ircam Centre Pompidou

Time Variable Tempo Detection and Beat Marking

Geoffroy Peeters (peeters@ircam.fr)
IRCAM - Analysis-Synthesis Team
SemanticHiFi I.S.T. European Project

Introduction

Goal:

- to detect time-variable tempo for music with and without percussion

Application:

- search in a database by tempo, meter
- beat synchronous processing (// pitch synchronous processing) -> audio summary
  beat synchronous mixing, beat slicing, segmentation into beat units
- music analysis

Huge number of proposals for this task:

- Scheirer, Goto, Gouyon, Dixon, Alonso, Laroche, Tzanetakis, Brown, Cemgil, Foote, Dannenberg, Klapuri, Paulus, ...
- For which music genre ? classical music, jazz music, … ?
  - Problems
    - Complexity of the rhythm,
    - No clear onsets,
    - Variations of the tempo

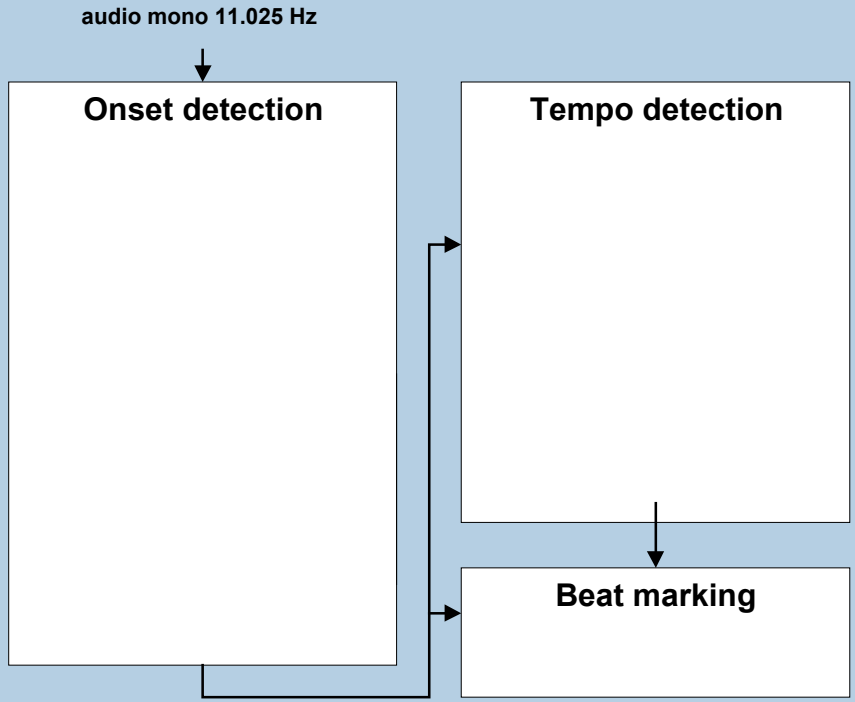
Two different kind of approaches

- signal energy along time (bank of filters) -> measure of periodicity
  - [Scheirer98]
- onsets detection -> onsets inter-distance -> IOIH -> most common periodicity
  - [Dixon01, Gouyon02]

Introduction

System overview

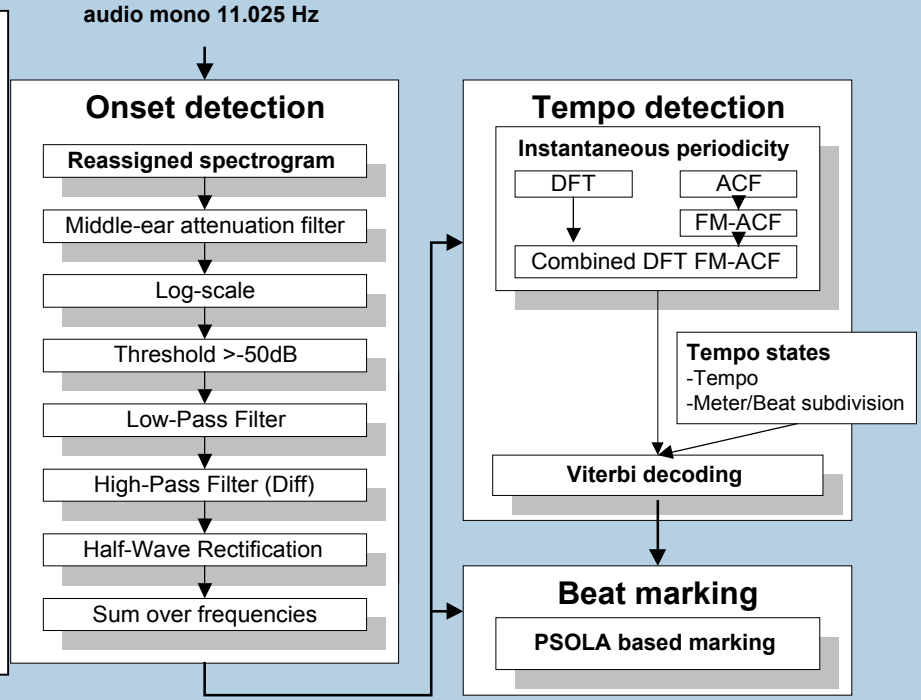
- Energy-onset function
  - ➔
- Tempo detection
  - periodicity measure
    - ➔
  - Tempo detection
    - ➔
- Beat marking
  - ➔
- Evaluation



Introduction

System overview

- Energy-onset function
  - reassigned spectral energy flux
- Tempo detection
  - periodicity measure
    - combined DFT / FM-ACF
  - Tempo detection
    - Viterbi tracking of “tempo states”
- Beat marking
  - PSOLA-based method
- Evaluation

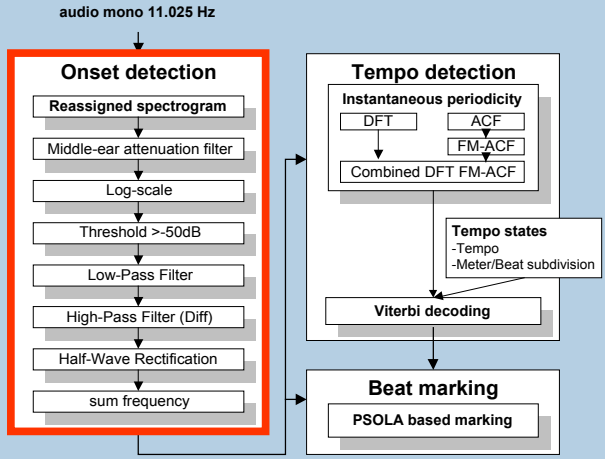


Onset-energy function

1. Onset-Energy Function

Introduction

- Energy function [Scheirer] or Discrete onset [Gouyon] ?
  - Missing/added onset, no onset, …
  - Continuous onset-energy function
- Observe the signal through something meaningful in terms of musical periodicity
  - spectrum similarity [Foote]
  - energy (bank of filters) [Scheirer98]
  - be able to detect note transitions at constant energy, slow attacks -> visible in a spectrogram
    - ➜ variations of the spectrogram along time (notes, …)
    - ➜ spectral energy flux [Laroche2003], Alonso [2004]

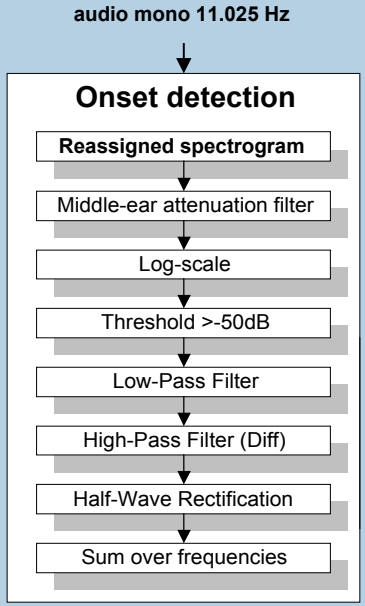


1. Onset-Energy Function

- Normal Spectral Energy Flux
  - Signal (mono, sr=11.025 Hz)
  - Spectrogram
  - Log-scale [Klapuri 1999]
  - Low-pass filter
  - High-pass filter
  - Half-Wave Rectified
  - Sum over frequencies

1. Onset-Energy Function

- Normal Spectral Energy Flux
- Reassigned Spectral Energy Flux
  - Signal (mono, sr=11.025 Hz)
  - Reassigned spectrum
    - hamming window 0.0928s./0.0464s.
    - Hop size 0.0058s.
  - Middle-ear filtering [Moore 1997]
  - Log-scale [Klapuri 1999]
  - Threshold of -50dB
  - Low-pass filter : elliptic filter of order 5, fc=10 Hz
    - remove fast varitions (noise, cymbals, …)
  - High-pass filter: [1,-1] differentiator
    - remove DC components, detect variations
  - Half-Wave Rectified
    - keep only the increasing (attack) parts
  - Sum over frequencies
  - $e(n=t_i)$ (sr=172 Hz)



1. Onset-Energy Function

Reassigned Spectrogram [Flandrin1999]

- reassigned the energy of the "bins" of the STFT to their center of gravity
- increase temporal and spectral resolution
  - avoids attack blurring
  - allows to better differentiate very close pitches

Method:

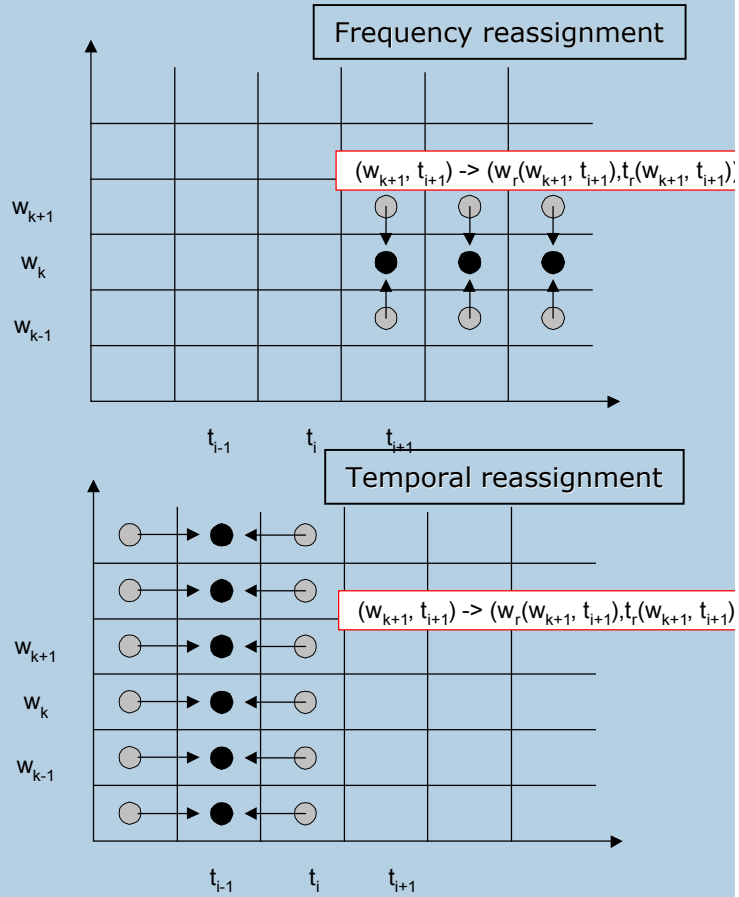
- reassigned the energy of the "bins" of the STFT to their center of gravity
  - Reassignment of frequency
    -> instantaneous frequency

$$\omega_r(x,t_i,\omega_k) = \omega_k - \Im\left\{\frac{STFT_{dh}(x,t_i,\omega_k)}{STFT_h(x,t_i,\omega_k)}\right\}$$

  - Reassignment of time
    -> group delay

$$t_r(x,t_i,\omega_k) = t_i + \Re\left\{\frac{STFT_{th}(x,t_i,\omega_k)}{STFT_h(x;t_i,\omega_k)}\right\}$$

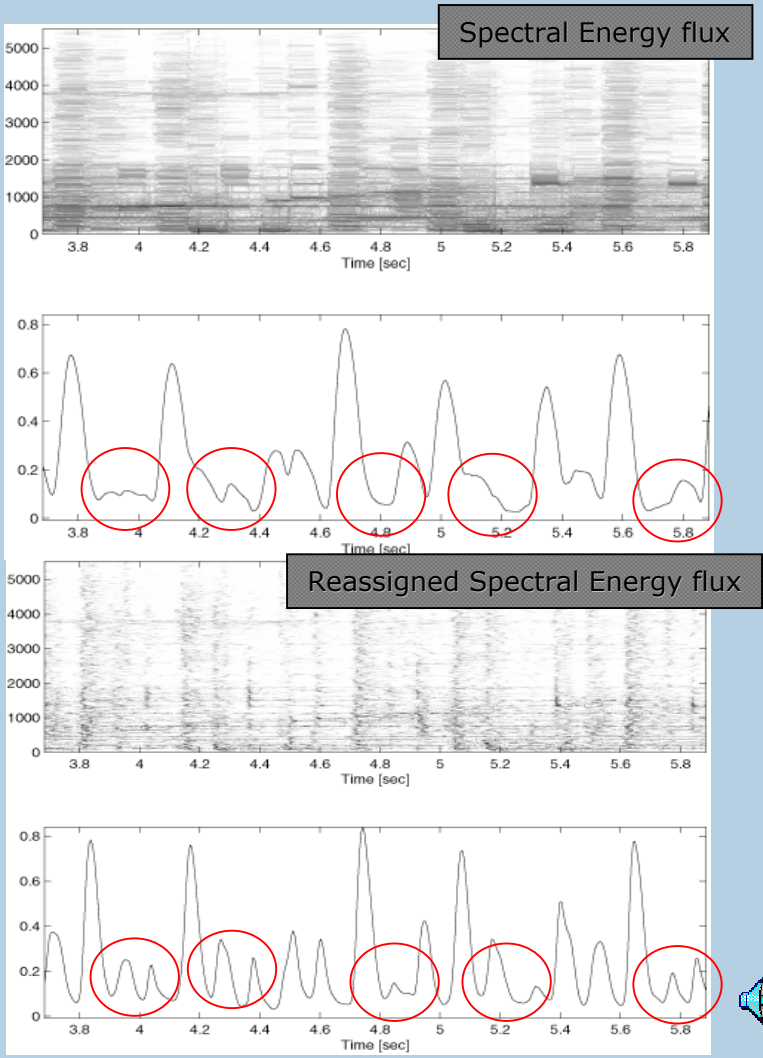
.



1. Onset-Energy Function

Results

- Carlinhos Brown “Pandeiro Deiro” from ISMIR 2004 test database



Tempo detection

2. Tempo detection

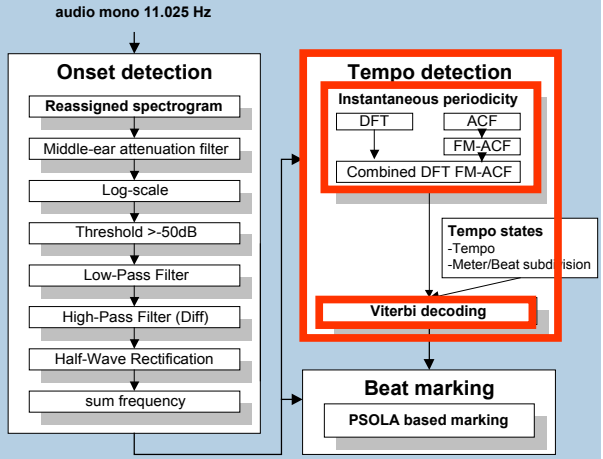
Tempo detection:

a) Periodicity Estimation

- estimate the dominant periodicities around a specific time

b) Tempo detection

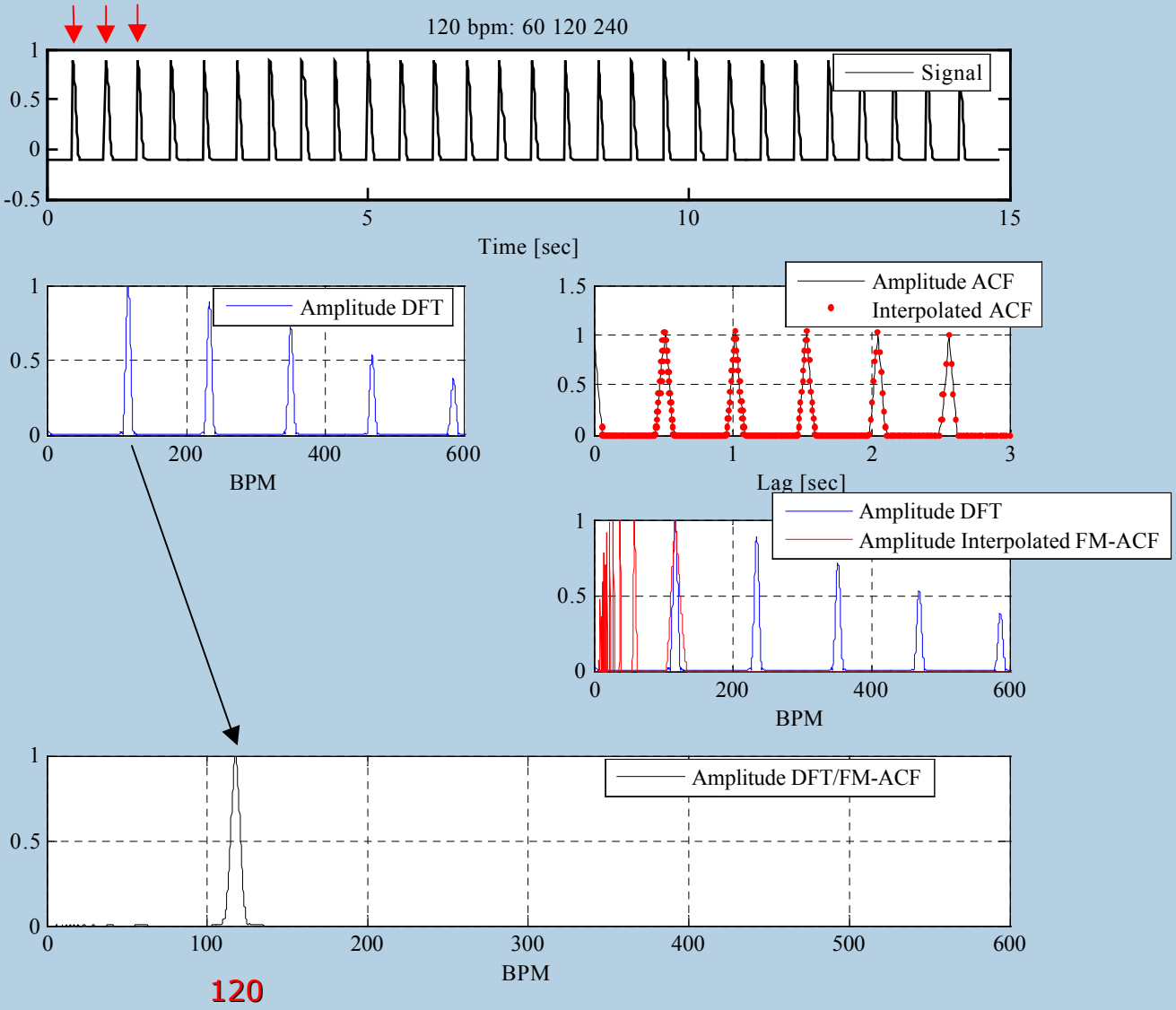
- estimate the tempo and meter/beat subdivision path that best explains the observed periodicities over time



Periodicity estimation

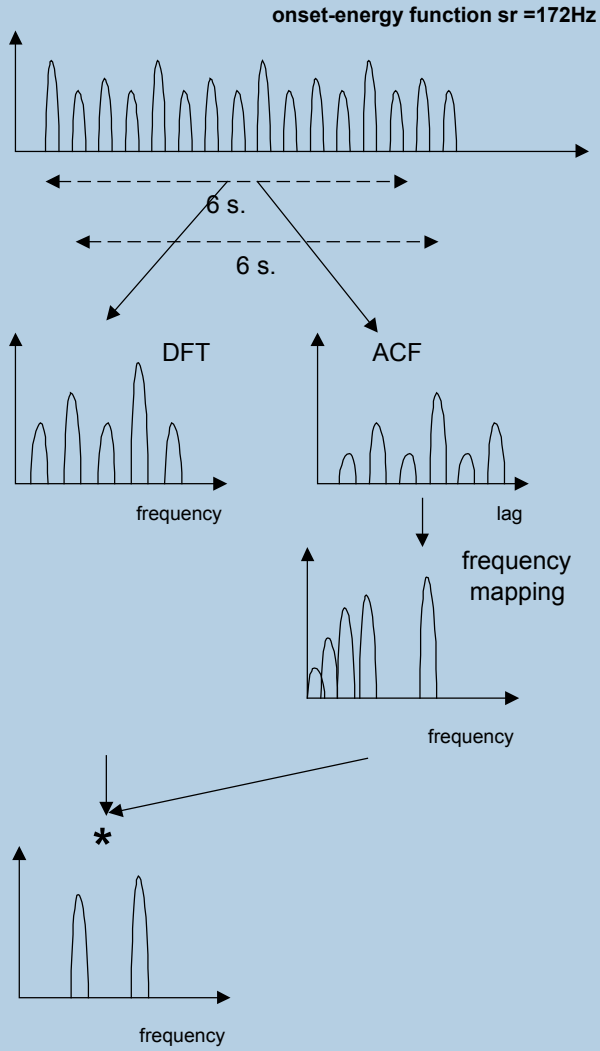
2a. Periodicity Estimation

- Periodicity estimation ?
  - Discrete Fourier Transform DFT $F(w_k,ti)$
    - frequency domain
    - DFT of e(n) => set of harmonically related frequency
    - depending on their relative amplitude -> difficult to decide which harmonic correspond to the tempo frequency
    - octave errors especially detrimental in the case of triple or compound meter -> can lead to musically insignificant frequencies
      - ➔ 4/3 <- 2 <- 4 (quarter note) -> 8 -> 12
  - AutoCorrelation Function ACF $A(l,ti)$
    - lag (time) domain
    - ACF of e(n) => set of periodically related lags
    - difficult to decide which period correspond to the tempo lag
  - Solution via models: Two-way mismatch, maximum likelihood, ...
  - Octave uncertainties of DFT and ACF occur in inverse domain
    Idea: combined both function
    - Keep only common periodicities

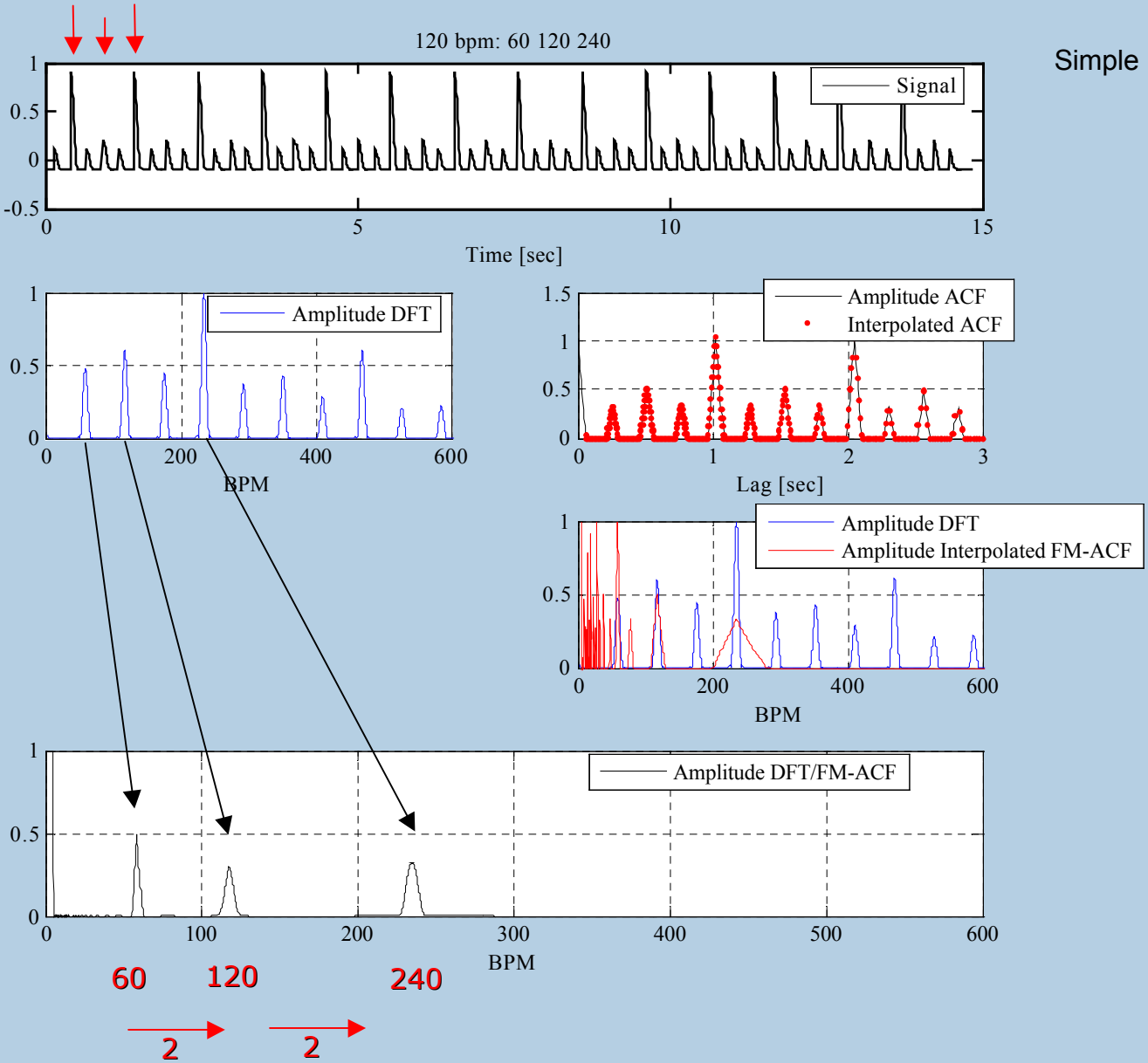


2a. Periodicity Estimation

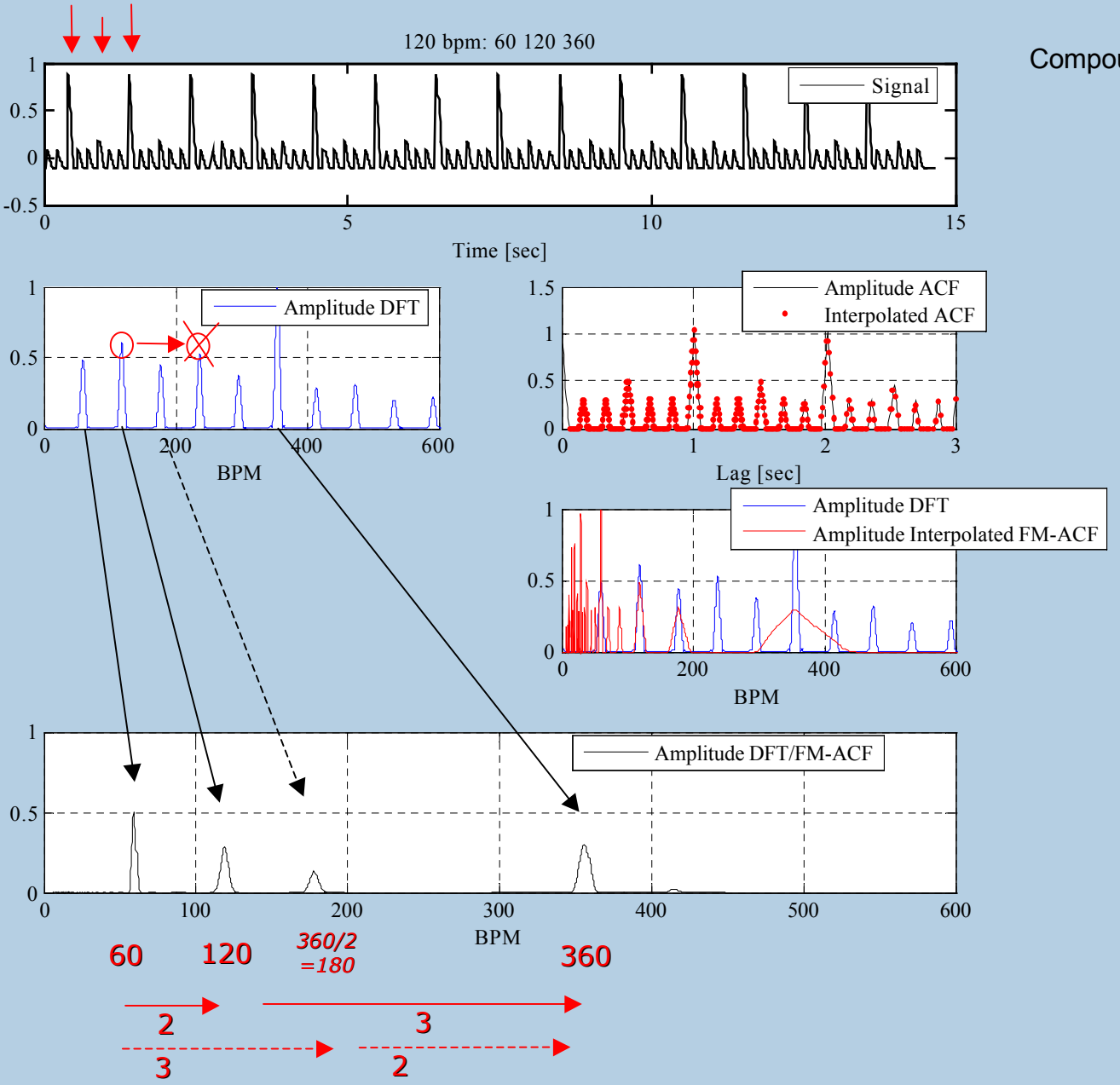
- Onset-energy signal 176Hz
- Frame analysis:
  - times t_i
  - window size 6 s., hop size 0.5 s.
- Discrete Fourier Transform DFT
  - $F(w_k,t_i)$
- AutoCorrelation Function ACF
  - $A(l,t_i)$ normalized in energy
  - Frequency-Mapped ACF
    - $A(l,t_i)$ -> $A(w_k,t_i)$
    - interpolation and sampling of $A(l,t_i)$ at sr/w_k
    - HWR (keep only positive correlation)
    - remark: decreasing frequency resolution
- Combined function: product function
  - $Y(w_k,t_i) = F(w_k,t_i)$. $A(w_k,t_i)$
- T/F matrix
  - $Y(w_k,t_i)$



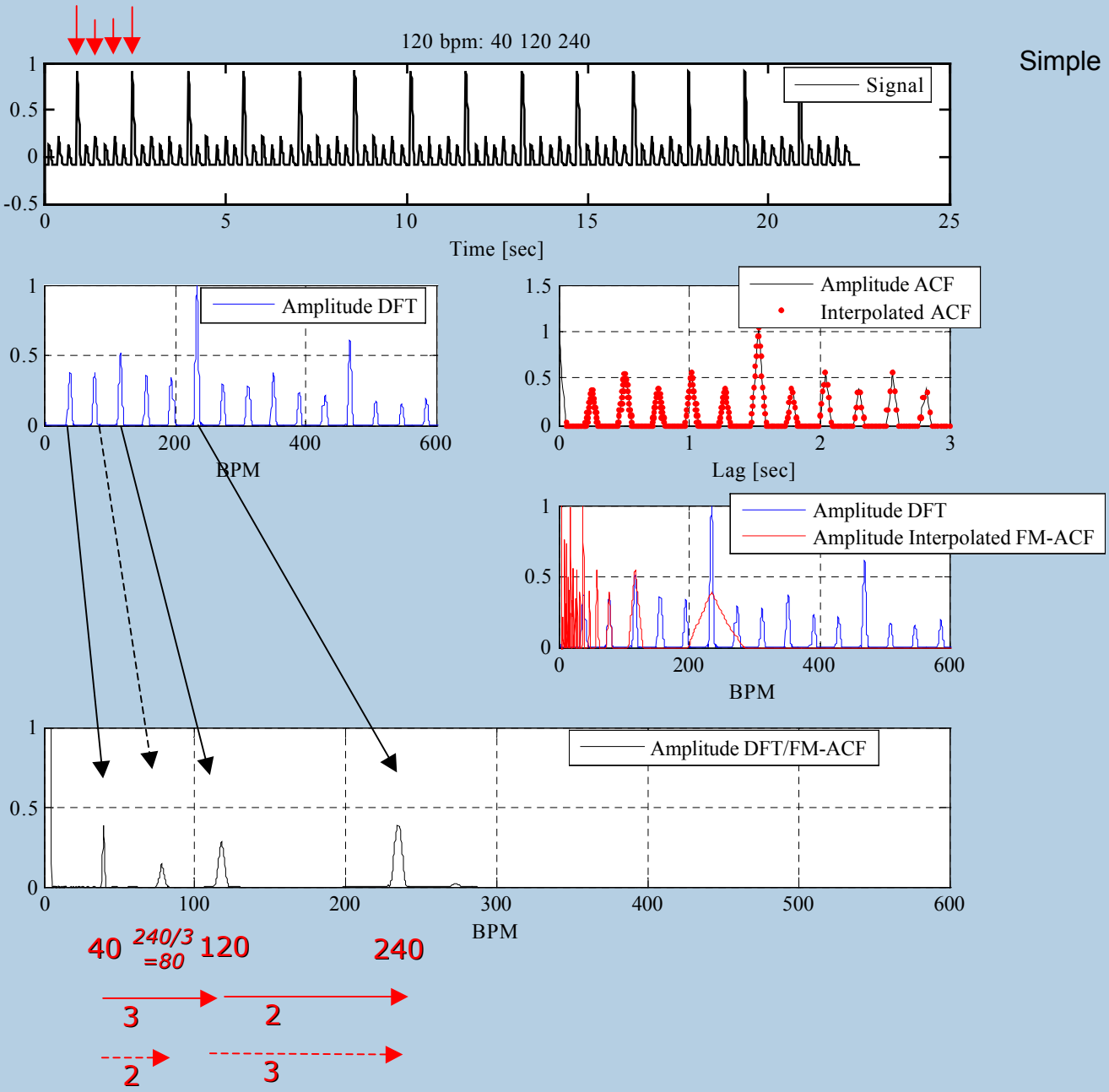
Simple meter in 2/4



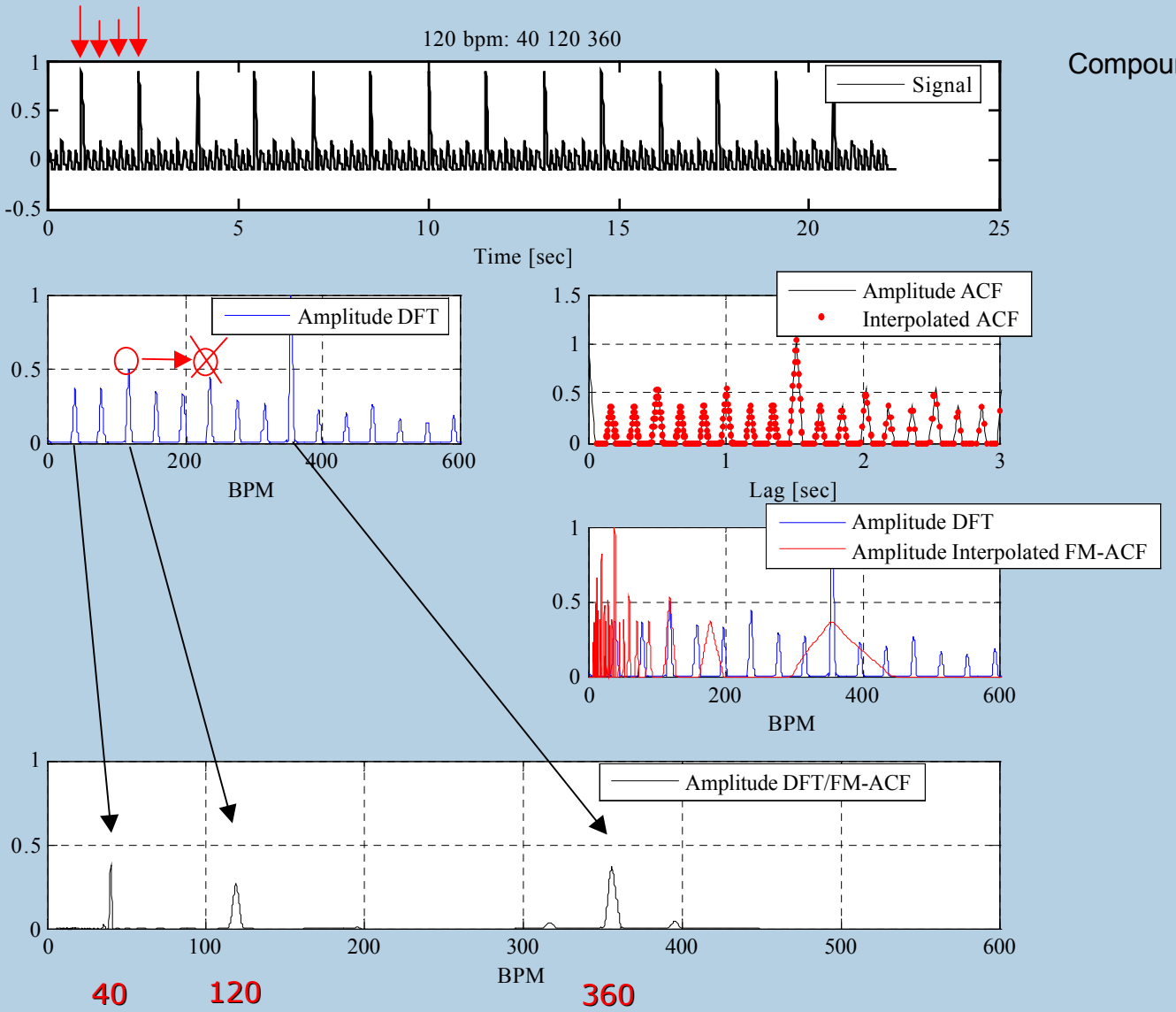
Compound meter in 2/4



Simple meter in 3/4



Compound meter in 3/4



2a. Periodicity Estimation

- Combined DFT / FM-ACF pro and cons
  - Pro:
    - allows to better distinguish the various periodicities relationships
    - without using a specific model
  - Cons:
    - still exist some ambiguous periodicities
    - decreasing frequency resolution of FM-ACF

Tempo and meter/beat subdivision estimation

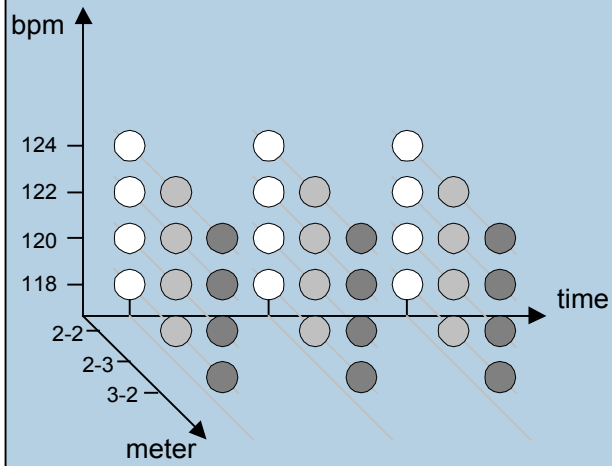
2b. Tempo Detection

Observed periodicities depend on

- Tempo b_i
- Rhythm characteristics
  - Meter/Beat Subdivision Templates (MBST) m_j
  - Three templates
    - (2-2) duple/simple meter 2/4
    - (2-3) duple/compound meter 6/8
    - (3-2) triple/simple meter 3/4

- Definition: Tempo states $S(i,j) = [b_i, m_j]$

- Estimate the tempo states S(i,j) path that best explains the observed periodicites $Y(w_k, t_i)$
  - Viterbi decoding algorithm



2b. Tempo Detection

Viterbi decoding algorithm

- Observation probabilities:
  - $p_{obs}(b_i,m_j) = p_{prior}(b_i,m_j)$. $p([b_i,m_j] \mid Y(w_k,t_i))$

  a) $p_{prior}(b_i,m_j)$
  - favors the detection of tempo in the range 50-150 bpm
  - doesn't favor any mbst

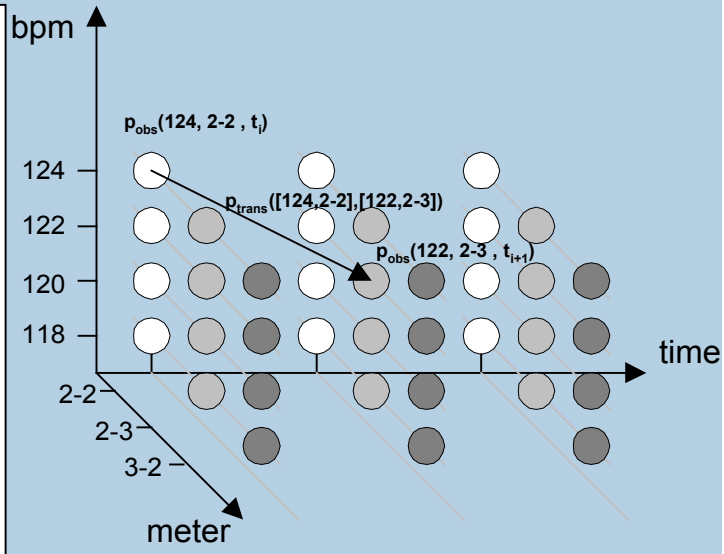
  $= p_{prior}(b_i) = N_{m=100,s=150}(b_i)$

  b) $p([b_i,m_j] \mid Y(w_k,t_i))$

  probability to observe a specific tempo b_i and mbst m_j

  - $p([x,(2-2)]|Y_n) = \frac{(\alpha Y_n(x/2)+Y_n(x)+\beta Y_n(2x))}{\sum_y Y_n(y)}$
  - $p([x,(2-3)]|Y_n) = \frac{(\alpha Y_n(x/2)+Y_n(x)+\beta Y_n(3x))}{\sum_y Y_n(y)}$
  - $p([x,(3-2)]|Y_n) = \frac{(\alpha Y_n(x/3)+Y_n(x)+\beta Y_n(2x))}{\sum_y Y_n(y)}$

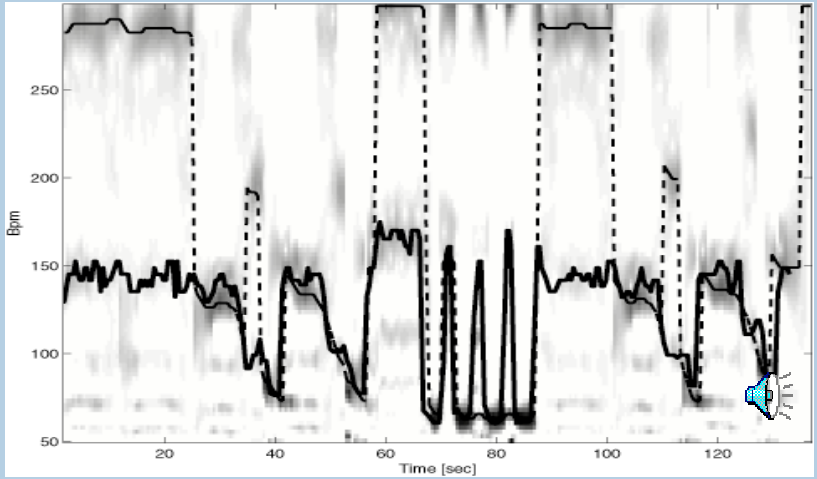
- Transition probabilities
  - $p_{trans}([b_i,m_j],[b_k,m_l])$
    - favors continuous tempo,
    - favors continuous mbst
  - $= p_{trans}([b_i,b_k])$ * $p_{trans}([m_j,m_l])$
    $= N_{m=bi,s=5}(b_k)$ * 0,1
- Initial probability



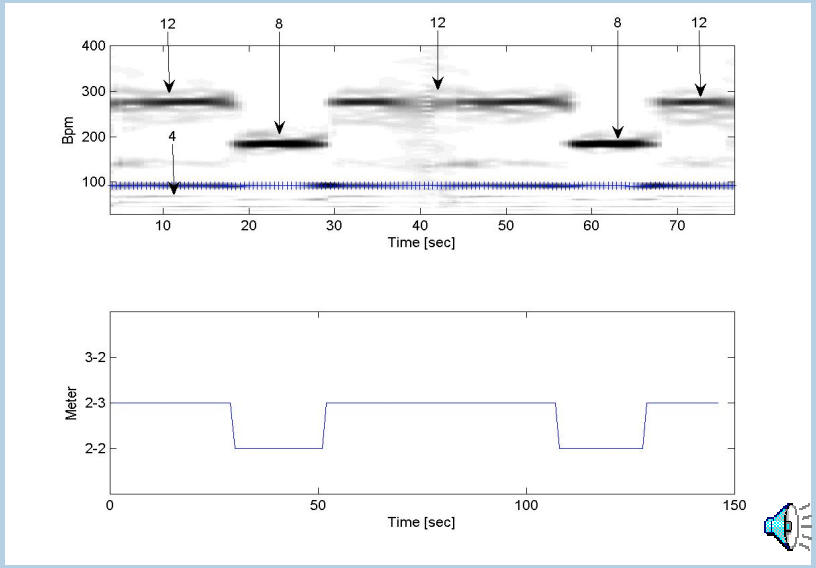
2b. Tempo Detection

Example:

- Brahms “Ungarische Tanze n5”



- Standard of Excellence accompaniment CD Book 2 All inst. Track 88. Looby Loo



Beat marking

3. Beat marking (not in the paper)

Method: PSOLA based

[Peeters, PHD 2001]

speech processing method

detect the glottal closure instant (GCI)

GCI are periodic

GCI causes a burst of energy

Search for local maxima

$$\Theta = [\theta_0, \theta_1, ..., \theta_i, ...]$$

$$I_i = \left[\theta_i - \frac{\text{Tbpm}_{i-1}}{\alpha}, \theta_i + \frac{\text{Tbpm}_i}{\alpha}\right]$$

Two constraints

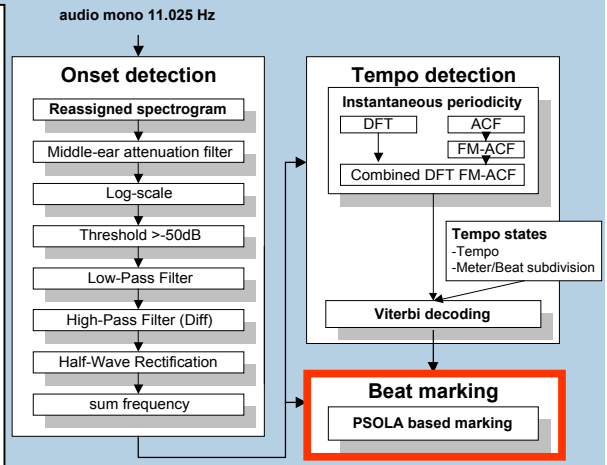
- local periodicity
- local maxima

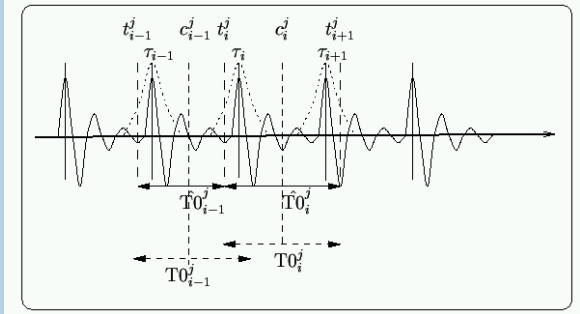
$$\begin{cases} \text{m}_i - \text{m}_{i-1} = \text{Tbpm}_{i-1} \\ \text{m}_{i+1} - \text{m}_i = \text{Tbpm}_i \\ \text{m}_i = \tau_i \end{cases}$$

- least-square solution

$$\epsilon = \sum_{i \in I} \left[((\text{m}_i - \text{m}_{i-1}) - \text{Tbpm}_{i-1})^2 + \beta(\text{m}_i - \tau_i)^2\right]$$

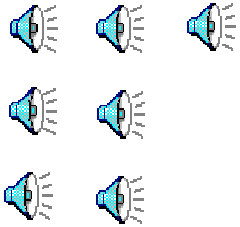
.





3. Beat marking (not in the paper)

Audio examples:



.

Evaluation

4. Evaluation

Databases

- Ballroom-dancer (ISMIR 2004) — 698 instances, 30 sec., ChaChaCha, Rumba, Quickstep, Waltz, ...
- RWC database — 182 Instances, 20 sec. Long Classical, Opera, Jazz, World, Pop, Rock
- Pop-rock hits database — 158 instances, 20 sec. Long Pop-Rock

Evaluation method

- accuracy 1 — ground truth tempo
- accuracy 2
  - 2:2 — 1/2 — 1 — 2
  - 3:2 — 1/3 — 1 — 2
  - 2:3 — 1/2 — 1 — 3

4. Evaluation

Comments:

- Ballroom-dancer
  - global 63%/92%
  - reminder ismir2004: 63% (Klapuri) / 92%(Anonymous) but different rules
  - accuracy 1: most errors in Jive, Quickstep, Rumba, Waltz (algorithm follows the tatum)
  - accuracy 2: most errors in Waltz (bad mbst estimation, onset difficult in slow chord trans.)
- RWC
  - global 79% (accuracy 2)
  - Classical Music: 70% (bad onset detection - slow chord transition / fuzzy tempo - difficult to detect manually)
  - Jazz Music: 83% (bad mbst estimation)
  - Music Genre: 85%
- Pop-rock hit's database
  - global 79% / 98%

Ballroom database

	ChaChaCha	Jive	Quick Step	Rumba	Samba	Tango	Viennese Waltz	Waltz	**Total**
# items	111	60	82	98	86	86	65	110	698
Accuracy 1	98%	53%	37%	54%	68%	95%	85%	44%	**63%**
Accuracy 2	100%	98%	91%	94%	93%	94%	91%	78%	**92%**

RWC database

	Classique	Jazz	Music Genre	**Total**
# items	78	59	61	**182**
Accuracy 1				
Accuracy 2	70%	83%	85%	**79%**

Poprock database

	Total
# items	158
Accuracy 1	**79%**
Accuracy 2	**98%**

4. Evaluation (not in the paper)

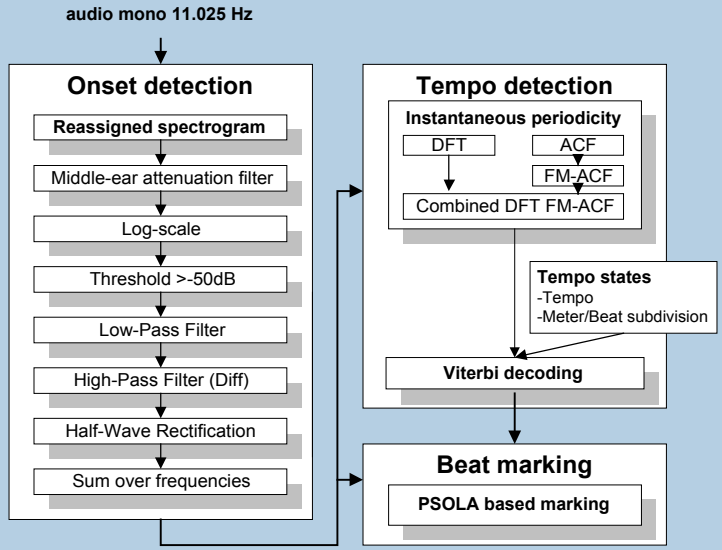
ISMIR2005

- **Goal**: The comparison and evaluation of current methods for the extraction of tempo from musical audio
- Estimate the two dominant perceptual tempi, their likelihood and first beat position
- **Dataset**: 140 wav files, 354 Megabytes

Rank	Participant	Score (std. de	At Least One	Both Tempos	At Least One	Both Phases (	Mean Absolut	Runtime (s	Machine
1	**Alonso, M.**	**0.689 (0.231)**	**95.00%**	55.71%	25.00%	5.00%	0.239	2875	G
2	***	0.675 (0.273)	90.71%	59.29%	32.14%	7.14%	0.222	1160	F
3	***	0.675 (0.272)	90.71%	59.29%	32.86%	6.43%	0.222	2621	F
4	***	0.670 (0.252)	92.14%	56.43%	40.71%	7.86%	0.311	3303	G
5	**Peeters, G.**	**0.656 (0.223)**	**95.71%**	47.86%	27.86%	4.29%	0.258	2159	R
6	***	0.649 (0.253)	92.14%	51.43%	37.14%	5.71%	0.305	2050	G
7	***	0.645 (0.294)	87.14%	55.71%	48.57%	10.71%	0.313	1357	G
8	***	0.644 (0.300)	86.43%	53.57%	37.14%	5.71%	0.230	1665	Y
9	***	0.628 (0.284)	86.43%	48.57%	26.43%	4.29%	0.224	1005	R
10	***	0.607 (0.287)	87.14%	47.14%	36.43%	6.43%	0.294	1388	R
11	***	0.597 (0.252)	90.71%	37.86%	30.71%	0.71%	0.239	70975	Y
12	***	0.583 (0.333)	80.71%	51.43%	28.57%	2.14%	0.223	180	B 0
13	***	0.538 (0.359)	71.43%	50.71%	28.57%	3.57%	0.295	7173	B 0

5. Conclusion and Future works

- Main proposals
  - reassigned spectral energy flux
  - combined DFT / FM-ACF
  - Viterbi decoding of tempo and meter/beat subdivision
  - PSOLA based marking
- Results/Evaluation
  - good starting point,
  - close to state of the art



5. Conclusion and Future works

- Tempo errors
  - onsets are difficult to detect (slow chord transition)
    - necessity to use other signal observations
  - bad estimation of mbst (2-2 is often confused with 2-3 in the presence of accentuated dotted-quarter note)
    - beat : dotted quarter note -> 8th not: 8th note triplet
  - Complex rhythm (jazz) not adequately represented by the mbst
    - extend the mbst
  - Classical music: rapid tempo change (ritardando), Jazz music: Syncopation
- Octave errors
  - beat is few emphasized -> follow the tatum
    - reduce the tempo range of prior probability
- Future works
  - onset evaluation (gain using reassigned spectral energy flux)
  - when no onset are present -> use another kind of observation
  - extend the templates to other kind of music (jazz, funk)