Lecture

Music Processing

Tempo and Beat Tracking

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

Introduction

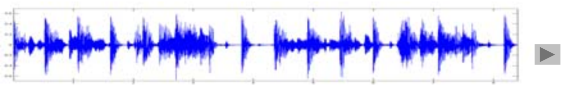
Basic beat tracking task:

Given an audio recording of a piece of music, determine the periodic sequence of beat positions.

“Tapping the foot when listening to music’’

Introduction

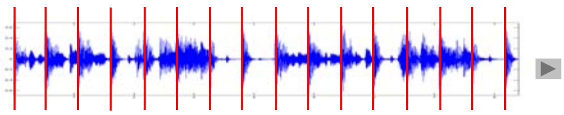
Example: Queen – Another One Bites The Dust



Time (seconds)

Introduction

Example: Queen – Another One Bites The Dust



Time (seconds)

Introduction

Example: Happy Birthday to you

Pulse level: **Measure**

Introduction

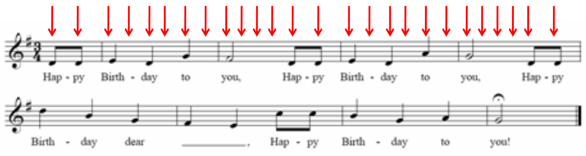
Example: Happy Birthday to you

Pulse level: **Tactus (beat)**

Introduction

Example: Happy Birthday to you

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



Introduction

Example: Chopin – Mazurka Op. 68-3

Pulse level: Quarter note

Tempo: ???

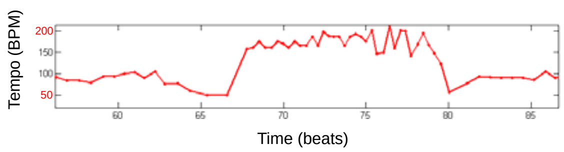
Introduction

Example: Chopin – Mazurka Op. 68-3

Pulse level: Quarter note

Tempo: 50-200 BPM

Tempo curve



Introduction

Example: Borodin – String Quartet No. 2

Pulse level: Quarter note

Tempo: 120-140 BPM (roughly)

Beat tracker without any prior knowledge

Beat tracker with prior knowledge on rough tempo range

Introduction

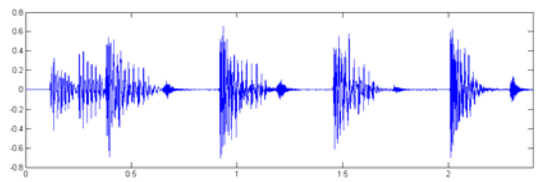
Challenges in beat tracking

- Pulse level often unclear
- Local/sudden tempo changes (e.g. rubato)
- Vague information
  (e.g., soft onsets, extracted onsets corrupt)
- Sparse information
  (often only note onsets are used)

Introduction

Tasks

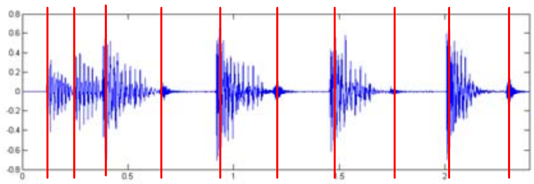
- Onset detection
- Beat tracking
- Tempo estimation



Introduction

Tasks

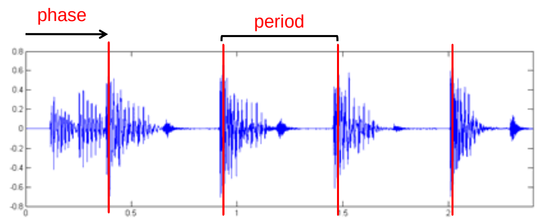
- Onset detection
- Beat tracking
- Tempo estimation



Introduction

Tasks

- Onset detection
- Beat tracking
- Tempo estimation



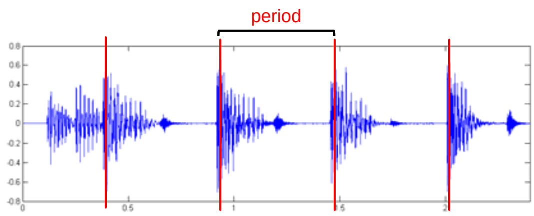
Introduction

Tasks

- Onset detection
- Beat tracking
- Tempo estimation

Tempo := 60 / period

Beats per minute (BPM)

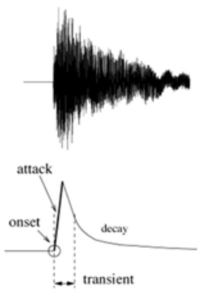


Onset Detection

- Finding start times of perceptually relevant acoustic events in music signal
- Onset is the time position where a note is played
- Onset typically goes along with a change of the signal's properties:
  - energy or loudness
  - pitch or harmony
  - timbre

Onset Detection

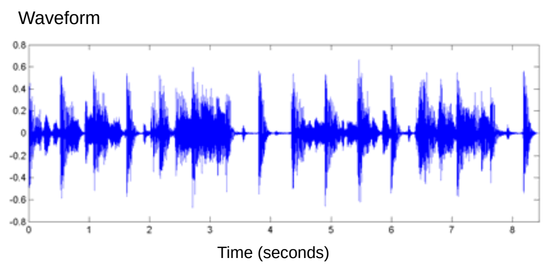
- Finding start times of perceptually relevant acoustic events in music signal
- Onset is the time position where a note is played
- Onset typically goes along with a change of the signal's properties:
  - energy or loudness
  - pitch or harmony
  - timbre



[Bello et al., IEEE-TASLP 2005]

Onset Detection (Energy-Based)

Steps

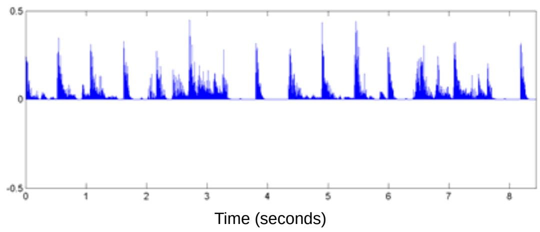


Onset Detection (Energy-Based)

Steps

1. Amplitude squaring

Squared waveform

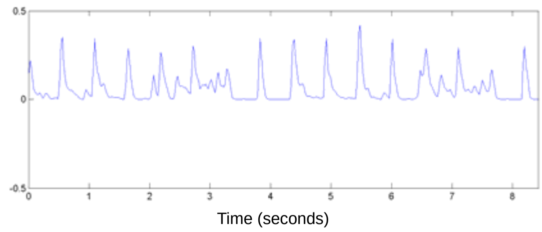


Onset Detection (Energy-Based)

Steps

1. Amplitude squaring
2. Windowing

Energy envelope



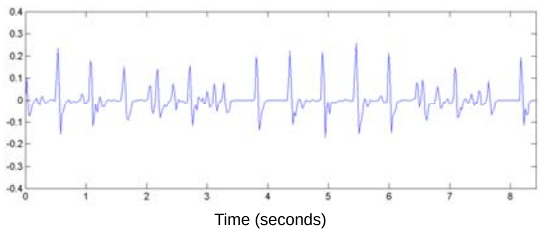
Onset Detection (Energy-Based)

Steps

1. Amplitude squaring
2. Windowing
3. Differentiation

Capturing energy changes

Differentiated energy envelope



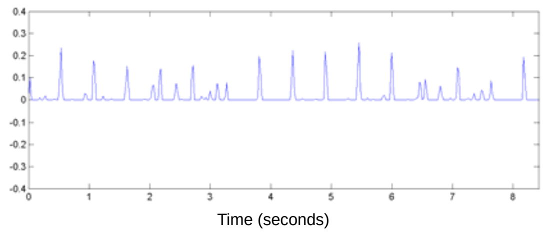
Onset Detection (Energy-Based)

Steps

1. Amplitude squaring
2. Windowing
3. Differentiation
4. Half wave rectification

Only energy increases are relevant for note onsets

Novelty curve

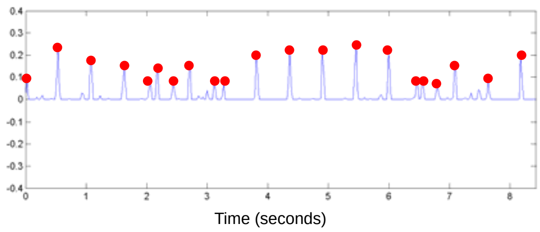


Onset Detection (Energy-Based)

Steps

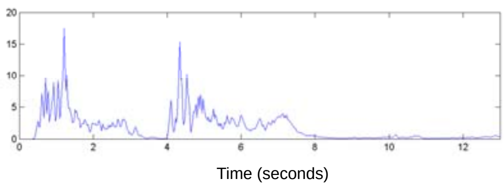
1. Amplitude squaring
2. Windowing
3. Differentiation
4. Half wave rectification
5. Peak picking

Peak positions indicate note onset candidates

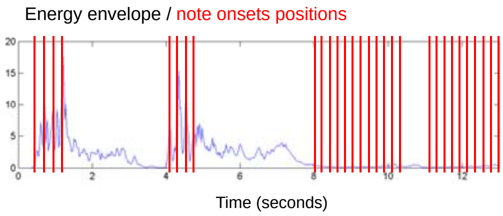


Onset Detection (Energy-Based)

Energy envelope



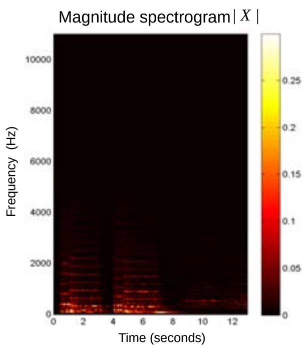
Onset Detection (Energy-Based)



Onset Detection

- Energy curves often only work for percussive music
- Many instruments such as strings have weak note onsets
- No energy increase may be observable in complex sound mixtures
- More refined methods needed that capture
  - changes of spectral content
  - changes of pitch
  - changes of harmony

Onset Detection (Spectral-Based)

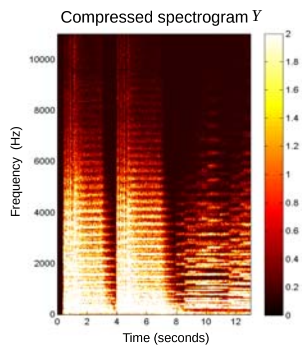


Steps:

1. Spectrogram

- *Aspects concerning pitch, harmony, or timbre are captured by spectrogram*
- *Allows for detecting local energy changes in certain frequency ranges*

Onset Detection (Spectral-Based)



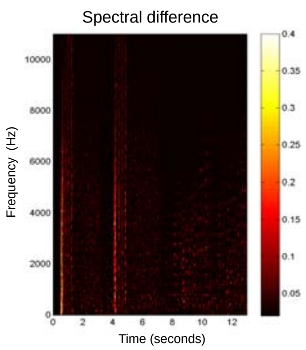
Steps:

1. Spectrogram
2. Logarithmic compression

$$Y = \log(1 + C \cdot |X|)$$

- *Accounts for the logarithmic sensation of sound intensity*
- *Dynamic range compression*
- *Enhancement of low-intensity values*
- *Often leading to enhancement of high-frequency spectrum*

Onset Detection (Spectral-Based)

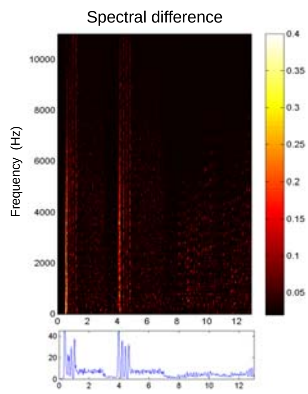


Steps:

1. Spectrogram
2. Logarithmic compression
3. Differentiation

- *First-order temporal difference*
- *Captures changes of the spectral content*
- *Only positive intensity changes considered*

Onset Detection (Spectral-Based)



Steps:

1. Spectrogram
2. Logarithmic compression
3. Differentiation
4. Accumulation

- *Frame-wise accumulation of all positive intensity changes*
- *Encodes changes of the spectral content*

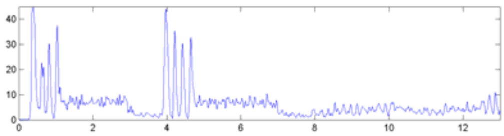
Novelty curve

Onset Detection (Spectral-Based)

Steps:

1. Spectrogram
2. Logarithmic compression
3. Differentiation
4. Accumulation

Novelty curve



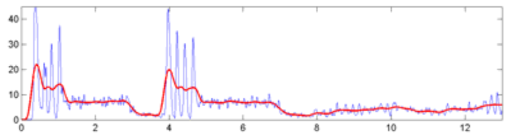
Onset Detection (Spectral-Based)

Steps:

1. Spectrogram
2. Logarithmic compression
3. Differentiation
4. Accumulation
5. Normalization

Novelty curve

Substraction of local average

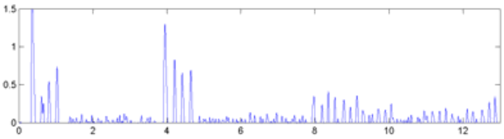


Onset Detection (Spectral-Based)

Steps:

1. Spectrogram
2. Logarithmic compression
3. Differentiation
4. Accumulation
5. Normalization

Normalized novelty curve

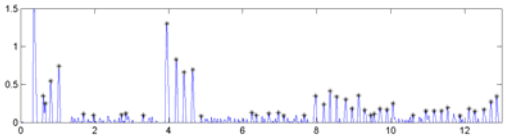


Onset Detection (Spectral-Based)

Steps:

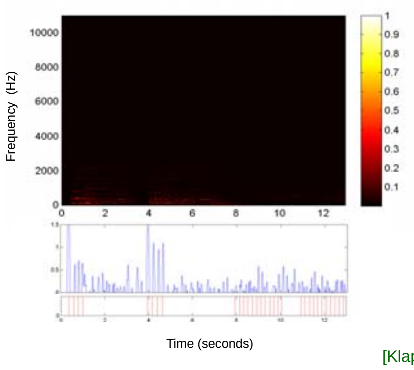
1. Spectrogram
2. Logarithmic compression
3. Differentiation
4. Accumulation
5. Normalization
6. Peak picking

Normalized novelty curve



Onset Detection (Spectral-Based)

Logarithmic compression is essential



$|X|$

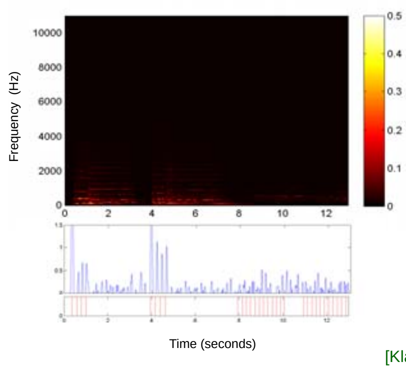
Novelty curve

Ground-truth onsets

[Klapuri et al., IEEE-TASLP 2006]

Onset Detection (Spectral-Based)

Logarithmic compression is essential



$Y = \log(1 + C \cdot |X|)$

C = 1

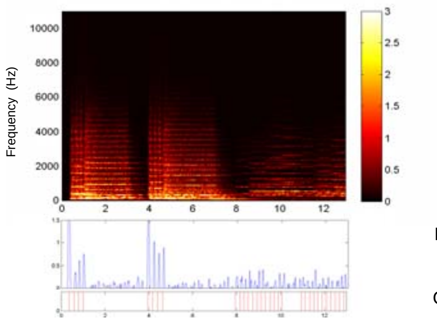
Novelty curve

Ground-truth onsets

[Klapuri et al., IEEE-TASLP 2006]

Onset Detection (Spectral-Based)

Logarithmic compression is essential



$Y = \log(1 + C \cdot |X|)$

C = 10

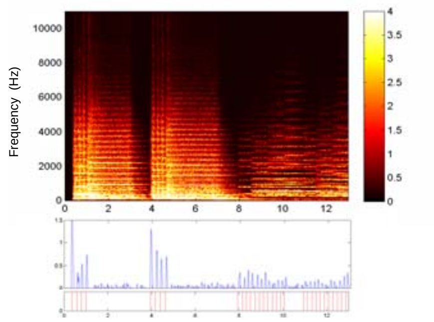
Novelty curve

Ground-truth onsets

[Klapuri et al., IEEE-TASLP 2006]

Onset Detection (Spectral-Based)

Logarithmic compression is essential



$Y = \log(1 + C \cdot |X|)$

C = 1000

Novelty curve

Ground-truth onsets

[Klapuri et al., IEEE-TASLP 2006]

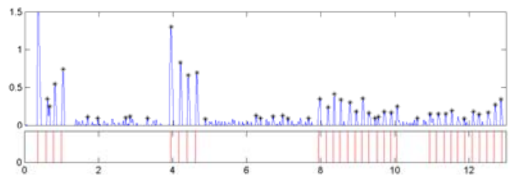
Onset Detection (Spectral-Based)

- Spectrogram $X = (X(t,k))_{t,k}$ $\quad t \in [1:T]$, $k \in [1:K]$
- Compressed Spectrogram $Y := \log(1 + C \cdot |X|)$ $\quad C > 1$
- Novelty curve $\Delta : [1 : T-1] \to \mathbb{R}$

$$\Delta(t) := \sum_{k=1}^{K} |Y(t+1,k) - Y(t,k)|_{\geq 0}$$

Onset Detection

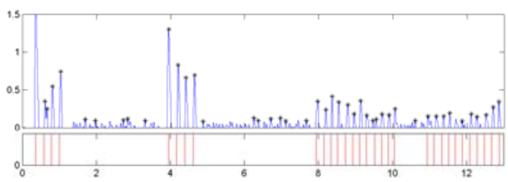
Peak picking



- Peaks of the novelty curve indicate note onset candidates

Onset Detection

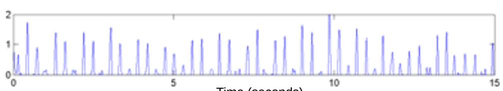
Peak picking



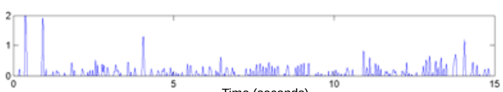
- Peaks of the novelty curve indicate note onset candidates
- In general many spurious peaks
- Usage of local thresholding techniques
- Peak-picking very fragile step in particular for soft onsets

Onset Detection

Shostakovich – 2nd Waltz



Borodin – String Quartet No. 2



Onset Detection

Drumbeat

Going Home

Lyphard melodie

Por una cabeza

Donau

Beat and Tempo

What is a beat?

- Steady pulse that drives music forward and provides the temporal framework of a piece of music
- Sequence of perceived pulses that are equally spaced in time
- The pulse a human taps along when listening to the music

[Parncutt 1994]
[Sethares 2007]
[Large/Palmer 2002]
[Lerdahl/ Jackendoff 1983]
[Fitch/ Rosenfeld 2007]

The term tempo then refers to the speed of the pulse.

Beat and Tempo

Strategy

- Analyze the novelty curve with respect to reoccurring or quasi-periodic patterns
- Avoid the explicit determination of note onsets (no peak picking)

Beat and Tempo

Strategy

- Analyze the novelty curve with respect to reoccurring or quasi-periodic patterns
- Avoid the explicit determination of note onsets (no peak picking)

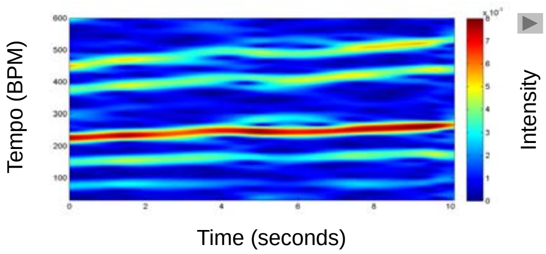
Methods

- Comb-filter methods
- Autocorrelation
- Fourier transfrom

[Scheirer, JASA 1998]
[Ellis, JNMR 2007]
[Davies/Plumbley, IEEE-TASLP 2007]
[Peeters, JASP 2007]
[Grosche/Müller, ISMIR 2009]
[Grosche/Müller, IEEE-TASLP 2011]

Tempogram

Definition: A tempogram is a time-tempo representation that encodes the local tempo of a music signal over time.



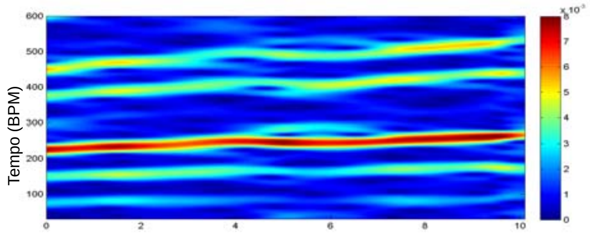
Tempogram (Fourier)

Definition: A tempogram is a time-tempo represenation that encodes the local tempo of a music signal over time.

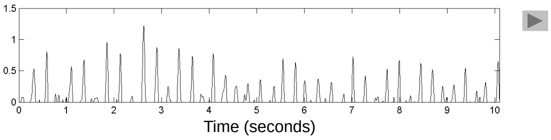
Fourier-based method

- Compute a spectrogram (STFT) of the novelty curve
- Convert frequency axis (given in Hertz) into tempo axis (given in BPM)
- Magnitude spectrogram indicates local tempo

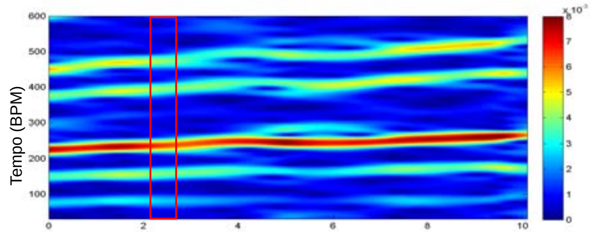
Tempogram (Fourier)



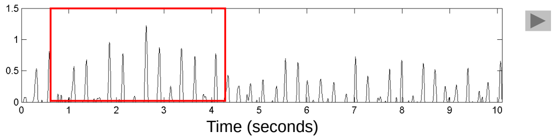
Novelty curve



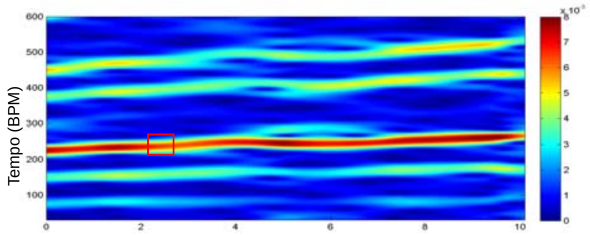
Tempogram (Fourier)



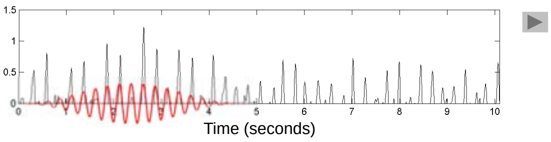
Novelty curve (local section)



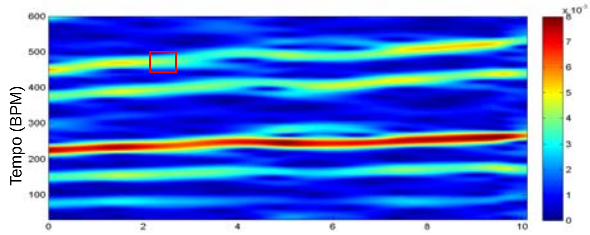
Tempogram (Fourier)



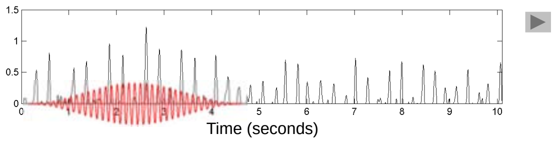
Windowed sinusoidal



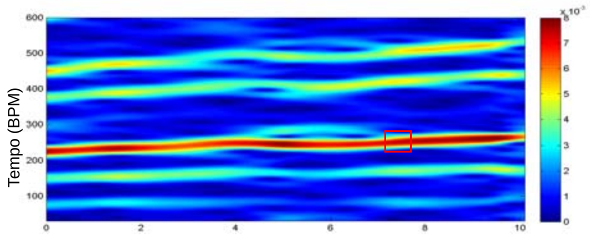
Tempogram (Fourier)



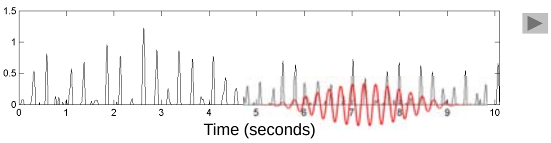
Windowed sinusoidal



Tempogram (Fourier)



Windowed sinusoidal



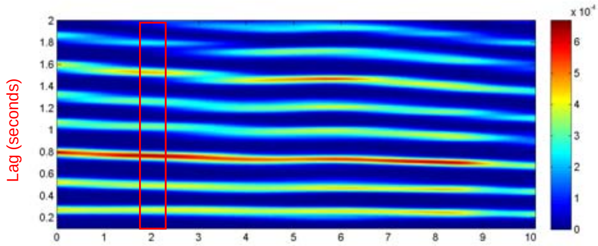
Tempogram (Autocorrelation)

Definition: A tempogram is a time-tempo represenation that encodes the local tempo of a music signal over time.

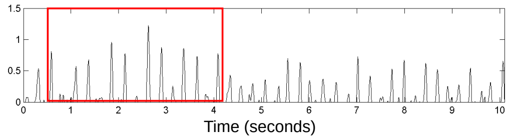
Autocorrelation-based method

- Compare novelty curve with time-lagged local sections of itself
- Convert lag-axis (given in seconds) into tempo axis (given in BPM)
- Autocorrelogram indicates local tempo

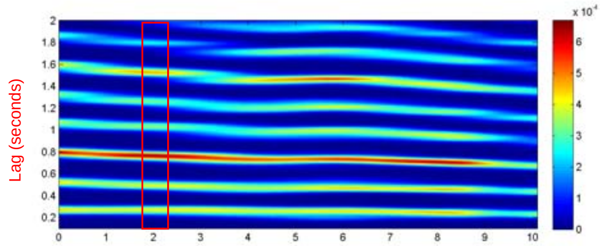
Tempogram (Autocorrelation)



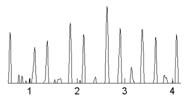
Novelty curve (local section)



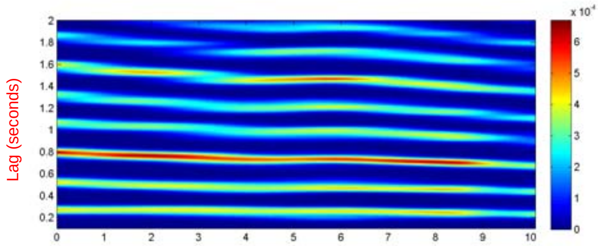
Tempogram (Autocorrelation)



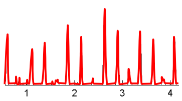
Windowed autocorrelation



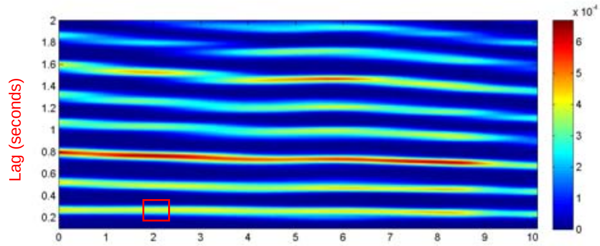
Tempogram (Autocorrelation)



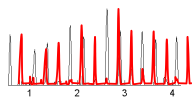
Lag = 0 (seconds)



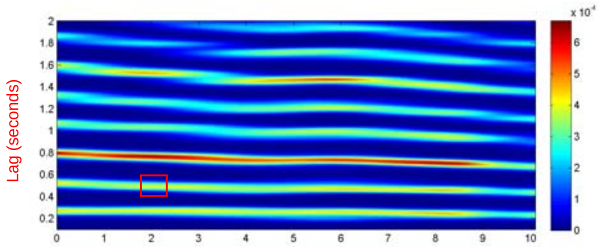
Tempogram (Autocorrelation)



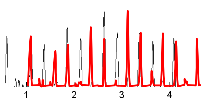
Lag = 0.26 (seconds)



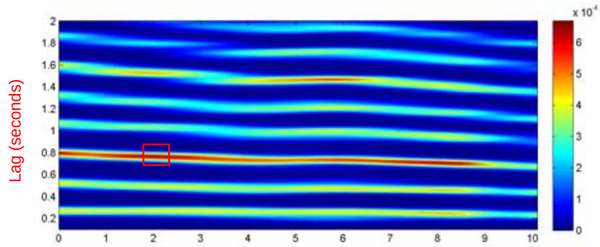
Tempogram (Autocorrelation)



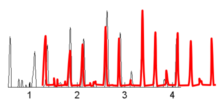
Lag = 0.52 (seconds)



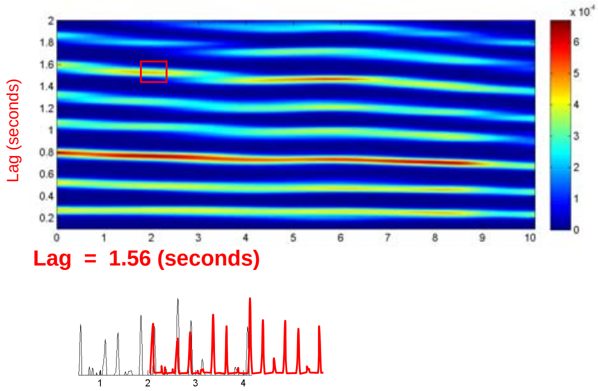
Tempogram (Autocorrelation)



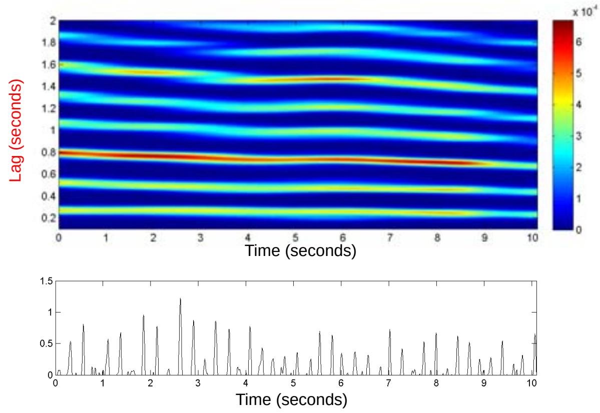
Lag = 0.78 (seconds)



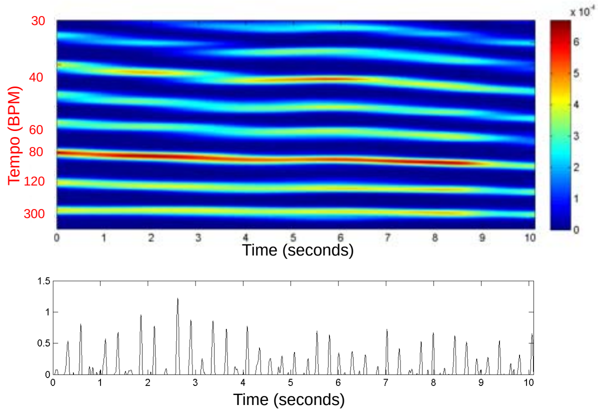
Tempogram (Autocorrelation)



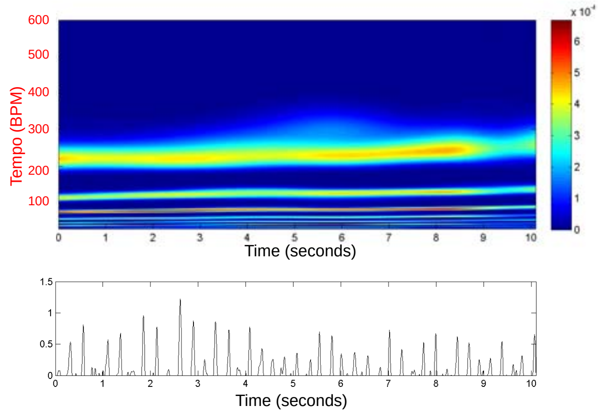
Tempogram (Autocorrelation)



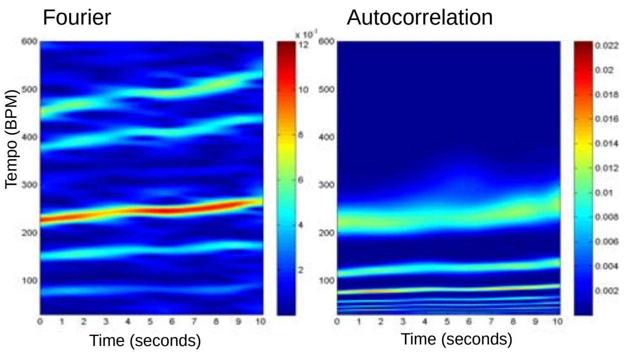
Tempogram (Autocorrelation)



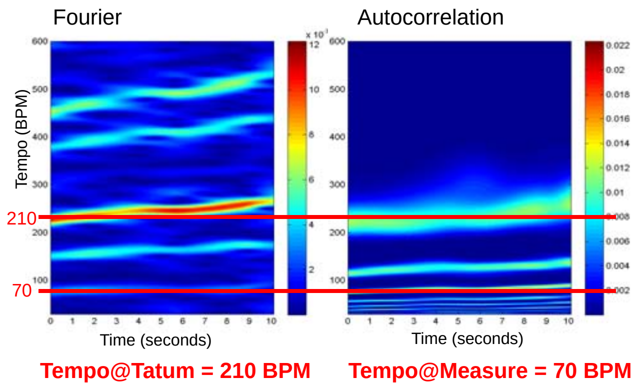
Tempogram (Autocorrelation)



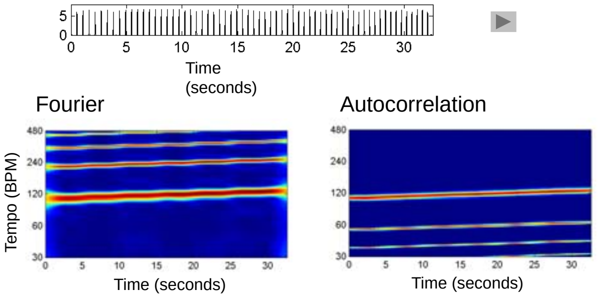
Tempogram



Tempogram



Tempogram



Emphasis of tempo harmonics (integer multiples)

Emphasis of tempo subharmonics (integer fractions)

[Peeters, JASP 2007][Grosche et al., ICASSP 2010]

Tempogram (Summary)

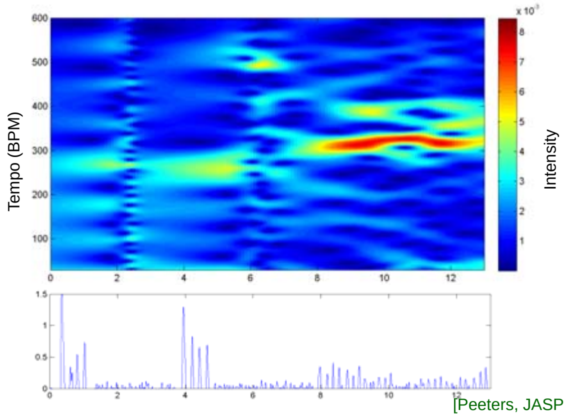
Fourier	Autocorrelation
Novelty curve is compared with sinusoidal kernels each representing a specific tempo	Novelty curve is compared with time-lagged local (windowed) sections of itself
Convert frequency (Hertz) into tempo (BPM)	Convert time-lag (seconds) into tempo (BPM)
Reveals novelty periodicities	Reveals novelty self-similarities
Emphasizes harmonics	Emphasizes subharmonics
Suitable to analyze tempo on tatum and tactus level	Suitable to analyze tempo on tactus and measure level

Beat Tracking

- Given the tempo, find the best sequence of beats
- Complex Fourier tempogram contains magnitude and phase information
- The magnitude encodes how well the novelty curve resonates with a sinusoidal kernel of a specific tempo
- The phase optimally aligns the sinusoidal kernel with the peaks of the novelty curve

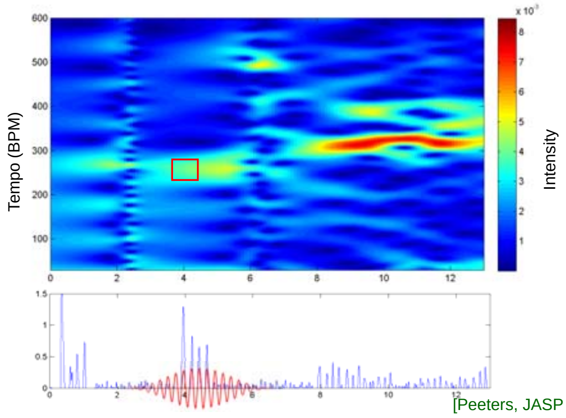
[Peeters, JASP 2005]

Beat Tracking



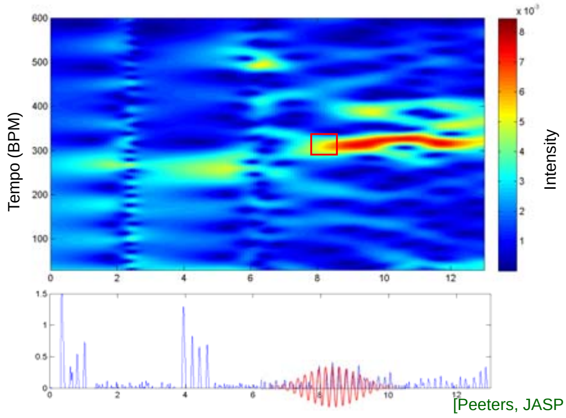
[Peeters, JASP 2005]

Beat Tracking



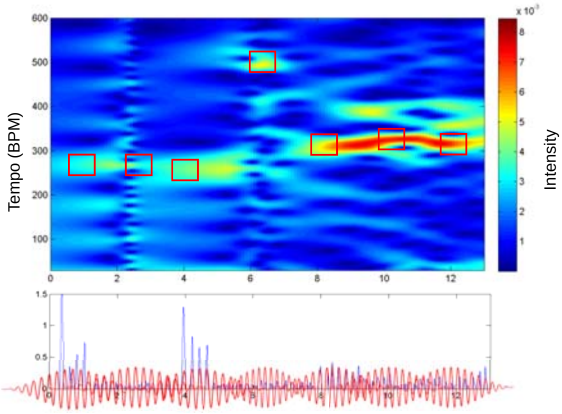
[Peeters, JASP 2005]

Beat Tracking

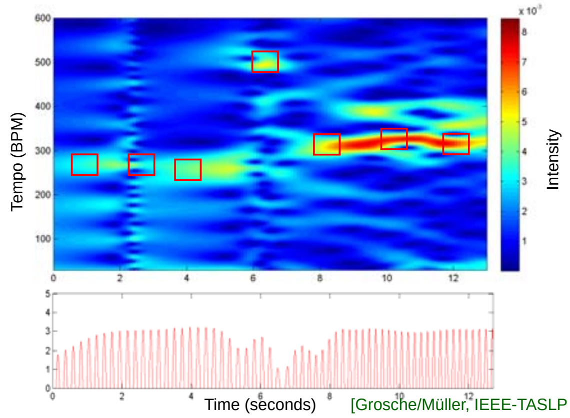


[Peeters, JASP 2005]

Beat Tracking



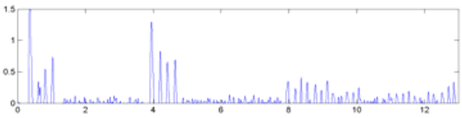
Beat Tracking



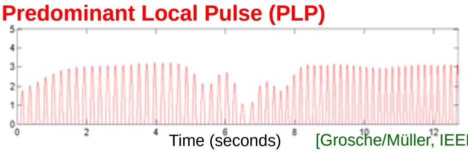
[Grosche/Müller, IEEE-TASLP 2011]

Beat Tracking

Novelty Curve



Predominant Local Pulse (PLP)



[Grosche/Müller, IEEE-TASLP 2011]

Beat Tracking

Novelty Curve

- Indicates note onset candidates
- Extraction errors in particular for soft onsets
- Simple peak-picking problematic

Predominant Local Pulse (PLP)

- Periodicity enhancement of novelty curve
- Accumulation introduces error robustness
- Locality of kernels handles tempo variations

[Grosche/Müller, IEEE-TASLP 2011]

Beat Tracking

- Local tempo at time t : $\tau_t \in \Theta$ Θ = [60:240] BPM
- Phase $\varphi_t := \frac{1}{2\pi}\arccos\left(\frac{\mathrm{Re}(\mathcal{T}(t,\tau_t))}{|\mathcal{T}(t,\tau_t)|}\right)$
- Sinusoidal kernel $\kappa_t : \mathbb{Z} \to \mathbb{R}$

$$\kappa_t(n) := W(n-t)\cos(2\pi(\tau_t/60 \cdot n - \varphi_t)) \qquad n \in \mathbb{Z}$$

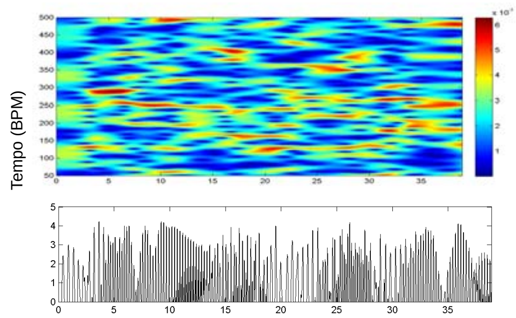
- Periodicity curve $\Gamma : [1:T] \to \mathbb{R}_{\geq 0}$

$$\Gamma(n) = \left| \sum_{t \in [1:T]} \kappa_t(n) \right|_{\geq 0} \qquad n \in [1:T]$$

[Grosche/Müller, IEEE-TASLP 2011]

Beat Tracking

Borodin – String Quartet No. 2

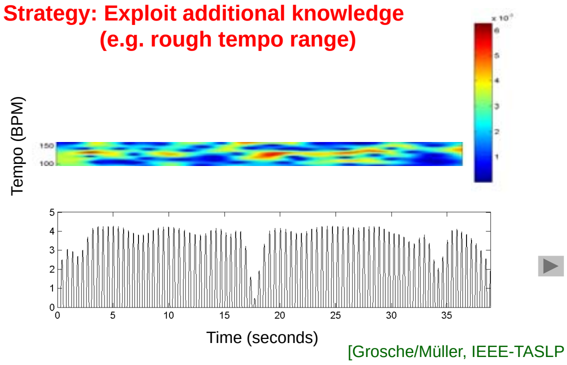


[Grosche/Müller, IEEE-TASLP 2011]

Beat Tracking

Borodin – String Quartet No. 2

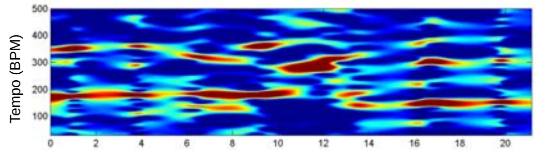
Strategy: Exploit additional knowledge (e.g. rough tempo range)



[Grosche/Müller, IEEE-TASLP 2011]

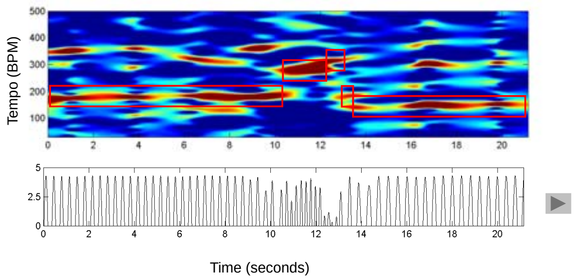
Beat Tracking

Brahms Hungarian Dance No. 5



Beat Tracking

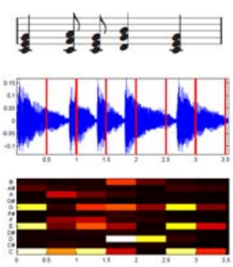
Brahms Hungarian Dance No. 5



Applications

- Feature design
  (beat-synchronous features, adaptive windowing)
- Digital DJ / audio editing
  (mixing and blending of audio material)
- Music classification
- Music recommendation
- Performance analysis
  (extraction of tempo curves)

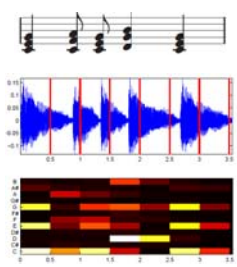
Application: Feature Design



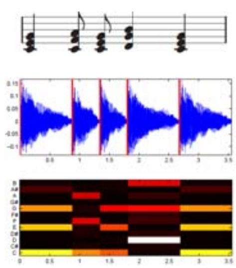
Fixed window size

[Ellis et al., ICASSP 2008] [Bello/Pickens, ISMIR 2005]

Application: Feature Design



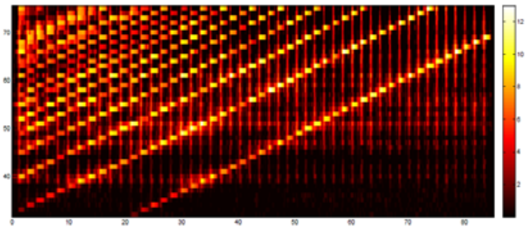
Fixed window size



Adaptive window size

[Ellis et al., ICASSP 2008] [Bello/Pickens, ISMIR 2005]

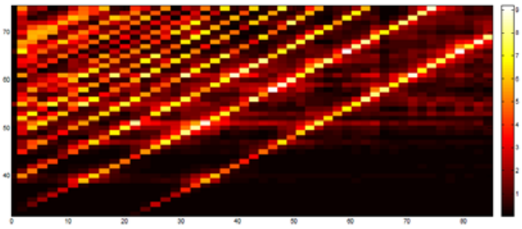
Application: Feature Design



Time
(seconds)

Fixed window size (100 ms)

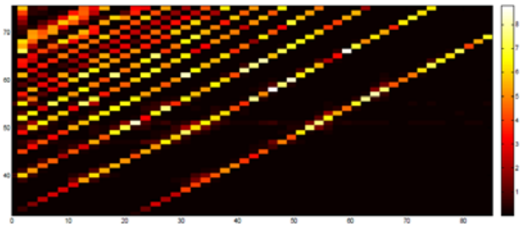
Application: Feature Design



Time
(seconds)

Adative window size (roughly 1200 ms)
Note onset positions define boundaries

Application: Feature Design



Time
(seconds)

Adative window size (roughly 1200 ms)
Note onset positions define boundaries

Denoising by excluding boundary neighborhoods

Application: Audio Editing (Digital DJ)

http://www.mixxx.org/

Application: Beat-Synchronous Light Effects

Summary

1. Onset Detection
   - Novelty curve (*something is changing*)
   - Indicates note onset candidates
   - Hard task for non-percussive instruments (strings)
2. Tempo Estimation
   - Fourier tempogram
   - Autocorrelation tempogram
   - Musical knowledge (tempo range, continuity)
3. Beat tracking
   - Find most likely beat positions
   - Exploiting phase information from Fourier tempogram