

Template-Based Estimation of Tempo: Using Unsupervised or Supervised Learning to create better spectral templates

Geoffroy Peeters

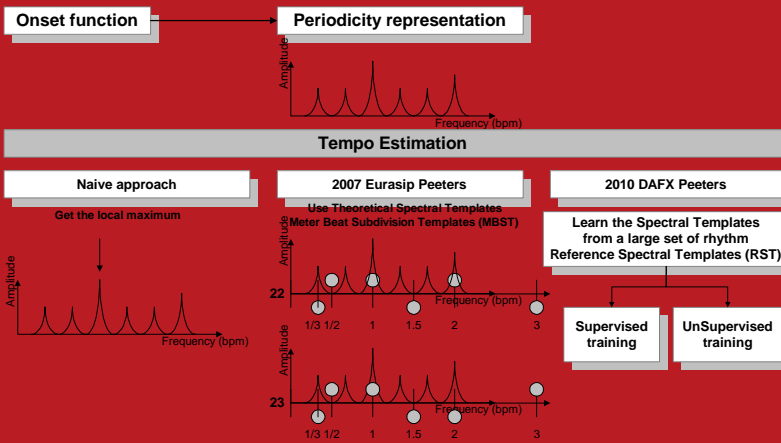
IRCAM - CNRS STMS

Introduction

Goal ?

Improve Tempo Estimation

Tempo Detection ?



Spectral Templates Representation [2]

Onset energy function: $o(n)$
STFT of $o(n)$: $Y_o(f_k, t_m)$
Tempo over time: $b(t_m)$

Sample and Tempo-normalized DFT = Spectral Template

- sampling $Y_o(f_k, t_m)$ at the frequencies $f_k = b(t_m) \cdot f_i$
- where $f_i = [1/4, 1/3, 1/2, 2/3, 3/4, 1, \dots, 8]$
- Result: $Y_o(l, t_m, b(t_m))$

Creating Reference Spectral Templates $m_j = Y_j(l)$

Unsupervised learning

- Fuzzy K-means algorithm applied to a training-set annotated into tempo b
- Clustering applied on vectors $[Y_o(l, b), b]$

Supervised Learning

- Mean over items of a specific rhythm class

Result: $m_j = Y_j(l)$ + Reference Tempo B_j

Hidden Markov Model (Viterbi Decoding)

State s_{ij} definition

- [tempo b_i , MBST m_j]
- [tempo b_i , Reference Spectral Template m_j]

Prior probability s_{ij}

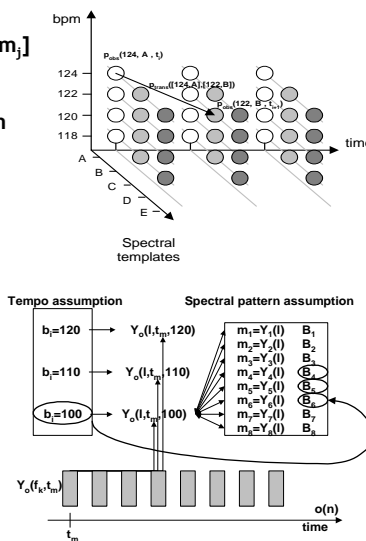
- Choice 1: favoring tempo around 120bpm
- Choice 2: uniform probability
- No preference on MBST or RST

Transition probability

- favors tempo continuity over time
- disadvantage change of MBST/RST

Observation probability

- Probability to observe $Y_o(f_k, t_m)$ given tempo b_i and RST m_j
- 1) Compute the $Y_o(l, t_m, b_i)$ corresponding to the assumption b_i
- 2) Select the m_j which have a reference tempo B_j close to b_i : $(\text{abs}(\log_2(b_i/B_j))) < 0.3785$
- 3) Compute the one-minus cosine distance between $Y_o(l, t_m, b)$ and $m_j = Y_j(l)$



Results

Test-Set: « Ballroom dancer »

Measures: Accuracy1/ Accuracy 2

Comparisons of:

- a) MBST b) ST Unsupervised c) ST Supervised
a) Prior centered on 120bpm b) No prior

| Tempo estimation | | | Acc1 | Acc2 |
|------------------------|-----------------|-----|--------------|--------------|
| MBST 22/23/32 | Prior tempo 120 | DFT | 65,0% | 89,4% |
| | No Prior | DFT | 44,0% | 87,1% |
| ST Unsupervised | Prior tempo 120 | DFT | 63,9% | 90,7% |
| | No Prior | DFT | 72,9% | 93,4% |
| ST Supervised | Prior tempo 120 | DFT | 62,5% | 89,1% |
| | No Prior | DFT | 75,2% | 94,8% |

Conclusion

Conclusion:

- Best results: **ST-Unsupervised** and **ST-Supervised** without prior tempo probability
- Improve upon previously published results (ismir04)
- Performances obtained with **ST-Unsupervised** = very promising
-> Does not necessitates a database annotaied into rhythm class
-> Can easily be used to other test-sets
- Future works: extending the set of m_j

Warning: Results indicated here are only indicative !
Training and test-set are the same in this preliminary experiment