

Sound event detection and rhythmic parsing of music signals

ISMIR Graduate School, October 4th-9th, 2004

Contents:

- Introduction
- Measuring the degree of *change* in music signals
- Onset detection
- Rhythmic pulse estimation
- Higher-level modeling for rhythmic parsing
- Demonstrations

1 Introduction

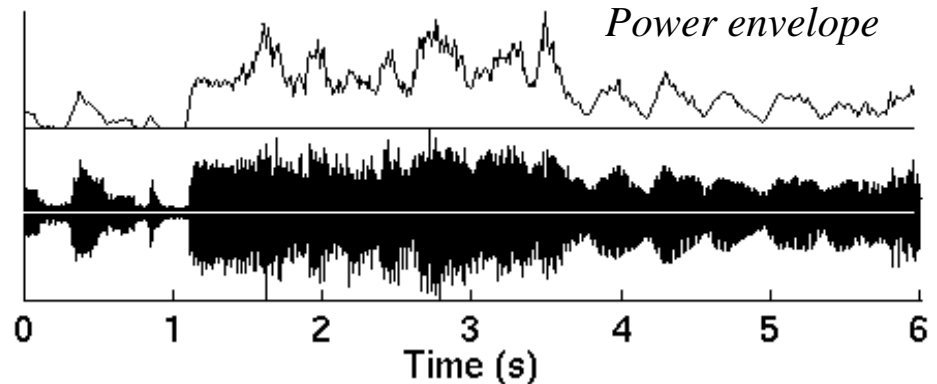
- *Onset detection* = Detection of the beginnings of discrete acoustic events in an acoustic signals
- *Rhythmic parsing* (= *musical meter analysis*)
 - detecting moments of musical stress in an acoustic signal and processing them so that underlying periodicities are discovered
 - e.g. tapping foot to music
 - rhythmically meaningful *segmentation* of musical signals at different time scales
- Applications
 - temporal framework for audio editing
 - audio/video synchronization
 - music segmentation for further analysis (e.g. transcription)

2 Measuring the *degree of change* in music

- *Moments of change* are important for onset detection and rhythmic parsing
 - Percept of an onset is caused by a noticeable change in the intensity, pitch or timbre of a sound
 - moments of musical stress (accents) are caused by the beginnings of sound events, sudden changes in loudness or timbre, harmonic changes
- *Perceptual change* should be estimated
 - to detect what humans detect and to ignore what humans ignore (music is changing all the time!)
 - to do musically meaningful rhythmic parsing









Measuring the degree of change in music

- Time-domain signal
→ some data reduction is needed
- *But*: the power envelope of a signal is not sufficient



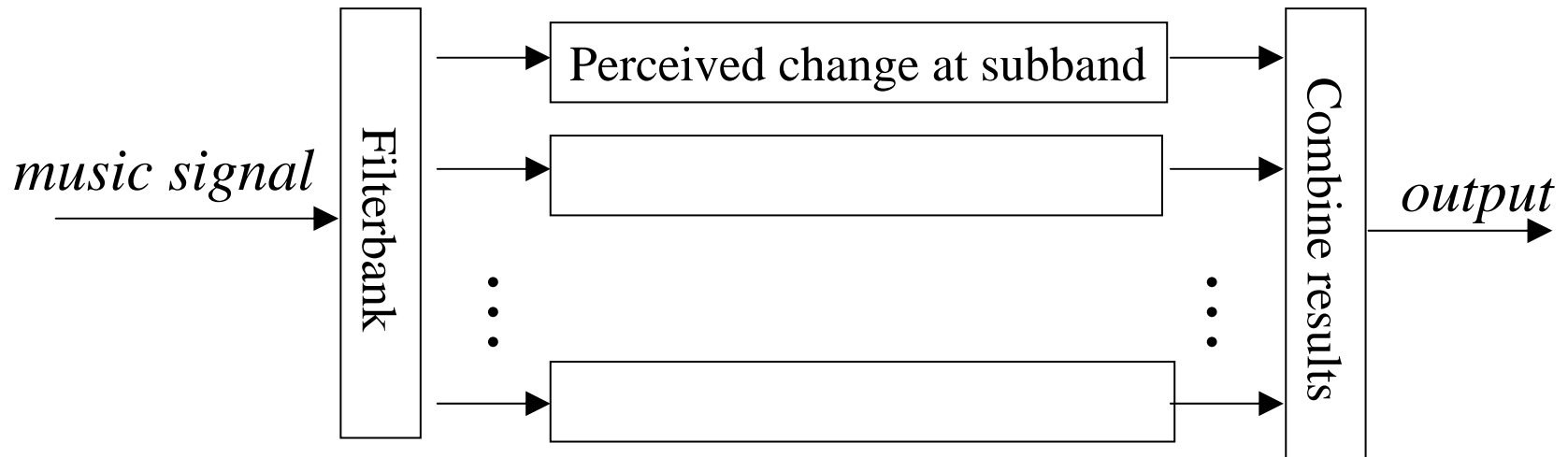
- Frequency selectivity of hearing: audibility of a change at each *critical band* is only affected by the spectral components within the same band
 - components within a single critical band may mask each other
 - but: if the frequency separation is sufficiently large, the masking component must be about million times louder than the other
- Need to *measure change independently at critical bands*, and then combine the results

Measuring the degree of change in music

- Scheirer's classical demonstration:
 - perceived rhythmic content of many music types remains the same if only the power envelopes of a few subbands are preserved and then used to modulate a white noise signal
 - one band is not enough
 - applies to music with "strong beat"
 - Let's repeat the experiment:
 - too much data reduction or not?
- | | | |
|---|---|-------------|
|  |  | Brentwood |
|  |  | Sambafrique |
|  |  | The Bells |
|  |  | Staring |

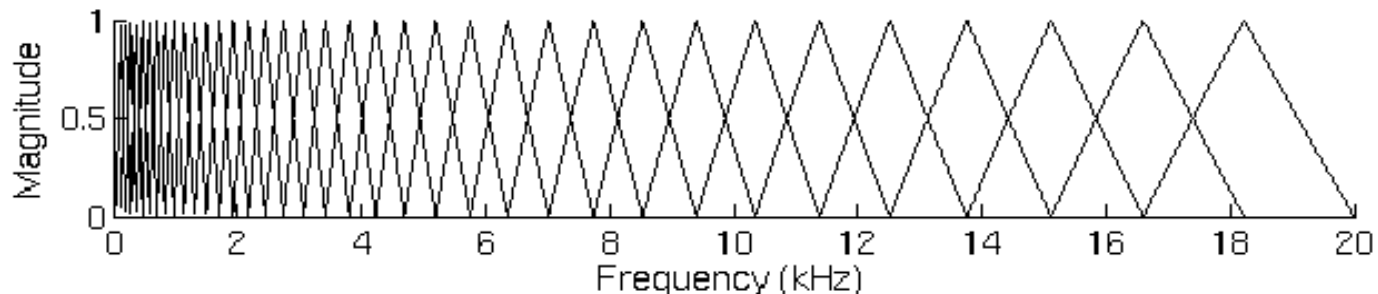
Measuring degree of change in music

Bandwise processing



■ Filterbank:

- Fourier transforms in successive 23ms time frames (50% overlap)
- in each frame, 36 triangular-response bandpass filters are simulated that are uniformly distributed on the Mel frequency scale (50Hz – 20kHz)



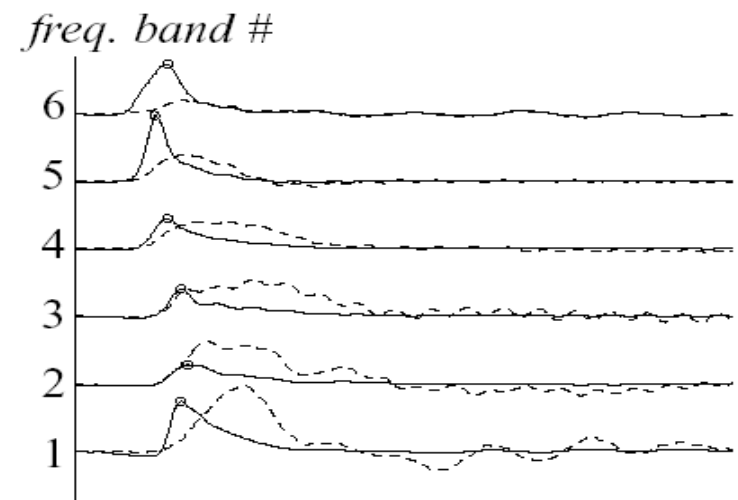
Measuring degree of change in music

Degree of change at each band

- Denote by $x_b(n)$ the *power envelope at critical band* $b=1,\dots,36$ as a function of time (frame index) n
- How to measure the degree of change at subbands? Differential?
 - For humans, the smallest detectable change in intensity, ΔI , is approximately proportional to the intensity I of the signal, the same amount of increase being more prominent in a quiet signal.
 - Audible ratio $\Delta I / I$ is approximately constant
- Thus it is reasonable to normalize the differential of power with power:

$$\frac{(d/dt)x_b(n)}{x_b(n)} = \frac{d}{dt} \ln[x_b(n)]$$

- **Figure** (piano onset):
 dashed line: $(d/dt) x_b(n)$
 solid line: $(d/dt) \ln[x_b(n)]$



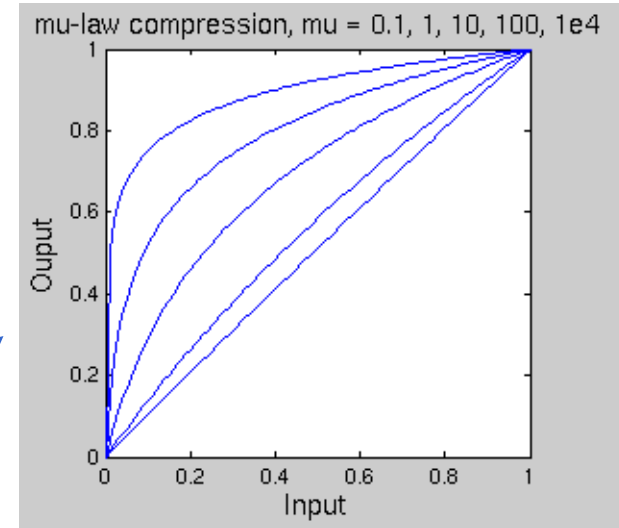
Measuring degree of change in music

Degree of change at each band

- A numerically robust way of calculating the logarithm is the *μ -law compression*,

$$y_b(n) = \frac{\ln[1 + \mu x_b(n)]}{\ln(1 + \mu)}$$

constant μ determines the degree of compression for $x_b(n)$ ($\mu=10\dots 10^4 / \sigma_x$)



- *Lowpass filter* the compressed power envelopes at 10Hz
 - denote resulting signal with $z_b(n)$
- *Differentiate, and retain only positive changes* ($\text{HWR}(x)=\max(x, 0)$):

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

- Form *weighted sum of* $z_b(n)$ *and* $z_b'(n)$:

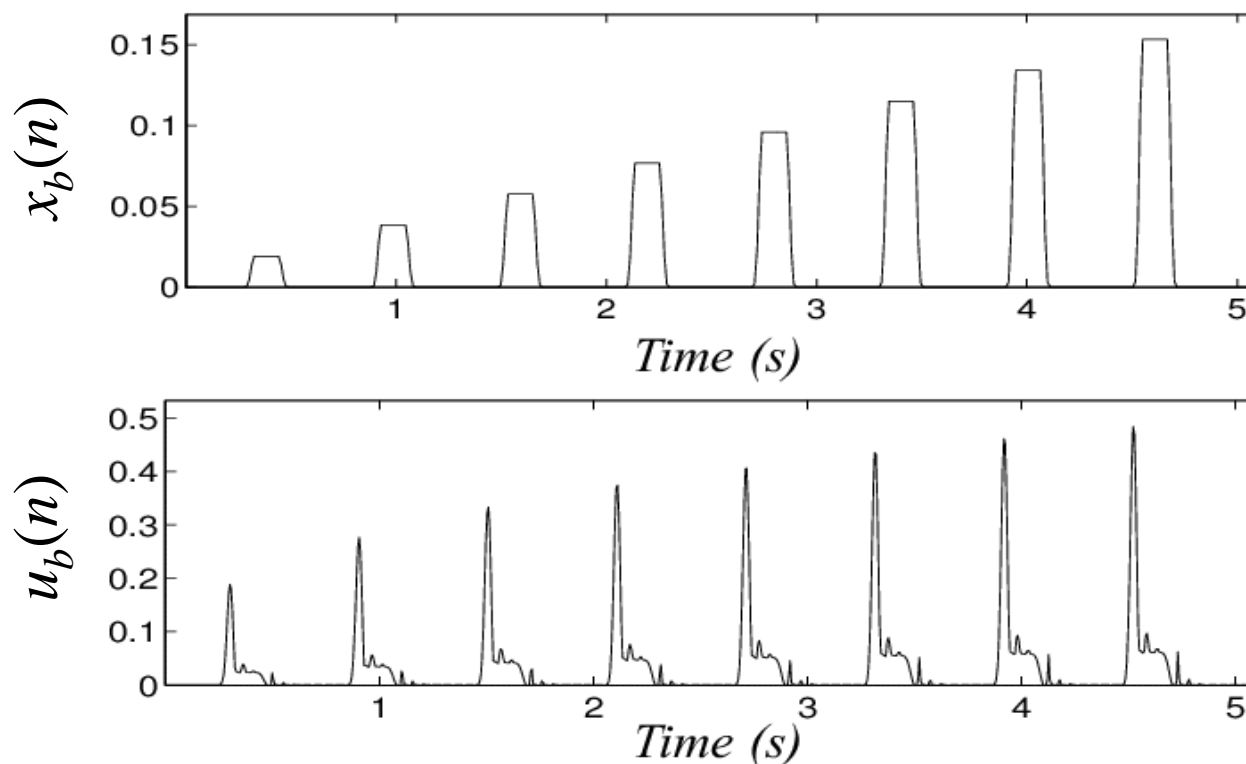
$$u_b(n) = (1-\lambda) z_b(n) + \lambda (f_r/f_{LP}) z_b'(n)$$

where $\lambda=0.8$ (or $0.6\dots 1.0$) balances between $z_b(n)$ and $z_b'(n)$, and (f_r/f_{LP}) is a normalizing constant

Measuring degree of change in music

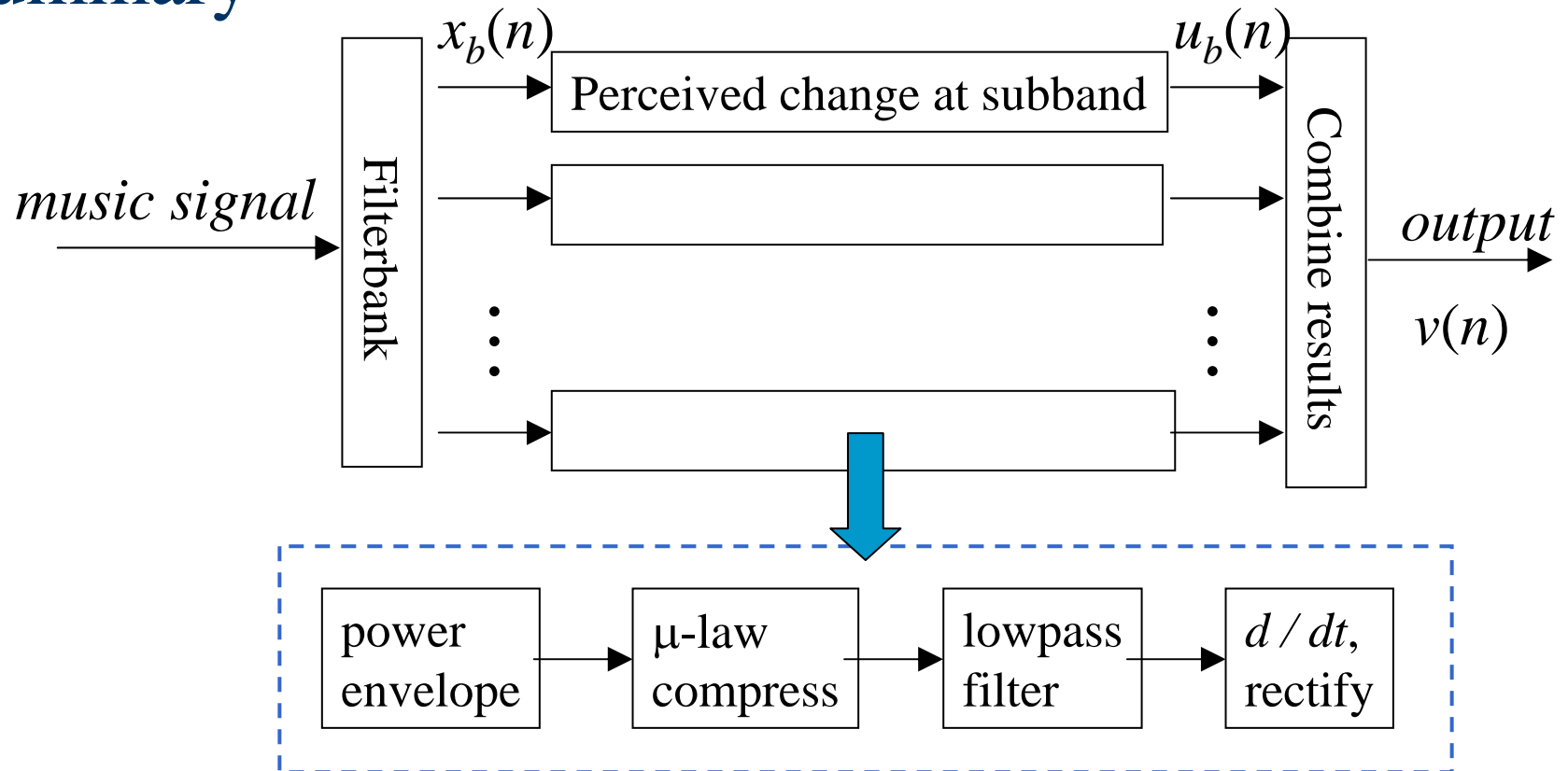
Degree of change at each band

- Figure: illustration of the dynamic compression and weighted differentiation steps for an artificial subband signal $x_b(n)$



Measuring degree of change in music

Summary

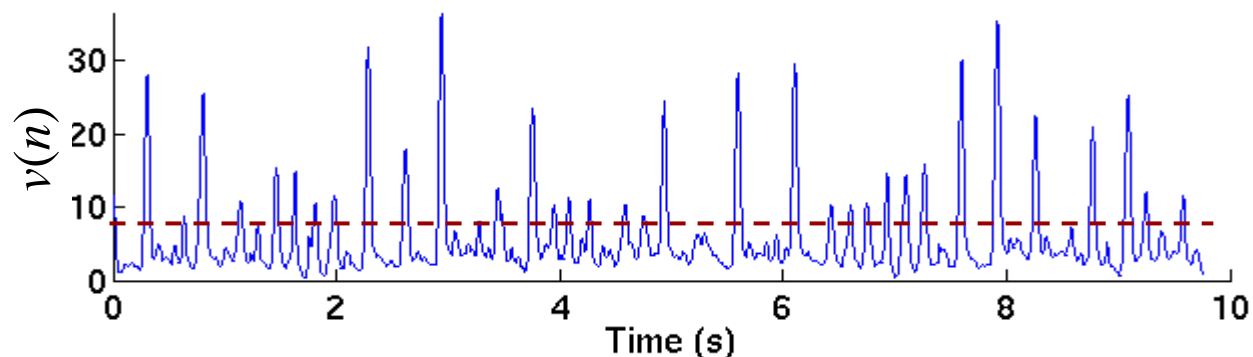


- **Finally:** sum across channels to estimate overall change

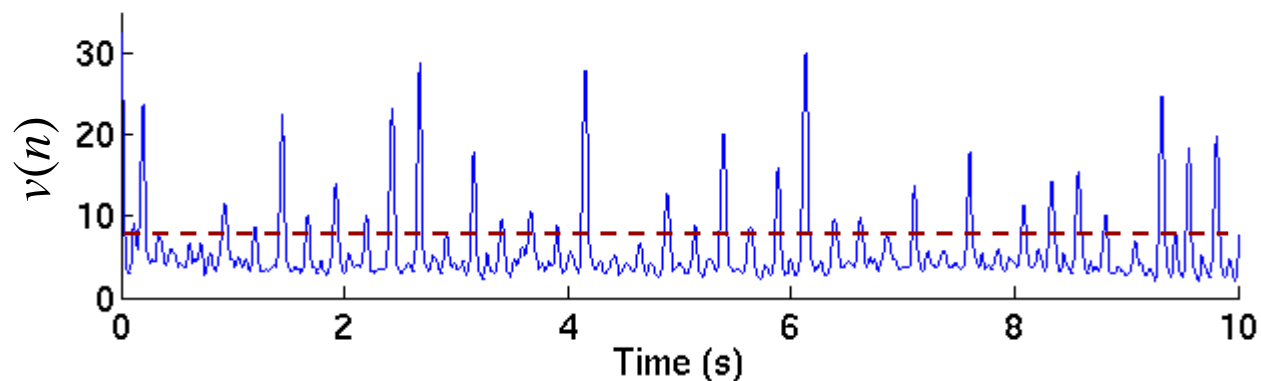
$$v(n) = \sum_{b=1}^{36} u_b(n)$$

Measured change signals

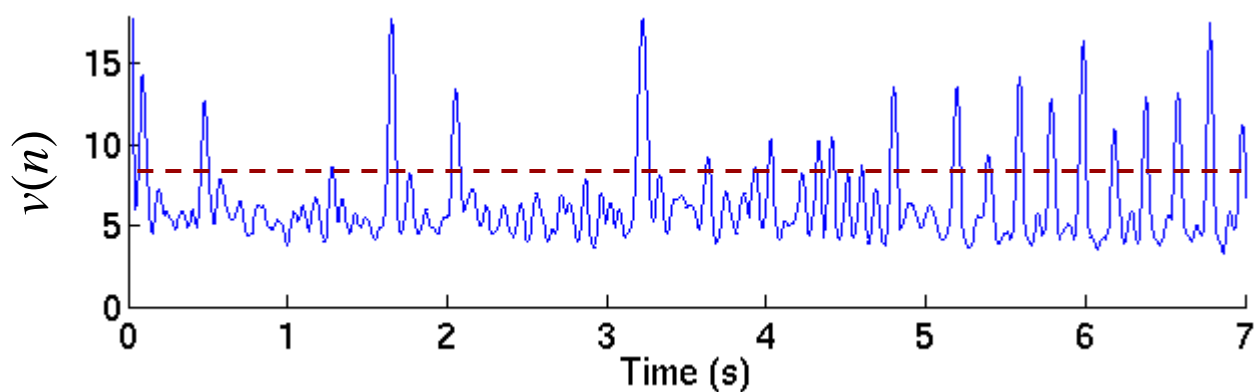
- Brentwood Jazz Quartet



- Lee Ritenour

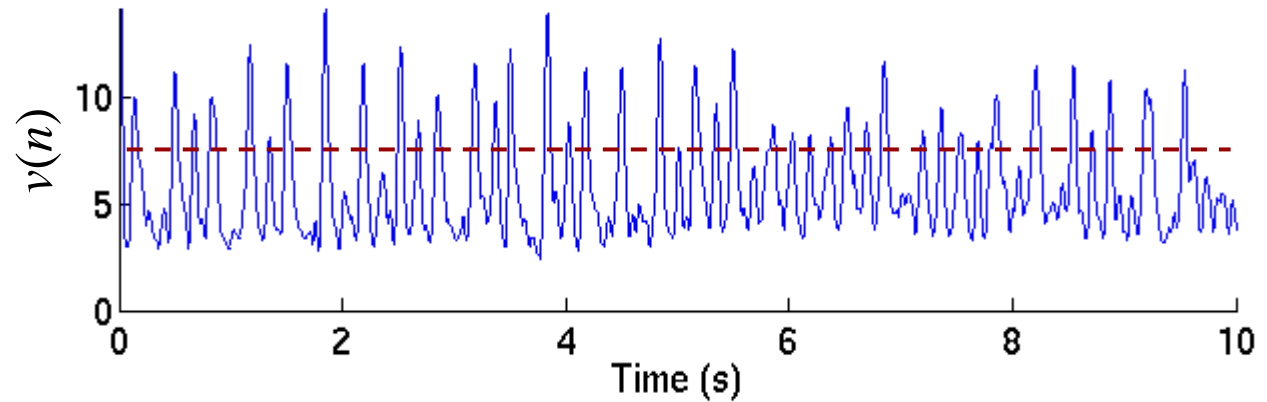


- Lynyrd Skynyrd

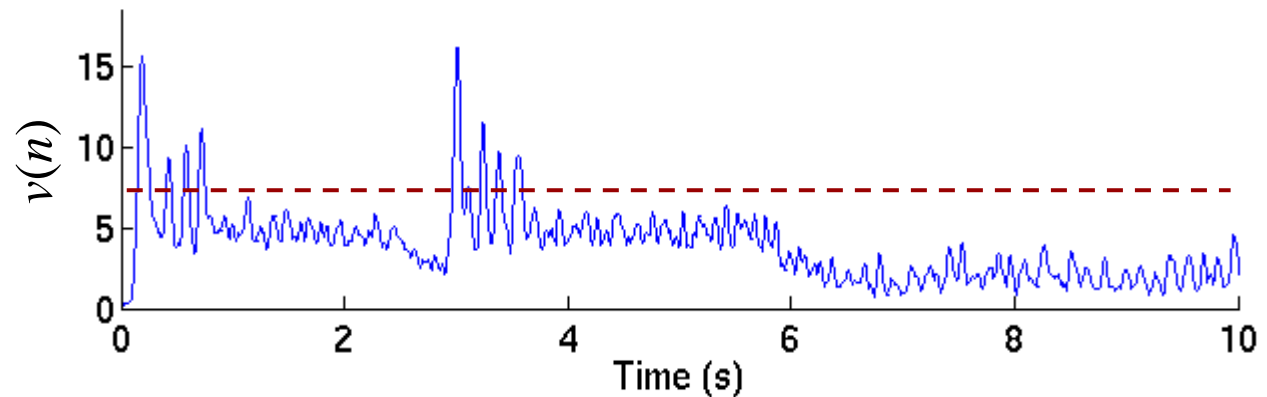


Measured change signals

■ Bach



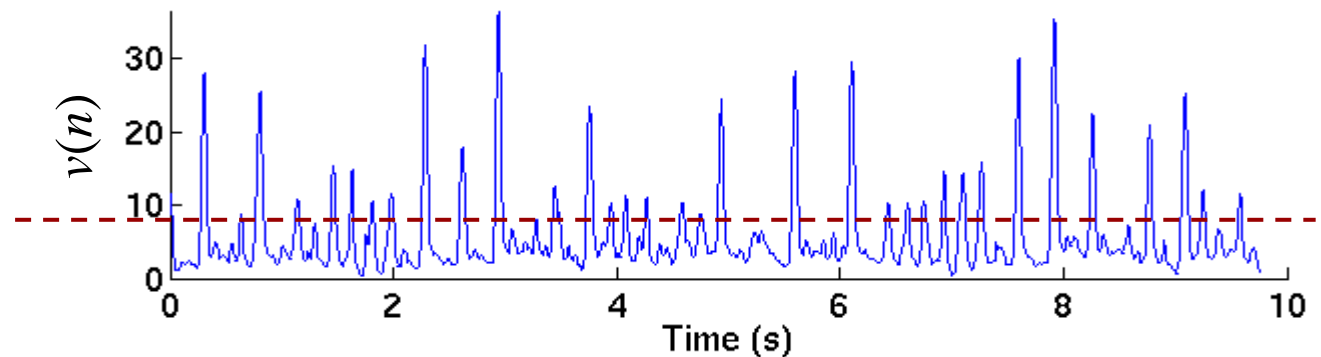
■ Beethoven



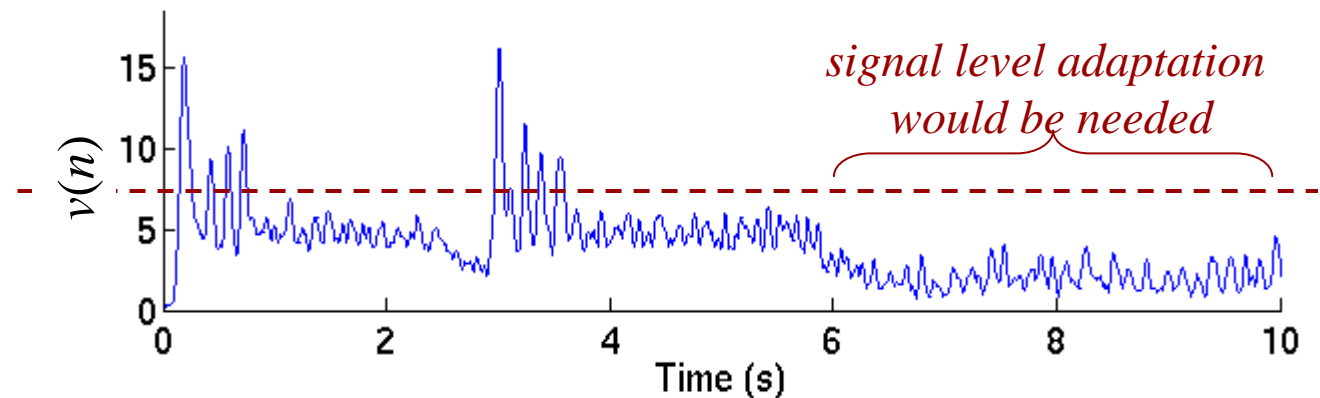
2 Onset detection

- Adaptive thresholding and peak picking
- Robust one-by-one detection of onsets is hard to attain!

- Brentwood:



- Beethoven

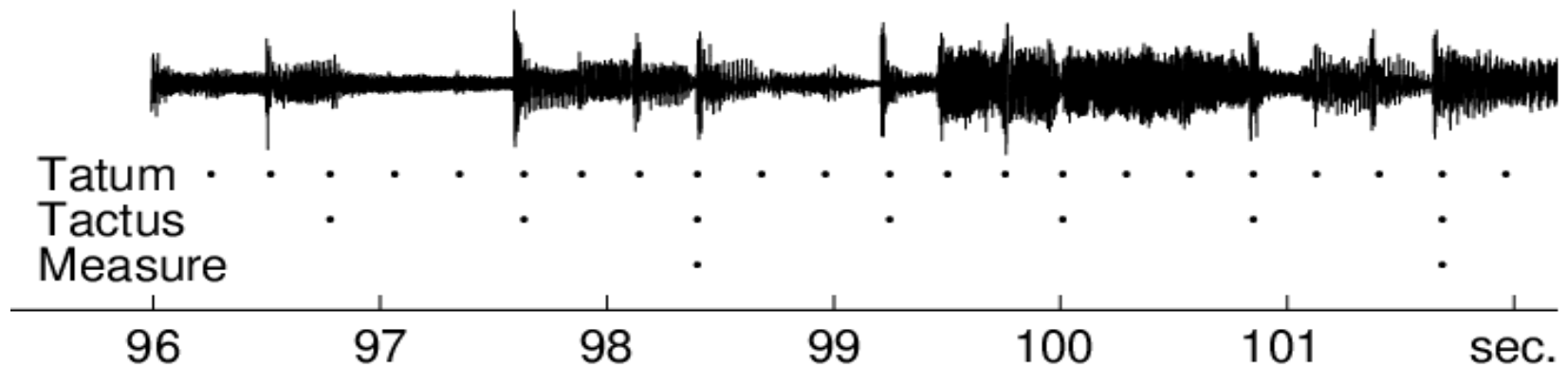


Onset detection

- Remark: there are several other approaches to onset detection, too
- Methods based on
 - complex-domain unpredictability
 - probabilistic modeling (model change detection)
 - changes at the output of an auditory model (pitch included)
 - independent component analysis
 - ...

3 Musical meter

- Characterizes the *temporal regularity* of a music signal
- **Figure:** Musical meter is hierarchical in structure
 - pulse sensations at different time scales
 - *tactus* level is the most prominent (“foot tapping rate”)
 - *tatum*: “time quantum” (fastest pulse)
 - *measure* pulse: related to harmonic change rate



Meter analysis

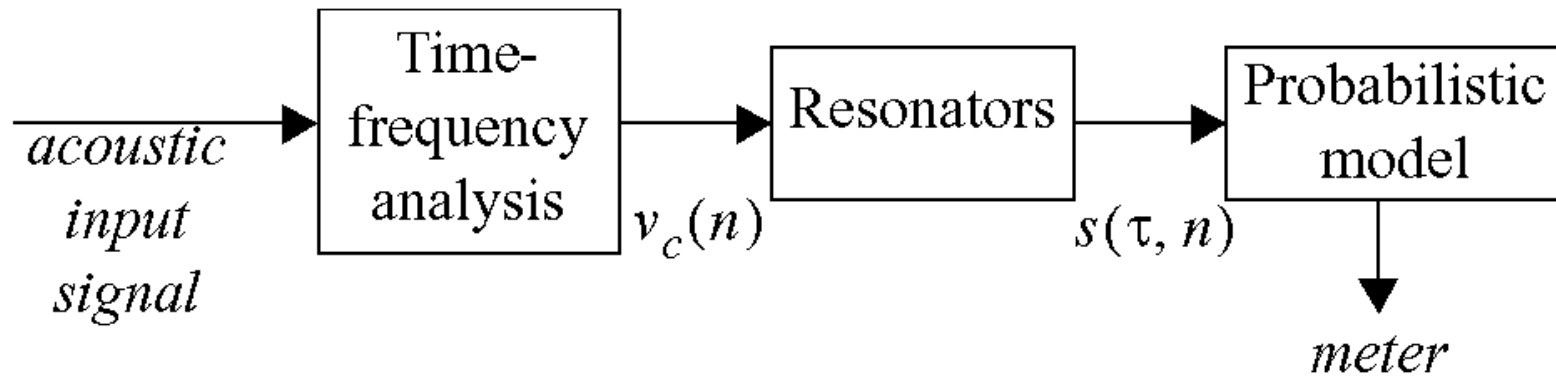
Existing methods

(*Early work*: Steedman77, LonguetHL82,84, LerdahlJ83, Lee85, PovelE85)

System	Input	Aim	Approach	Evaluation material
Parncutt 1994	score	meter	rule-based	artificial synthesized patterns
Brown 1994	score	meter	autocorrelation	classical scores
Rosenthal 1992	MIDI	meter	rule-based	92 piano performances
LargeK 1994	MIDI	meter	oscillators	few example pieces
Temperley 1999	MIDI	meter, quant.	rule-based	source codes available
Dixon 2001	MIDI, audio	tactus	rule-based	source codes available
CemgilK 2000-03	MIDI	tactus, quant.	probabilistic model	expressive performances
Raphael 2001	MIDI, audio	tactus, quant.	probabilistic model	expressive performances
GotoM 1995-97	audio	meter	DSP	85 pieces, pop, 4/4 time
Scheirer 1998	audio	tactus	DSP	"strong beat", sources available
Laroche 2001	audio	tactus, swing	probabilistic model	steady-tempo music → demos
SetharesS 2001	audio	meter	DSP	steady-tempo music → examples
GouyonHC 2002	audio	tatum pulse	DSP	57 drum tracks, steady tempo
Klapuri 2003	audio	meter	DSP + probabilistic	474 pieces, all music types

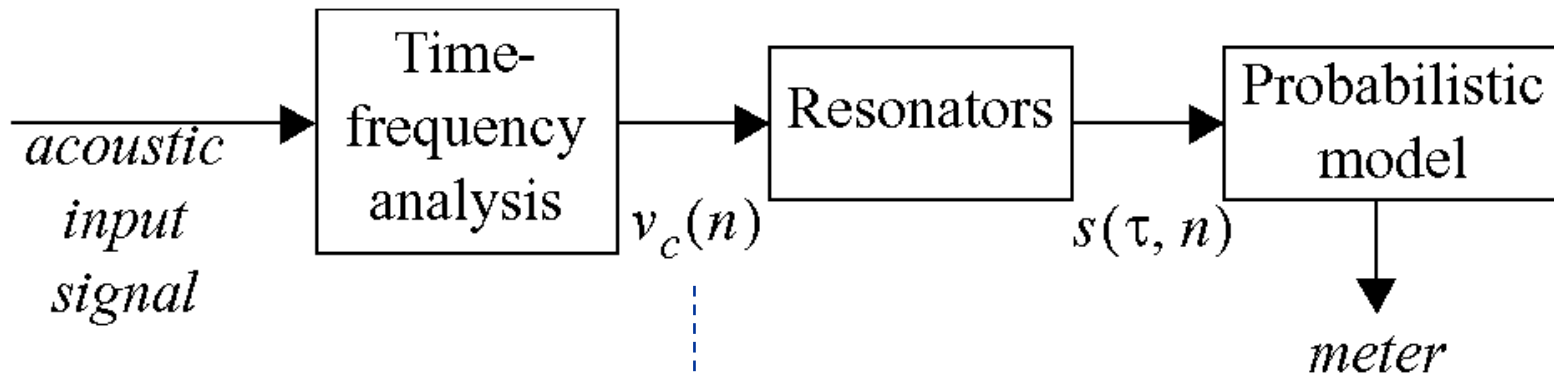
Meter analysis

Overview of the TUT method



Meter analysis

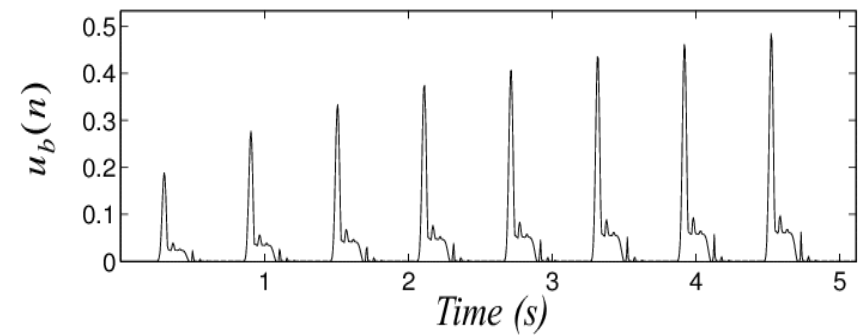
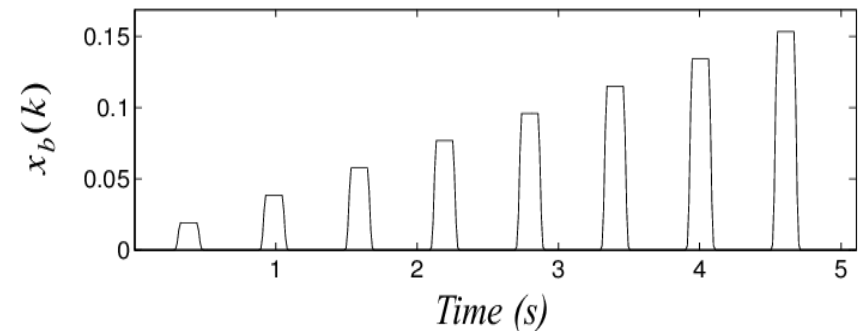
3.1 Degree of accent (change)



Accent signals (degree of change)

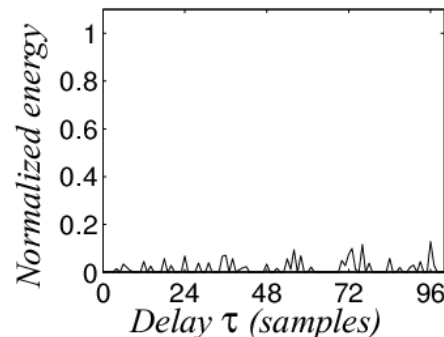
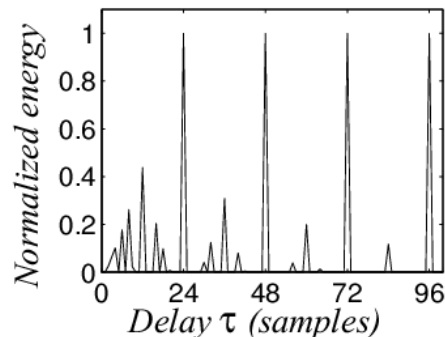
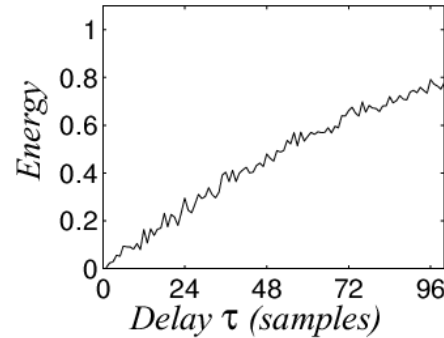
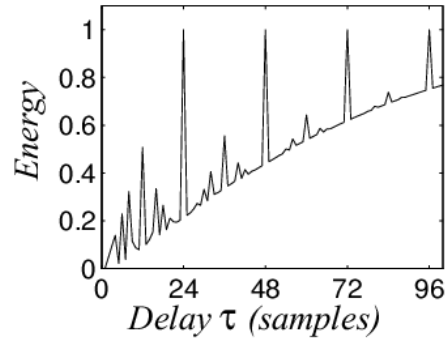
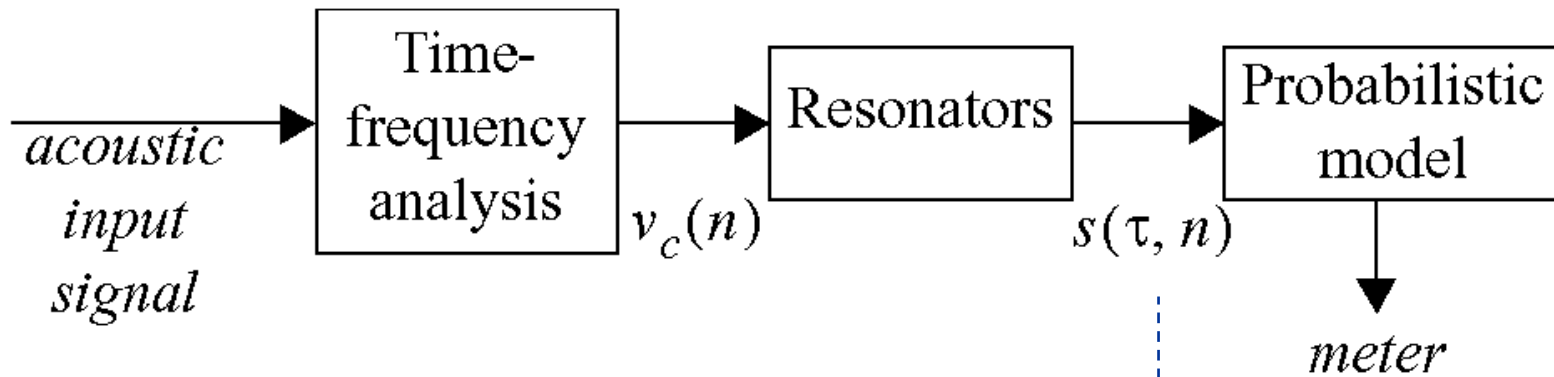
- Degree of accentuation as a function of time at *four frequency channels*

$$v_c(n) = \sum_{b=(c-1)M+1}^{cM} u_b(n)$$



Meter analysis

3.2 Metrical pulse saliencies



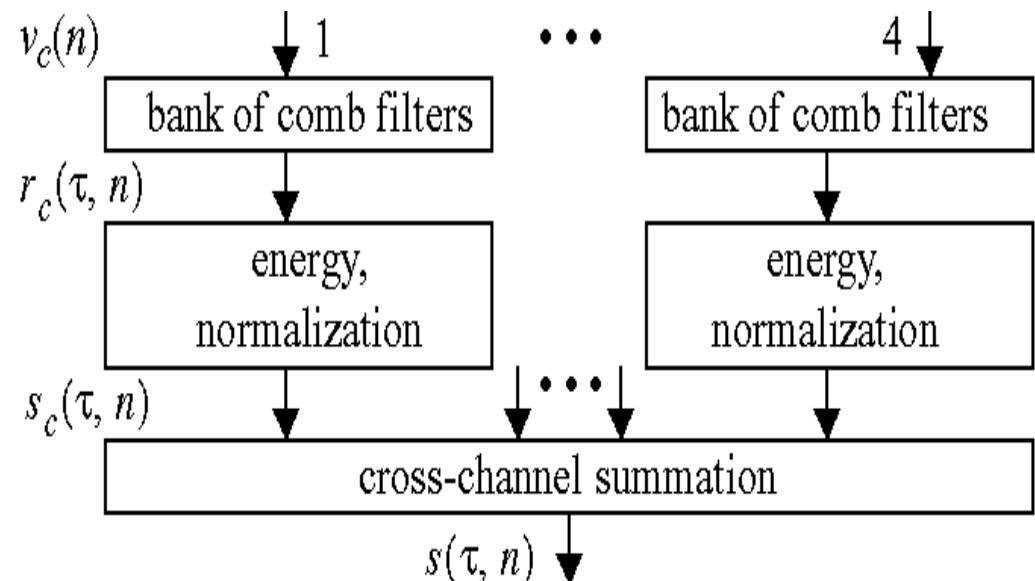
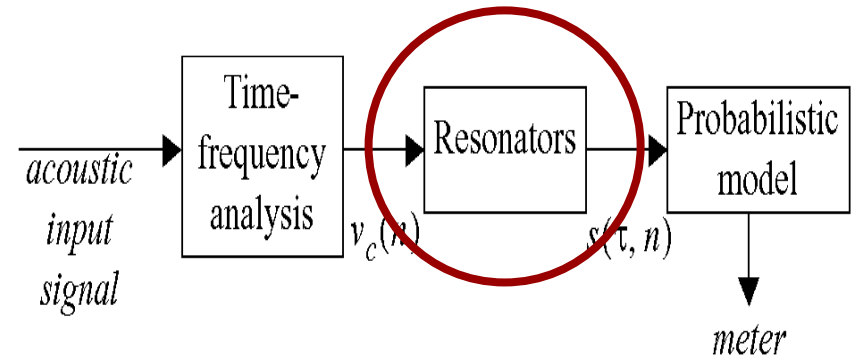
Metrical pulse saliencies (weights)

- Saliencies of different metrical pulses τ at time n (resonator energies)

Meter analysis

Bank of comb filter resonators

- Pulse saliences are obtained by analyzing the periodicity of the four accent signals
- **Comb filters** were found very suitable for this purpose
 - originally proposed by Scheirer

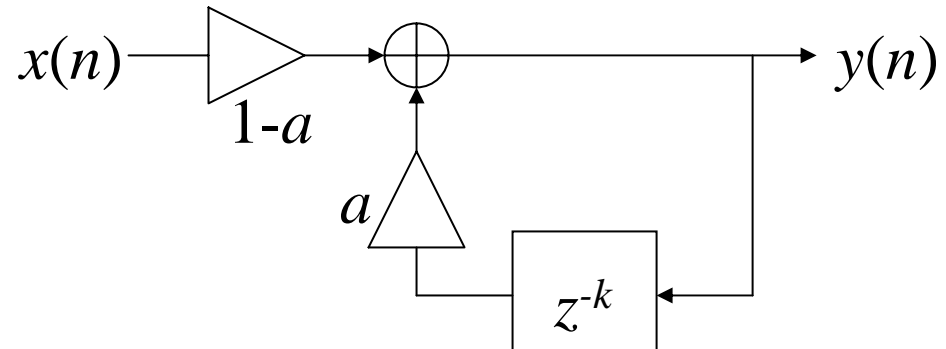


Meter analysis

Comb filters

- Transfer function

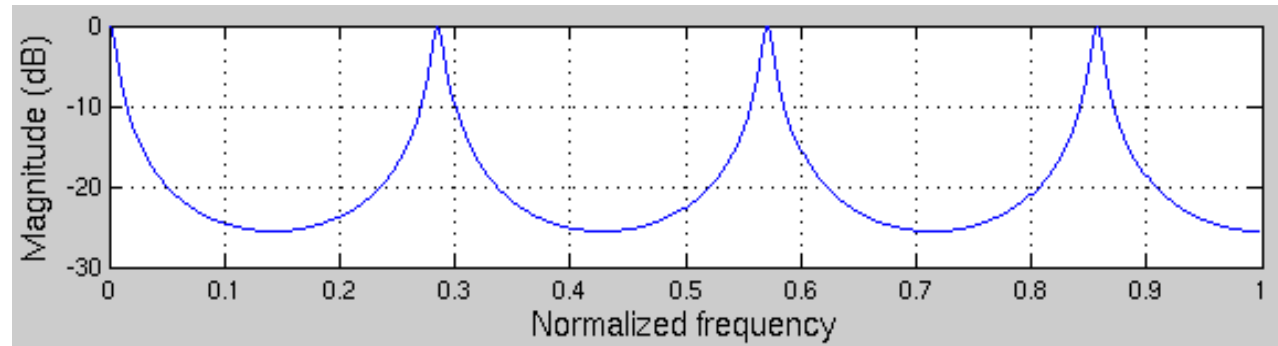
$$H(z) = \frac{1-a}{1-az^{-k}}$$



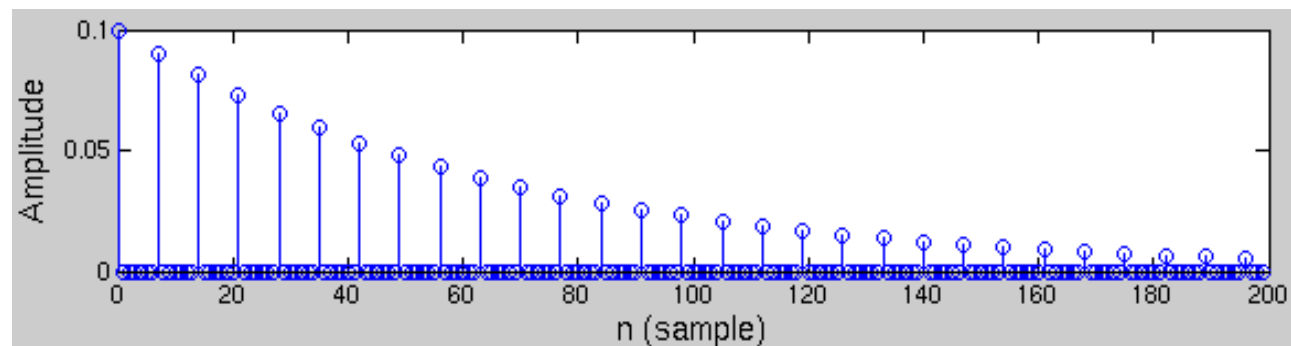
- Magnitude response

$$a = 0.9$$

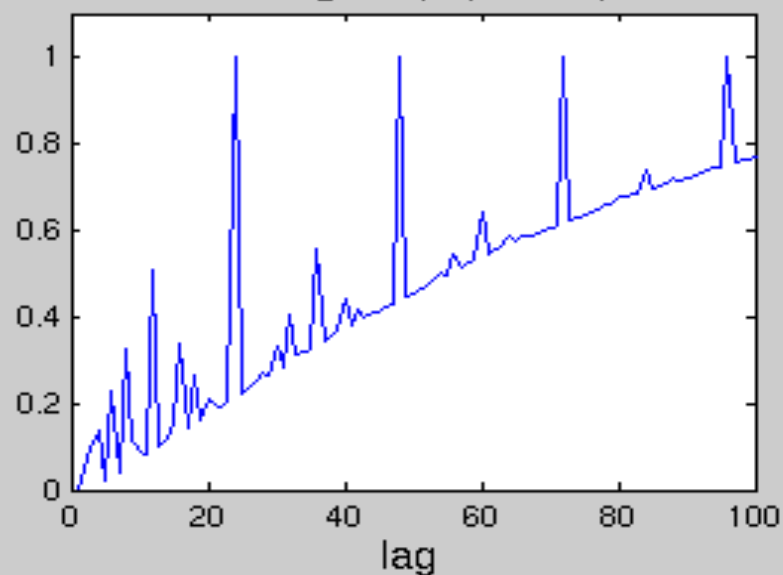
$$k = 7$$



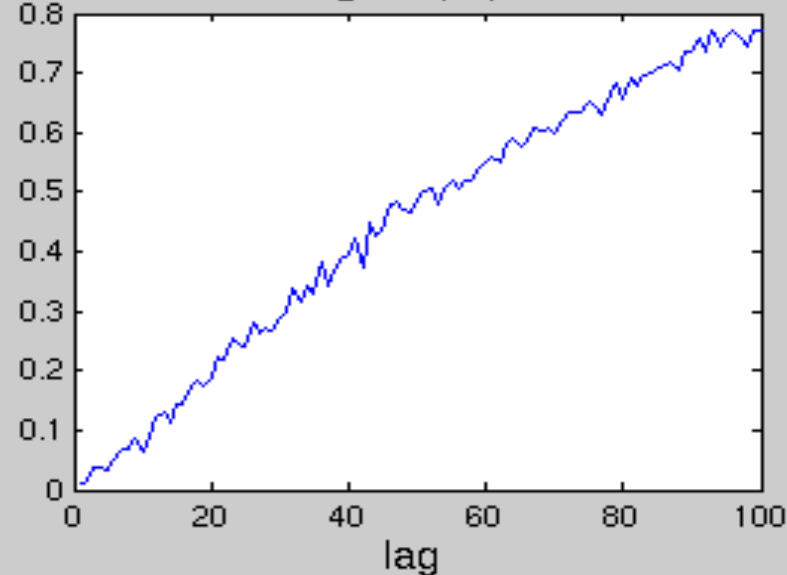
- Impulse response:



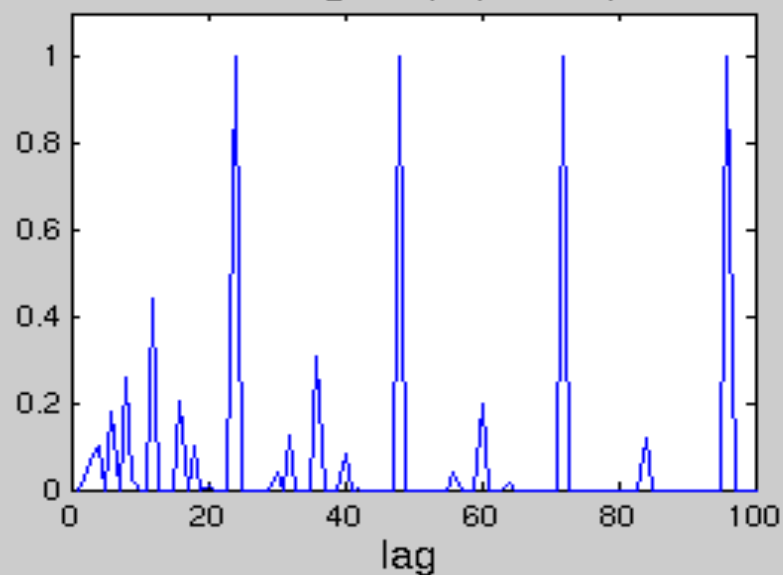
Resonator energies (input: impulse train)



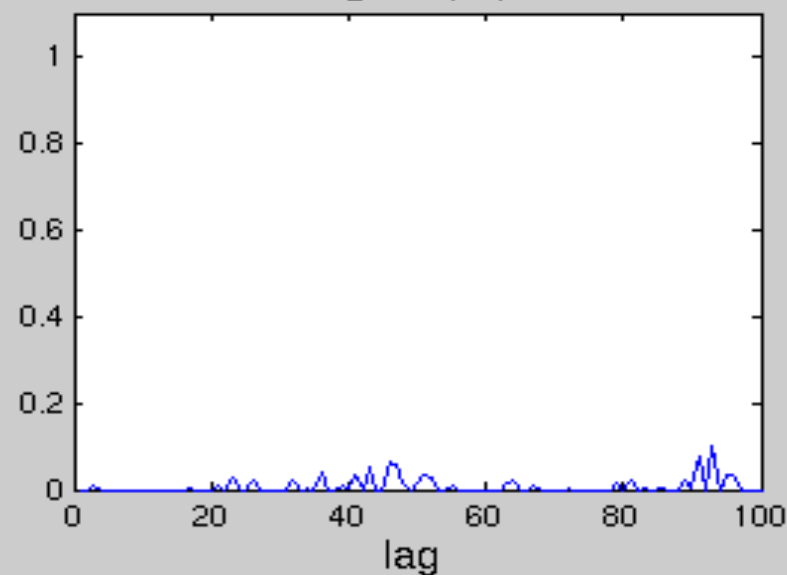
Resonator energies (input: white noise)



Normalized energies (input: impulse train)

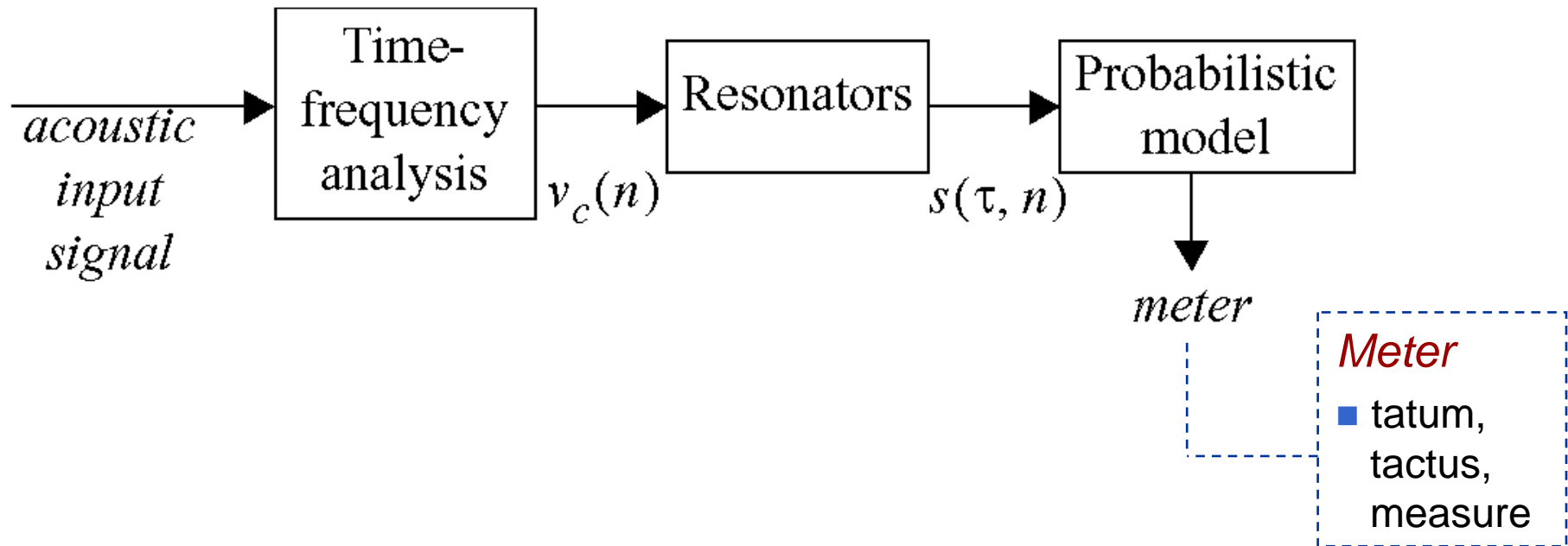


Normalized energies (input: white noise)



Meter analysis

3.3 Higher-level modeling



*Finds pulse **periods** first and then **phases** only for the winning periods*

Pulse periods

HMM

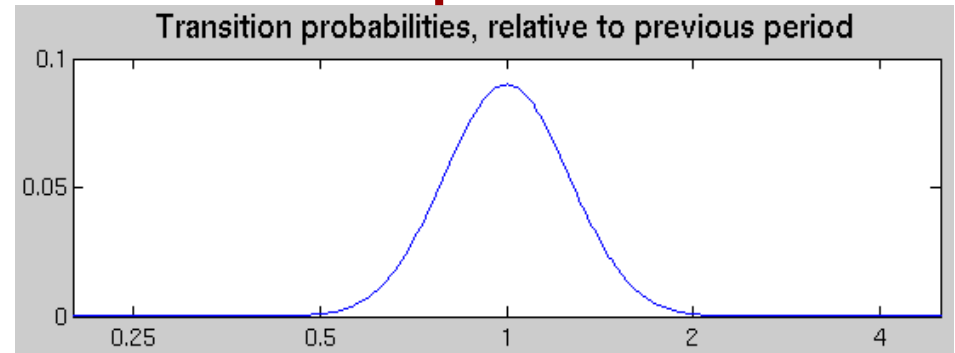
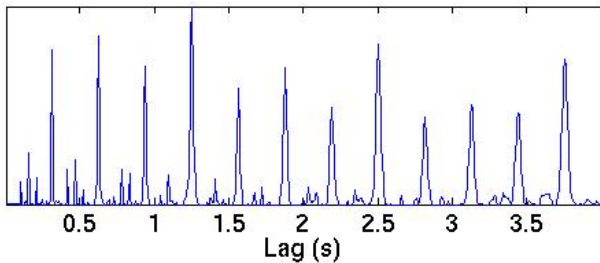
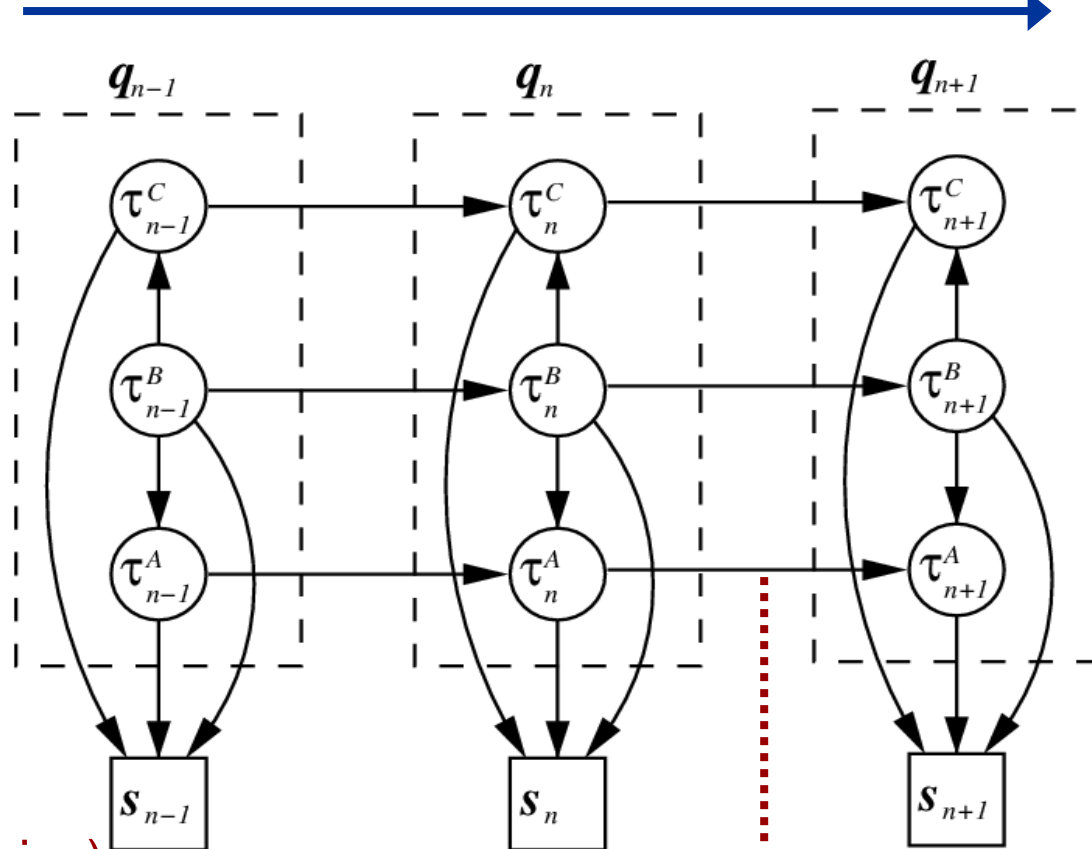
Time

Measure:

Tactus:

Tatum:

Observation:
(comb filter energies)



Meter analysis

Probabilistic model

- Values of the three hidden variables (pulse periods) are jointly defined using a *state variable* $\mathbf{q}_n = [j, k, l]$
 - equivalent to $\tau^A = j, \tau^B = k, \tau^C = l$
- Hidden-state process is a first-order Markov process
 - initial state distribution $P(\mathbf{q}_1)$
 - transition probabilities $P(\mathbf{q}_n | \mathbf{q}_{n-1})$
 - state-conditional observation densities $P(s_n | \mathbf{q}_n)$
- Joint probability density of a state sequence $Q = (\mathbf{q}_1 \mathbf{q}_2 \dots \mathbf{q}_N)$ and observation sequence $O = (s_1 s_2 \dots s_N)$:

$$p(Q, O) = P(\mathbf{q}_1) P(s_1 | \mathbf{q}_1) \prod_{n=2}^N P(\mathbf{q}_n | \mathbf{q}_{n-1}) p(s_n | \mathbf{q}_n)$$

- A remaining problem is to find reasonable estimates for the model parameters

Meter analysis

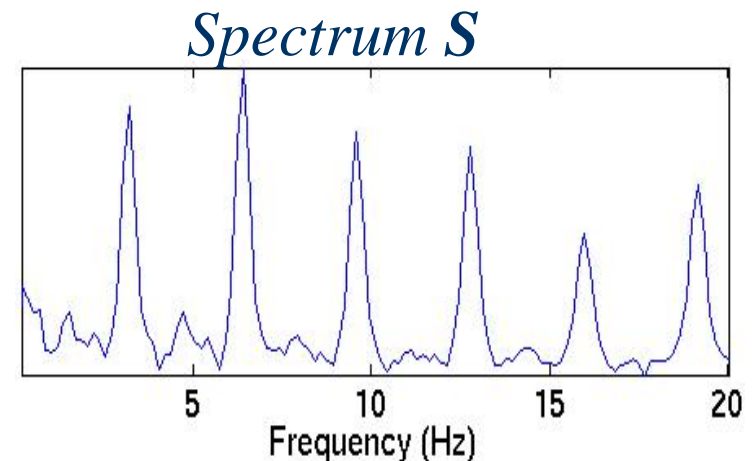
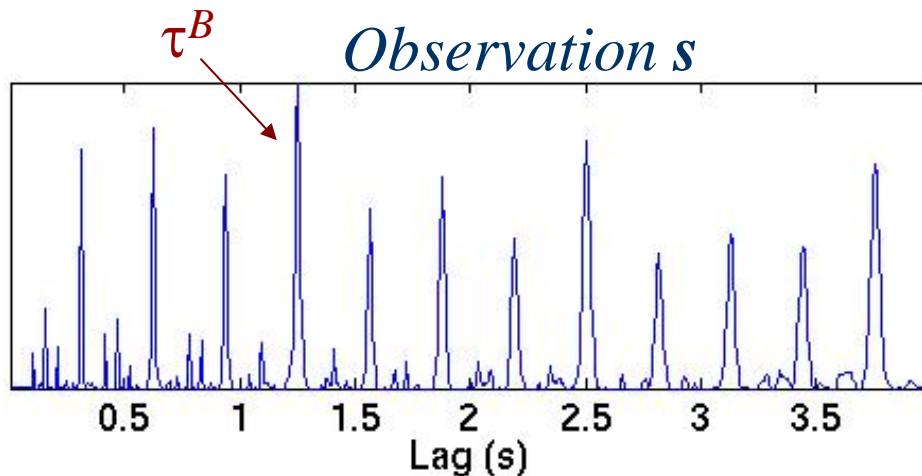
Observation likelihoods

- By making a series of assumptions we arrive at the following approximation for $P(s_n | \mathbf{q}_n)$:

$$P(s_n | \mathbf{q}_n=[j, k, l]) = s_n(k) s_n(l) \underbrace{S_n(1/j)}$$

where $S_n(1/\tau)$ is the *Fourier transform* of $s_n(\tau)$

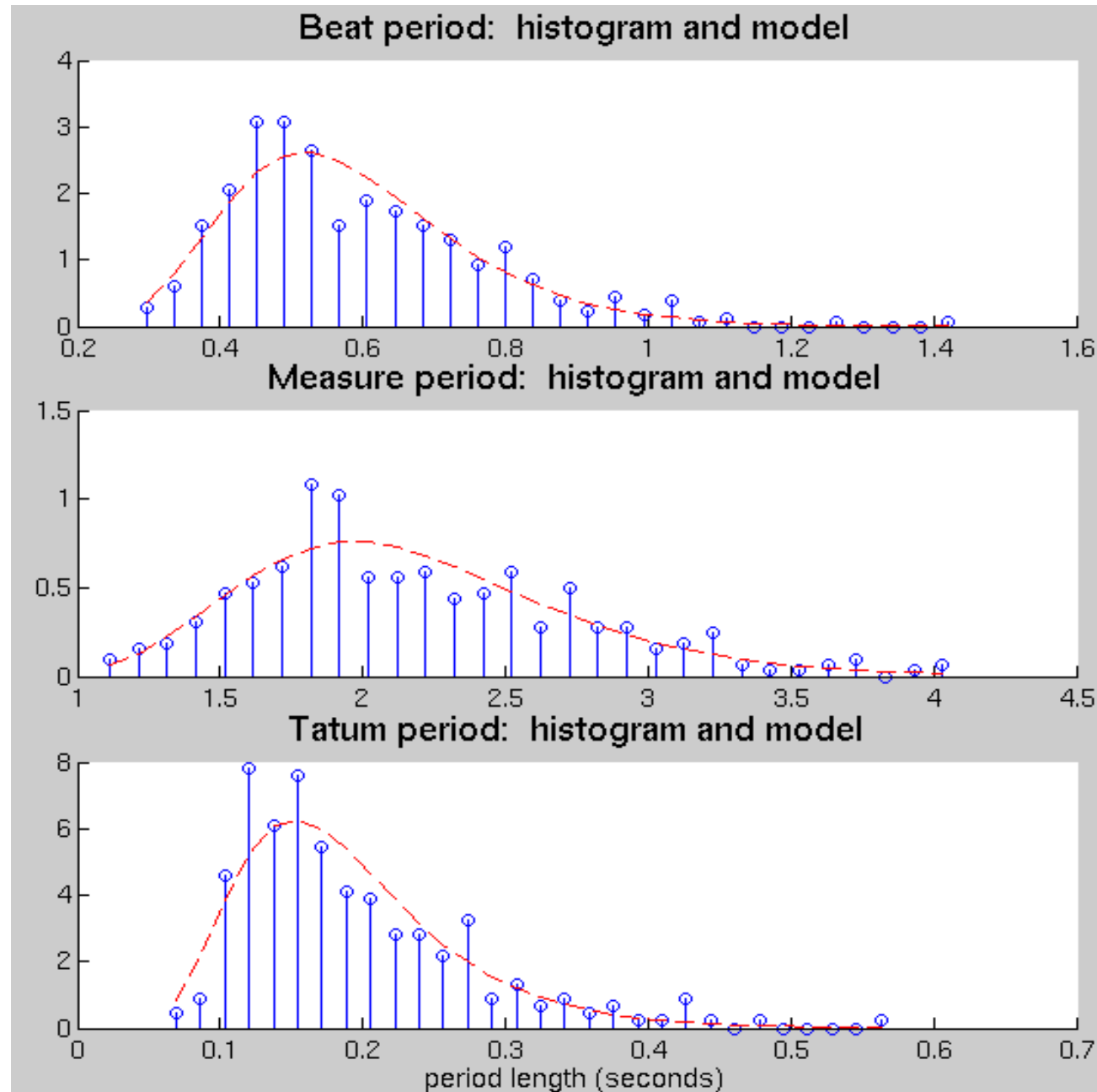
- By definition, other pulse periods are integer multiples of the tatum period \rightarrow overall $s(\tau, n)$ gives information about tatum



Meter analysis

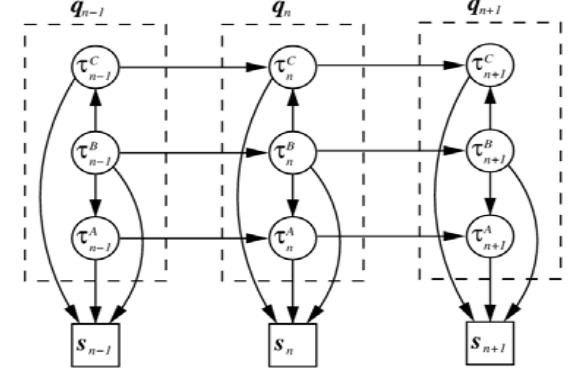
Priors

- Period priors: 2-parameter log-normal distribution
 - suggested by Parncutt (1994)



Meter analysis

Prior probability for meter



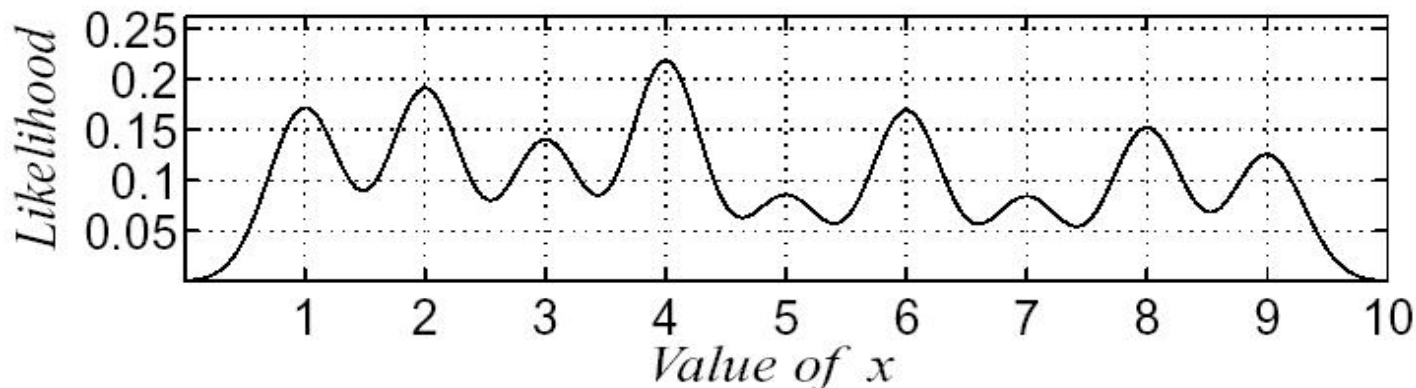
- Assumptions of our model can be written as

$$P(q_n | q_{n-1}) = P(\tau_n^B | \tau_{n-1}^B) P(\tau_n^A | \tau_n^B, \tau_{n-1}^A) P(\tau_n^C | \tau_n^B, \tau_{n-1}^C)$$

- Conditional probabilities are presented as a product

$$P(\tau_n^C | \tau_n^B, \tau_{n-1}^C) = P(\tau_n^C | \tau_{n-1}^C) \frac{P(\tau_n^C, \tau_n^B | \tau_{n-1}^C)}{P(\tau_n^C | \tau_{n-1}^C) P(\tau_n^B | \tau_{n-1}^C)} \approx P(\tau_n^C) f\left(\frac{\tau_n^C}{\tau_n^B}\right)$$

- Gaussian mixture density $f(x)$ models the relation dependencies of simultaneous periods independent of their frequencies of occurrence
 - realizes a *tendency towards binary and ternary integer relations*



Meter analysis

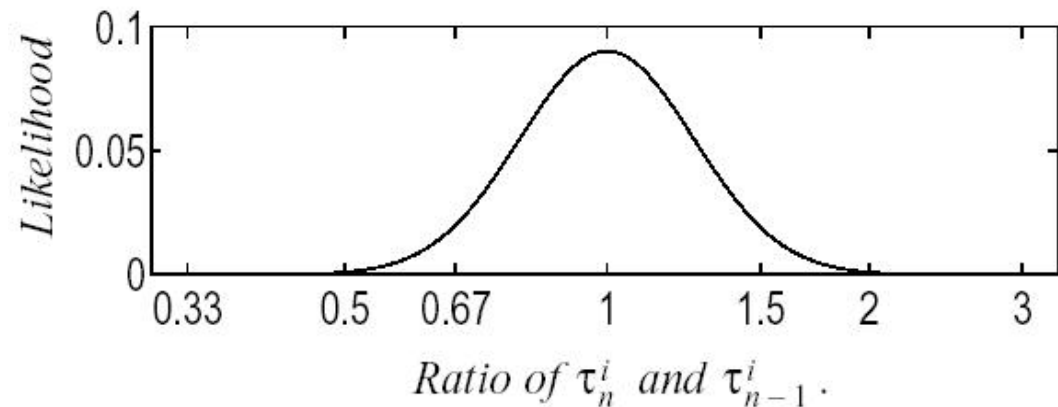
Transition probabilities

- Transition probabilities

- modeled as a product of the *a priori* probability for a certain period and a term which describes tendency to remain in a certain period

$$P(\tau_n | \tau_{n-1}) = P(\tau_n) \frac{P(\tau_n, \tau_{n-1})}{P(\tau_n)P(\tau_{n-1})} \approx P(\tau_n) g\left(\frac{\tau_n}{\tau_{n-1}}\right)$$

- Function g implements Normal distribution as a function of the relative change in the period



$$g\left(\frac{\tau_n}{\tau_{n-1}}\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2} \left[\ln\left(\frac{\tau_n}{\tau_{n-1}}\right)\right]^2\right\}$$

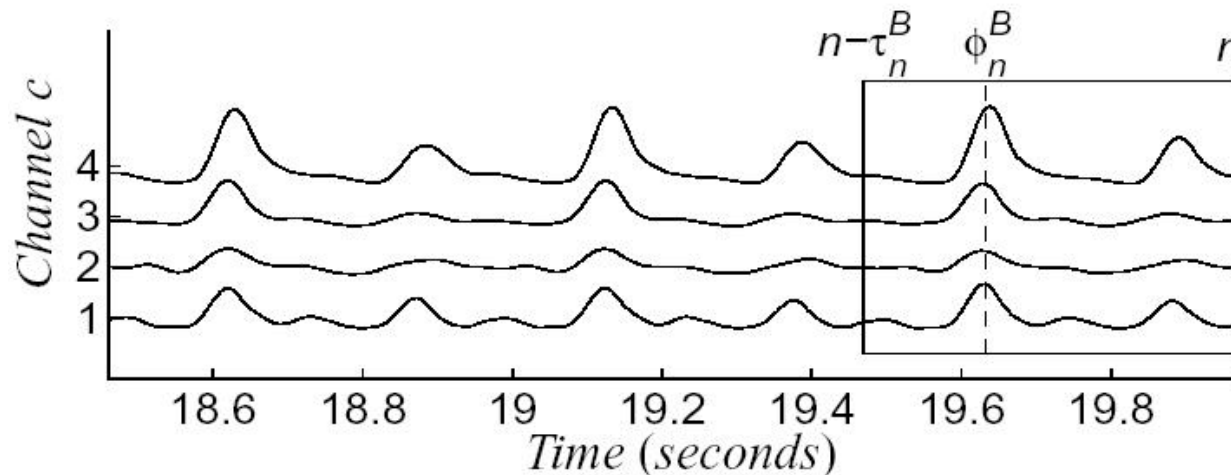
Meter analysis

Finding state sequence

- The most likely sequence of meter estimates can be found using Viterbi algorithm
 - *causal algorithm*: meter estimate at time n is determined according to the end-state of the best partial path at that time
 - *noncausal* meter estimates after seeing a complete sequence of observations can be computed using backward decoding

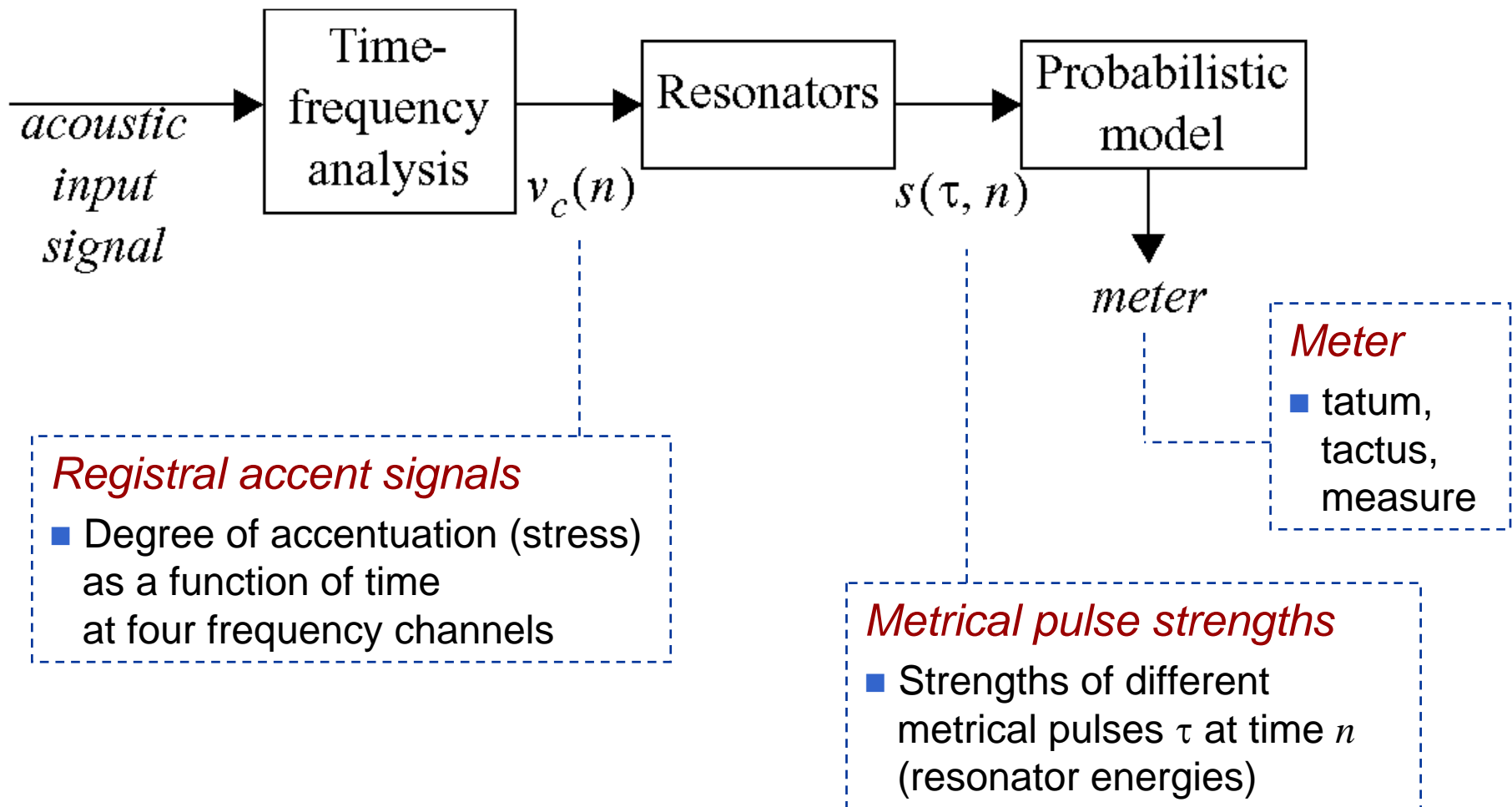
Phase estimation (beat locations)

- The model finds periods first and then the phases for the different levels of the meter
 - *phase estimation* is based on the last τ outputs of the resonator of the winning period (i.e., the filter state)
 - probabilistic modeling for phase is very similar to that for period estimation, but is estimated separately for beat/measure/tatum



Meter analysis

Summary



Demonstrations

- <http://www.cs.tut.fi/~klap/iiro/meter/>