-based Audio Features

Example: Chromatic scale

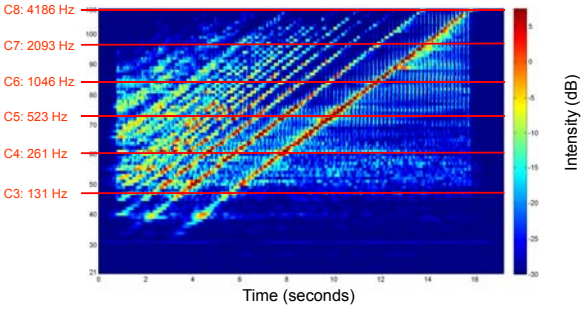
Spectrogram



Chroma-based Audio Features

Example: Chromatic scale

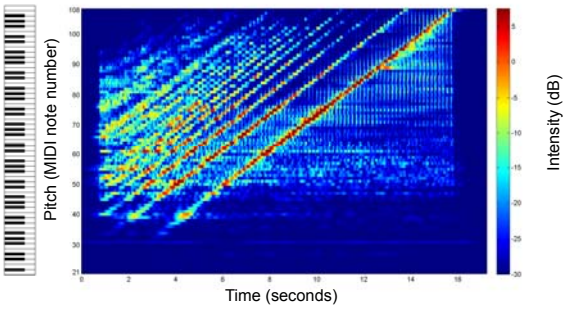
Log-frequency spectrogram



Chroma-based Audio Features

Example: Chromatic scale

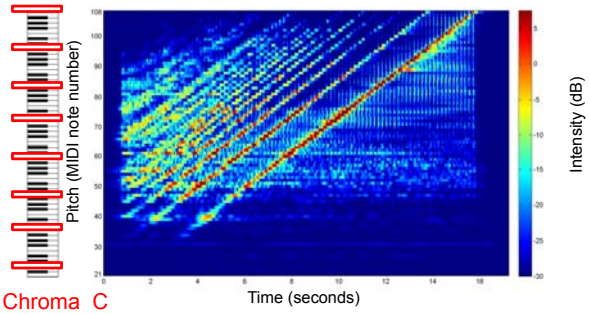
Log-frequency spectrogram



Chroma-based Audio Features

Example: Chromatic scale

Log-frequency spectrogram

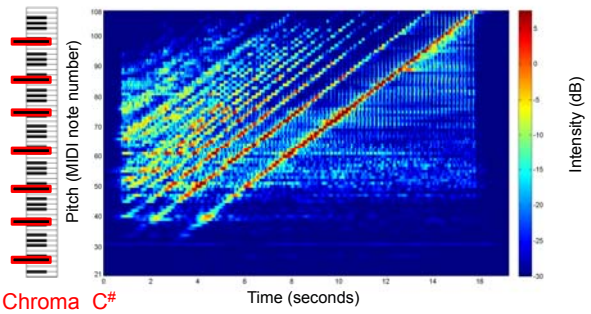


Chroma C

Chroma-based Audio Features

Example: Chromatic scale

Log-frequency spectrogram

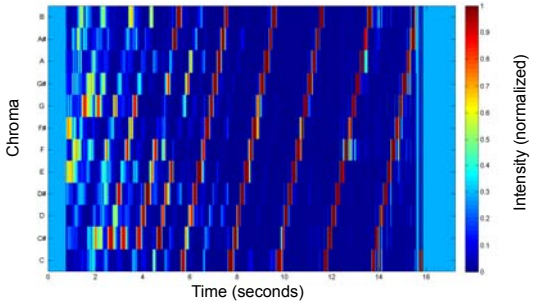


Chroma C#

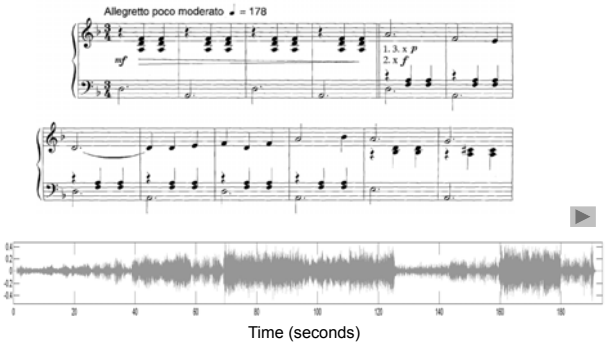
Chroma-based Audio Features

Example: Chromatic scale

Chroma representation (normalized, Euclidean)

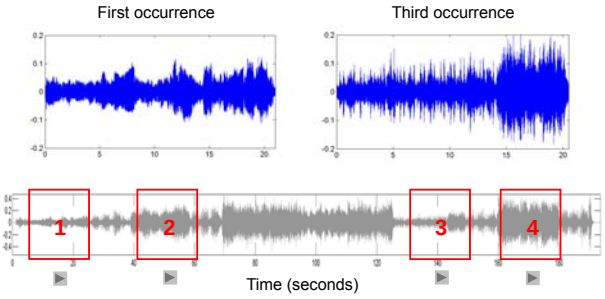


Motivation: Audio Matching

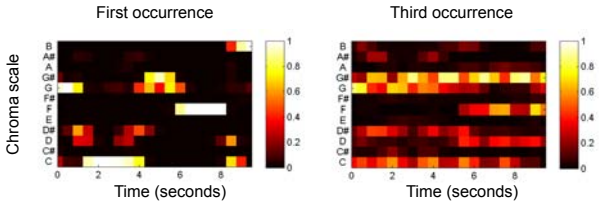


Motivation: Audio Matching

Four occurrences of the main theme

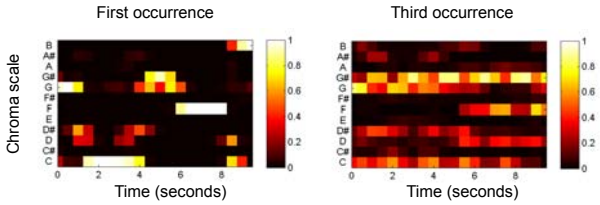


Chroma Features



Chroma Features

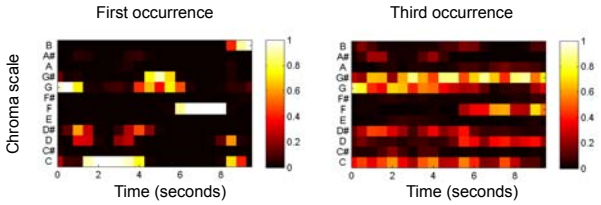
How to make chroma features more robust to timbre changes**?**



M. Müller and S. Ewert
Towards Timbre-Invariant Audio Features for Harmony-Based Music.
IEEE Trans. on Audio, Speech & Language Processing, Vol. 18, No. 3, pp. 649-662, 2010.

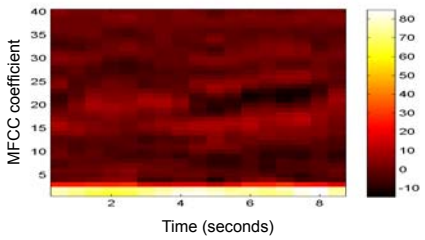
Chroma Features

How to make chroma features more robust to timbre changes**?**
Idea: Discard timbre-related information

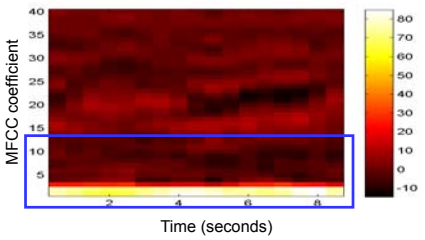


M. Müller and S. Ewert
Towards Timbre-Invariant Audio Features for Harmony-Based Music.
IEEE Trans. on Audio, Speech & Language Processing, Vol. 18, No. 3, pp. 649-662, 2010.

MFCC Features and Timbre

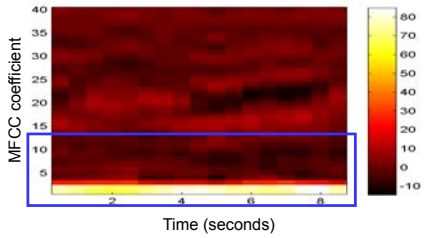


MFCC Features and Timbre



Lower MFCCs ⟷ Timbre

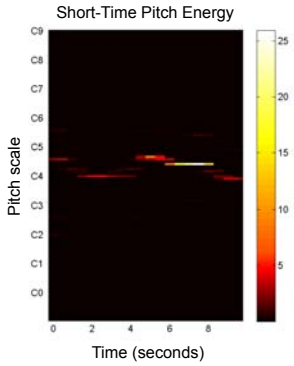
MFCC Features and Timbre



Lower MFCCs ⟷ Timbre

Idea: Discard lower MFCCs to achieve timbre invariance

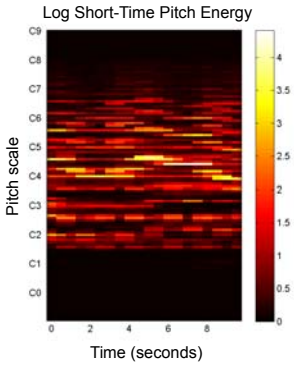
Enhancing Timbre Invariance



Steps:

1. Log-frequency spectrogram

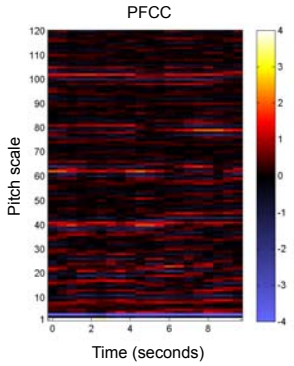
Enhancing Timbre Invariance



Steps:

1. Log-frequency spectrogram
2. Log (amplitude)

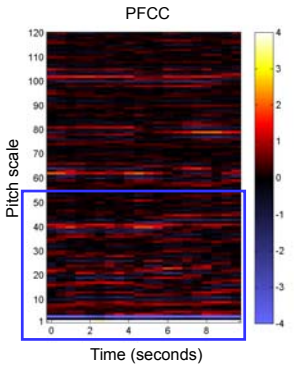
Enhancing Timbre Invariance



Steps:

1. Log-frequency spectrogram
2. Log (amplitude)
3. DCT

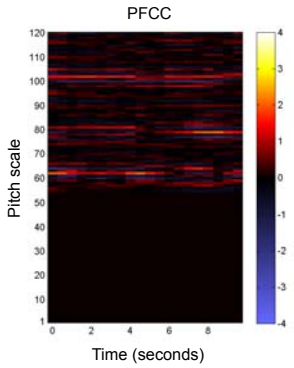
Enhancing Timbre Invariance



Steps:

1. Log-frequency spectrogram
2. Log (amplitude)
3. DCT
4. Discard lower coefficients [1:n-1]

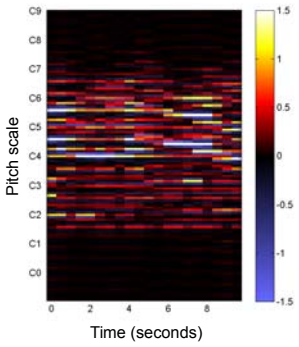
Enhancing Timbre Invariance



Steps:

1. Log-frequency spectrogram
2. Log (amplitude)
3. DCT
4. Keep upper coefficients [n:120]

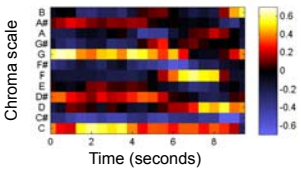
Enhancing Timbre Invariance



Steps:

1. Log-frequency spectrogram
2. Log (amplitude)
3. DCT
4. Keep upper coefficients [n:120]
5. Inverse DCT

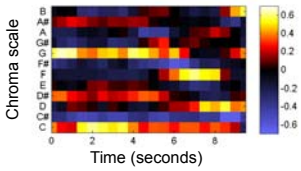
Enhancing Timbre Invariance



Steps:

1. Log-frequency spectrogram
2. Log (amplitude)
3. DCT
4. Keep upper coefficients [n:120]
5. Inverse DCT
6. Chroma & Normalization

Enhancing Timbre Invariance



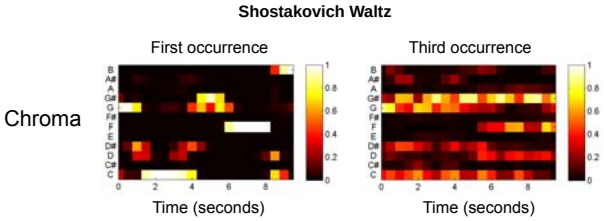
Steps:

1. Log-frequency spectrogram
2. Log (amplitude)
3. DCT
4. Keep upper coefficients [n:120]
5. Inverse DCT
6. Chroma & Normalization

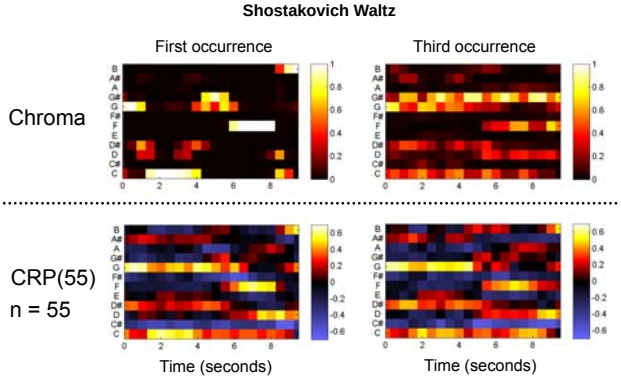
CRP(n)

Chroma DCT-**R**educed Log-**P**itch

Chroma versus CRP



Chroma versus CRP



Audio Analysis

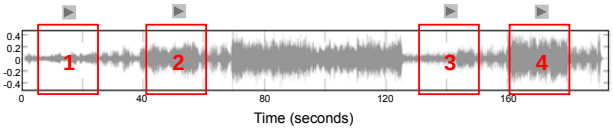
Idea:

Use “Audio Matching” for analyzing and understanding audio & feature properties:

- Relative comparison
- Compact
- Intuitive
- Quantitative evaluation

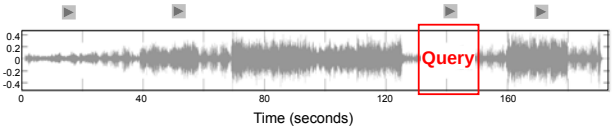
Audio Analysis

Example: Shostakovich, Waltz (Yablonsky)



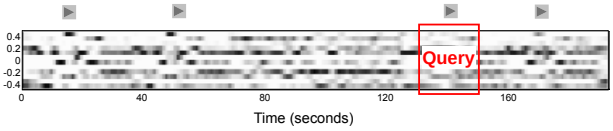
Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)



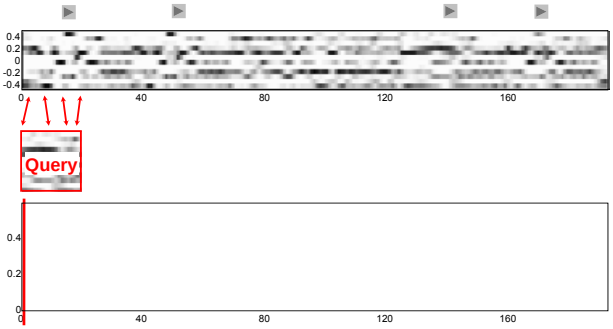
Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)



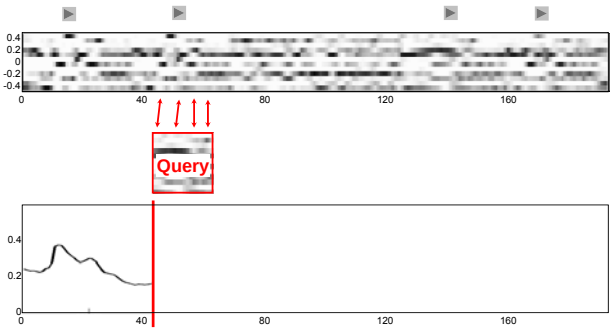
Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)



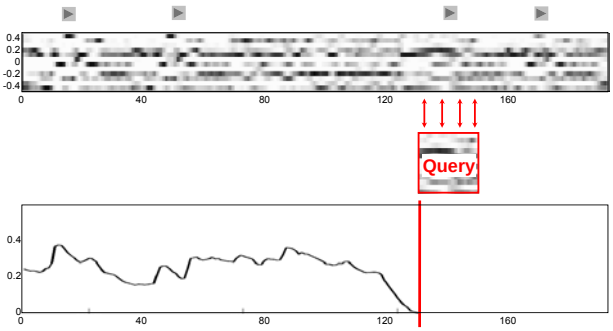
Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)



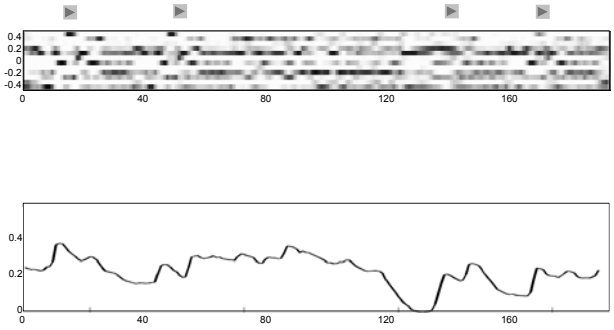
Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)



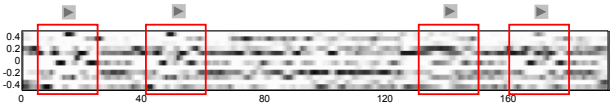
Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)

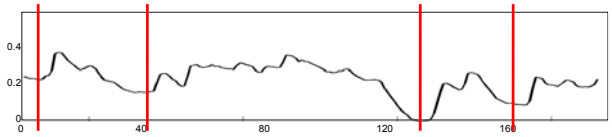


Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)



Expected matching positions (should have local minima)

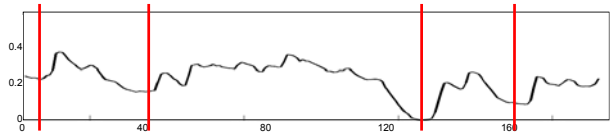


Audio Analysis

Idea:

- Use matching curve for analyzing feature properties

Expected matching positions (should have local minima)

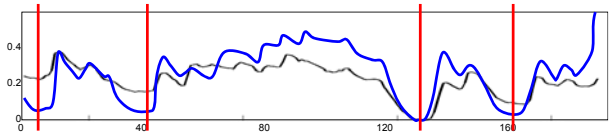


Audio Analysis

Idea:

- Use matching curve for analyzing feature properties
- Example: Chroma feature of higher timbre invariance

Expected matching positions (should have local minima)

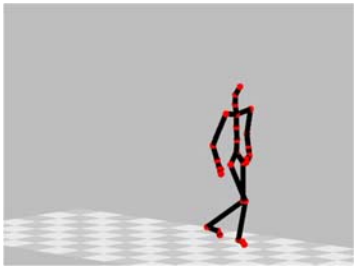


Overview

- Audio Features based on Chroma Information
  Application: Audio Matching
- Motion Features based on Geometric Relations
  Application: Motion Retrieval
- Musically Informed Audio Decomposition
  Application: Audio Editing

Motion Capture Data

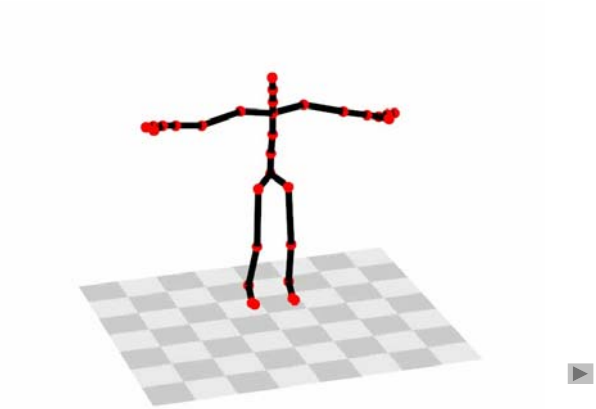
- 3D representations of motions
- Computer animation
- Sports
- Gait analysis



Motion Capture Data

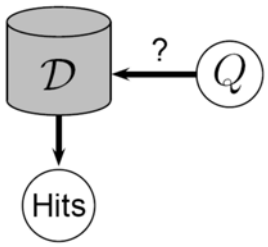
Optical System

Motion Capture Data

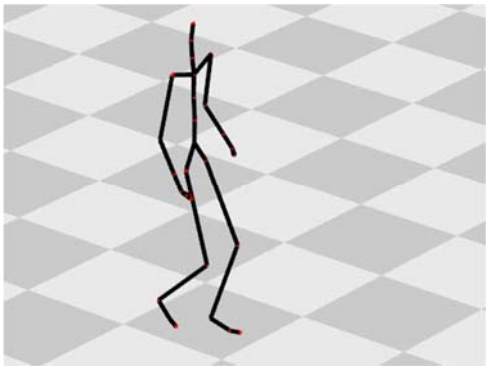


Motion Retrieval

- $\mathcal{D}$ = MoCap database
- Q = query motion clip
- Goal: find all motion clips in $\mathcal{D}$ similar to Q

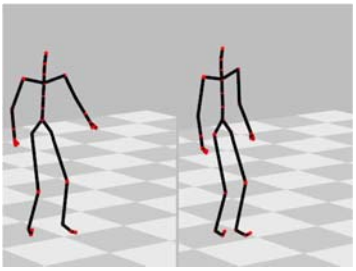


Motion Retrieval



Motion Retrieval

- Numerical similarity vs. logical similarity
- Logically related motions may exhibit significant spatio-temporal variations



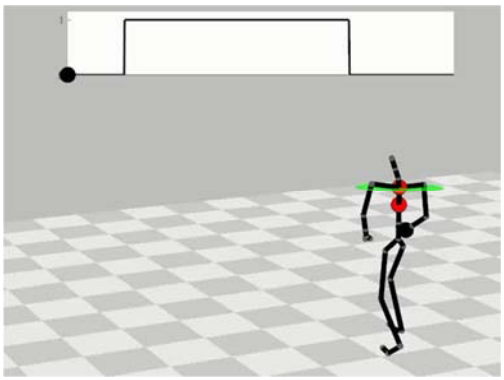
Relational Features

- Exploit knowledge of kinematic chain
- Express geometric relations of body parts
- Robust to motion variations

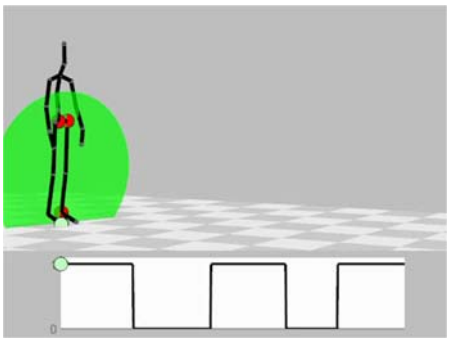
Meinard Müller, Tido Röder, and Michael Clausen
Efficient content-based retrieval of motion capture data.
ACM Transactions on Graphics (SIGGRAPH), vol. 24, pp. 677-685, 2005.

Meinard Müller and Tido Röder
Motion templates for automatic classification and retrieval of motion capture data.
Proceedings of the 2006 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA), Vienna, Austria, pp. 137-146, 2006.

Relational Features

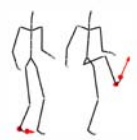


Relational Features

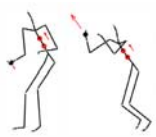


Relational Features

Right knee bent?

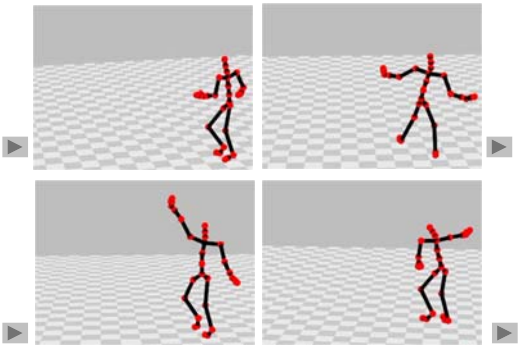


Right foot fast?

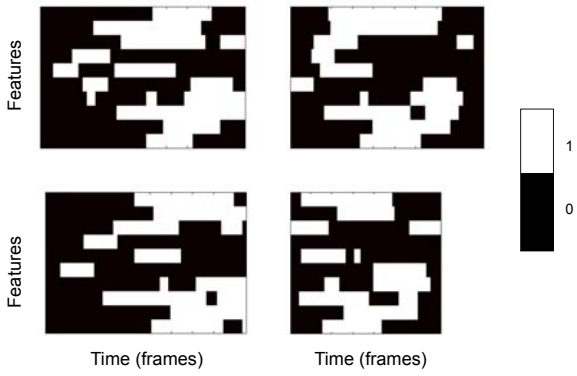


Right hand moving upwards?

Motion Templates (MT)

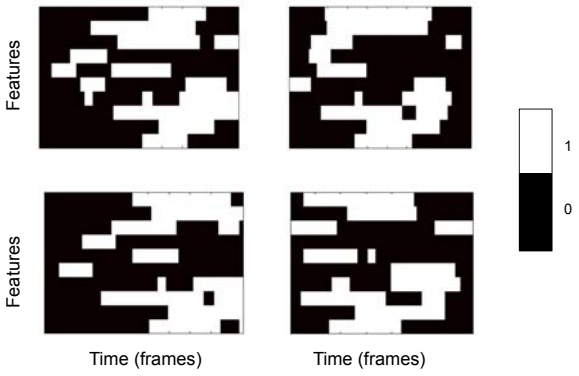


Motion Templates (MT)



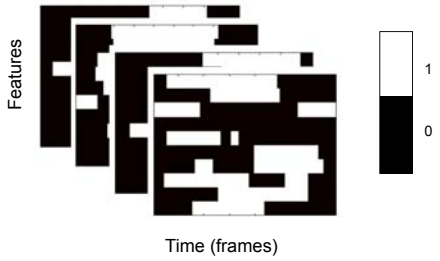
Motion Templates (MT)

Temporal alignment



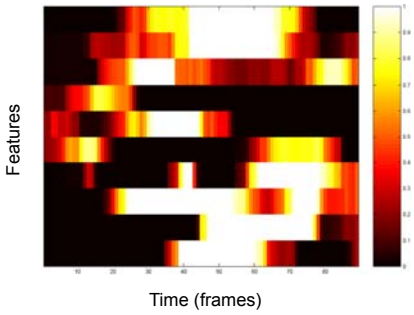
Motion Templates (MT)

Superimpose templates



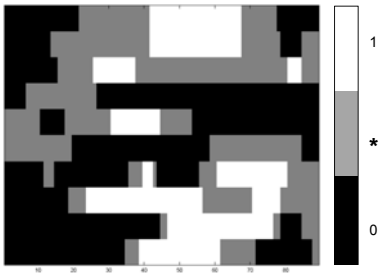
Motion Templates (MT)

Average template



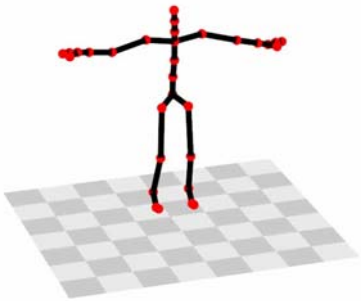
Motion Templates (MT)

Quantized template

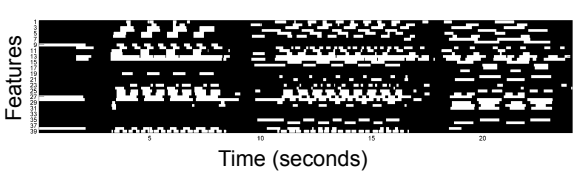


- Gray areas indicate inconsistencies / variations
- Achieve invariance by disregarding gray areas

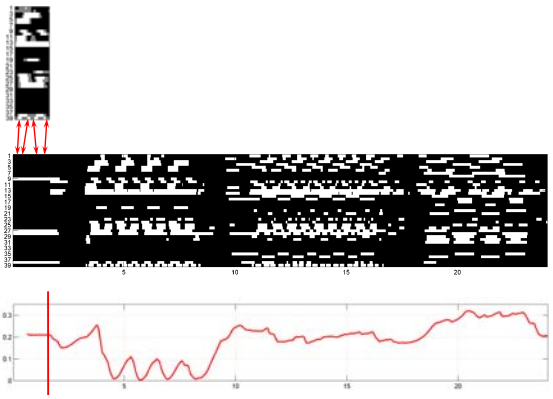
MT-based Motion Retrieval



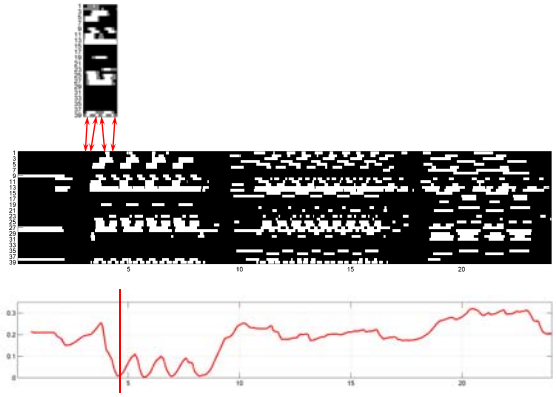
MT-based Motion Retrieval



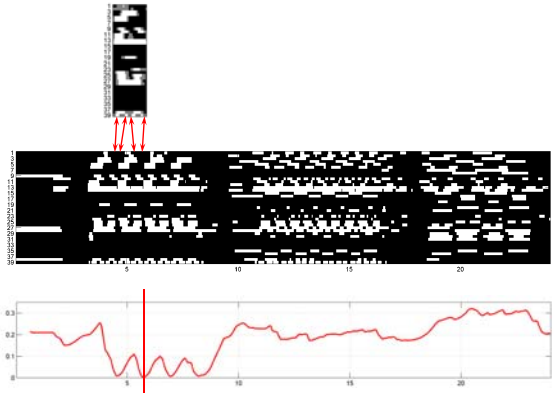
MT-based Motion Retrieval: Jumping Jack



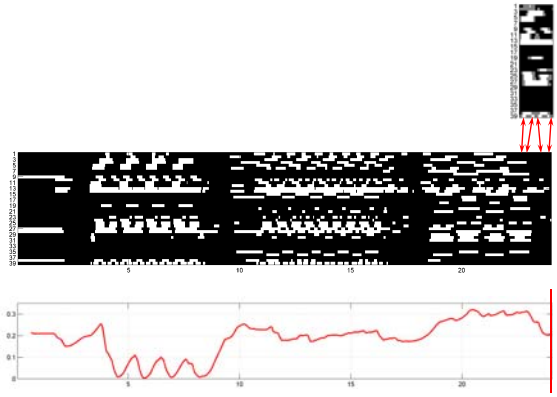
MT-based Motion Retrieval: Jumping Jack



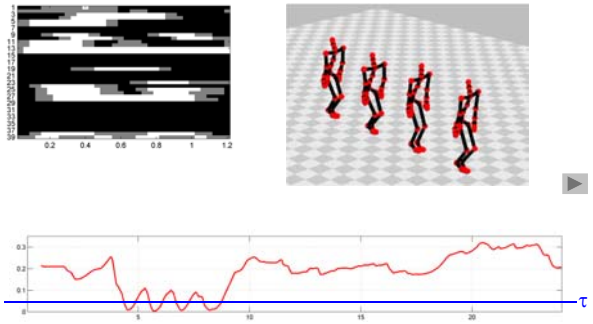
MT-based Motion Retrieval: Jumping Jack



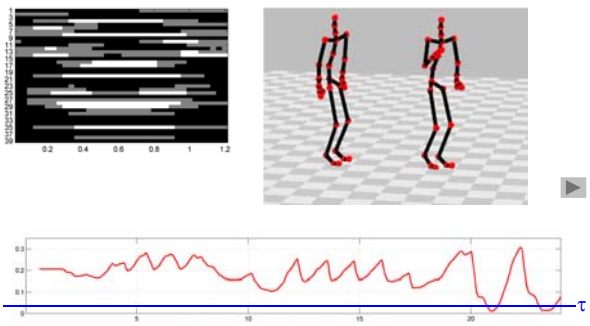
MT-based Motion Retrieval: Jumping Jack



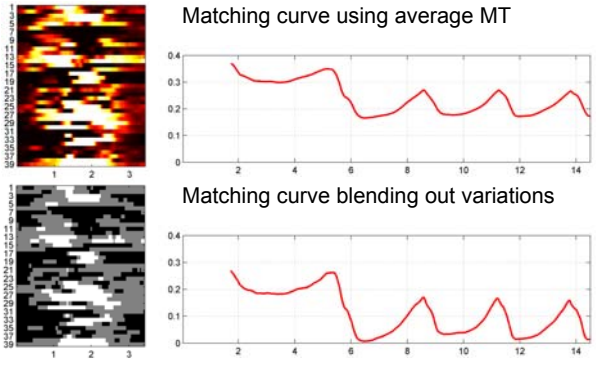
MT-based Motion Retrieval: Jumping Jack



MT-based Motion Retrieval: Elbow-To-Knee



MT-based Motion Retrieval: Cartwheel



Conclusions (Motion)

Informed Feature Representations

- Exploit model assumptions
  - Equal-tempered scale
  - Kinematic chain
- Deal with variances on feature level
  - Enhancing timbre invariance
  - Relational features
  - Quantized motion templates
- Consider requirements for specific application
  - Explicit information often not required
  - Mid-level features

Features with explicit meaning.

Makes subsequent steps more robust and efficient!

Avoid making problem harder as it is.

Overview

- Audio Features based on Chroma Information
  Application: Audio Matching
- Motion Features based on Geometric Relations
  Application: Motion Retrieval
- Musically Informed Audio Decomposition
  Application: Audio Editing

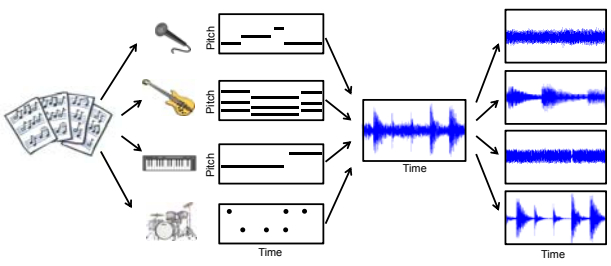
Musically Informed Audio Decomposition

- Extraction of main melody
- Separation of drum track
- Separation of instrumental voices
- Decomposition into individual note events
- Harmonic-percussive separation

Exploit musical knowledge to support decomposition process

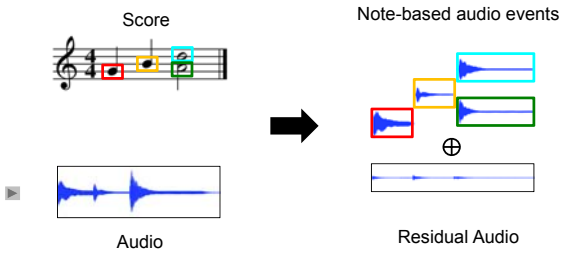
Score-Informed Source Separation

Exploit musical score to support separation process



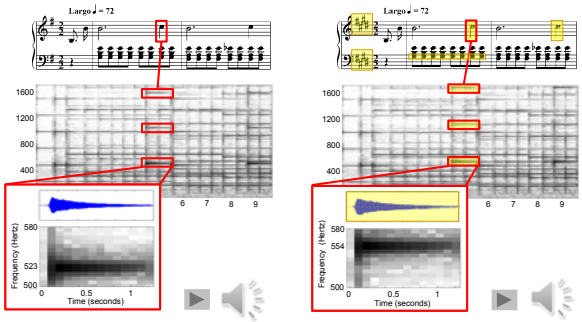
Score-Informed Audio Decomposition

Parameterize audio signal using score's note events

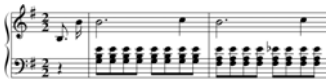


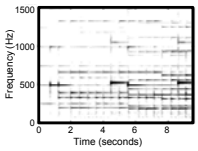
Score-Informed Audio Decomposition

Application: Audio editing

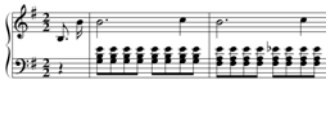


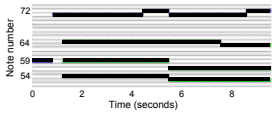
NMF-based Audio Decomposition

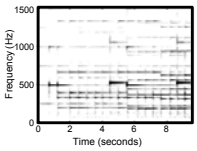




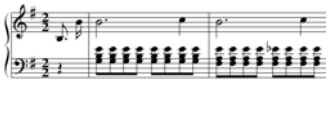
NMF-based Audio Decomposition

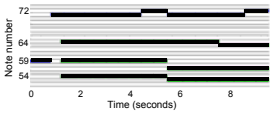


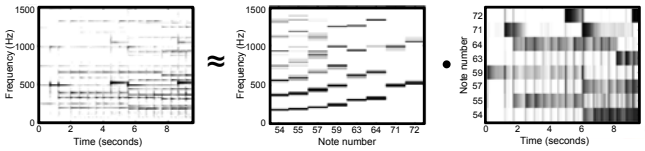




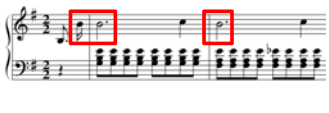
NMF-based Audio Decomposition

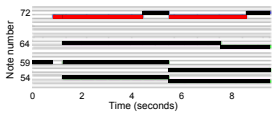


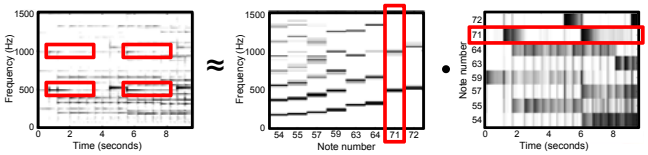




NMF-based Audio Decomposition



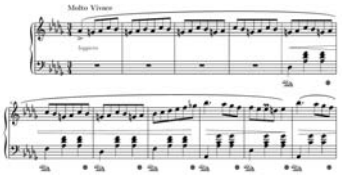




Score-Informed Audio Decomposition

Application: Separating left and right hands for piano

Chopin, Waltz Op. 64, No. 1

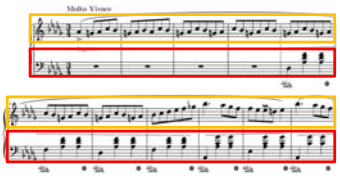


Original

Score-Informed Audio Decomposition

Application: Separating left and right hands for piano

Chopin, Waltz Op. 64, No. 1



Original

Left/right hand

Right hand

Left hand

Audio Decomposition

Works reasonable

Audio Decomposition

Much more difficult

Audio Decomposition

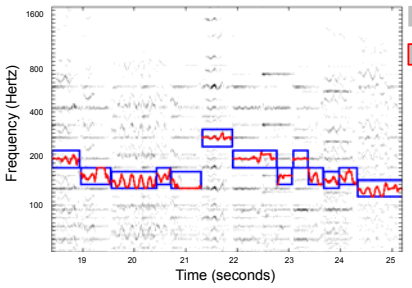
Related problems:

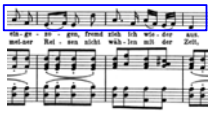
- F0 estimation
- Melody tracking
- Human voice
- Vibrato

Much more difficult

F0 Estimation

Score-informed F0



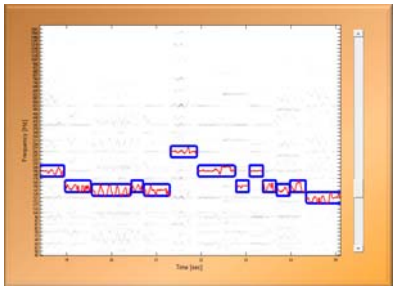


Original audio

F0 estimation

Score-Informed Source Separation

Application: Voice separation and editing

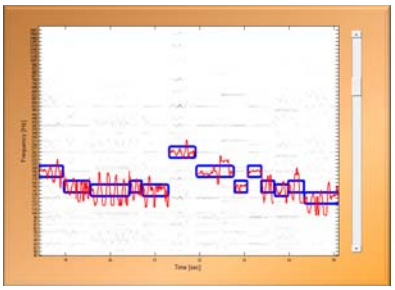


Original audio

Separated voice

Score-Informed Source Separation

Application: Voice separation and editing



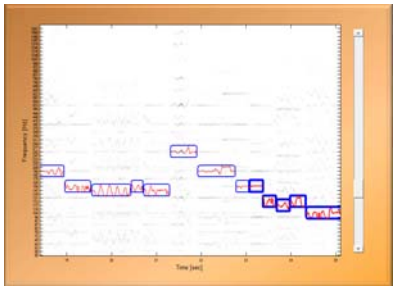
Original audio

Separated voice

Amplified vibrato

Score-Informed Source Separation

Application: Voice separation and editing



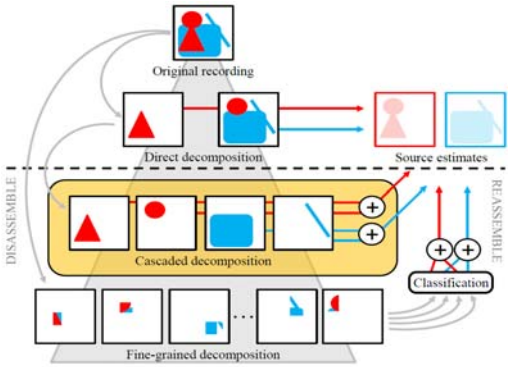
Original audio

Separated voice

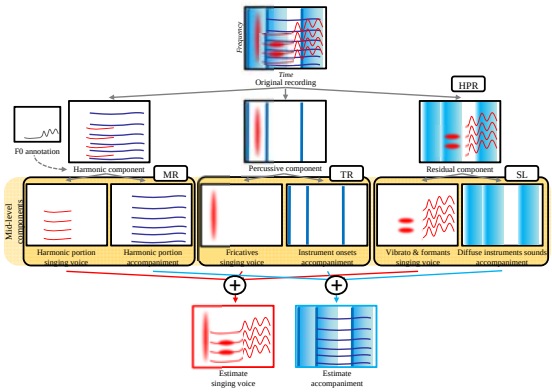
Amplified vibrato

Pitch shift

Cascaded Audio Decomposition

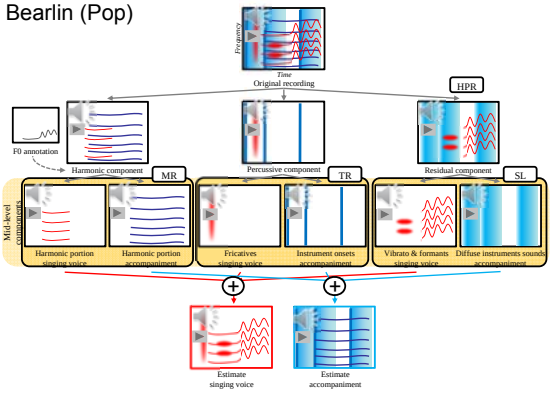


Cascaded Audio Decomposition



Cascaded Audio Decomposition

Bearlin (Pop)



Textbook

A First Course on Music Processing

Audio, Analysis, Algorithms, Applications

- Approx. 500 pages
- Approx. 300 figures
- Exercises
- To appear: End of 2015

Springer

Chapter		Music Processing Scenario
1		Music Represenations
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