Analysis of the Meter of Musical Signals (Klapuri et al)

Douglas Eck
IFT6080 Winter 2008

Main ideas from abstract

***Abstract*— A method is decribed which analyzes the basic pattern of beats in a piece of music, the musical meter. The analysis is performed jointly at three different time scales: at the temporally atomic *tatum* pulse level, at the *tactus* pulse level which corresponds to the tempo of a piece, and at the musical *measure* level. Acoustic signals from arbitrary musical genres are considered. For the initial time-frequency analysis, a new technique is proposed which measures the degree of musical accent as a function of time at four different frequency ranges. This is followed by a bank of comb filter resonators which extracts features for estimating the periods and phases of the three pulses. The features are processed by a probabilistic model which represents primitive musical knowledge and uses the low-level observations to perform joint estimation of the tatum, tactus, and measure pulses. The model takes into account the temporal dependencies between successive estimates and enables both causal and noncausal analysis. The method is validated using a manually annotated database of 474 music signals from various genres. The method works robustly for different types of music and improves over two state-of-the-art reference methods in simulations.**

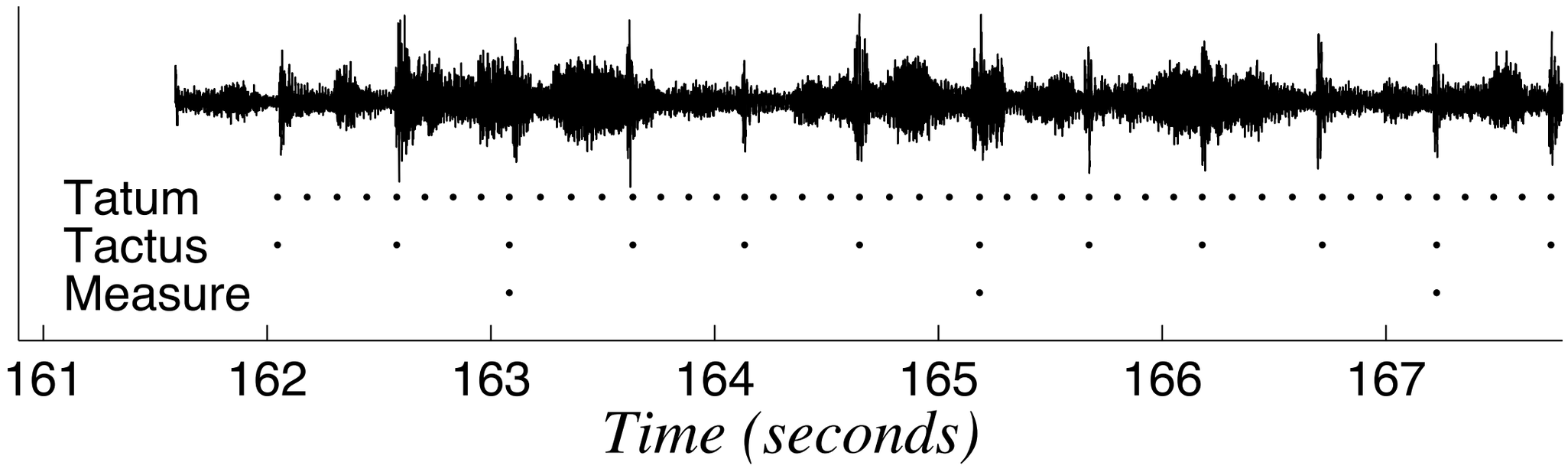
Joint estimation of 3 levels of MH

Comb filterbank

New onset detection

Probabilistic model

Musical levels



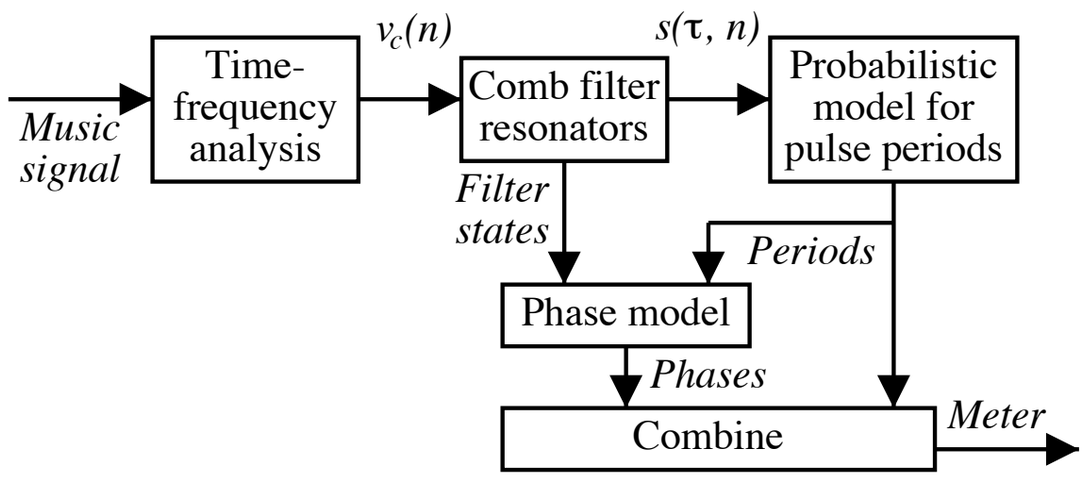
- *Tatum*: shortest durations regularly encountered (from jazz drummer Art Tatum)
- *Tactus*: Most prominent beat level
- *Measure*: A slower changing level with integer harmonic relationship to the tactus

Previous work

This paper has good previous work section. Some papers of interest:

- Large & Kolen: oscillator model from last week
- Parncutt: interesting meter inference model
- Brown: early use of autocorrelation
- Cemgil & Kappen: discussed on Wednesday
- Goto et. al: instrument-specific beat tracking
- Scheirer: discussed today
- Raphael: generative model to be discussed later

Method summary



- Time freq analysis (onset detection)
- Comb filterbank (similar to autocorrealtion)
- Prob. model for finding periods
- Periods drive a secondary phase-extraction module
- Period and phase information drive meter extraction

Onset detector

- Bello et al paper from two weeks ago gives more recent approaches
- Broadband approach: Scheirer did period detection independently for 5 independent bands
- Narrowband approach: Goto uses many narrower bands but sums their differences before doing period detection
- Tradoff between too few subbands (harmonic changes do not show up) and too many subbands (measuring periodicity at each subband is no longer appropriate)

Onset detector ctd.

Parameterized a model such that Goto and Scheirer are special cases

uLaw log transform y(k) of signal is interpolated and LP filtered using a Butterworth filter yielding smooth $z_b(n)$

$$y_b(k) = \frac{\ln(1 + \mu x_b(k))}{\ln(1 + \mu)},$$

Difference z_b'(n) is calculated from half-wave-rectified $z_b(n)$

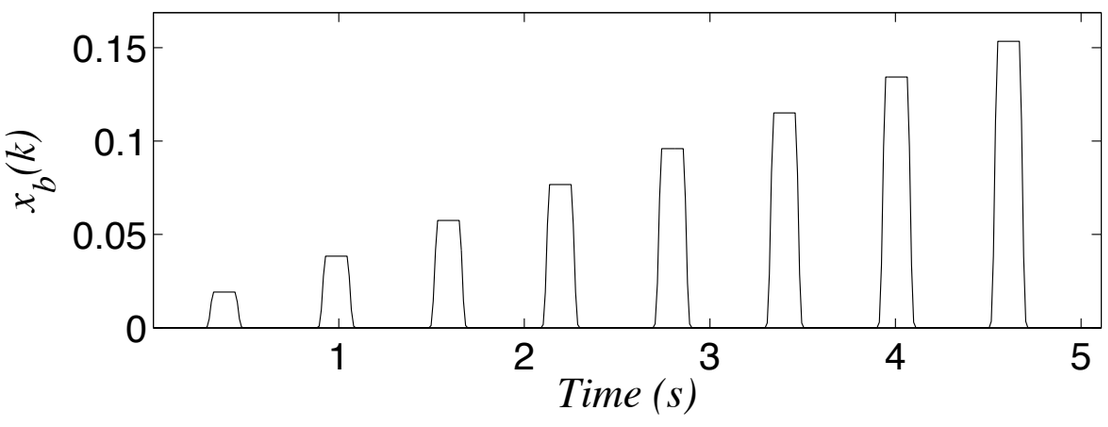
$$z_b{}'(n) = \mathrm{HWR}(z_b(n) - z_b(n-1)),$$

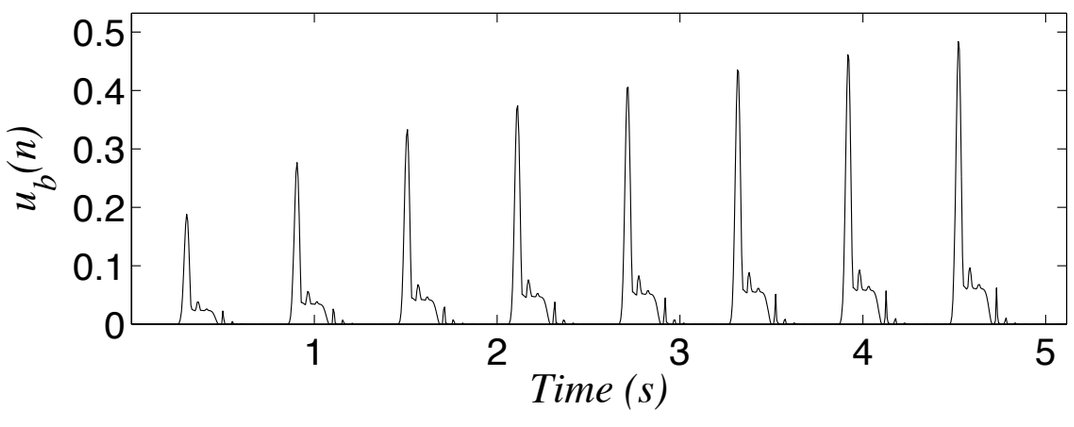
Weighted average $u_b(n)$ is taken from z_b'(n) and $z_b(n)$

$$u_b(n) = (1 - \lambda) z_b(n) + \lambda \frac{f_r}{f_{\mathrm{LP}}} z_b{}'(n),$$

Adjacent $u_b(n)$ values are summed to make $v_c(n)$

$$v_c(n) = \sum_{b=(c-1)m_0+1}^{cm_0} u_b(n), \quad c = 1, \ldots, c_0.$$





Comb filter bank

Output of comb filter with delay γ for $v_c(n)$
[Note similarity to autocorrelation.]

$$r_c(\tau, n) = \alpha_\tau r_c(\tau, n-\tau) + (1-\alpha_\tau) v_c(n),$$

A bank of such resonators was applied, with τ getting values from 1 to $\tau_{\max}$, where $\tau_{\max} = 688$ corresponds to four seconds. The computational complexity of one resonator is $O(1)$ per input sample, and the overall resonator filterbank requires of the order $c_0 f_r \tau_{\max}$ operations per second, which is not too demanding for real-time applications.

Instantaneous energies $\hat{r}_c(\tau, n)$ of each comb filter in channel c at time n are calculated as

$$\hat{r}_c(\tau, n) = \frac{1}{\tau} \sum_{i=n-\tau+1}^{n} r_c(\tau, i)^2. \tag{7}$$

These are then normalized to obtain

$$s_c(\tau, n) = \frac{1}{1-\gamma(\alpha_\tau)} \left(\frac{\hat{r}_c(\tau, n)}{\hat{v}_c(n)} - \gamma(\alpha_\tau) \right), \tag{8}$$

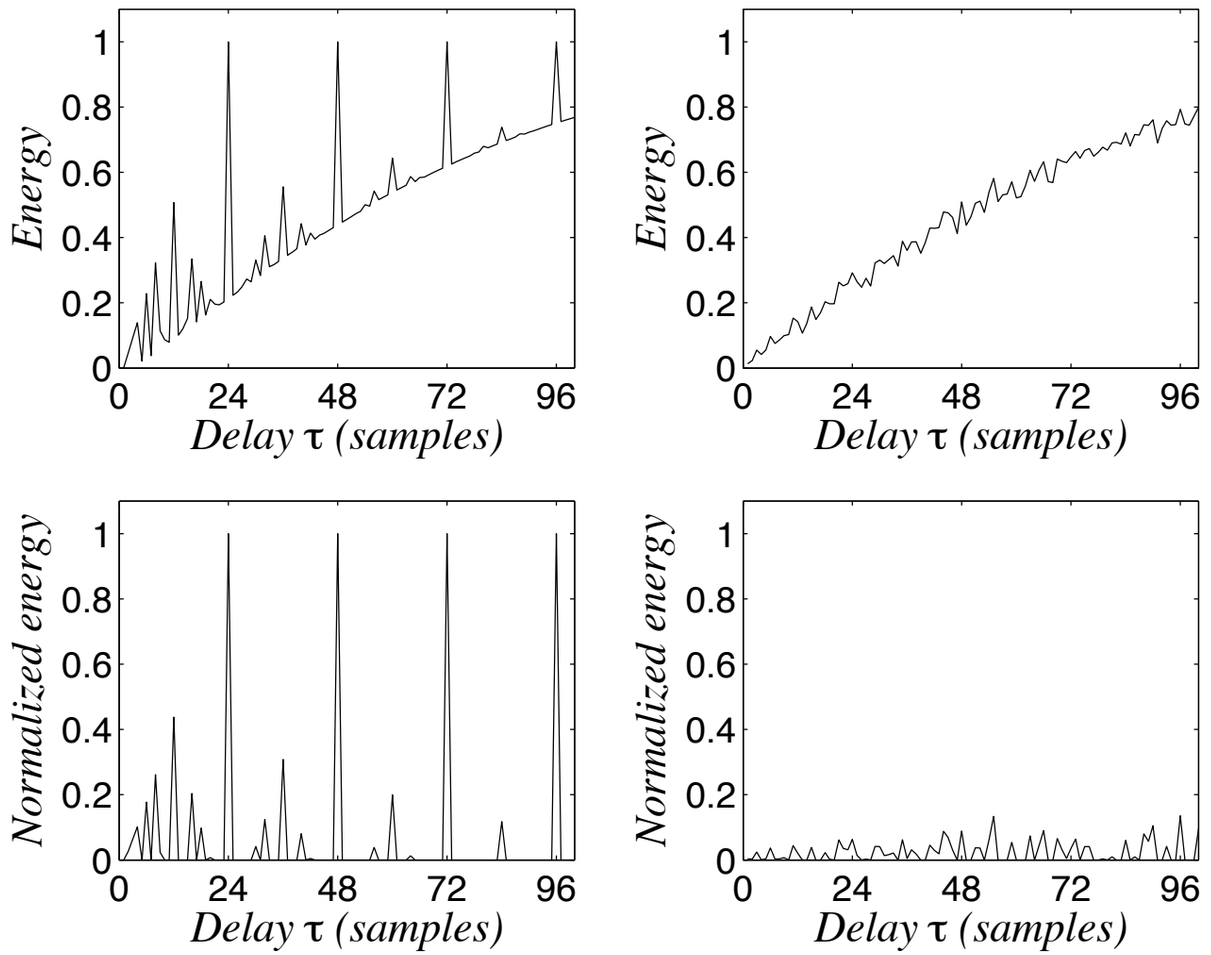


Fig. 4. Resonator energies for an impulse train with a period-length of 24 samples (left) and for white noise (right). Upper panels show the energies $\hat{r}_c(\tau, n)$ and the lower panels normalized energies $s_c(\tau, n)$.

Metrical salience

- Metrical salience function
  (observation for generative model)

$$s(\tau, n) = \sum_{c=1}^{c_0} s_c(\tau, n).$$

- Calculating salience of *tatum* via discrete power spectrum S(*f,n*) which performs a DCT using half-hanning emphasis window

$$S(f, n) = f \left| \frac{1}{\tau_{\max}} \sum_{\tau=1}^{\tau_{\max}} \left(s(\tau, n) \zeta(\tau) e^{-i2\pi f(\tau-1)/\tau_{\max}} \right) \right|^2$$

$$\zeta(\tau) = 0.5(1 - \cos(\pi(\tau - 1 + \tau_{\max})/\tau_{\max})).$$

- The rationale behind calculating the discrete Fourier transform (DFT) in (10) is that, by definition, other pulse periods are integer multiples of the tatum period. Thus the overall function $s(\tau, n)$ contains information about the tatum and this is conveniently gathered for each tatum-frequency candidate f using the DFT as in (10). For comparison, Gouyon et al. [24] used an inter-onset-interval histogram and Maher's two-way mismatch procedure [34] served the same purpose. Their idea was to find a tatum period which best explained the multiple harmonically related peaks in the histogram. Frequencies above 20 Hz can be discarded from $S(f, n)$, since tatum frequencies faster than this are very rare.

Probabilistic model for pulse periods

- Simplifying assumption: we can calculate period independent of phase (compare APM).
- Use HMM (Hidden Markov Model) to observer the energies of resonators $s(\gamma,n)$ denoted s_n
- Unobserved: tatum period, tactus period and measure period, respectively: τ_n^{A}, τ_n^{B}, τ_n^{C}.
- Define meter state as holding values of all three unobserved states: $\boldsymbol{q}_n = [j,k,l]$ equivalent to:

$$\tau_n^{\mathrm{A}} = j,\ \tau_n^{\mathrm{B}} = k,\ \text{and}\ \tau_n^{\mathrm{C}} = l.$$

Defining the HMM

The joint probability density of a state sequence $Q = (\boldsymbol{q}_1 \boldsymbol{q}_2 \ldots \boldsymbol{q}_N)$ and observation sequence $O = (\boldsymbol{s}_1 \boldsymbol{s}_1 \ldots \boldsymbol{s}_N)$ can be written as

$$p(Q, O) = P(\boldsymbol{q}_1) p(\boldsymbol{s}_1 | \boldsymbol{q}_1) \prod_{n=2}^{N} P(\boldsymbol{q}_n | \boldsymbol{q}_{n-1}) p(\boldsymbol{s}_n | \boldsymbol{q}_n), \quad (12)$$

where the term $P(\boldsymbol{q}_n | \boldsymbol{q}_{n-1})$ can be decomposed as

$$P(\boldsymbol{q}_n | \boldsymbol{q}_{n-1}) = P(\tau_n^{\mathrm{B}} | \boldsymbol{q}_{n-1}) P(\tau_n^{\mathrm{A}} | \tau_n^{\mathrm{B}}, \boldsymbol{q}_{n-1}) P(\tau_n^{\mathrm{C}} | \tau_n^{\mathrm{B}}, \tau_n^{\mathrm{A}}, \boldsymbol{q}_{n-1})$$

It is musically meaningful to assume that

$$P(\tau_n^{\mathrm{C}} | \tau_n^{\mathrm{B}}, \tau_n^{\mathrm{A}}, \boldsymbol{q}_{n-1}) = P(\tau_n^{\mathrm{C}} | \tau_n^{\mathrm{B}}, \boldsymbol{q}_{n-1}), \quad (14)$$

This and other assumptions yield:

$$P(\boldsymbol{q}_n | \boldsymbol{q}_{n-1}) = P(\tau_n^{\mathrm{B}} | \tau_{n-1}^{\mathrm{B}}) P(\tau_n^{\mathrm{A}} | \tau_n^{\mathrm{B}}, \tau_{n-1}^{\mathrm{A}}) P(\tau_n^{\mathrm{C}} | \tau_n^{\mathrm{B}}, \tau_{n-1}^{\mathrm{C}}). \quad (15)$$

Using the same assumptions, $P(\boldsymbol{q}_1)$ is decomposed and simplified as

$$P(\boldsymbol{q}_1) = P(\tau_1^{\mathrm{B}}) P(\tau_1^{\mathrm{A}} | \tau_1^{\mathrm{B}}) P(\tau_1^{\mathrm{C}} | \tau_1^{\mathrm{B}}). \quad (16)$$

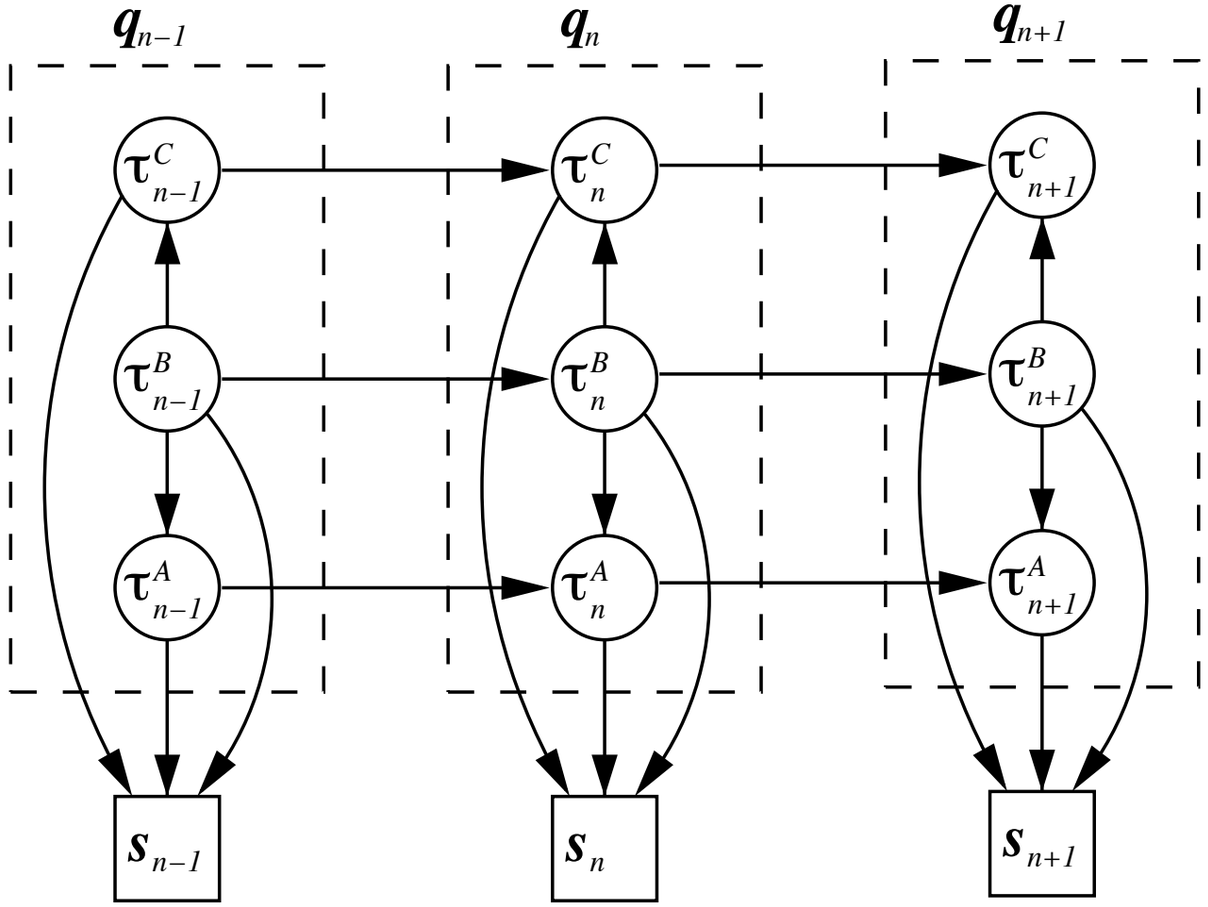


Fig. 5. Hidden markov model for the temporal evolution of the tatum, beat, and measure pulse periods.

Estimating the parameters

$$P(\tau_n^{\mathrm{C}}|\tau_n^{\mathrm{B}},\tau_{n-1}^{\mathrm{C}}) = P(\tau_n^{\mathrm{C}}|\tau_{n-1}^{\mathrm{C}})\frac{P(\tau_n^{\mathrm{C}},\tau_n^{\mathrm{B}}|\tau_{n-1}^{\mathrm{C}})}{P(\tau_n^{\mathrm{C}}|\tau_{n-1}^{\mathrm{C}})P(\tau_n^{\mathrm{B}}|\tau_{n-1}^{\mathrm{C}})},$$

$$P(\tau_n^i|\tau_{n-1}^i) = P(\tau_1^i)\frac{P(\tau_n^i,\tau_{n-1}^i)}{P(\tau_n^i)P(\tau_{n-1}^i)} \approx P(\tau_1^i) f\left(\frac{\tau_n^i}{\tau_{n-1}^i}\right), \tag{20}$$

where $i \in \{\mathrm{A},\mathrm{B},\mathrm{C}\}$. The function f,

$$f\left(\frac{\tau_n^i}{\tau_{n-1}^i}\right) = \frac{1}{\sigma_1\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma_1^2}\left(\ln\left(\frac{\tau_n^i}{\tau_{n-1}^i}\right)\right)^2\right], \tag{21}$$

Prior probabilities for tactus period lengths, $P(\tau^{\mathrm{B}})$, have been measured from actual data by several authors [12], [35], [36]. As suggested by Parncutt [12], we apply the two-parameter lognormal distribution

$$p(\tau^i) = \frac{1}{\tau^i\sigma^i\sqrt{2\pi}}\exp\left[-\frac{1}{2(\sigma^i)^2}\left(\ln\left(\frac{\tau^i}{m^i}\right)\right)^2\right], \tag{22}$$

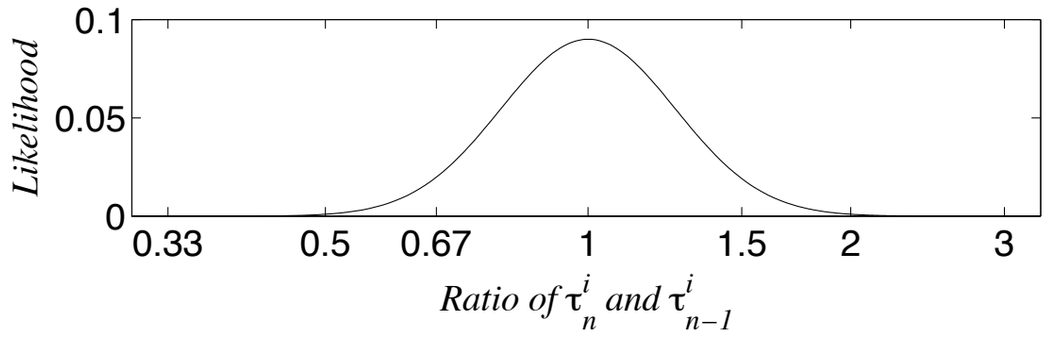


Fig. 6. The likelihood function $f(\tau_n^i/\tau_{n-1}^i)$ which describes the tendency that the periods are slowly-varying.

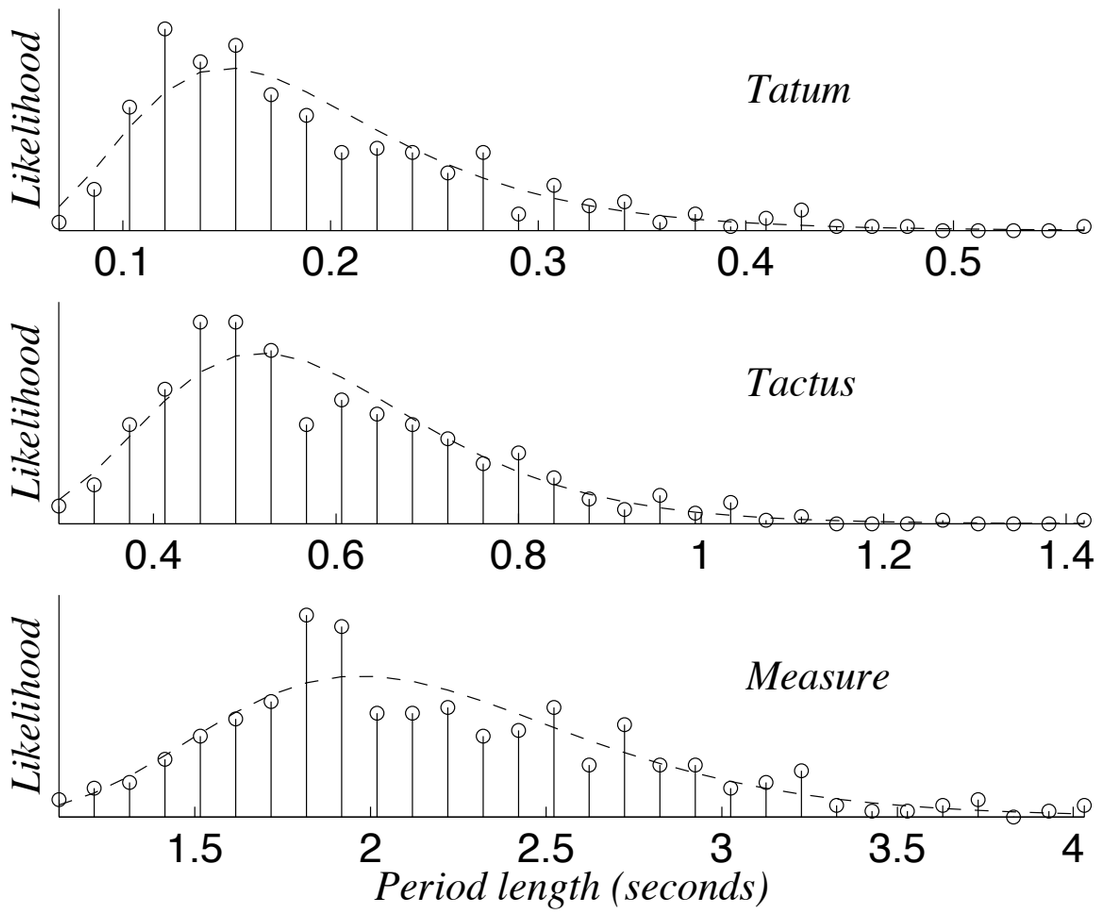


Fig. 7. Period-length histograms and the corresponding lognormal distributions for tatum, tactus, and measure pulses.

Relating different levels of MH

The relation dependencies of simultaneous periods are modeled as follows. We model the latter terms in (18)–(19) as

$$\frac{P(\tau_n^{\mathrm{A}}, \tau_n^{\mathrm{B}} | \tau_{n-1}^{\mathrm{A}})}{P(\tau_n^{\mathrm{A}} | \tau_{n-1}^{\mathrm{A}}) P(\tau_n^{\mathrm{B}} | \tau_{n-1}^{\mathrm{A}})} \approx g\left(\frac{\tau^{\mathrm{B}}}{\tau^{\mathrm{A}}}\right), \tag{23}$$

$$\frac{P(\tau_n^{\mathrm{C}}, \tau_n^{\mathrm{B}} | \tau_{n-1}^{\mathrm{C}})}{P(\tau_n^{\mathrm{C}} | \tau_{n-1}^{\mathrm{C}}) P(\tau_n^{\mathrm{B}} | \tau_{n-1}^{\mathrm{C}})} \approx g\left(\frac{\tau^{\mathrm{C}}}{\tau^{\mathrm{B}}}\right), \tag{24}$$

where $g(x)$ is a Gaussian mixture density of the form

$$g(x) = \sum_{l=1}^{9} w_l N(x; l, \sigma_2), \tag{25}$$

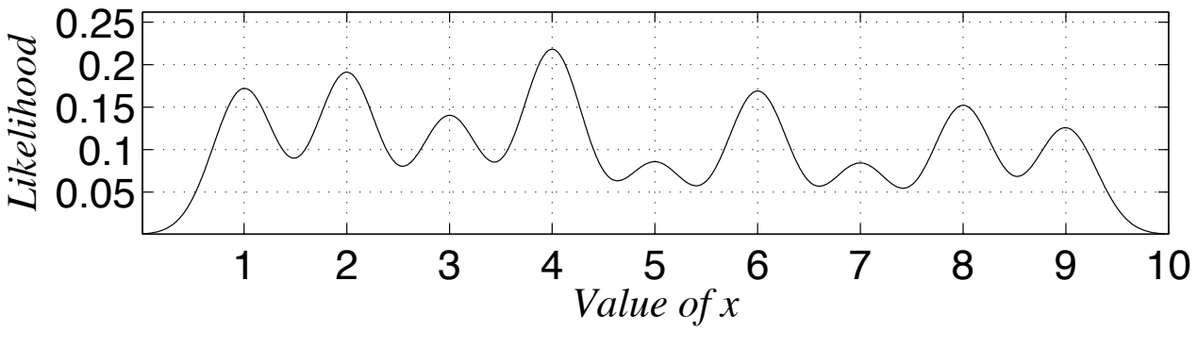


Fig. 8. Distribution $g(x)$ which models the relation dependencies of simultaneous periods (see (25)).

The weights w_l were obtained by first assigning them values according to a musical intuition. Then the dynamic range of the weights was found by raising them to a common power which was varied between 0.1 and 10. The value which performed best in small-scale simulations was selected. Finally, small adjustments to the values were made.

Finding optimal sequence

- Viterbi search

3) Finding the optimal sequence of period estimates: Now we must obtain an estimate for the unobserved state variables given the observed resonator energies and the model parameters. We do this by finding the most likely sequence of state variables $Q = (\boldsymbol{q}_1 \boldsymbol{q}_2 \ldots \boldsymbol{q}_N)$ given the observed data $O = (\boldsymbol{s}_1 \boldsymbol{s}_1 \ldots \boldsymbol{s}_N)$. This can be straighforwardly computed using the Viterbi algorithm widely applied in speech recognition [38]. Thus, we seek the sequence of period estimates,

$$\hat{Q} = \arg\max_{Q} \left(p(Q, O) \right), \tag{26}$$

where $p(Q, O)$ denotes the joint probability density of the hidden and observed variables (see (12)).

- State space too large for full Viterbi. Instead used beam serarch carrying 5 best search candidates forward in time.

Phase estimation

- Tactus and measure done independently from fiterbank outputs
- Fit 2nd HMM to filterbank outputs.

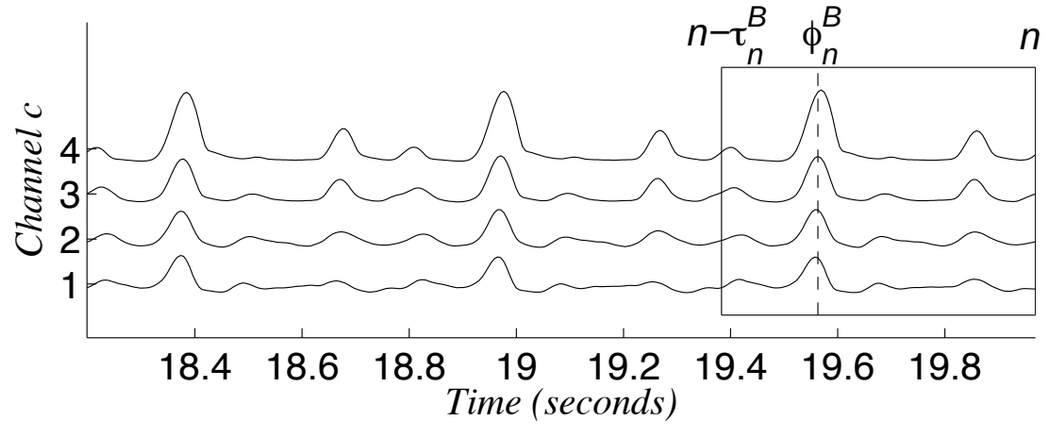


Fig. 9. The rectangle indicates the observation matrix R_n^{B} for tactus phase estimation at time n (here period τ_n^{B} is 0.51 s.). Dashed line shows the correct phase in this case.

State-conditional observation likelihoods $p(R_n^{\mathrm{B}}|\hat{\varphi_n^{\mathrm{B}}})$ for the tactus pulse are approximated as

$$p(R_n^{\mathrm{B}}|\varphi_n^{\mathrm{B}} = j) \propto \sum_{c=1}^{c_0}(c_0 - c + 2)(R_n^{\mathrm{B}})_{c,j}, \qquad (27)$$

where $c = 1$ corresponds to the lowest-frequency channel.

- Measure level used template matching

The first pattern can be summarized as "low, loud, –, loud", and the second as "low, –, loud, –". The two patterns are combined into a single vector to perform phase estimation according to whichever pattern matches better to the data

$$h_n^{(1,2)}(l) = \max\left(h_n^{(1)}(l), h_n^{(2)}(l)\right). \qquad (31)$$

The state-conditional observation likelihoods are then defined as

$$p(R_n^{\mathrm{C}}|\varphi_n^{\mathrm{C}} = j) \propto h_n^{(1,2)}(j - (n - \hat{\tau}_n^{\mathrm{C}} + 1)). \qquad (32)$$

- Error measure for Viterbi:

normally distributed as a function of a *prediction error* e which measures the deviation of φ_n^i from the predicted next beat occurence time given the previous beat time φ_{n-1}^i and the period $\hat{\tau}_n^i$:

$$P(\varphi_n^i|\varphi_{n-1}^i) = \frac{1}{\sigma_3\sqrt{2\pi}} \exp\left(-\frac{e^2}{2\sigma_3^2}\right), \qquad (33)$$

where

$$e = \frac{1}{\hat{\tau}_n^i}\left\{\left[\left(\left|\varphi_n^i - \varphi_{n-1}^i\right| + \frac{\hat{\tau}_n^i}{2}\right) \bmod \hat{\tau}_n^i\right] - \frac{\hat{\tau}_n^i}{2}\right\}, \qquad (34)$$

Results

TABLE I

STATISTICS OF THE EVALUATION DATABASE.

	# Pieces with annotated pulses		
Genre	Tatum	Tactus	Measure
Classical	69	84	0
Electronic / dance	47	66	62
Hip hop / rap	22	37	36
Jazz / blues	70	94	71
Rock / pop	114	124	101
Soul / RnB / funk	42	54	46
Unclassified	12	15	4
Total	376	474	320

TABLE II

TACTUS ANALYSIS PERFORMANCE (%) OF DIFFERENT METHODS.

	Continuity required			Individual estimates		
Method	Correct	Accept d/h	Period c.	Correct	Accept d/h	Period c.
Causal	57	**68**	74	63	78	76
Noncausal	59	**73**	74	64	80	75
Scheirer [20]	27	**31**	30	48	69	57
Dixon [16]	7	**26**	10	15	53	25
O+Dixon	12	**39**	15	22	63	30

TABLE III

METER ANALYSIS PERFORMANCE OF THE PROPOSED METHOD.

		Continuity required			Individual estimates		
Method	Pulse	Correct	Accept d/h	Period	Correct	Accept d/h	Period
Causal	Tatum	44	**57**	62	51	**72**	65
	Tactus	57	**68**	74	63	78	76
	Measure	42	**48**	**78**	43	51	81
Non-causal	Tatum	45	**63**	62	52	**74**	65
	Tactus	59	**73**	74	64	80	75
	Measure	46	**54**	**79**	47	55	81

Results ctd

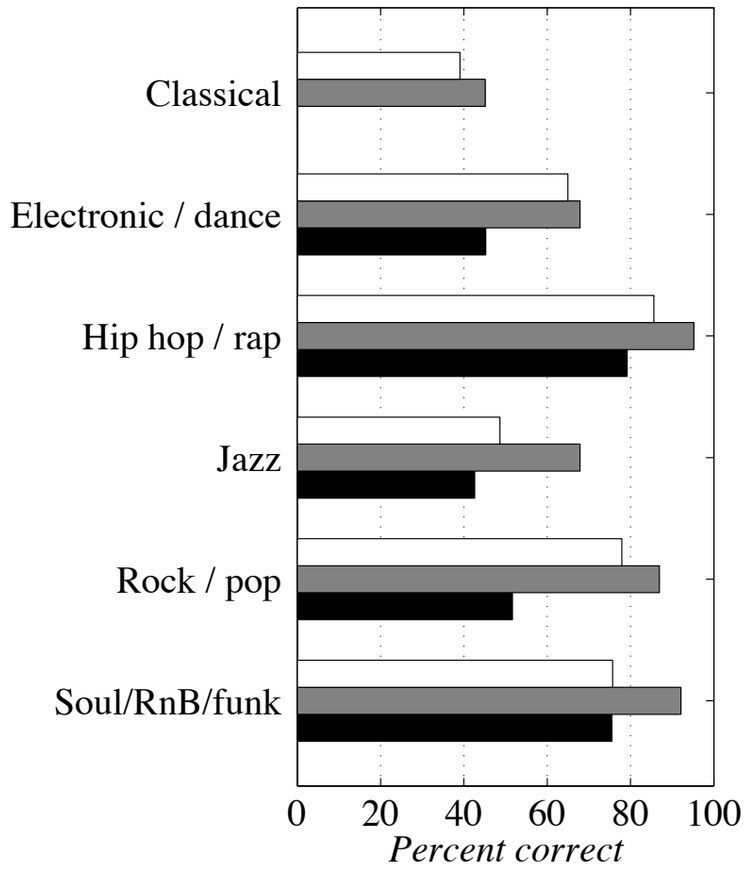


Fig. 10. Performance of the proposed causal system within different musical genres. The "accept d/h" (continuity required) percentages are shown for the tatum (white), tactus (gray), and measure pulses (black).

TABLE IV
METER ANALYSIS PERFORMANCE (%) FOR DIFFERENT SYSTEM CONFIGURATIONS.

	Continuity required, accept d/h			Individual estimates, accept d/h		
Method	Tatum	Tactus	Measure	Tatum	Tactus	Measure
0. Baseline	**63**	**73**	**54**	74	80	55
1. No joint estim.	**58**	**68**	**49**	71	75	50
2. No temporal proc.	**45**	**54**	**31**	72	77	50
3. Neither of the two	**41**	**50**	**25**	70	72	44