Tempo and Beat Estimation of Music Signals

Miguel A. ALONSO, Bertrand DAVID and Gaël RICHARD

{malonso,bedavid,grichard}@tsi.enst.fr

École Nationale Supérieure des Télécommunications (ENST)

Paris, France

Presentation content

- Introduction
- Description of the algorithm
- Performance analysis
- Sound examples
- Conclusions

Introduction

- automatic music analysis is an active research area

Introduction

- automatic music analysis is an active research area
- beat-tracking is an essential part of this field
  - important for many audio applications
  - approach dealing with music recordings
  - automatic estimation is difficult for a broad variety of music

Introduction

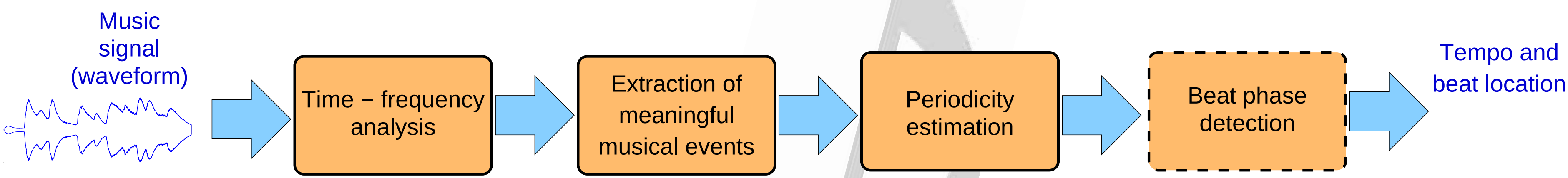
- automatic music analysis is an active research area
- beat-tracking is an essential part of this field
  - important for many audio applications
  - approach dealing with music recordings
  - automatic estimation is difficult for a broad variety of music
- the proposed system aims at various musical genres

Introduction

- automatic music analysis is an active research area
- beat-tracking is an essential part of this field
  - important for many audio applications
  - approach dealing with music recordings
  - automatic estimation is difficult for a broad variety of music
- the proposed system aims at various musical genres
- most algorithms are based on the same general architecture

Beat-tracking

- General architecture



- Introduction
- **Description of the algorithm**
- Performance analysis
- Sound examples
- Conclusions

Salient features detection

- objective : *detect the most salient features of the music signal (note onsets, attacks, cord changes, etc.)*
- robust detection for polyphonic music is a difficult task
- events are easier to detect if the signal is decomposed in frequency channels

Salient features detection

- objective : *detect the most salient features of the music signal (note onsets, attacks, cord changes, etc.)*
- robust detection for polyphonic music is a difficult task
- events are easier to detect if the signal is decomposed in frequency channels
- motivated by previous work, we define the *Spectral Energy Flux (SEF)* $E(f,k)$ of an audio signal

Spectral energy flux (1/3)

- a discrete time audio signal $x(n)$ is transformed into the frequency domain

$$\tilde{X}(f,m) = \sum_{n=0}^{N-1} w(n)x(n+mM)e^{-j2\pi fn}$$

Spectral energy flux (1/3)

- a discrete time audio signal $x(n)$ is transformed into the frequency domain

$$\tilde{X}(f,m) = \sum_{n=0}^{N-1} w(n)x(n+mM)e^{-j2\pi f n}$$

- the SEF is defined as an approximation to the derivative of the signal frequency content with respect to time

$$E(f,k) = \sum_m h(m-k)\, G(f,m)$$

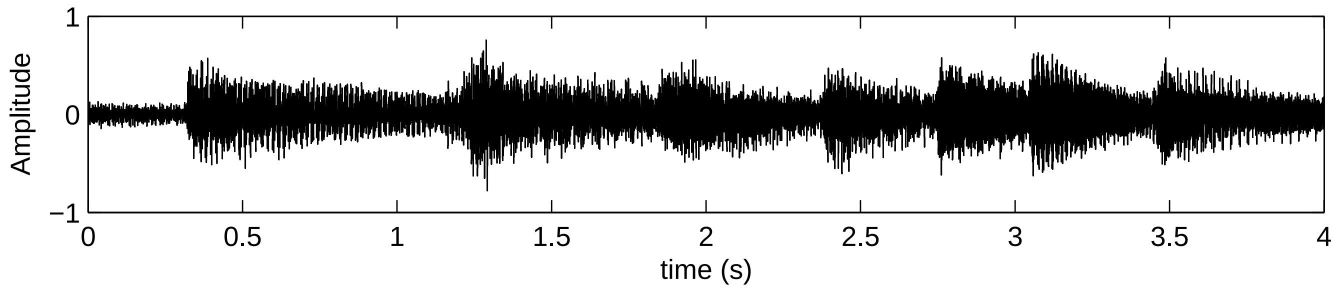
where $h(m)$ approximates a differentiator filter with $H(e^{j2\pi f}) \simeq j2\pi f$ and the transformation $G(f,m) = \mathcal{F}\{|\tilde{X}(f,m)|\}$ is obtained via a two step process: a low-pass filtering and a non-linear compression of $|\tilde{X}(f,m)|$

Spectral energy flux (2/3)

- Piano example

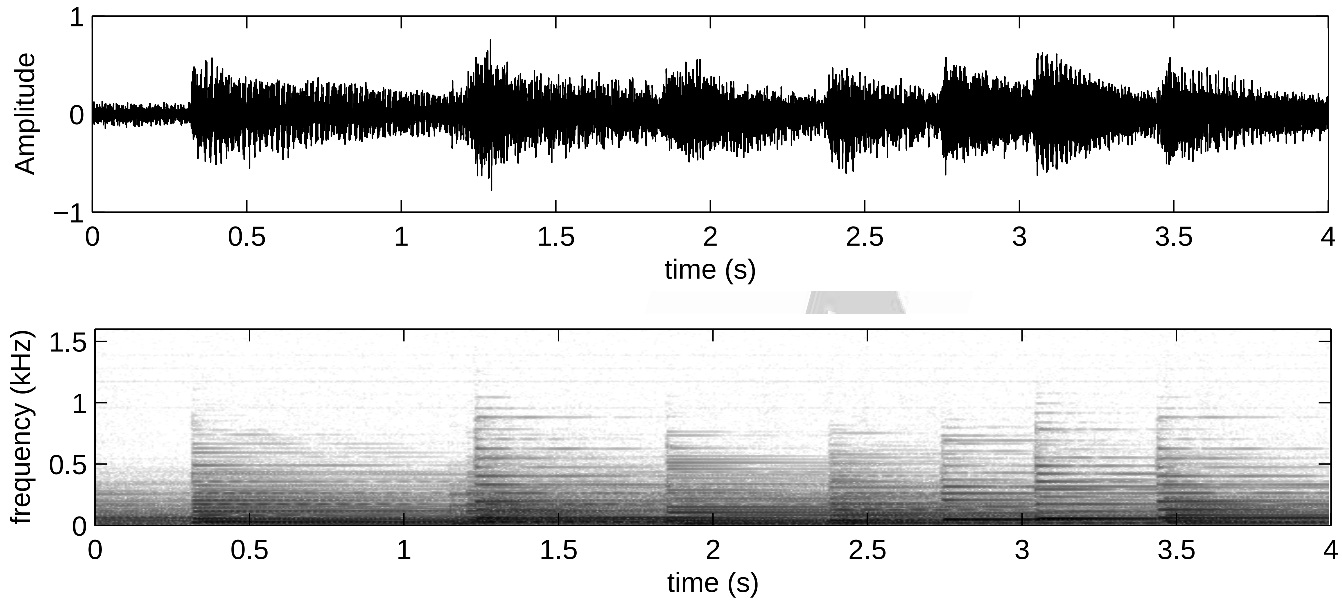
Spectral energy flux (2/3)

- Piano example



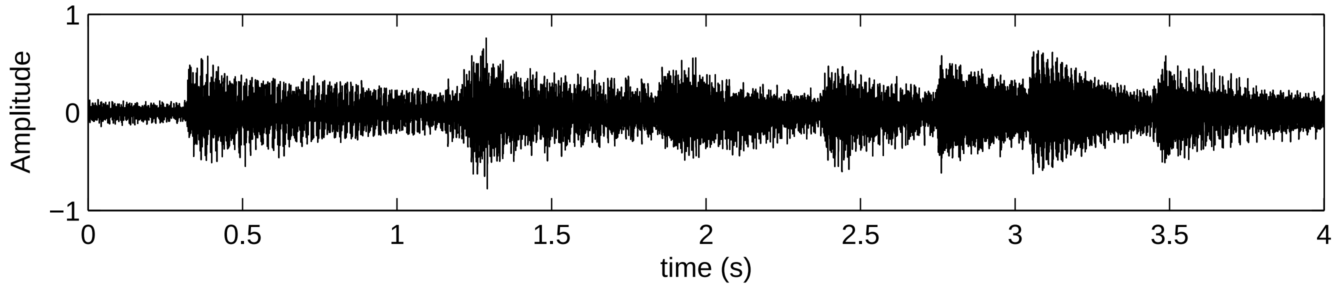
Spectral energy flux (2/3)

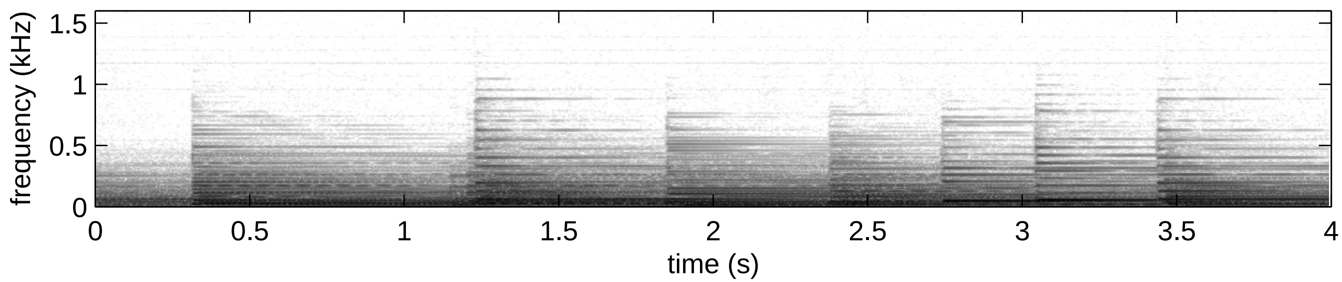
- Piano example

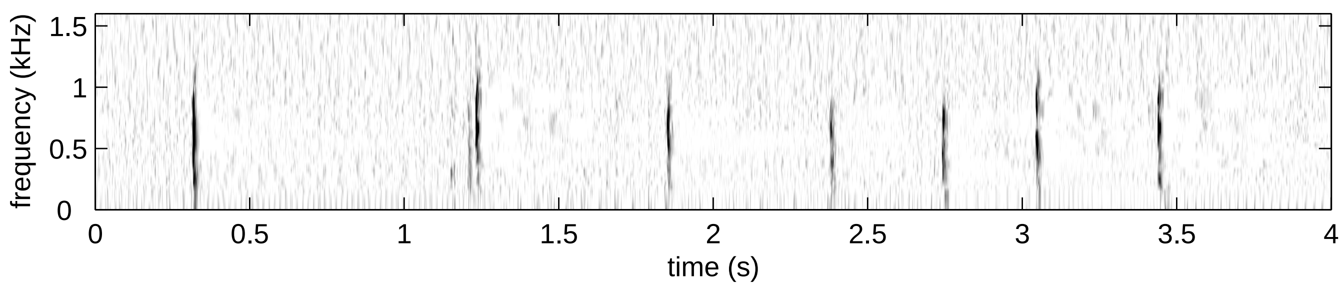


Spectral energy flux (2/3)

- Piano example

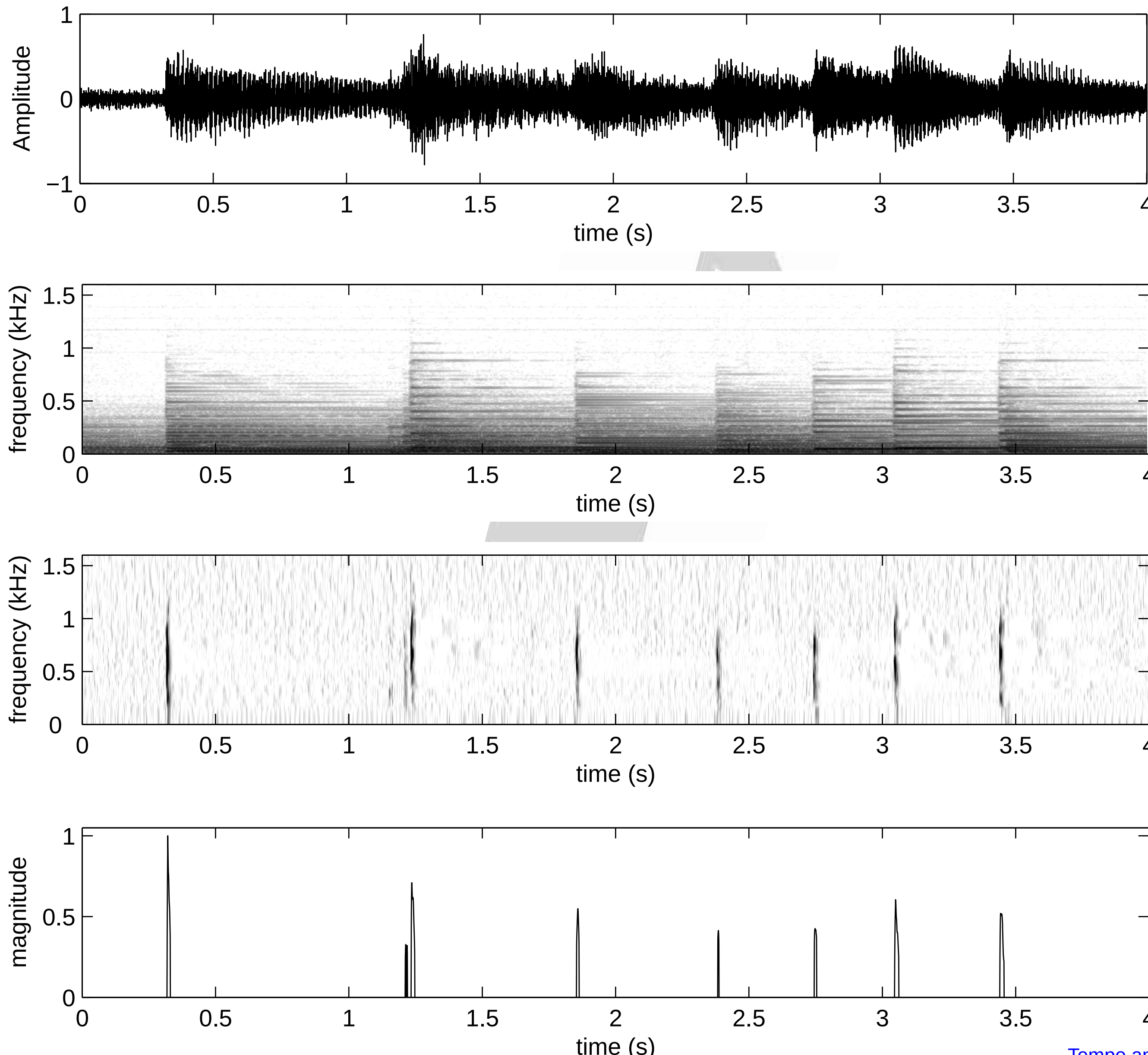






Spectral energy flux (2/3)

- Piano example

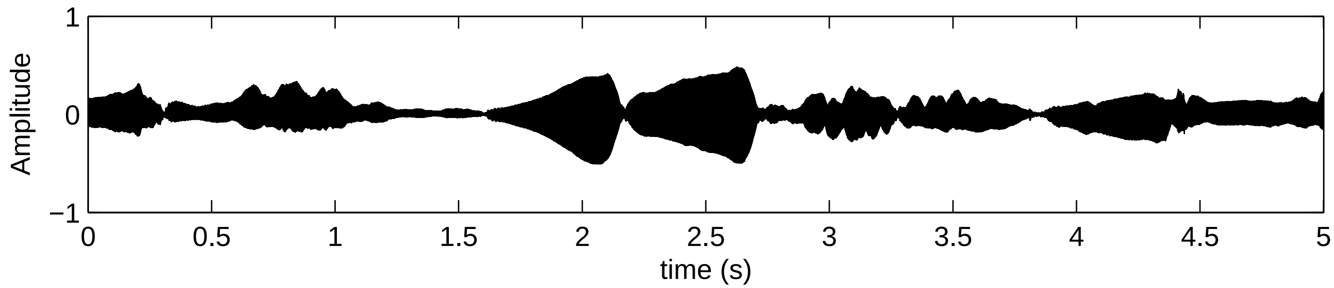


Spectral energy flux (3/3)

- Violin example

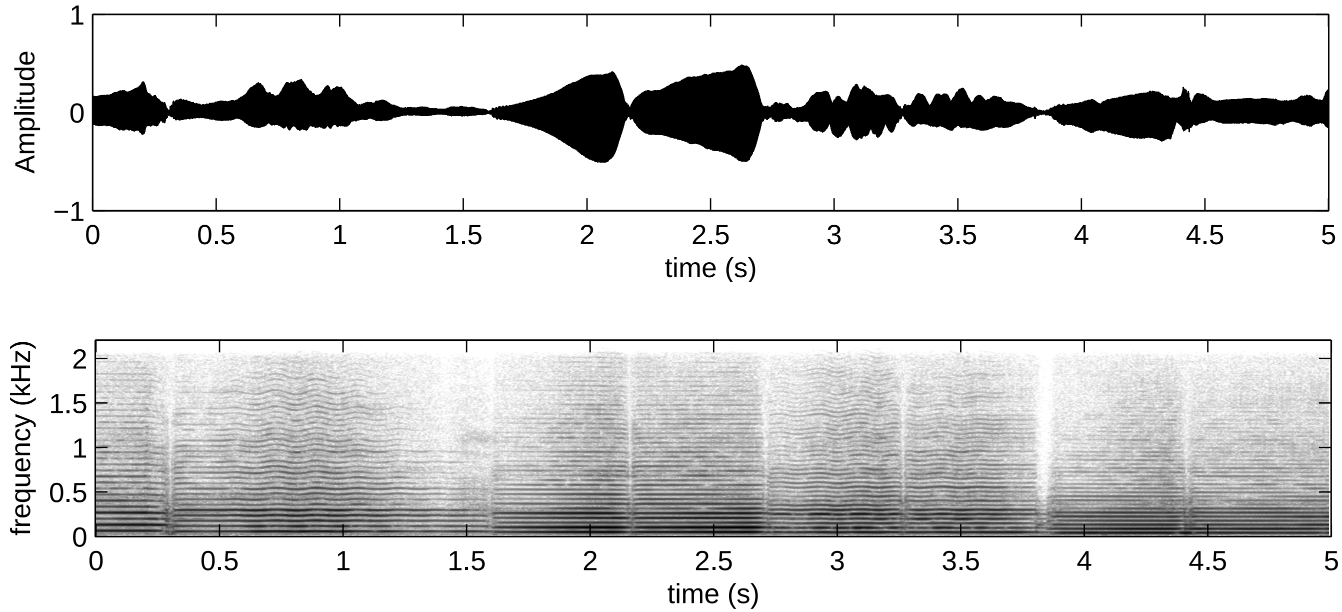
Spectral energy flux (3/3)

- Violin example



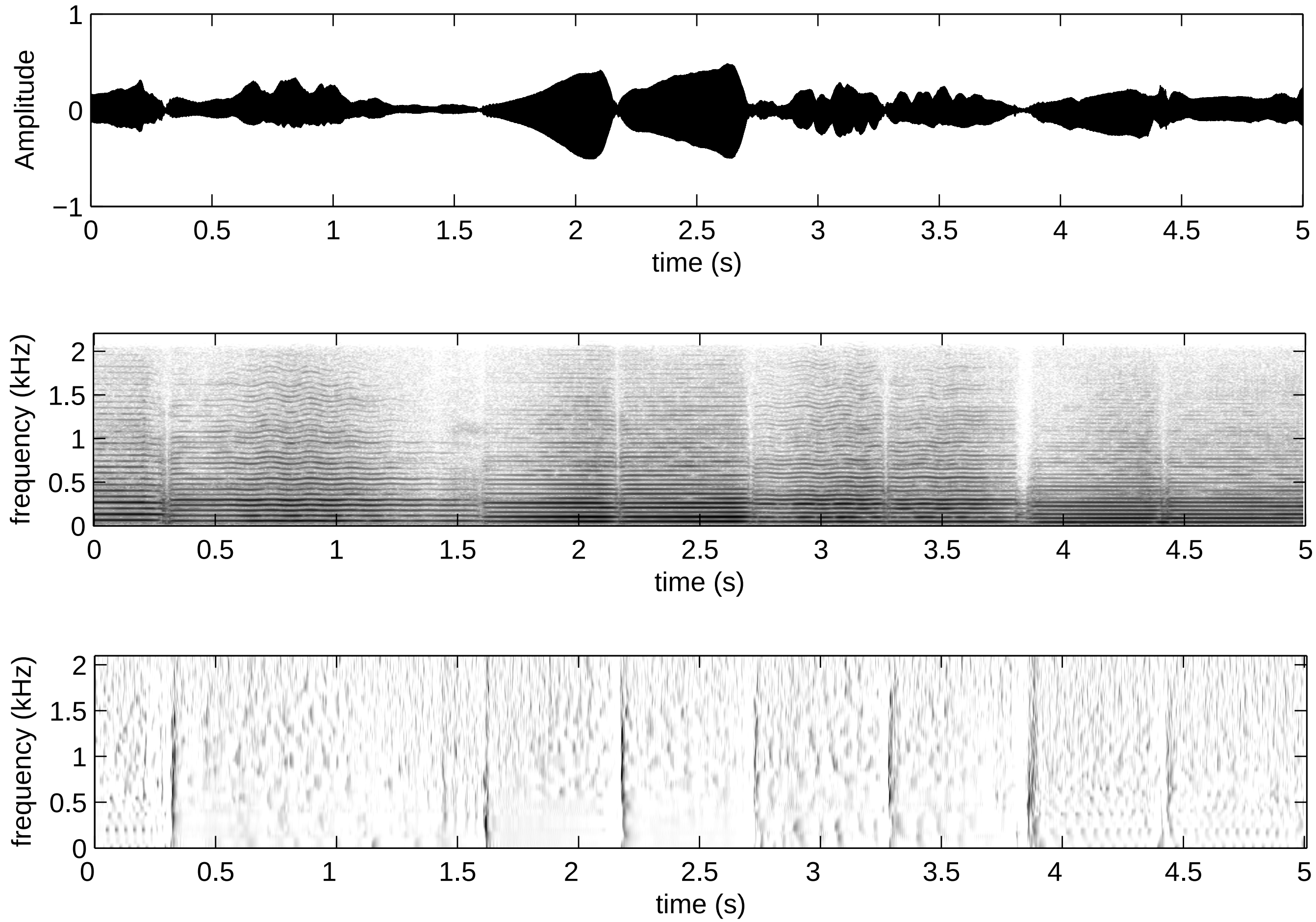
Spectral energy flux (3/3)

- Violin example



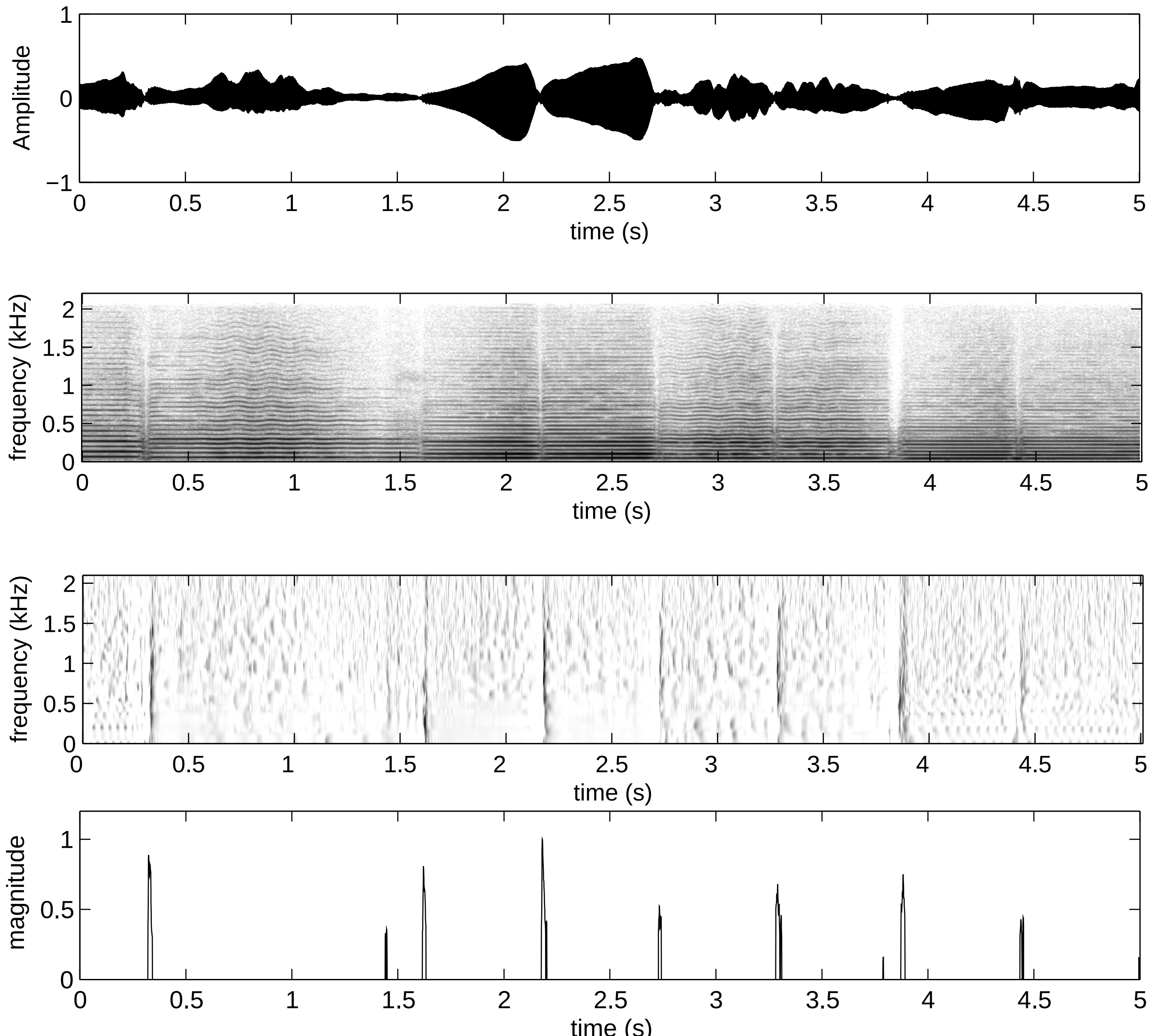
Spectral energy flux (3/3)

- Violin example



Spectral energy flux (3/3)

- Violin example



Periodicity estimation

- two different methods are used

Periodicity estimation

- two different methods are used
  - *spectral product*

$$S(e^{j2\pi f}) = \prod_{m=1}^{M} |P(e^{j2\pi m f})| \quad \text{for } f < \frac{1}{2M}$$

where $P(e^{j2\pi f})$ is the Fourier transform of $p(k)$, the output of the onset detection stage

Periodicity estimation

- two different methods are used
  - *spectral product*

$$S(e^{j2\pi f}) = \prod_{m=1}^{M} |P(e^{j2\pi m f})| \quad \text{for } f < \frac{1}{2M}$$

where $P(e^{j2\pi f})$ is the Fourier transform of $p(k)$, the output of the onset detection stage

  - *autocorrelation function*

$$r(\tau) = \sum_k p(k+\tau)p(k)$$

Periodicity estimation

- two different methods are used
  - *spectral product*

$$S(e^{j2\pi f}) = \prod_{m=1}^{M} |P(e^{j2\pi m f})| \quad \text{for } f < \frac{1}{2M}$$

where $P(e^{j2\pi f})$ is the Fourier transform of $p(k)$, the output of the onset detection stage

  - *autocorrelation function*

$$r(\tau) = \sum_k p(k+\tau)p(k)$$

- the tempo $\mathbb{T}$ of the segment under analysis is obtained

Beat location (1/2)

- method based on a comb filter

Beat location (1/2)

- method based on a comb filter
  - a pulse-train $q(k)$ of tempo $\mathbb{T}$ is correlated with $p(k)$

Beat location (1/2)

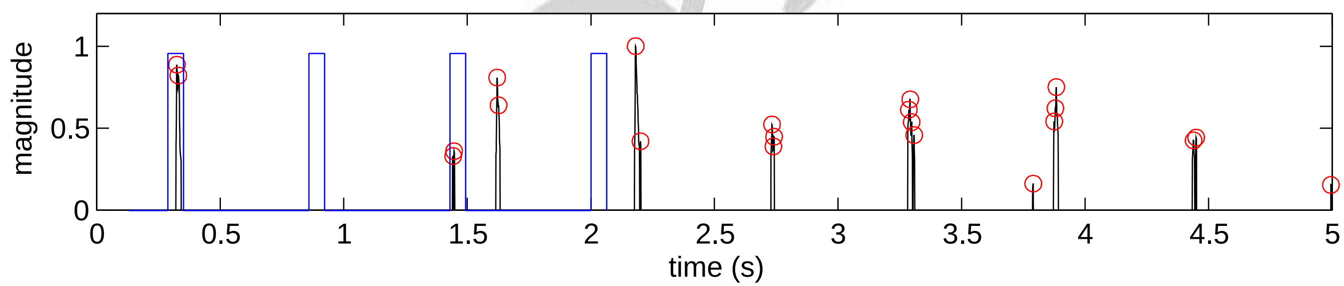
- method based on a comb filter
  - a pulse-train $q(k)$ of tempo $\mathbb{T}$ is correlated with $p(k)$
  - low complexity operation, it is evaluated only at the indices corresponding to the maxima of $p(k)$

Beat location (1/2)

- method based on a comb filter
  - a pulse-train $q(k)$ of tempo $\mathbb{T}$ is correlated with $p(k)$
  - low complexity operation, it is evaluated only at the indices corresponding to the maxima of $p(k)$
  - find the time index (t_0) where the cross-correlation is maximal

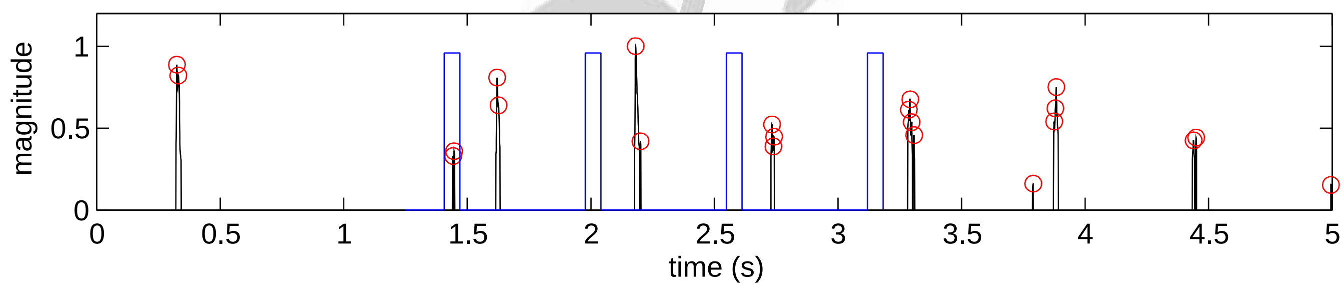
Beat location (1/2)

- method based on a comb filter
  - a pulse-train $q(k)$ of tempo $\mathbb{T}$ is correlated with $p(k)$
  - low complexity operation, it is evaluated only at the indices corresponding to the maxima of $p(k)$
  - find the time index (t_0) where the cross-correlation is maximal



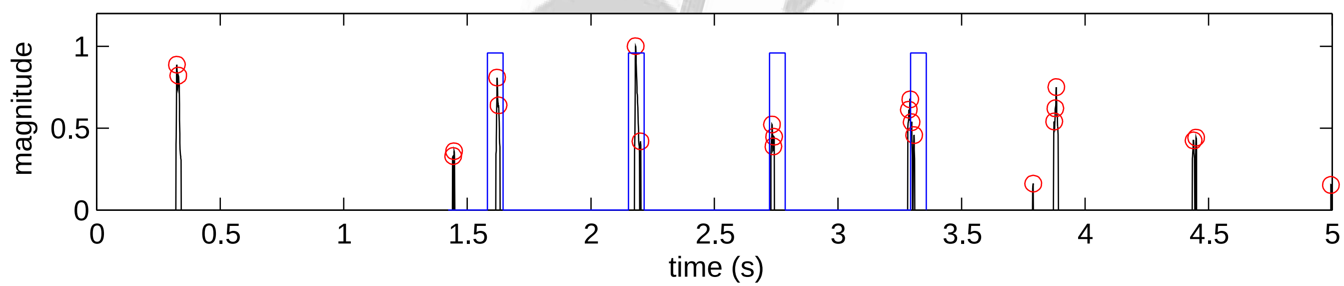
Beat location (1/2)

- method based on a comb filter
  - a pulse-train $q(k)$ of tempo $\mathbb{T}$ is correlated with $p(k)$
  - low complexity operation, it is evaluated only at the indices corresponding to the maxima of $p(k)$
  - find the time index (t_0) where the cross-correlation is maximal



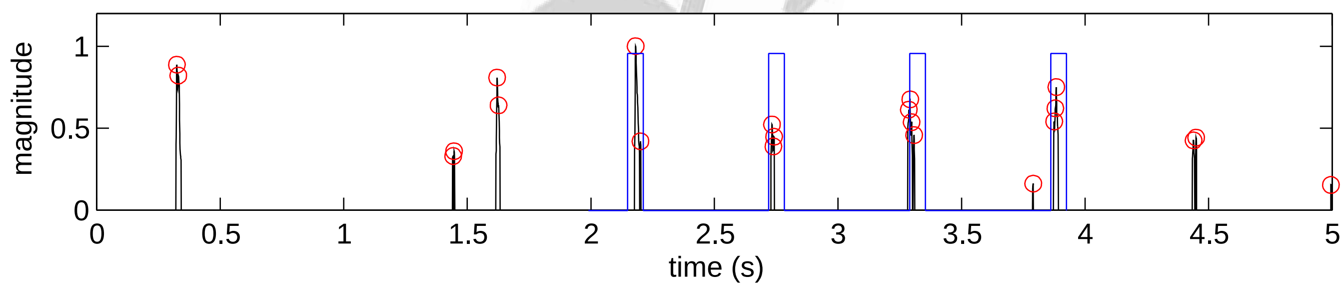
Beat location (1/2)

- method based on a comb filter
  - a pulse-train $q(k)$ of tempo $\mathbb{T}$ is correlated with $p(k)$
  - low complexity operation, it is evaluated only at the indices corresponding to the maxima of $p(k)$
  - find the time index (t_0) where the cross-correlation is maximal



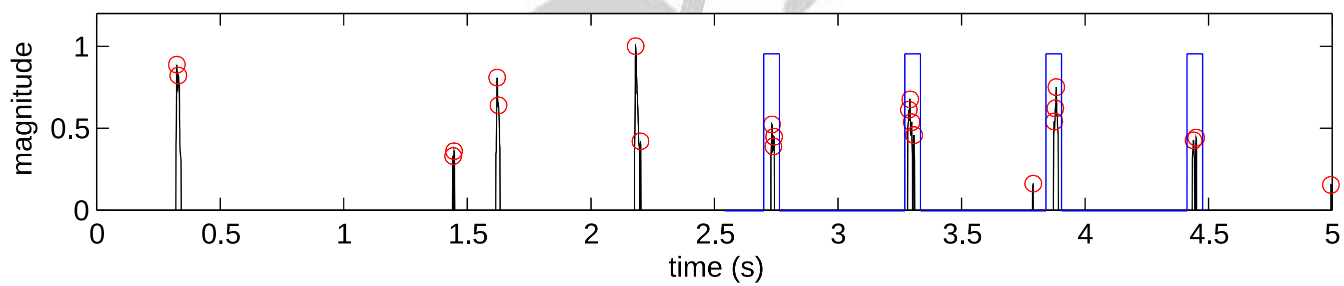
Beat location (1/2)

- method based on a comb filter
  - a pulse-train $q(k)$ of tempo $\mathbb{T}$ is correlated with $p(k)$
  - low complexity operation, it is evaluated only at the indices corresponding to the maxima of $p(k)$
  - find the time index (t_0) where the cross-correlation is maximal



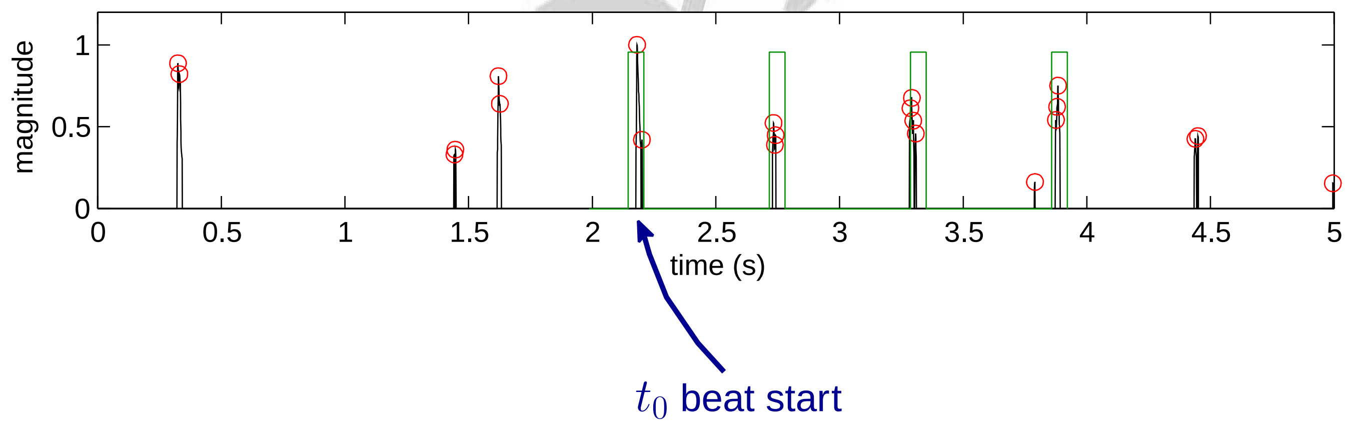
Beat location (1/2)

- method based on a comb filter
  - a pulse-train $q(k)$ of tempo $\mathbb{T}$ is correlated with $p(k)$
  - low complexity operation, it is evaluated only at the indices corresponding to the maxima of $p(k)$
  - find the time index (t_0) where the cross-correlation is maximal



Beat location (1/2)

- method based on a comb filter
  - a pulse-train $q(k)$ of tempo $\mathbb{T}$ is correlated with $p(k)$
  - low complexity operation, it is evaluated only at the indices corresponding to the maxima of $p(k)$
  - find the time index (t_0) where the cross-correlation is maximal



Beat location (2/2)

- for the successive beats in the analysis window

Beat location (2/2)

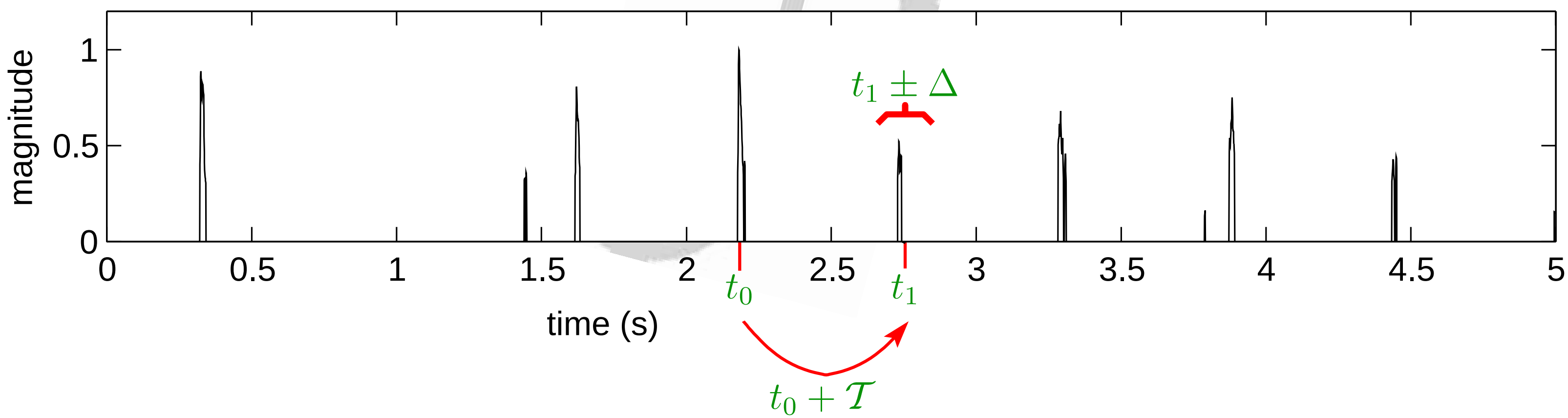
- for the successive beats in the analysis window
  - one beat period $\mathcal{T}$ is added to t_0 or to the last beat location, i.e., $t_i = \lfloor t_{i-1} + \mathcal{T} \rceil$

Beat location (2/2)

- for the successive beats in the analysis window
  - one beat period $\mathcal{T}$ is added to t_0 or to the last beat location, i.e., $t_i = \lfloor t_{i-1} + \mathcal{T} \rfloor$
  - a corresponding peak in $p(k)$ is searched at time $t_i \pm \Delta$

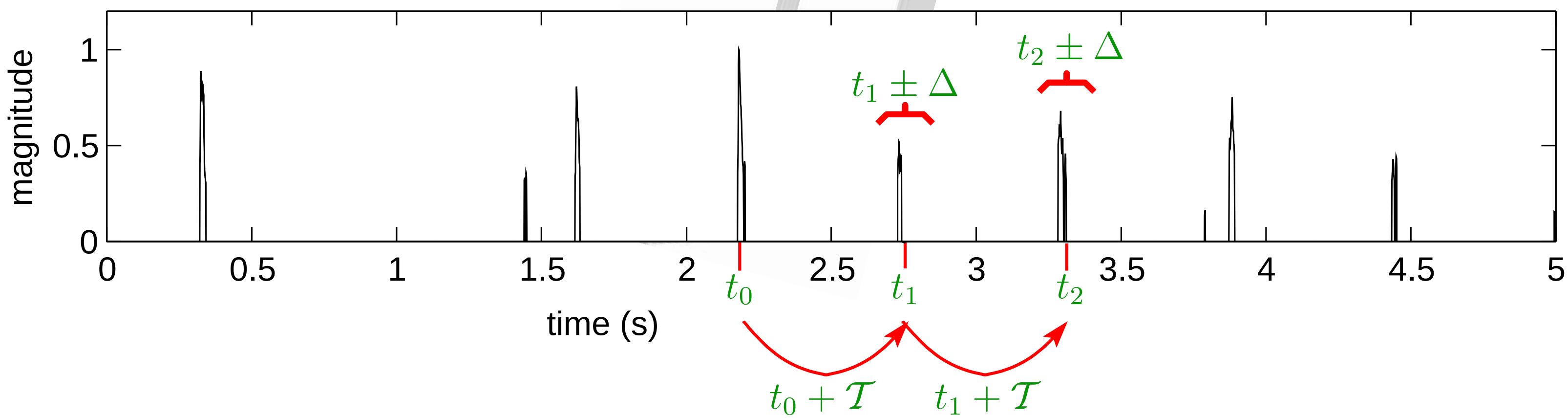
Beat location (2/2)

- for the successive beats in the analysis window
  - one beat period $\mathcal{T}$ is added to t_0 or to the last beat location, i.e., $t_i = \lfloor t_{i-1} + \mathcal{T} \rfloor$
  - a corresponding peak in $p(k)$ is searched at time $t_i \pm \Delta$



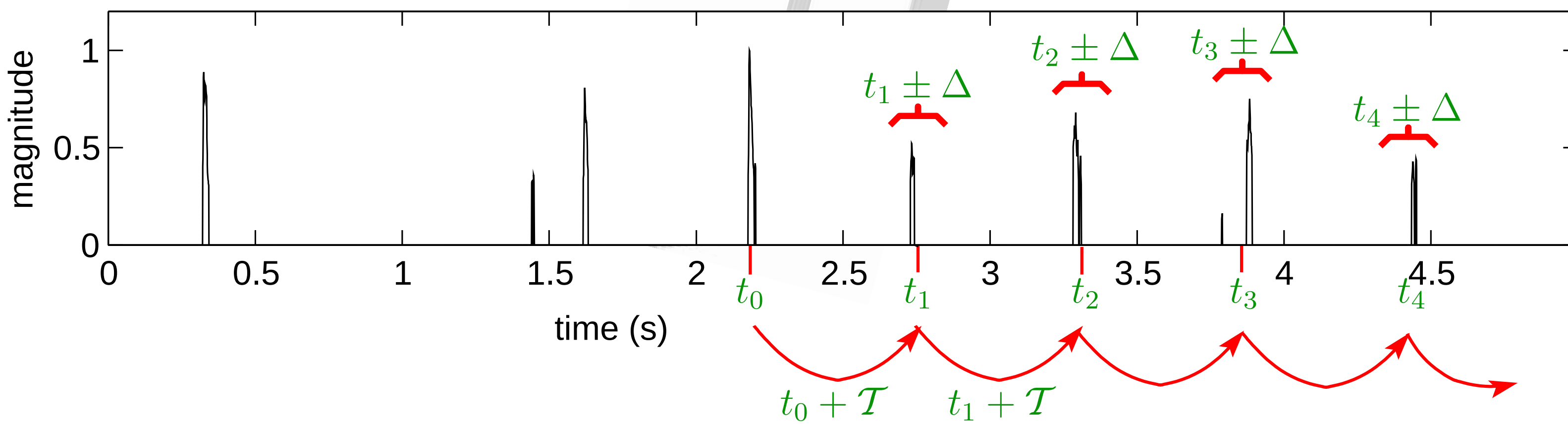
Beat location (2/2)

- for the successive beats in the analysis window
  - one beat period $\mathcal{T}$ is added to t_0 or to the last beat location, i.e., $t_i = \lfloor t_{i-1} + \mathcal{T} \rfloor$
  - a corresponding peak in $p(k)$ is searched at time $t_i \pm \Delta$



Beat location (2/2)

- for the successive beats in the analysis window
  - one beat period $\mathcal{T}$ is added to t_0 or to the last beat location, i.e., $t_i = \lfloor t_{i-1} + \mathcal{T} \rfloor$
  - a corresponding peak in $p(k)$ is searched at time $t_i \pm \Delta$



Beat location (2/2)

- for the successive beats in the analysis window
  - one beat period $\mathcal{T}$ is added to t_0 or to the last beat location, i.e., $t_i = \lfloor t_{i-1} + \mathcal{T} \rfloor$
  - a corresponding peak in $p(k)$ is searched at time $t_i \pm \Delta$
  - when the last beat occurs its location is stored

Beat location (2/2)

- for the successive beats in the analysis window
  - one beat period $\mathcal{T}$ is added to t_0 or to the last beat location, i.e., $t_i = \lfloor t_{i-1} + \mathcal{T} \rfloor$
  - a corresponding peak in $p(k)$ is searched at time $t_i \pm \Delta$
  - when the last beat occurs its location is stored
- the tempo of the new analysis window is calculated

Beat location (2/2)

- for the successive beats in the analysis window
  - one beat period $\mathcal{T}$ is added to t_0 or to the last beat location, i.e., $t_i = \lfloor t_{i-1} + \mathcal{T} \rfloor$
  - a corresponding peak in $p(k)$ is searched at time $t_i \pm \Delta$
  - when the last beat occurs its location is stored
- the tempo of the new analysis window is calculated
  - if $\mathbb{T}_{\text{new}}$ differs by less than 10% from $\mathbb{T}_{\text{old}}$, new peaks are searched in the same way (using $\mathcal{T}_{\text{new}}$)

Beat location (2/2)

- for the successive beats in the analysis window
  - one beat period $\mathcal{T}$ is added to t_0 or to the last beat location, i.e., $t_i = \lfloor t_{i-1} + \mathcal{T} \rfloor$
  - a corresponding peak in $p(k)$ is searched at time $t_i \pm \Delta$
  - when the last beat occurs its location is stored
- the tempo of the new analysis window is calculated
  - if $\mathbb{T}_{\mathrm{new}}$ differs by less than 10% from $\mathbb{T}_{\mathrm{old}}$, new peaks are searched in the same way (using $\mathcal{T}_{\mathrm{new}}$)
  - otherwise, a new beat phase is calculated and peaks are searched using $\mathcal{T}_{\mathrm{new}}$

- Introduction
- Description of the algorithm
- **Performance analysis**
- Sound examples
- Conclusions

Performance analysis

- evaluation using a corpus of 489 musical excerpts

Performance analysis

- evaluation using a corpus of 489 musical excerpts
- wide variety of instruments, dynamic range, etc.

Performance analysis

- evaluation using a corpus of 489 musical excerpts
- wide variety of instruments, dynamic range, etc.
- wide diversity of musical genres

Performance analysis

- evaluation using a corpus of 489 musical excerpts
- wide variety of instruments, dynamic range, etc.
- wide diversity of musical genres

Genre	Pieces	Percentage
classical	137	28.0 %
jazz	79	16.2 %
latin	37	7.6 %
pop	40	8.2 %
rock	44	9.0 %
reggae	30	6.1 %
soul	24	4.9 %
rap, hip-hop	20	4.1 %
techno	23	4.7 %
other	55	11.2 %
total	**489**	**100 %**

Performance analysis

- evaluation using a corpus of 489 musical excerpts
- wide variety of instruments, dynamic range, etc.
- wide diversity of musical genres
- tempi in the 50 to 200 BPM range

Performance analysis

- evaluation using a corpus of 489 musical excerpts
- wide variety of instruments, dynamic range, etc.
- wide diversity of musical genres
- tempi in the 50 to 200 BPM range
- the tempo of each musical piece was manually annotated

Performance analysis

- evaluation using a corpus of 489 musical excerpts
- wide variety of instruments, dynamic range, etc.
- wide diversity of musical genres
- tempi in the 50 to 200 BPM range
- the tempo of each musical piece was manually annotated
  - the musician listens to a musical excerpt using headphones

Performance analysis

- evaluation using a corpus of 489 musical excerpts
- wide variety of instruments, dynamic range, etc.
- wide diversity of musical genres
- tempi in the 50 to 200 BPM range
- the tempo of each musical piece was manually annotated
  - the musician listens to a musical excerpt using headphones
  - while listening, he/she taps the tempo

Performance analysis

- evaluation using a corpus of 489 musical excerpts
- wide variety of instruments, dynamic range, etc.
- wide diversity of musical genres
- tempi in the 50 to 200 BPM range
- the tempo of each musical piece was manually annotated
  - the musician listens to a musical excerpt using headphones
  - while listening, he/she taps the tempo
  - the tapping signal is recorded and the reference tempo ($\mathbb{T}_{\mathrm{R}}$) is extracted from it

Evaluation protocol

- it is generally difficult to define the "*correct beat*" in an objective way

Evaluation protocol

- it is generally difficult to define the “*correct beat*” in an objective way
- people have a tendency to tap at different metrical levels

Evaluation protocol

- it is generally difficult to define the "*correct beat*" in an objective way
- people have a tendency to tap at different metrical levels
- this problem also occurs in automatic tempo estimation
  - $\mathbb{T}$ estimated by the algorithm is labeled as correct if there is a less than 5% disagreement from $\mathbb{T}_{\mathrm{R}}$, as follows:

$$0.95\alpha\mathbb{T} < \mathbb{T}_{\mathrm{R}} < 1.05\alpha\mathbb{T} \text{with } \alpha \in \{\tfrac{1}{2}, 1, 2\}$$

Results (1/3)

- the proposed algorithm was compared to our own previous work on tempo estimation
- it was also compared to our own implementation of the methods proposed by Paulus[1] and Scheirer[2]

Results (1/3)

- the proposed algorithm was compared to our own previous work on tempo estimation
- it was also compared to our own implementation of the methods proposed by Paulus[1] and Scheirer[2]
- overall recognition rate for the evaluated systems

Method	Recognition rate
Paulus	56.3 %
Scheirer	67.4 %
SP .	63.2 %
AC .	73.6 %
SP using SEF.	84.0 %
AC using SEF	89.7 %

Results (2/3)

- the performance of these methods by musical genre is

Method Genre	Paulus %	Scheirer %	SP %	AC %	SP-SEF %	AC-SEF %
classical	46.0	46.2	48.2	70.8	71.5	82.4
jazz	57.0	70.9	62.0	69.8	78.4	86.0
latin	70.3	81.1	62.1	70.3	91.8	94.5
pop	57.5	70.0	75.0	85.7	92.5	92.5
rock	40.9	84.1	61.3	84.4	81.8	88.6
reggae	76.7	86.7	86.6	76.9	96.6	100
soul	50.0	87.5	70.8	76.7	100	100
rap	75.0	85.0	75.0	56.5	100	100
techno	69.6	56.3	65.2	95.0	95.6	100
other	61.8	69.1	74.5	66.7	89.0	90.9

Results (3/3)

- due to the unavailability of *beat-labeled* audio tracks, the beat location was evaluated at a subjective level
- during the evaluation, we found that the algorithm produces erroneous results when
  - processing signals with long fading-in attacks
  - many instruments play simultaneously, their “spectral mixture” lacks of stable regions leading to false onsets
  - tempo varies too quickly in short time segments or if there are large beat gaps in the signal

[1] Paulus J. and Klapuri A.,“*Measuring the similarity of rhythmic patterns*”, Proceedings of the ISMIR, 2002.

[2] Scheirer, E.D., “*Tempo and beat analysis of acoustic music signals*”, JASA, January 1998.

Sound examples

- example rock
- example country music
- example soul
- example salsa
- example guitarre
- example jazz 1
- example jazz 2
- example musique classique 1
- example musique classique 2

Conclusions

- efficient beat tracking algorithm for audio recordings

Conclusions

- efficient beat tracking algorithm for audio recordings
- the concept of Spectral Energy Flux was used to derive an onset detector

Conclusions

- efficient beat tracking algorithm for audio recordings
- the concept of Spectral Energy Flux was used to derive an onset detector
  - effective for a large range of audio signals
  - straightforward to implement
  - relatively low computational cost

Conclusions

- efficient beat tracking algorithm for audio recordings
- the concept of Spectral Energy Flux was used to derive an onset detector
  - effective for a large range of audio signals
  - straightforward to implement
  - relatively low computational cost
- the performance was evaluated on a large database containing 489 musical pieces

Conclusions

- efficient beat tracking algorithm for audio recordings
- the concept of Spectral Energy Flux was used to derive an onset detector
  - effective for a large range of audio signals
  - straightforward to implement
  - relatively low computational cost
- the performance was evaluated on a large database containing 489 musical pieces
- global success rate of 89.7%

Conclusions

- efficient beat tracking algorithm for audio recordings
- the concept of Spectral Energy Flux was used to derive an onset detector
  - effective for a large range of audio signals
  - straightforward to implement
  - relatively low computational cost
- the performance was evaluated on a large database containing 489 musical pieces
- global success rate of 89.7%
- the system works off-line
  - non-causality issues should be solved before a real time implementation