Beat-tracking using a Probabilistic Framework and Linear Discriminant Analysis

Geoffroy.peeters@ircam.fr
Ircam Sound Analysis-Synthesis Team
CNRS STMS

Introduction

- Applications of beat-tracking
  - Beat-synchronous analysis
    - … score alignment, cover version identification
  - Beat-synchronous processing
    - … time-stretching, beat-shuffling, beat-slicing
  - Music analysis
    - … prior for pitch estimation, onset estimation
  - Visualization
    - … time-grid in audio sequencers

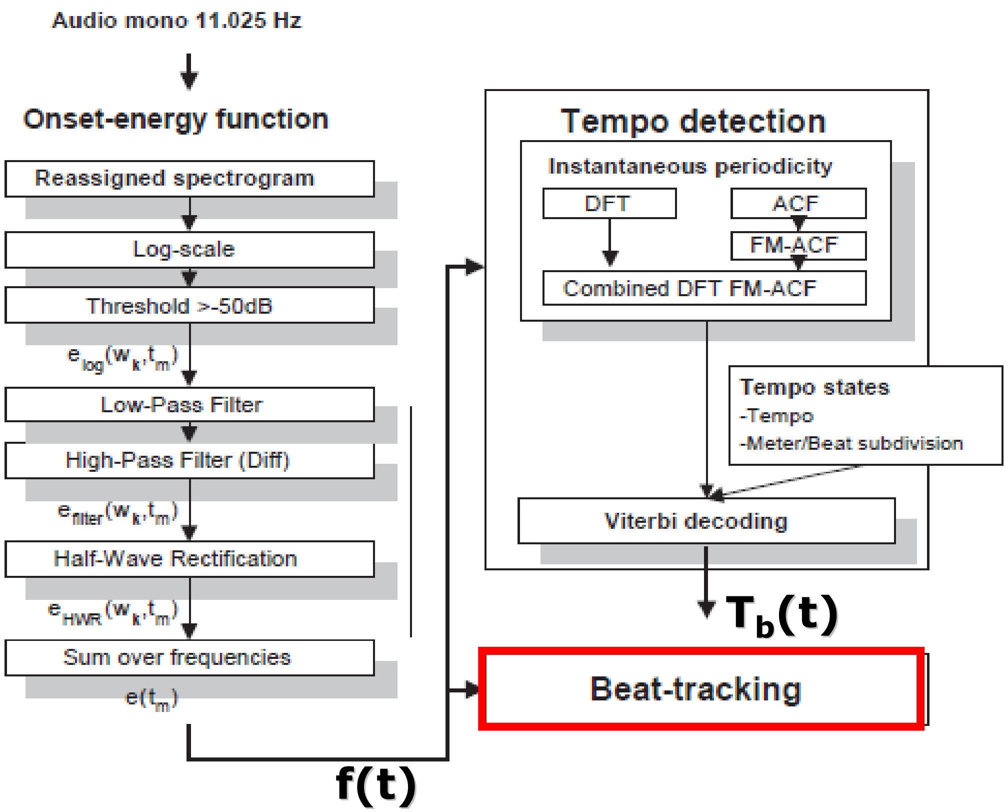
- State-of-the-art
  - Current results far from being perfect
    - See last Audio Beat Tracking contest (MIREX-2006) : P-score: 0.575
      - … Good results for most rock, pop or dance music tracks (except highly compressed dance music tracks: hi-hat problem)
      - … Difficulties for classical (tempo variation), jazz (syncopation), world music
        + Western modern music= Drum'n'bass or R'n'B

Introduction

- Content:
  - Two new approaches for beat-tracking
    - given tempo/ meter as input
  - Algorithm 1: based on P-sola/ GCI location algorithm
  - Algorithm 2: Probabilistic framework
    - Formulation: inverse Viterbi
    - Prior probability
    - Observation probability
      - … use of beat-template: LDA training of the best beat-templates
    - Transition probability
    - Decoding
  - Beat-templates evaluation
  - Large-scale evaluation on four test-sets

Input parameters of our system

- Previous tempo/ meter estimation algorithm



[Peeters, G. (2007). "Template-based estimation of time-varying tempo." EURASIP Journal on Advances in Signal Processing **2007**(Special Issue on Music Information Retrieval Based on Signal Processing): Article ID 67215, 14 pages.]

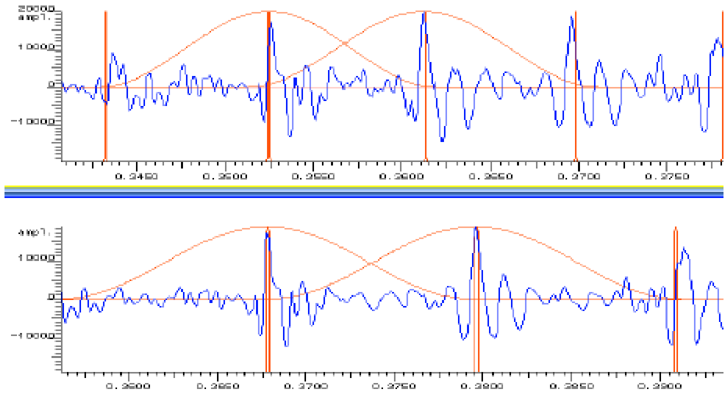
Introduction

- Content:
  - Two new approaches for beat-tracking
    - given tempo/ meter as input
  - Algorithm 1: based on P-sola/ GCI location algorithm
  - Algorithm 2: Probabilistic framework
    - Formulation: inverse Viterbi
    - Prior probability
    - Observation probability
      - use of beat-template: LDA training of the best beat-templates
    - Transition probability
    - Decoding
  - Beat-templates evaluation
  - Large-scale evaluation on four test-sets

Beat-tracking algorithm 1: P-sola method

ircam Centre Pompidou

Introduction

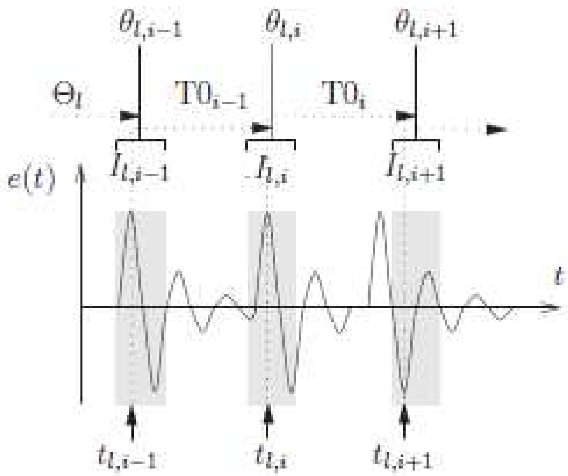


- P-sola
  - Pitch-Synchronous Over-Lap and Add
  - used for speech pitch-shifting or time-stretching
- P-sola Analysis
  - Detect the Glottal Closure Instant (GCI)
- Characteristics of the GCI
  - (a) GCI close to the local maxima of the energy function
  - (b) Inter-distance between successive GCIs close to the local pitch period $T_0(t)$
- Looks close to the problem of Beat-tracking
  - (a) Beat-markers close to the local maxima of the onset-energy-function f(t)
  - (b) Inter-distance between successive beat-markers equal to the local tempo $T_b(t)$
- Idea: adapt a P-sola analysis algorithm to the beat-marking case

  [Peeters, G. (2001). Modeles et modelisation du signal sonore adaptes a ses caracteristiques locales. Analyse/Synthese. Ircam, Paris, France, Universite Paris VI.]

Beat-tracking algorithm 1: P-sola method

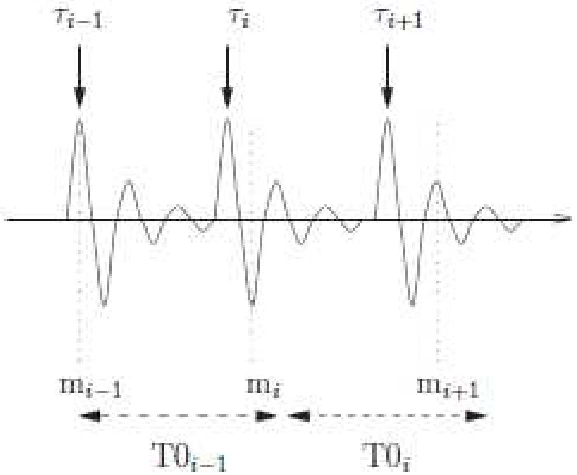
- First stage
  Local maxima detection
  - See paper for details



- Second stage
  Least-square optimization
  - Markers should satisfy
    - (a) Markers m_i close to local maxima T_i
    - (b) Inter-distance between m_i and m_{i+1} equal to local tempo-period Tb_i

$$\begin{cases} (a) : \mathrm{m}_i = \tau_i \\ (b) : \mathrm{m}_i - \mathrm{m}_{i-1} = \mathrm{Tb}_{i-1} \\ (b) : \mathrm{m}_{i+1} - \mathrm{m}_i = \mathrm{Tb}_i \end{cases}$$

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



Beat-tracking algorithm 1: P-sola method

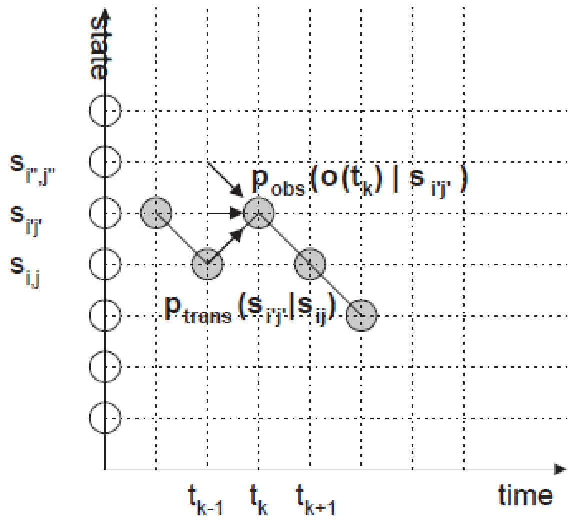
- Problems of P-sola algorithm and motivations for a probabilistic framework
  - First stage:
    - Binary decision
      - … a time is a local maximum of $f(t)$ or not.
      - … only one local maximum per period T_b is estimated.
    - Consequences
      - … If the estimated local maximum is not the one corresponding to the beat positions, the marking will be incorrect
      - … If there is no local maximum in the signal (for example a part of a track without any onset such as a beat in the middle of a silence part), the algorithm also fails
    - Solution: have several candidates for the local maxima and associated probabilities
  - Second stage:
    - no adaptive weighting β between the constraints
      - … (a) "close-to-local maxima" and (b) "inter-distance close to local period"
    - Constant weight β over time
    - Ideally: if a part of a track has no clear onsets, the periodicity constraint should be favored
  - Solutions:
    - probabilities associated to the times and to the transitions between times
  - Formulation
    - HMM ?

Introduction

- Content:
  - Two new approaches for beat-tracking
    - given tempo/ meter as input
  - Algorithm 1: based on P-sola/ GCI location algorithm
  - Algorithm 2: Probabilistic framework
    - Formulation: inverse Viterbi
    - Prior probability
    - Observation probability
      - use of beat-template: LDA training of the best beat-templates
    - Transition probability
    - Decoding
  - Beat-templates evaluation
  - Large-scale evaluation on four test-sets

Beat-tracking algorithm 2: Viterbi/ LDA

- HMM/ Viterbi decoding
  - Hidden states s_i
  - Probabilities:
    - Pinit(s_i)
    - Pobs($o(t)|s_i$)
    - Ptrans($s_i|s_{i+1}$)
  - Viterbi:
    - find the best succession of hidden states over time



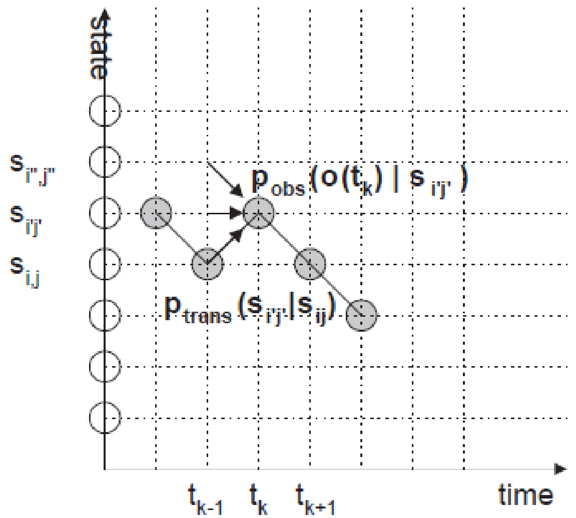
- HMM/ Viterbi for beat-tracking V1
  - Hidden states s_i ?
    - Beat/ non-beat status of a time
  - Viterbi:
    - Decode beat/ non-beat status over time
  - Problem:
    - We want to use the transition probability for representing tempo constraint i.e. the distance between two successive beats is the tempo periood
    - Not possible to do it with a first order HMM (need to go back to the first previous beat, N-previous states)
  - Solution:
    - define state s_i as "time ti is a beat"
  - New formulation
    - Possible to use the transition probability between successive states, because only beats are states

Beat-tracking algorithm 2: Viterbi/ LDA

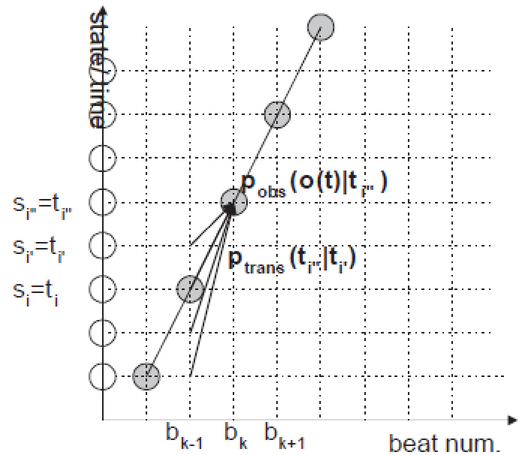
Quaero

ircam Centre Pompidou

- HMM/ Viterbi decoding
  - Hidden states s_i
  - Probabilities:
    - Pinit(s_i)
    - Pobs($o(t)|s_i$)
    - Ptrans($s_i|s_{i+1}$)
  - Viterbi:
    - find the best succession of hidden states over time

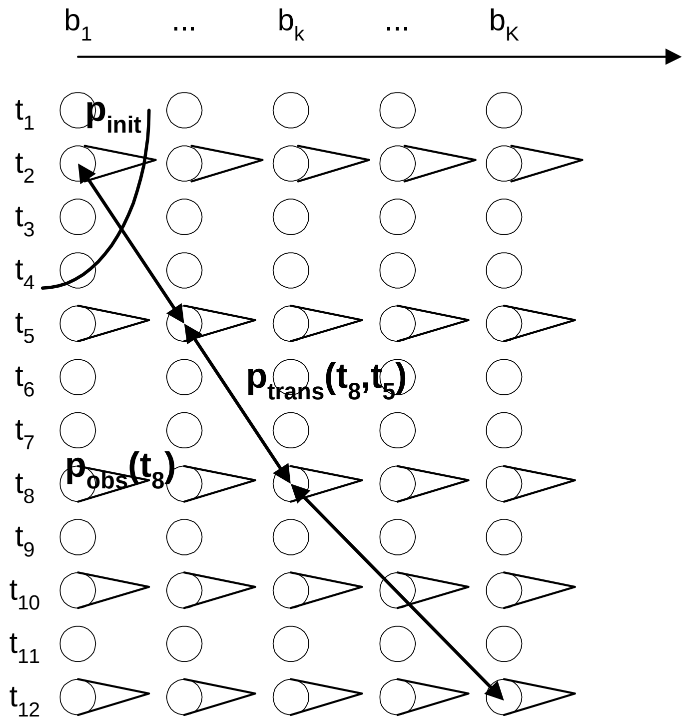


- HMM/ Viterbi for beat-tracking V2
  - Hidden states ?
    - State s_i is defined as « time t_i is a beat »
  - Viterbi ? Problem !
    - if states are beats then we need to decode beats over time. But beats are times !
  - Solution:
    - decode beats (times) over beat-numbers b_k (b_k is a monolithically increasing function)
  - Inverse Viterbi:
    - Look for the succession of s_i (t_i) that best explains the beat number succession b_k



Beat-tracking algorithm 2: Viterbi/ LDA

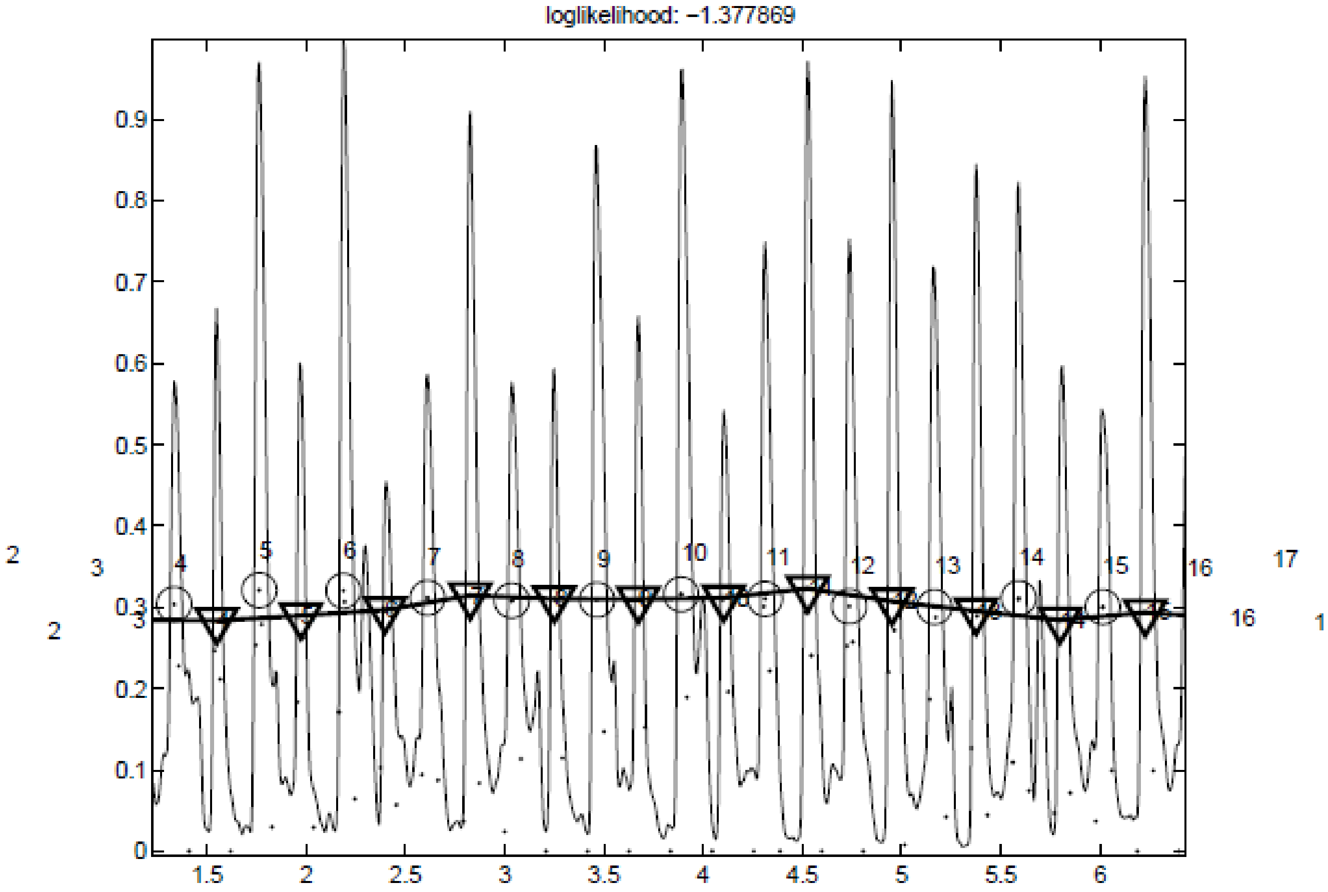
- Probabilistic formulation
  - Hidden State s_i= « a specific time t_i is a beat »
    - Initial probability $p_{init}(s_i)$
      - … Represents the probability to be in hidden state s_i ("t_i is a beat") at the beginning of the decoding
      - … Favors t_i to be a time close to the beginning of the track
    - Emission probability $p_{obs}(\underline{o}(t)|s_i)$
      - … Represents the probability to observe $o(t)$ given a specific state s_i (given that "t_i is a beat")
      - … Correlation with a Beat-templates at tempo Tb(t)
      - … Optimization
    - Transition probability $p_{trans}(s_i'|s_i)$
      - … Represents the probability to transit from state s_i (or "t_i is beat") to state s_i' (or "t_i' is the next beat")
      - … Left-Right HMM



  - Inverse Viterbi decoding
    - provides the best succession of states hence the best succession of "time t_i is a beat"
    - Forward/ Modified Backward algorithm
    - Optimization

Beat-tracking algorithm 2: Viterbi/ LDA

- Examples of decoding:



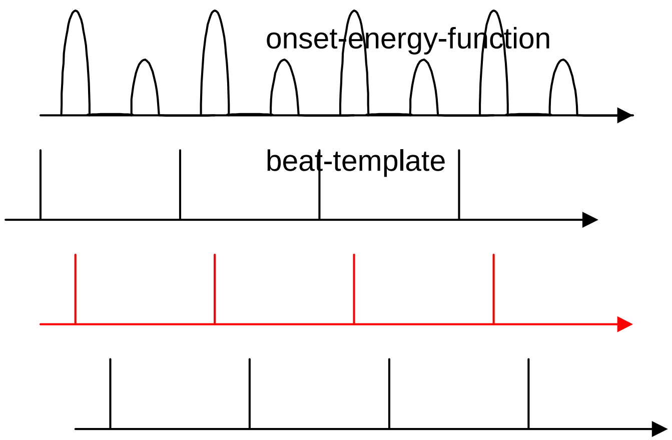
Introduction

- **Content:**
  - Two new approaches for beat-tracking
    - given tempo/ meter as input
  - Algorithm 1: based on P-sola/ GCI location algorithm
  - Algorithm 2: Probabilistic framework
    - Formulation: inverse Viterbi
    - Prior probability
    - Observation probability
      - use of beat-template: LDA training of the best beat-templates
    - Transition probability
    - Decoding
  - Beat-templates evaluation
  - Large-scale evaluation on four test-sets

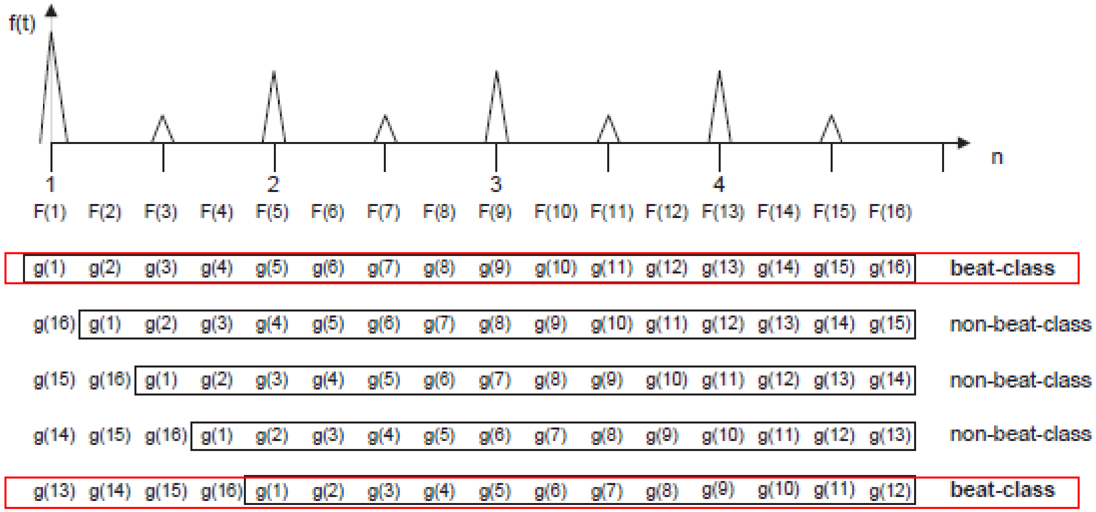
Beat-tracking algorithm 2: Viterbi/ LDA

Learning the best beat-template by LDA

- Learning the best beat-template by LDA
  - Beat-template must be chosen such as
    - A) have the maximum correlation with the local signal when t_i is a beat-position
    - B) provide the largest discrimination between the correlation values when t_i is a beat-position and a non-beat position



Beat-tracking algorithm 2: Viterbi/ LDA
Learning the best beat-template by LDA



- Notations:
  - F(n)
    - function obtained by sampling the local values of f(t, t in [ti, ti + 4Tb]) by N value
  - g(1)...g(N)
    - the discrete sequence of values of the beat-template representing a one-bar duration beat-pattern.
    - g(1)= downbeat position, g(1 + jN/4), j in [1, 2, 3] other beat positions.

Beat-tracking algorithm 2: Viterbi/ LDA
Learning the best beat-template by LDA

- We look for the beat-template (the values of g(n), n in [1,N])
  - which maximizes the correlation with F(n) when t_i is a beat-position
  - which minimizes it when t_i is not a beat-position:

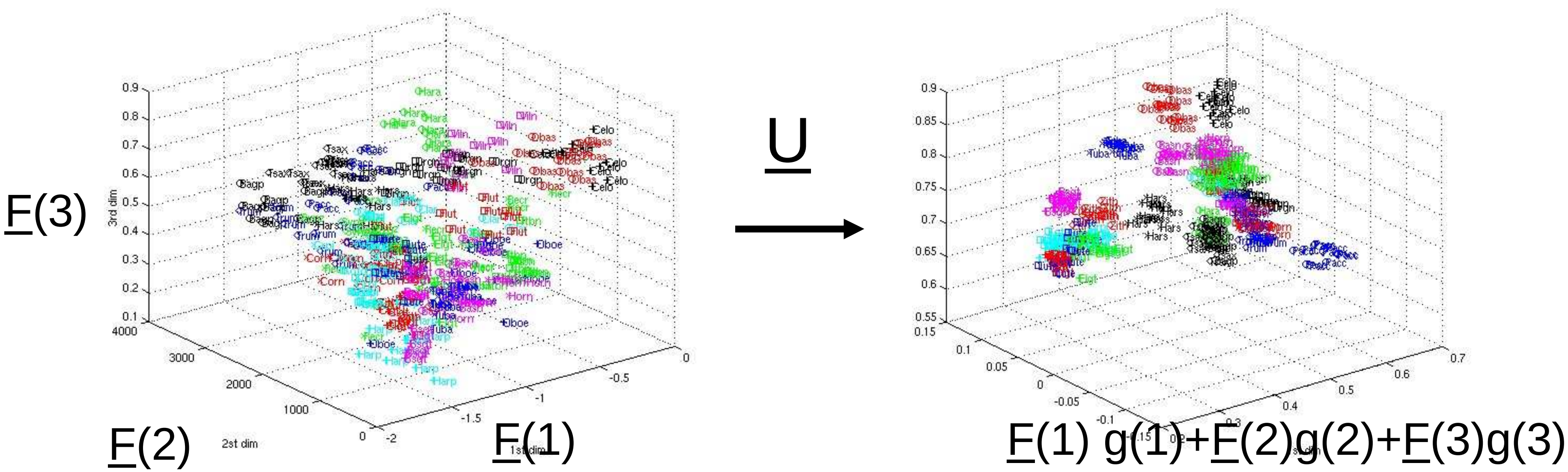
- F(1+j)g(1)+F(2+j)g(2)+...+F(N +j)g(N)
  - maximum value for j in [0,N/4, 2N/4, 3N/4]
  - minimum value for all the other values of j

  - Looks close to the problem of finding the best weights g(n) to apply to the dimensions F(n) of multi-dimensional observations in order to maximize class separation

  - Can be solved using Linear Discriminant Analysis (LDA)
    - The weights are the *g*(*n*),
    - The dimensions of the feature vectors = successive values of *F*(*n*)
    - The classes = "beat" and "non-beat"

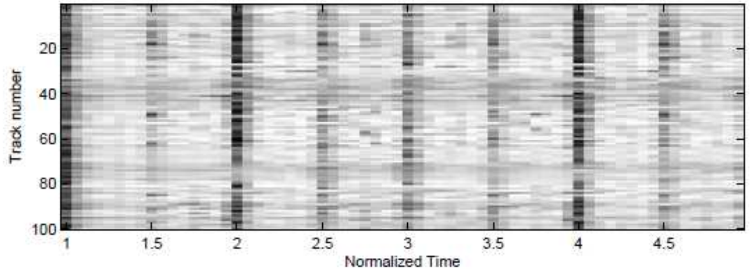
Beat-tracking algorithm 2: Viterbi/ LDA
Learning the best beat-template by LDA

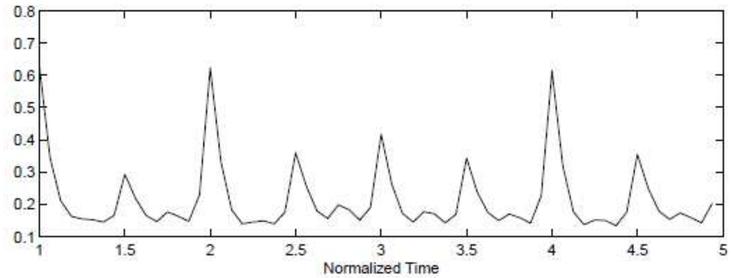
- Linear Discriminant Analysis
  - Compute the matrix U
    - such that after transformation of the features by this matrix, the ratio of the Between-Class-Inertia and the Total-Inertia is maximized
    - Solution= eigen vectors of $\underline{\underline{T}}^{-1}\underline{\underline{B}}\,\underline{u} = \lambda \underline{u}$



Beat-tracking algorithm 2: Viterbi/ LDA
Learning the best beat-template by LDA

- LDA observations to learn from:
  Two-classes "beat" and "non-beat"
  - Using a test-set annotated into beat and down-beat positions
    - Using annotations for beat
    - and circular permutation of the annotation for the beat
    - See paper for details

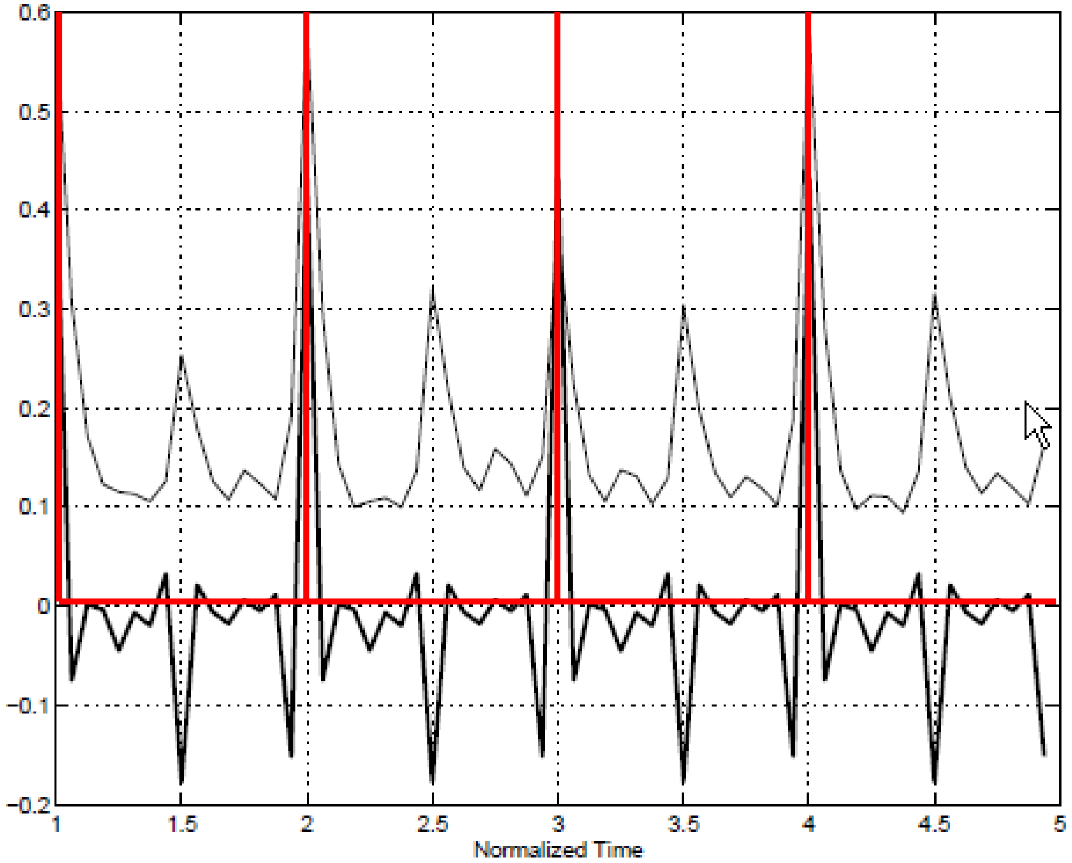




- LDA Solution:
  For a two classes problem
  - only one column remains in U
  - = the weights to apply to F(n) in order to maximize class separation
  - = beat-template g(n).

Beat-tracking algorithm 2: Viterbi/ LDA
Learning the best beat-template by LDA

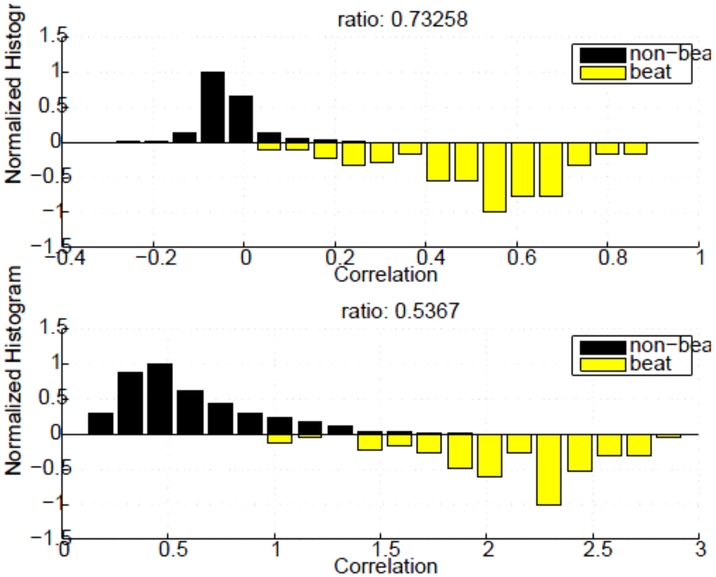
- Examples:
  - on RWC Popular-Music test-set



- Content:
  - Two new approaches for beat-tracking
    - given tempo/ meter as input
  - Algorithm 1: based on P-sola/ GCI location algorithm
  - Algorithm 2: Probabilistic framework
    - Formulation: inverse Viterbi
    - Prior probability
    - Observation probability
      - use of beat-template: LDA training of the best beat-templates
    - Transition probability
    - Decoding
  - Beat-templates evaluation
  - Large-scale evaluation on four test-sets

Evaluation

- Beat-templates comparison
  - What ? validate the assumption that
    - LDA-trained beat-templates provide a better discrimination between the “beat” and “non-beat” classes than usual beat-templates
  - How ?
    - compute the values of the correlation between $f(t)$ and $g(t)$ when using the LDA-trained or the usual beat-templates for $g(t)$
    - compute the ratio r of the Between-Class-Inertia to the Total-Inertia
    - the larger this ratio is, the best the separation is between the two classes beat and non-beat



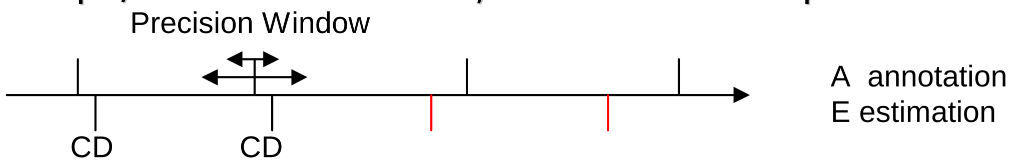
		Test-set			
		PopRock	RWC-Popular	RWC-Jazz	RWC-Classical
Training	PopRock	**0.71**	0.71	0.53	0.37
	RWC-Popular	0.69	**0.73**	0.53	0.38
	RWC-Jazz	0.64	0.66	**0.61**	0.45
	RWC-Classical	0.60	0.64	0.56	**0.49**
	Normal template	0.48	0.54	0.35	0.23

Introduction

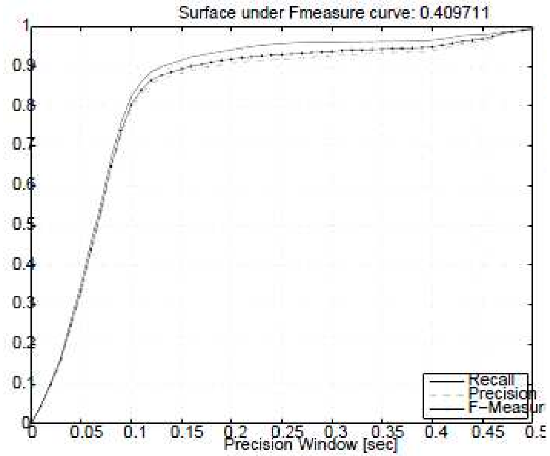
- Content:
  - Two new approaches for beat-tracking
    - given tempo/ meter as input
  - Algorithm 1: based on P-sola/ GCI location algorithm
  - Algorithm 2: Probabilistic framework
    - Formulation: inverse Viterbi
    - Prior probability
    - Observation probability
      - … use of beat-template: LDA training of the best beat-templates
    - Transition probability
    - Decoding
  - Beat-templates evaluation
  - <u>Large-scale evaluation on four test-sets</u>

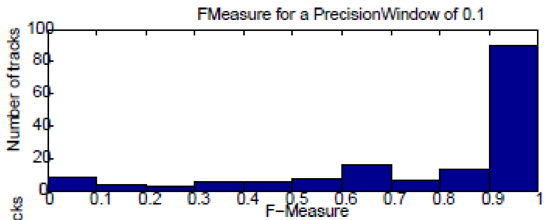
Evaluation

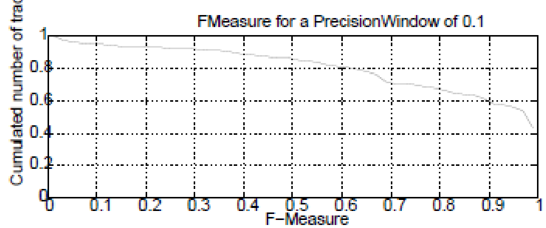
- How to compare ?
  - Measure the performances of the beat-tracking system BUT using [Peeters2007] for tempo/ meter estimation; → measure the performances of the whole system



  - Precision Window (PW)
    - expressed as a percentage of the smallest period in a track
  - Recall(PW) =CD(PW)/A
  - Precision(PW) =CD(PW)/D
  - FMeasure(PW) =2 R(PW)P(PW) / (R(PW)+P(PW))
  - Area Under Curve (AUC) F-measure/ PW
  - Histogram of the values of the F-measure(PW=0.1) for all the track of a given test-set
    - indicates the percentage of tracks having a specific Fmeasure(PW=0.1).
  - Cumulated-histogram
    - indicates the percentage of tracks having "at least" a specific Fmeasure(PW=0.1)
  - Percentage of tracks with F-measure(PW=0.1) >50%
  - Area Under Curve (AUC) of the cumulated-histogram







Evaluation

- Results and discussion
  - Variations among test-set:
    - performances are best
      - for the PopRock extract (FMeas=0.93) and RWC-Popular-Music (FMeas=0.85) test-sets
      - than for the more complex Jazz rhythms (FMeas=0.59) or the time-variable tempi of Classical music (FMeas=0.43).
  - P-sola against Viterbi:
    - Considering all criteria (all the columns of the table) and all test-sets, the Viterbi method leads systematically to better results than the P-sola one
    - Statistical T-Student test
      - H0 hypothesis
        the average Fmeasure(PW=0.1) are equal for the P-sola and Viterbi
      - H1 hypothesis
        they are different
      - Results:
        For the test-sets RWC Popular-Music and RWC Popular-Jazz we can reject the null hypothesis at a 5% significance level

	Method	Tau	Sigma	Beat-template	Recall PW=0.1	Precision PW=0.1	F-Meas. PW=0.1	AUC FMeas/PW	%Track F-Meas(PW=0.1)>0	AUC Percent/ Cumul FMeas
Poprock	P-sola				,91	,88	,89	,44	,92	,88
	Viterbi	32	0.05	LDA-shared	,93	,90	,91	,44	,96	,90
	Viterbi	8	0.05	LDA-shared	,94	,91	,92	,45	,96	,91
	Viterbi	8	0.02	LDA-shared	,94	,91	,91	,45	,96	,90
	Viterbi	8	0.05	LDA-sam	**,95**	**,92**	**,93**	**,45**	**,96**	**,91**
	Viterbi	8	0.05	LDA-all	,94	,91	,92	,45	,96	,91
	Viterbi	8	0.05	Usual	,94	,90	,91	,44	,95	,90
Popular	P-sola				,78	,73	,75	,38	,81	,74
	Viterbi	32	0.05	LDA-shared	,87	,83	,84	,42	,90	,83
	Viterbi	8	0.05	LDA-shared	,88	,83	,85	,42	,91	,84
	Viterbi	8	0.02	LDA-shared	,88	,84	,85	,42	,91	,84
	Viterbi	8	0.05	LDA-sam	**,88**	**,83**	**,85**	**,42**	**,91**	**,84**
	Viterbi	8	0.05	LDA-all	,88	,83	,84	,42	,89	,84
	Viterbi	8	0.05	Usual	,88	,84	,85	,42	,90	,85
Jazz	P-sola				,51	,42	,45	,30	,36	,33
	Viterbi	32	0.05	LDA-shared	,64	,53	,57	,33	,60	,47
	Viterbi	8	0.05	LDA-shared	,64	,53	,57	,33	,56	,48
	Viterbi	8	0.02	LDA-shared	,65	,54	,58	,33	,60	,50
	Viterbi	8	0.05	LDA-sam	,63	,52	,56	,33	,60	,47
	Viterbi	8	0.05	LDA-all	,64	,53	,57	,33	,62	,49
	Viterbi	8	0.05	Usual	**,66**	**,55**	**,59**	**,34**	**,68**	**,53**
Classical	P-sola				,48	,35	,38	,25	,25	,33
	Viterbi	32	0.05	LDA-shared	,52	,36	,41	,26	,42	,36
	Viterbi	8	0.05	LDA-shared	,53	,37	,42	,27	,42	,38
	Viterbi	8	0.02	LDA-shared	,51	,37	,41	,27	,36	,31
	Viterbi	8	0.05	LDA-sam	,52	,38	,41	,27	,42	,39
	Viterbi	8	0.05	LDA-all	,52	,36	,40	,26	,41	,38
	Viterbi	8	0.05	Usual	**,54**	**,38**	**,43**	**,27**	**,42**	**,41**

Evaluation

- Results and discussion
  - Best parameters for the Viterbi decoding algorithms
    - On average (over the test-sets)
    - Slight improvement obtained whith
      - … *T* = 8 : large horizon for reassignment
      - … *s* = 0.05: allows more marker discontinuities
    - All beat-template methods give very close results
      - … except for the Jazz-Music and Classical-Music where, surprisingly, the usual beat-template performs slightly better

	Method	Tau	Sigma	Beat-template	Recall PW=0.1	Precision PW=0.1	F-Meas. PW=0.1	AUC FMeas/PW	%Track F-Meas(PW=0.1)>0	AUC Percent/ Cumul FMeas
Poprock	P-sola				,91	,88	,89	,44	,92	,88
	Viterbi	32	0.05	LDA-shared	,93	,90	,91	,44	,96	,90
	Viterbi	8	0.05	LDA-shared	,94	,91	,92	,45	,96	,91
	Viterbi	8	0.02	LDA-shared	,94	,91	,91	,45	,96	,90
	Viterbi	8	0.05	LDA-sam	**,95**	**,92**	**,93**	**,45**	**,96**	**,91**
	Viterbi	8	0.05	LDA-all	,94	,91	,92	,45	,96	,91
	Viterbi	8	0.05	Usual	,94	,90	,91	,44	,95	,90
Popular	P-sola				,78	,73	,75	,38	,81	,74
	Viterbi	32	0.05	LDA-shared	,87	,83	,84	,42	,90	,83
	Viterbi	8	0.05	LDA-shared	,88	,83	,85	,42	,91	,84
	Viterbi	8	0.02	LDA-shared	,88	,84	,85	,42	,91	,84
	Viterbi	8	0.05	LDA-sam	**,88**	**,83**	**,85**	**,42**	**,91**	**,84**
	Viterbi	8	0.05	LDA-all	,88	,83	,84	,42	,89	,84
	Viterbi	8	0.05	Usual	,88	,84	,85	,42	,90	,85
Jazz	P-sola				,51	,42	,45	,30	,36	,33
	Viterbi	32	0.05	LDA-shared	,64	,53	,57	,33	,60	,47
	Viterbi	8	0.05	LDA-shared	,64	,53	,57	,33	,56	,48
	Viterbi	8	0.02	LDA-shared	,65	,54	,58	,33	,60	,50
	Viterbi	8	0.05	LDA-sam	,63	,52	,56	,33	,60	,47
	Viterbi	8	0.05	LDA-all	,64	,53	,57	,33	,62	,49
	Viterbi	8	0.05	Usual	**,66**	**,55**	**,59**	**,34**	**,68**	**,53**
Classical	P-sola				,48	,35	,38	,25	,25	,33
	Viterbi	32	0.05	LDA-shared	,52	,36	,41	,26	,42	,36
	Viterbi	8	0.05	LDA-shared	,53	,37	,42	,27	,42	,38
	Viterbi	8	0.02	LDA-shared	,51	,37	,41	,27	,36	,31
	Viterbi	8	0.05	LDA-sam	,52	,38	,41	,27	,42	,39
	Viterbi	8	0.05	LDA-all	,52	,36	,40	,26	,41	,38
	Viterbi	8	0.05	Usual	**,54**	**,38**	**,43**	**,27**	**,42**	**,41**

Conclusion and future works

- Results and discussion:
  - Use of the proposed Viterbi method allows to improve the beat-tracking estimation for all test-sets.
    - Considering the difficulty of beat-tracking for Jazz and Classical music, this result is particularly important.
    - Recall and Precision values obtained for the Jazz (R=0.66 and P=0.55) and Classical (R=0.54 and P=0.38) test-sets
      - … Large part of the errors are insertions errors (estimation of twice the correct tempo)
      - … Results could be better if using better tempo estimation as input
  - Use of LDA-trained beat-templates over usual beat-templates
    - Better when comparing discrimination power
    - For beat tracking
      - … Slightly improve the results for the PopRock extract test-set.
      - … Not the case for the Jazz and Classical test-sets.
    - Why ?
      - … LDA-trained beat-templates assumes tracks with a specific constant rhythm pattern.
        - This is the case for pop-rock music
        - This is not the case for Jazz and Classical music.
  - Viterbi method
    - Very promising: can be easily extended, add new types of observation probabilities