Assessment of state-of-the-art meter analysis systems with an extended meter description model

Matthias Varewyck, Jean-Pierre Martens
{Matthias.Varewyck,Jean-Pierre.Martens} @elis.UGent.be
Department of Electronics and Information Systems
Ghent University (Belgium)

AIM: determine and compare the abilities of 4 state-of-the-art automatic meter analysis algorithms of musical audio

METER DESCRIPTION MODEL

Basic assumptions

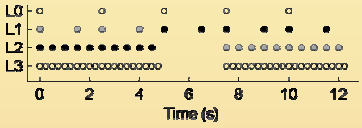
There exists a quasi periodic pattern marked by main temporal accents with mean inter timestamp interval (ITI) in [1.5s 6s] ~ bar transitions
1. Secondary accents characterize perceivable periodic patterns at smaller time scales. They are subject to a set of restrictions.
2. Secondary accents can be represented by timestamps on a uniform grid. Grid changes coincide with bar transitions.

Restrictions on secondary accents

Timestamps on one level L_λ are copied to next lower level $L_{\lambda+1}$
- Extra timestamps on $L_{\lambda+1}$ obtained by dividing each ITI on L_λ into 2 or 3 equal parts (division ratio is fixed during each bar)
- For ONE level L_λ, both divisions can occur ➔ two different ITI's on $L_{\lambda+1}$

Salience per bar & per level

Represents the attentional strength of evoked temporal pattern
- Major salience on level L_λ
- Minor saliences allowed on $L_{\lambda-1}$ and/or $L_{\lambda+1}$



EVALUATION METHODOLOGY

Compare two sequences of timestamps
- hypothesized timestamps H_m (m=1..M)
- annotated timestamps A_n (n=1..N) (different schemes for selecting them from full multilevel annotation)

Use tolerance on allowed time differences of A_n ➔ deletions, insertions

Selection schemes

1. Bar independent level
   - **Bar level** = highest level
   - **Beat level** = level with highest average salience
   - **Best tracked level** = level best supporting hypotheses
2. Bar dependent level
   - **Tatum level** = lowest level
   - **Correct beat level** = level with highest salience

Time tolerance

- Based on human tapping experiment: insert missing fourth event in isochronic sequence of length six (Friberg & Sundberg)
- Generally used tolerances: 10% (bar) (Klapuri) 17.5% (non bar) (Goto)
  ➔ Tolerance = 5 * min(6ms, 0.025 * ITI)

Time lag compensation

Aim: compensate small but consistent time lags between H_m and A_n that originate from the algorithm itself (D_0) and from different characteristics of different onset types (spread around D_0)
Procedure:
1. Compute algorithm dependent time lag D_0
2. Determine song dependent time lag D in [D_0 - 20ms, D_0 + 20ms]
3. Subtract D from H_m (m = 1..M)

EXPERIMENTAL FRAMEWORK

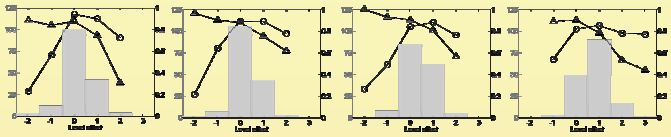
- 30s excerpts from 161different songs, partly used before in MIREX
- Max number of annotated levels/song: min = 2, mean = 4.5, max = 6
- Four meter analysis algorithms evaluated: KLAPURI, DAVIES, DIXON and ELLIS
  ➔ Data available at https://speech.elis.ugent.be/ (see downloads)

EXPERIMENTAL RESULTS

1. Beat tracking abilities

a) P, R, F : Precision, Recall, F-rate with respect to best tracked level
N_{beat} : number of times best tracked level coincides with beat level

Algorithm	D_0 (ms)	P	R	F	N_{beat}
KLAPURI	0	0.89	0.80	0.843	100
DAVIES	-8	0.87	0.83	0.850	106
DIXON	-20	0.83	0.85	0.841	85
ELLIS	20	0.83	0.76	0.793	49



➔ DIXON and KLAPURI show similar P,R & F but different N_{beat}
➔ ELLIS and DIXON tend to prefer level below the beat level

b) P, R, F: Precision, Recall, F-rate with respect to correct beat level (columns 2-4)

	Beat tracker (FE+BE)			FE+STIU		
Algorithm	P	R	F	P	R	F
KLAPURI	0.72	0.75	0.731	0.48	0.63	0.534
DAVIES	0.70	0.79	0.742	0.48	0.68	0.564
DIXON	0.62	0.82	0.709	0.42	0.51	0.463
ELLIS	0.48	0.78	0.595	0.45	0.65	0.533

➔ Same ranking, but much smaller P's
~ tendency of trackers to prefer level below the beat level

2. Role of front-end and back-end

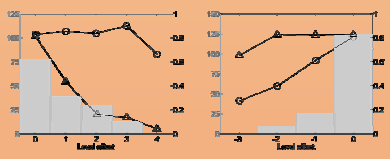
- Replace normal back-end by simple timestamp induction unit (STIU) to extract timestamps from accent functions
  1. Peak-valley search (left-to-right): 2 control parameters
     - ΔT = time difference between successive peaks
     - δ_A = relative drop in amplitude necessary to search for next peak
  2. Peak rejection: TH on peak amplitude
- Select STIU with best [ΔT, δ_A, TH] for each algorithm
- Compare results of FE+BE to results of FE+STIU (table: columns 5-7)

➔ All except ELLIS' back-end very superior to STIU
➔ DIXON's front-end inferior (larger ΔF in column 7 than in column 4)

3. Bar & tatum tracking abilities

Assess KLAPURI outputs with respect to bar and tatum levels

Level	D0 (ms)	P	R	F	M/N
Bar	5.8	0.45	0.45	0.451	0.997
Tatum	0	0.75	0.70	0.725	0.933



➔ Only 45% of the bar onsets detected correctly
➔ Low F & M/N ≈ 1 → bar onsets often out of phase
➔ High P for levels below bar level (Fig) → in phase with lower level

CONCLUSIONS

- Most algorithms prefer to track beat level or level just below the beat
- Bar onset tracking inadequate for applications in which phase of bar-onsets is important (e.g. mixing)
- Taking inter-level dependencies into account in meter analysis (as in KLAPURI) does not lead to higher beat tracking performance