CSC475 Music Information Retrieval

Rhythm Analysis

George Tzanetakis

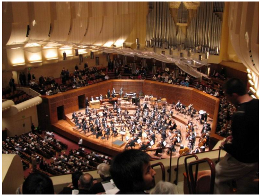
University of Victoria

2014

Table of Contents I

The importance of rhythm and music

- Music making and listening is part of every human culture that we know
- Why do we listen do music ?
  - Is it evolutionary cheesecake ?
  - What would an alien think while observing a Radiohead concert ? the San Fransisco Symphony Orchestra ? Justin Beiber ?
- Music without any underlying rhythmic structure is rare.



Why rhythm ?

- One of our defining characteristics as species is our expanded perception of time into past and future
- Listening to music = doing brain push-ups for your time perception
- A cave painting abstracts a static view of what our senses perceive during a hunt
- A dance with pounding drums abstracts our entire mental state over time during a hunt

Terminology

Rhythm: the hierarchical temporal structure of music
Pulse: semi-regular sequence of accents
Rhythm: hierarchical grouping of pulses with higher metrical levels correspond to longer time divisions
Tempo: the frequency of the semi-regular sequence of pulses that a human would "tap" to.
Beats: the pulses corresponding to the tempo
Tatum: the shortest regular unit of pulse spacing
Durations: quarter notes, eight notes

Tasks

Tempo estimation: "global" tempo

Beat Tracking: time-varying semi-regular sequence of beats

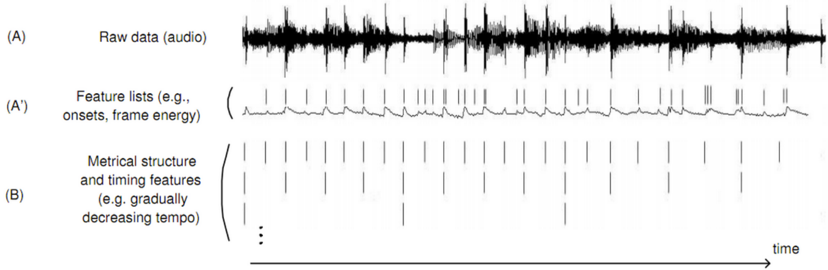
Rhythm description: a feature vector characterizing the rhythm (tempo variant or tempo invariant)

Drum transcription: a drum score i.e time onsets of different drum sounds

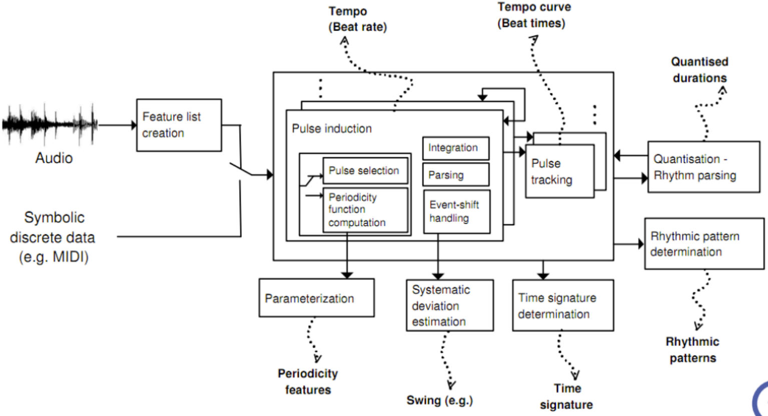
Rhythm pattern detection: basic rhythmic patterns and their locations in time

For each of these tasks there are several variants depending on: causal, real-time (predictive), various constraints and assumptions based on domain of application. Figures mostly from my work but the underlying concepts apply to most rhythm analysis algorithms.

The big picture



Functional Units



Extension of (Gouyon and Dixon, 2005b)

A representation and two simple ideas

Onset Strength Signal

A regular function of the audio signal with higher values around rhythmic accents

Regularity

Rhythmic events tend to occur semi-regularly with minor and locally smooth changes.

Self-similarity

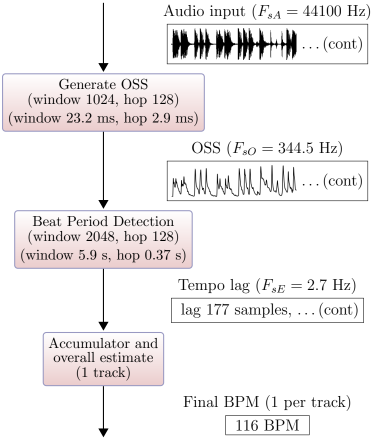
The OSS signal tends to be mostly self-similar at lags corresponding to the basic rhythm periodicities

Most rhythm analysis algorithms use some form of OSS and leverage one or both of these two basic ideas.

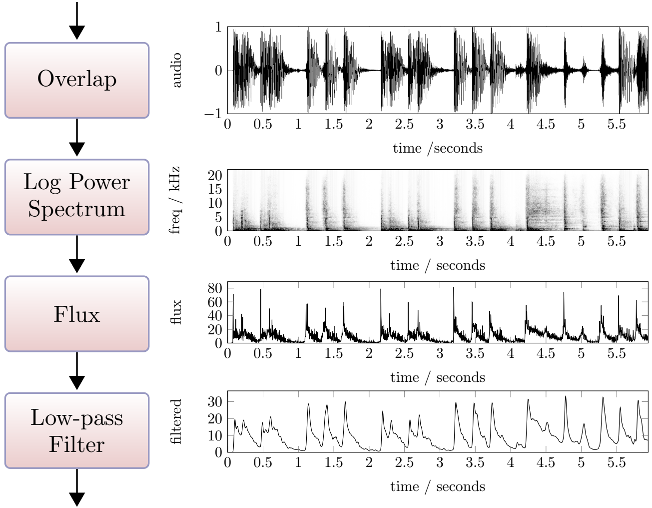
Table of Contents I

Tempo Estimation

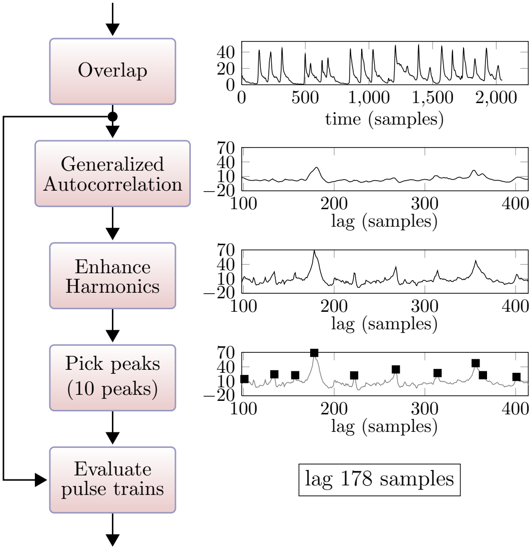
Figures are based on a simple and effective tempo estimation method proposed Tzanetakis and Percival.



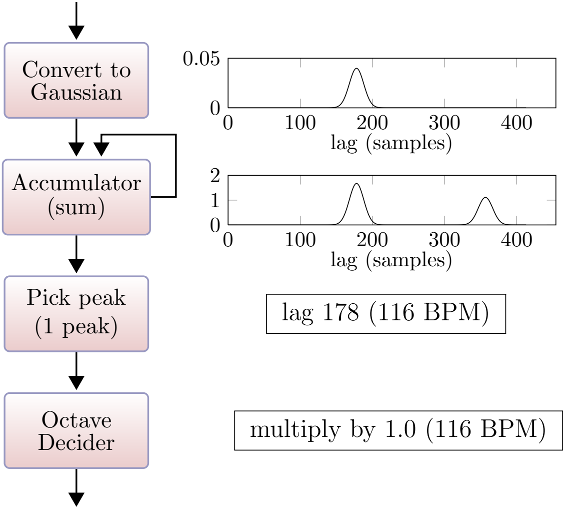
Onset Strength Signal



Detection of periods of interest



Accumulate and decide octave



Octave decision via machine learning

Main idea

Calculate features of the rhythm profile and train octave decision (0.5, 1, 2.0) using a tempo annotated train set.

$$F_1 = \sum_{i=0}^{L-10} C(i), \qquad F_2 = \sum_{i=0.5L-10}^{0.5L+10} C(i), \qquad F_3 = L \qquad (1)$$

TABLE I
CONFUSION MATRIX OF STEM−SVM3 TRAINING.

0.5	1.0	2.0	
74	134	0	0.5
30	1853	305	1.0
0	408	902	2.0

Tempo estimation evaluation

Accuracy 1 is the percent of estimates which are within 4% BPM of the ground-truth tempo, and *Accuracy 2* is the percent of estimates which are within 4% of a multiple of $\frac{1}{3}, \frac{1}{2}, 1, 2,$ or 3 times the ground-truth tempo.

Audio Tempo Estimation		
SubID	**Participants**	**P-Score**
FW1	Fu-Hai Frank Wu	0.78
FW2	Fu-Hai Frank Wu	0.76
GP1	Geoffroy Peeters	0.75
FK2	Florian Krebs	0.75

Tempo estimation results

University of Victoria Computer Science

Accuracy 1 results

	files	stem algorithms svm3	heur	base	echonest	gkiokas	zplane	klapuri	ibt	qm_vamp	scheirer
ACM MIRUM*	1410	**73.3**	70.6	55.6–	72.1	72.7	70.1	68.9	63.0–	63.9–	56.5–
BALLROOM*	698	65.6	66.3	40.8–	**89.8**+	63.2	66.9	64.9	64.3	66.9	58.2–
GTZAN GENRES*	999	**78.3**	74.6	63.4–	72.5–	71.7–	68.9–	70.5–	61.0–	58.8–	57.7–
HAINSWORTH	222	69.8	71.6	41.0–	72.1	64.4	69.8	71.6	**72.5**	68.0	53.2–
ISMIR04_SONGS	465	61.1	58.7	57.0	**63.2**	57.0	56.3	58.1	46.7–	43.0–	46.0–
SMC_MIRUM	217	27.6	19.4	19.4	18.9	**35.0**	18.4	18.0	17.5	12.4–	13.8–
Dataset average		62.6	60.2	46.2	64.8	60.7	58.4	58.6	54.2	52.2	47.6
Total average	4011	69.1	66.7	52.4–	71.4	66.5	64.8–	64.7–	58.9–	58.2–	53.4–

Accuracy 2 results

	files	stem algorithms svm3	heur	base	echonest	gkiokas	zplane	klapuri	ibt	qm_vamp	scheirer
ACM MIRUM*	1410	97.1	97.3	96.8	94.9	**98.0**	93.8	96.9	93.2	93.0	82.3–
BALLROOM*	698	95.0	92.8	93.1	96.3	**98.0**	94.8	92.8	90.3	90.8	81.2–
GTZAN GENRES*	999	94.7	**95.1**	95.0	91.6	93.9	89.1	92.5	87.0–	87.7–	77.6–
HAINSWORTH	222	**86.9**	**86.9**	**86.9**	84.2	84.7	82.4	84.2	82.0	77.5	69.4–
ISMIR04_SONGS	465	86.7	87.3	88.0	86.0	**91.0**	82.6	89.5	76.6–	79.8	72.0–
SMC_MIRUM	217	45.6	46.5	47.0	34.1	**51.6**	31.8–	41.9	36.9	30.9–	25.8–
Dataset average		84.3	84.3	84.5	81.2	86.2	79.1	83.0	77.6	76.6	68.0
Total average	4011	91.6	91.5	91.4	89.4	92.9	87.5–	90.6	85.5–	85.6–	76.0–

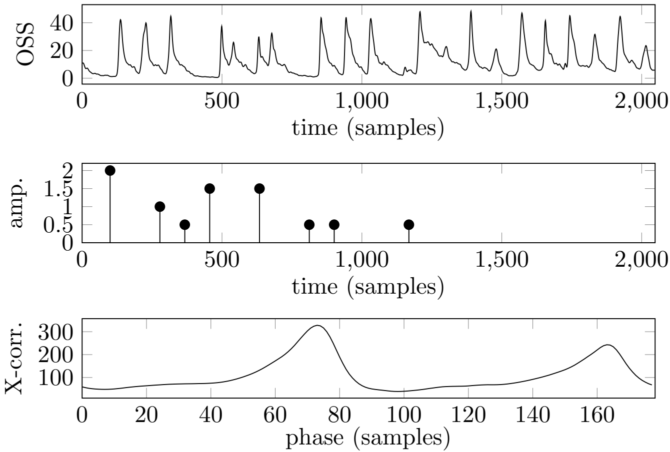
Table of Contents I

Beat Tracking

Given one or more period estimates and either the discrete onset locations or the onset strength signal identity the subset of onsets that correspond to the beat.

- Cross-correlation with pulse sequences
- Adaptive oscillators
- Histogramming of inter-onset intervals
- Multiple agent approaches
- Dynamic programming
- Probabilistic models

Cross-correlation with pulse sequences



Multiple agents

From the INESC Beat Tracker (IBT) Oliveira et al, 2012:

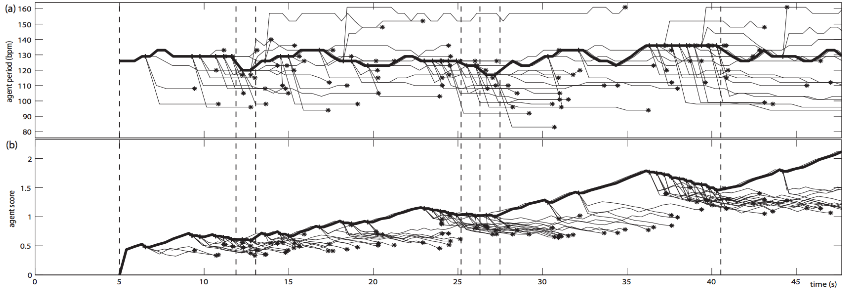


Fig. 3: Evolution of the best agents' proprieties (bold line) and children (branches) while causally tracking the beats of a piece of "Lost in the Flood" by Bruce Springsteen, in the `single` induction mode of operation: (a) Best agents' period; (b) Best agents' score. The starting points of the branches represent moments at which children were created from the current best agent, the asterisks the moments at which the same children were killed, and the vertical dashed lines represent changes of the best agent.

Oscillating filter approaches

Adaptive oscillator is excited by the OSS, and, hopefully, the oscillator resonates at the frequency of the beat. Example oscillator:

$$o(n) = 1 + \tanh \alpha(\cos 2\pi\phi(n) - 1),$$

where $o(n)$ defines an output waveform with pulses at beat locations with width tuned by *alpha* and phase $\phi(n)$.

Alternatively multiple fixed frequency oscillators can be utilized and the phase of the highest scoring one can be used to estimate the beat locations.

Note: Dynamic programming approaches will be discussed later in the context of beat tracking in Afro-Cuban music.

Probilistic Models I

University of Victoria Computer Science

Basic Idea

Model the rhythm process using a generative, explicit underlying probabilistic model controlled by parameters. The algorithm then needs to estimate the parameters of the model using standard esimation procedures such as the Kalman filter, Markov Chain Monte Carlo (MCMC), and particle filtering algorithms.

Probilistic Models II

Representative example from Cemgil, Hainsworth: sequential update of a tempo process characterized by a tempo and phase at each onset location.

$$\theta_n = \mathbf{\Phi}_n(\gamma_n)\theta_{n-1} + v_n,$$
$$s_n = \mathbf{H}_n\theta_n + \epsilon_n$$

s_n is the set of observed onset times, while θ_n is the tempo process at iteration (observed onset) n and can be expanded as:

$$\theta_n = \begin{bmatrix} \rho_n \\ \Delta_n \end{bmatrix}, \mathbf{H}_n = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

ρ_n is the predicted time of the nth obsergvation s_n, and Δ_n is the beat period in seconds.

Beat Tracking Evaluation

All beat tracking measures assume availability of a ground truth sequence of beats and a predicted sequence of beats.

- F-measure (Dixon) with ± 70 milliseconds tolerance
- Gaussian error function (Cemgil):

$$W(x) = e^{-x^2/2\sigma_e^2}$$

- P-Score cross-correlation with sequence of pulses at 10ms resolution
- Goto - statistics of beats and corresponding predicted beats
- Continuity-based evaluation - regions of successive tracked beats

Beat Tracking Results

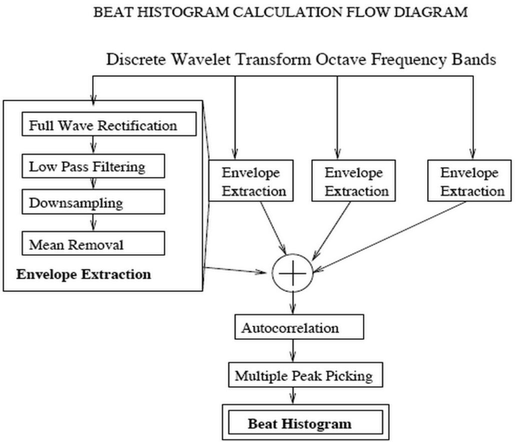
Audio Beat Tracking

SubID	Participants	F-Measure MCK	MAZ	SMC
FK1	Florian Krebs	0.57	0.58	0.40
FW3	Fu-Hai Frank Wu, Jyh-Shing Roger Jang	0.52	0.49	0.34
FW4	Fu-Hai Frank Wu, Jyh-Shing Roger Jang	0.52	0.48	0.34
FW5	Fu-Hai Frank Wu	0.43	0.67	0.35
GKC2	Aggelos Gkiokas, Vassilis Katsouros, George Carayannis	0.50	0.42	0.37
GP2	Geoffroy Peeters	0.50	0.49	0.36
GP3	Geoffroy Peeters	0.50	0.47	0.35
GP4	Geoffroy Peeters	0.50	0.56	0.37
KB1	Florian Krebs, Sebastian Böck	0.54	0.52	0.41
KFRO1	Maksim Khadkevich, Thomas Fillon, Gaël Richard, Maurizio Omologo	0.51	0.35	0.33
ODGR1	João Lobato Oliveira, Matthew Davies, Fabien Gouyon, Luis Paulo Reis	0.51	0.42	0.31

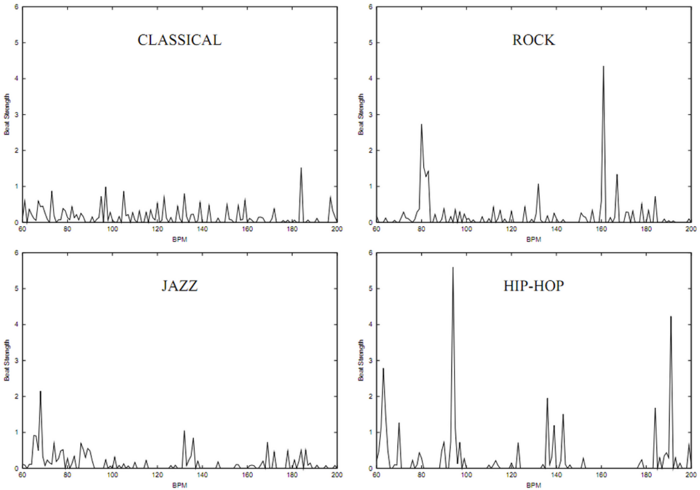
Table of Contents I

Wavelets and Beat Histograms

From Tzanetakis and Cook, 2002:



Beat Histograms



Classification results

	Genres	Classical	Jazz
RND	10	25	16
PHF(5)	23	40	26
BHF(6)	28	39	31
STFT(9)	45	78	58
MFCC(10)	47	61	56
FULL(30)	59	77	61

Table of Contents I

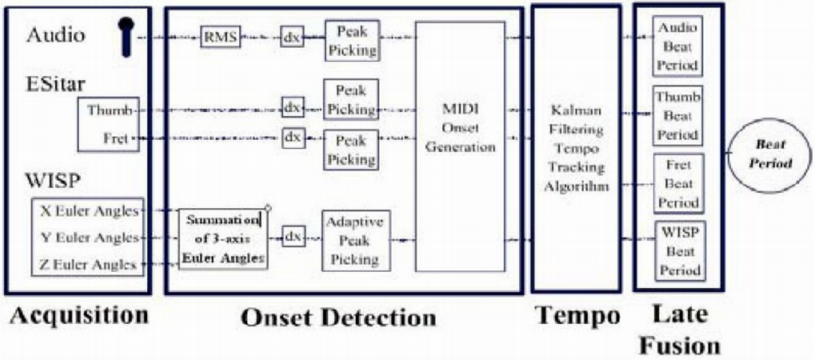
Human perception and computer extraction of beat strength

(Tzanetakis & Cook, Dafx 2002)

- A different rhythm "dimension" than tempo
- 1D sorting of music at the same tempo
- User study showed that listeners can consistently rank songs along the "Beat Strength" dimension
- The overall energy and/or the main peak height of the Beat Histogram correlate nicely ith Beat Strength
- Low values (speech, classical music), medium (pop, country), high (heavy metal, techno)

Table of Contents I

Multimodal analysis of Sitar Performance



Kalman filtering for real-time tracking

$$x_k = \begin{pmatrix} \tau_k \\ \Delta_k \end{pmatrix} = \begin{pmatrix} 1 & \gamma_k \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \tau_{k-1} \\ \Delta_{k-1} \end{pmatrix} + w_k$$

The state consists of current onset time and period and depends on the previous state, a switching variable for different rhythmic units, and noise models.

Mulimodal Beat Tracking Results

University of Victoria Computer Science

Signal	Tempo (BPM)			
	80	100	120	140
Audio	46%	85%	86%	80%
Fret	27%	27%	57%	56%
Thumb	35%	62%	75%	65%
WISP	50%	91%	69%	53%
LATE FUSION:				
Audio/*WISP*/Thumb/Fret	45%	83%	89%	84%
Audio/*WISP*/Thumb	55%	88%	90%	82%
Audio/ *WISP*	58%	88%	89%	72%
Audio/Thumb	57%	88%	90%	80%
WISP/Thumb	47%	95%	78%	69%