Periodicity detection

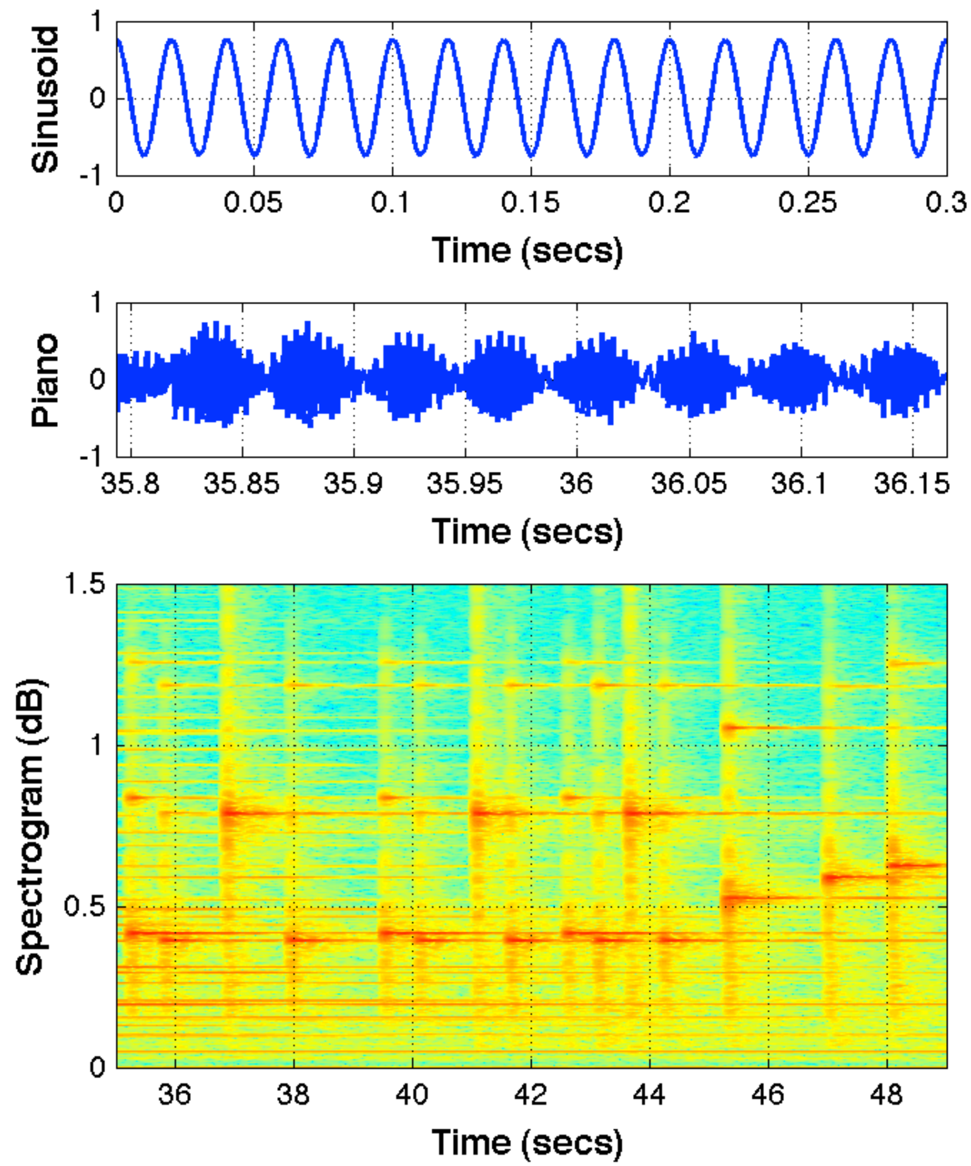
Juan Pablo Bello
EL9173 Selected Topics in Signal Processing: Audio Content Analysis
NYU Poly

Periodicity detection

- Formally, a periodic signal is defined as:

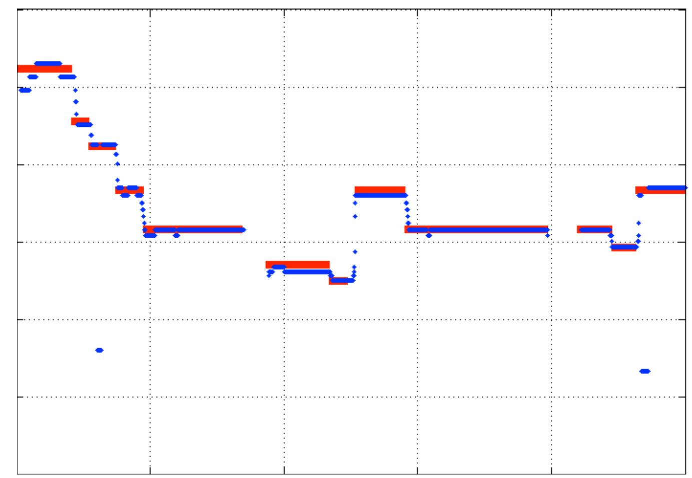
$$x(t) = x(t + T_0), \forall t$$

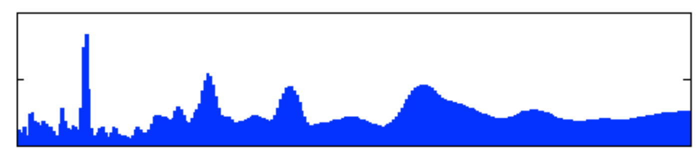
- Detect the fundamental period/frequency (and phase)

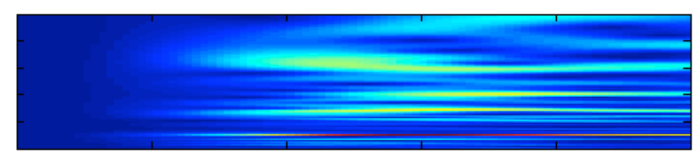


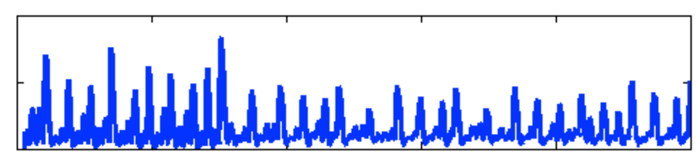
Applications

- At short (pitch) and long (rhythm) time scales:
    - pitch-synchronous analysis
    - voice/sound identification
    - prosodic analysis
    - bioacoustics
    - melodic analysis
    - note transcription
    - beat tracking, segmentation



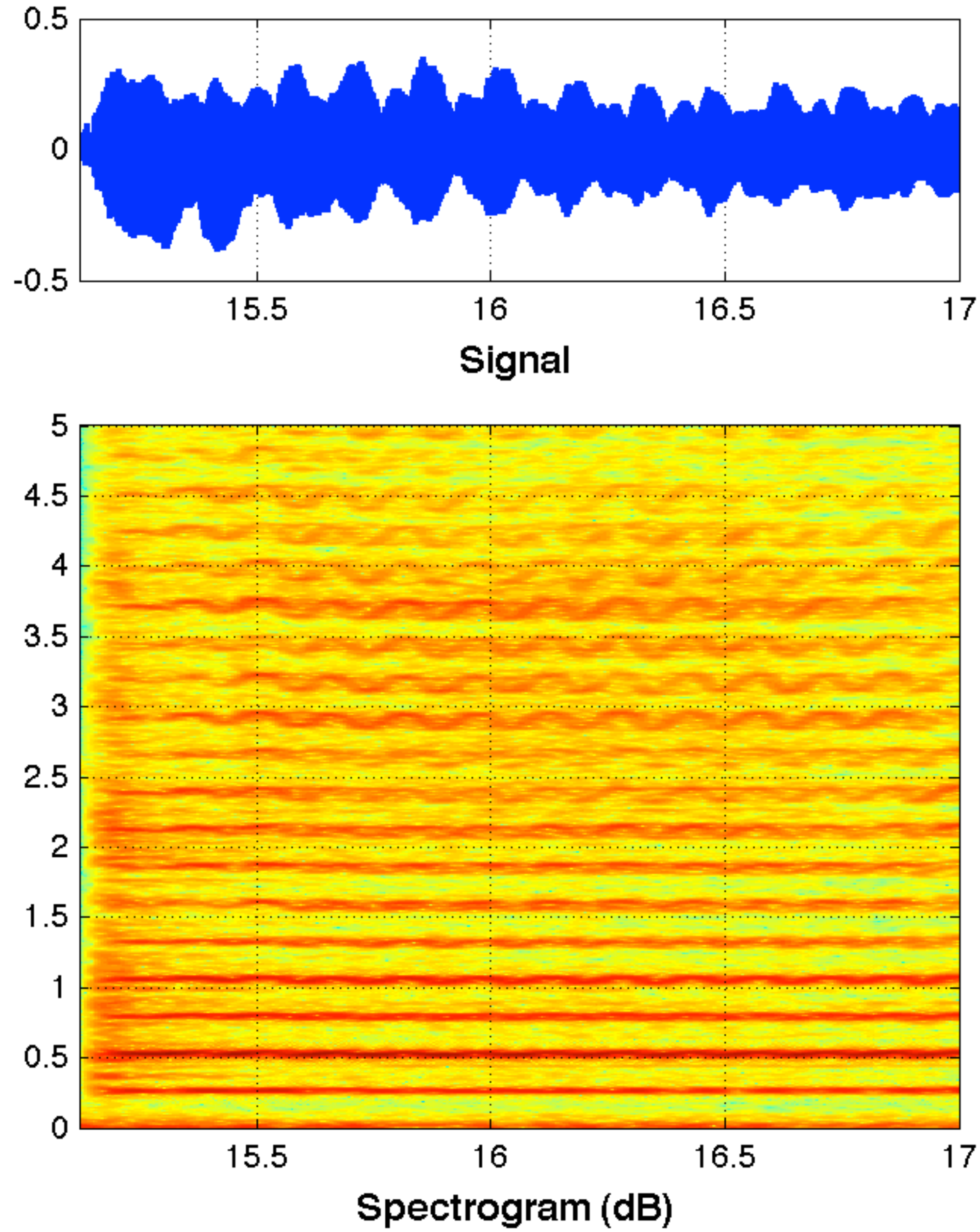






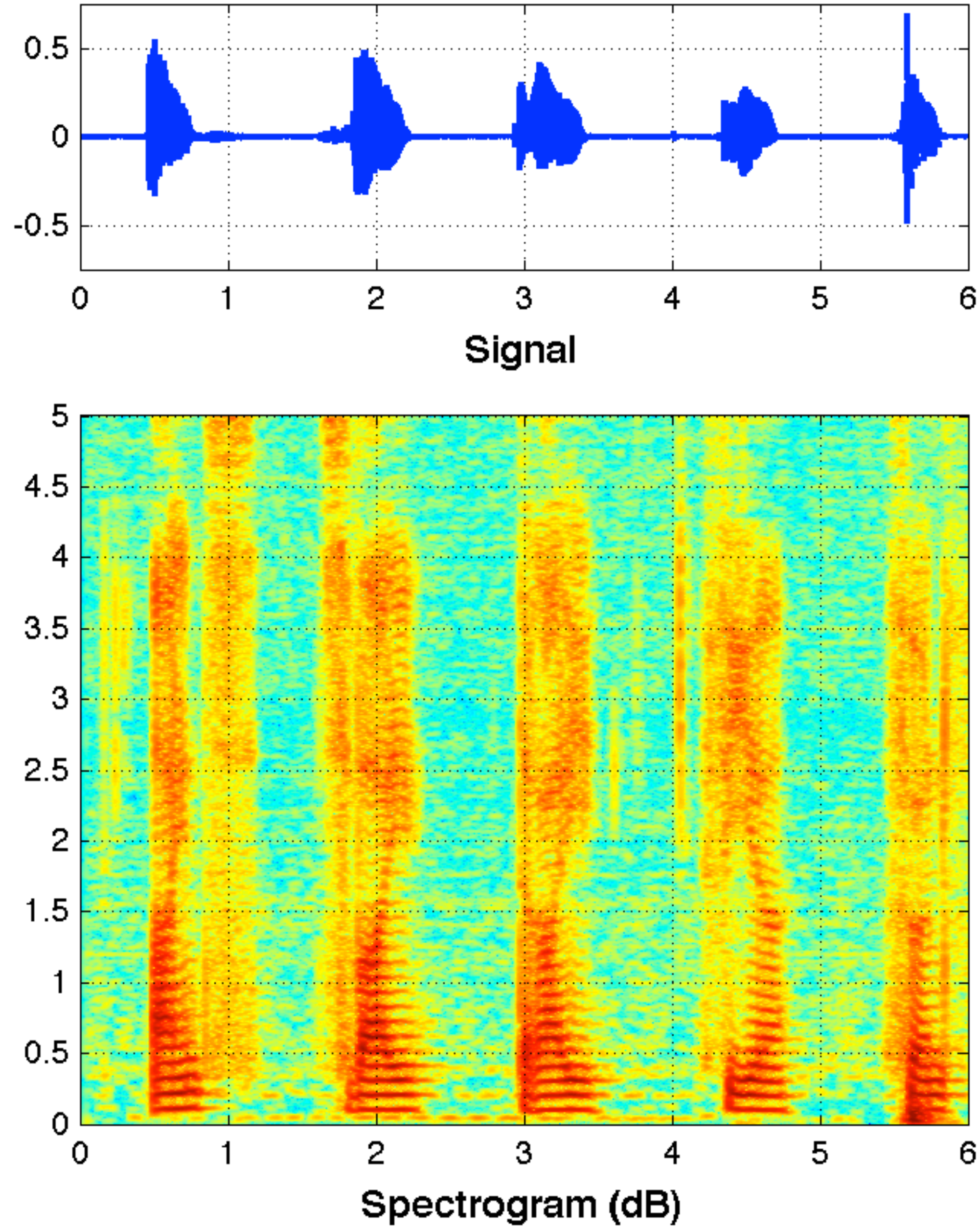
Difficulties

- Quasi-periodicities, temporal variations
- Multiple periodicities associated with f0
- transients and noise
- Polyphonies: information overlap



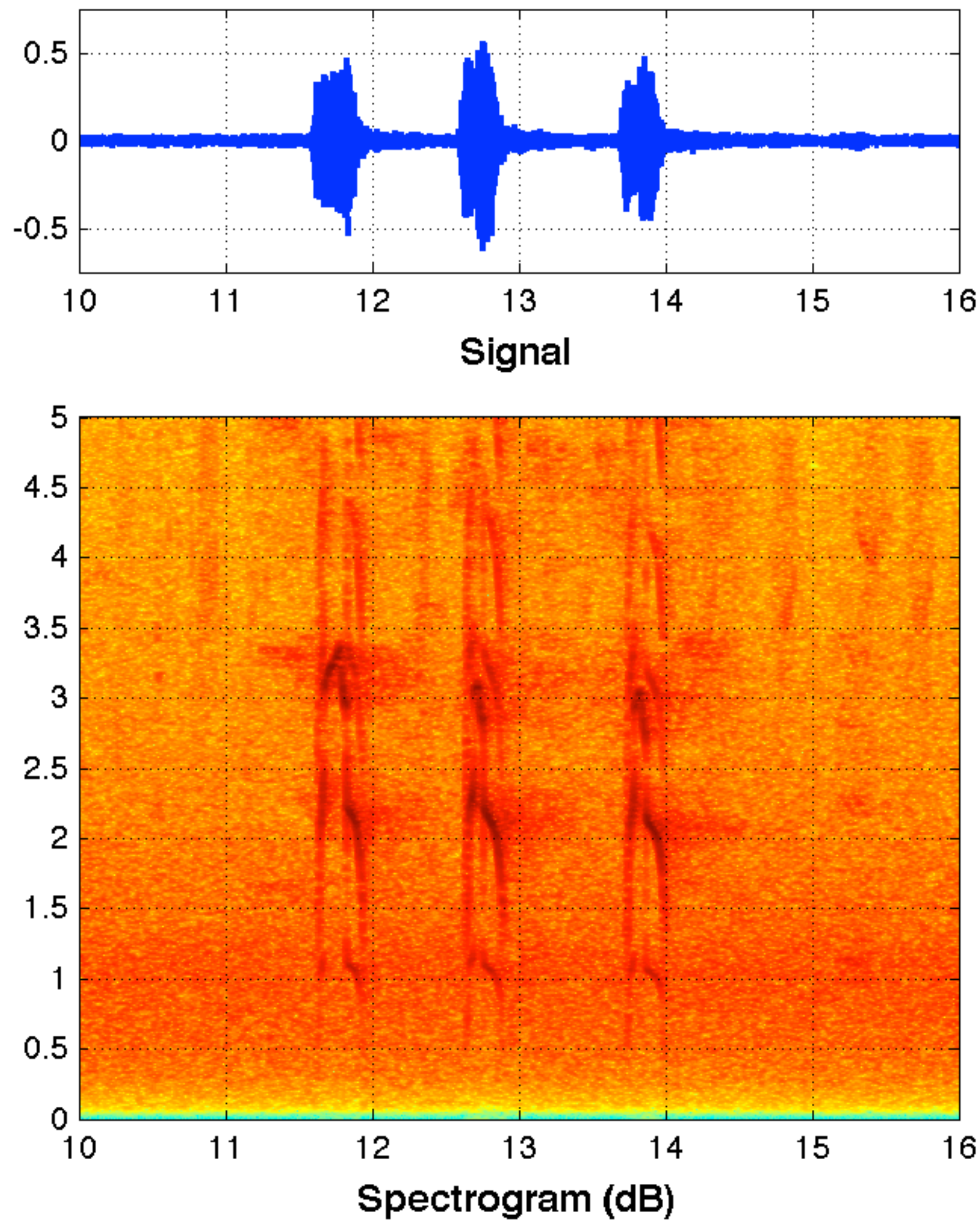
Difficulties

- Quasi-periodicities, temporal variations
- Multiple periodicities associated with f0
- transients and noise
- Polyphonies: information overlap



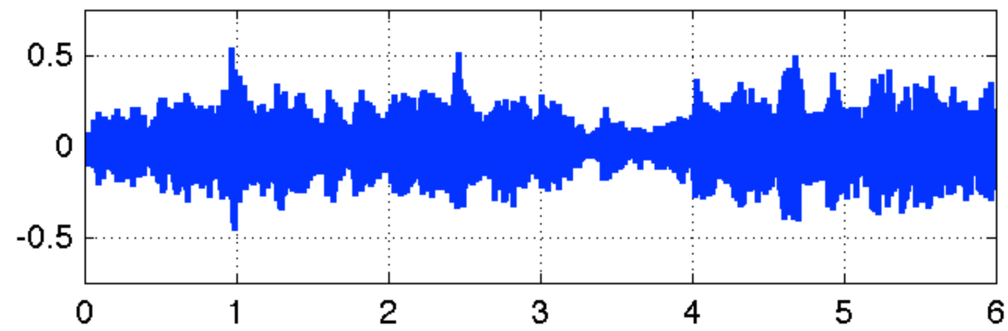
Difficulties

- Quasi-periodicities, temporal variations
- Multiple periodicities associated with f0
- transients and noise
- Polyphonies: information overlap

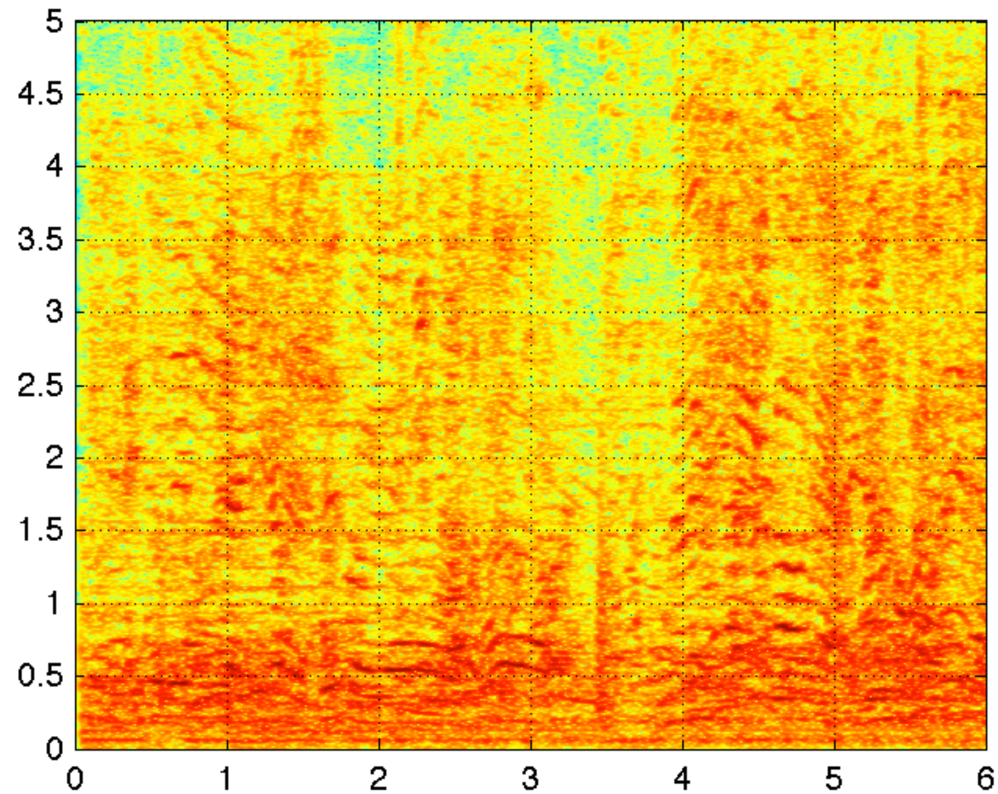


Difficulties

- Quasi-periodicities, temporal variations
- Multiple periodicities associated with f0
- transients and noise
- Polyphonies: information overlap



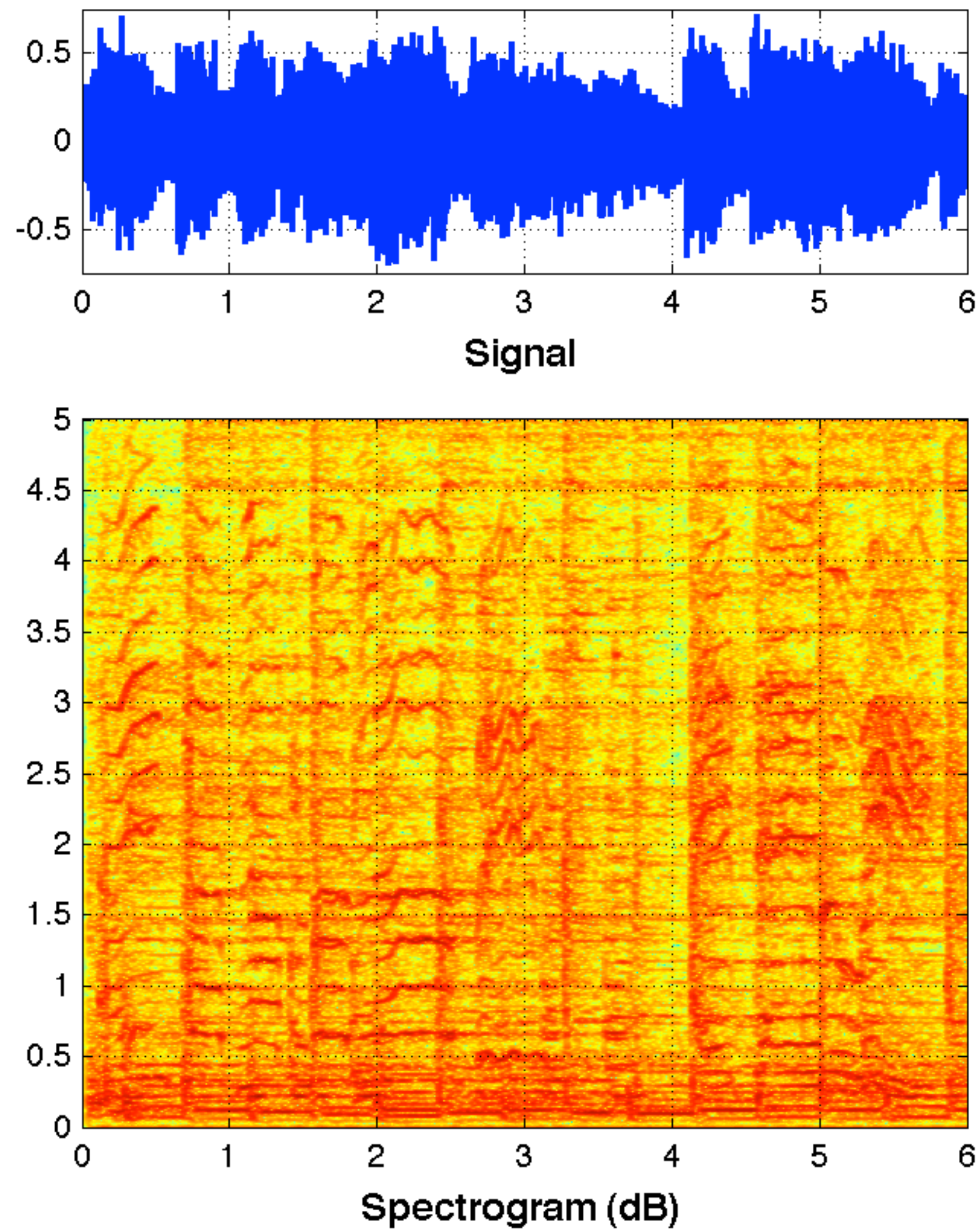
Signal



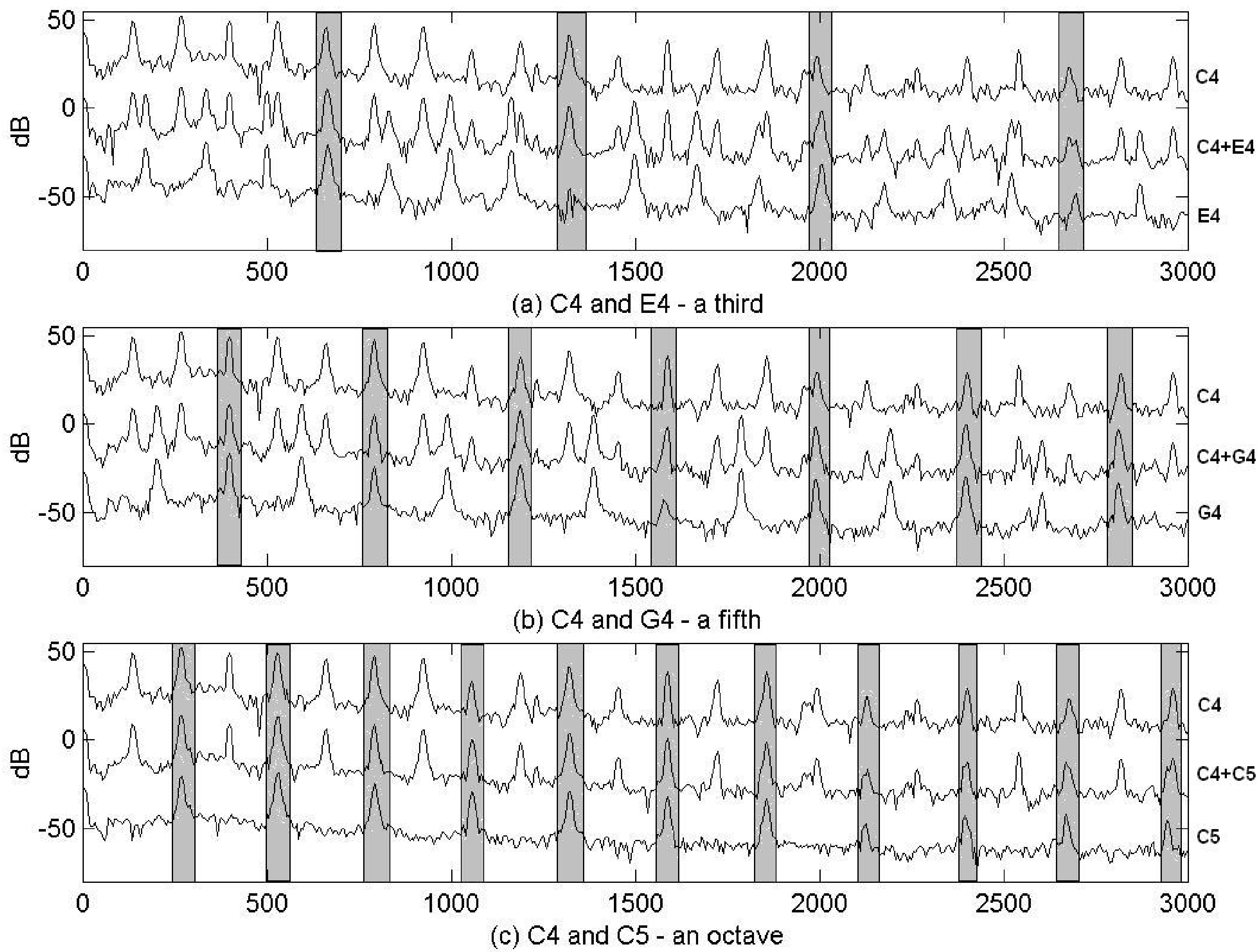
Spectrogram (dB)

Difficulties

- Quasi-periodicities, temporal variations
- Multiple periodicities associated with f0
- transients and noise
- Polyphonies: information overlap

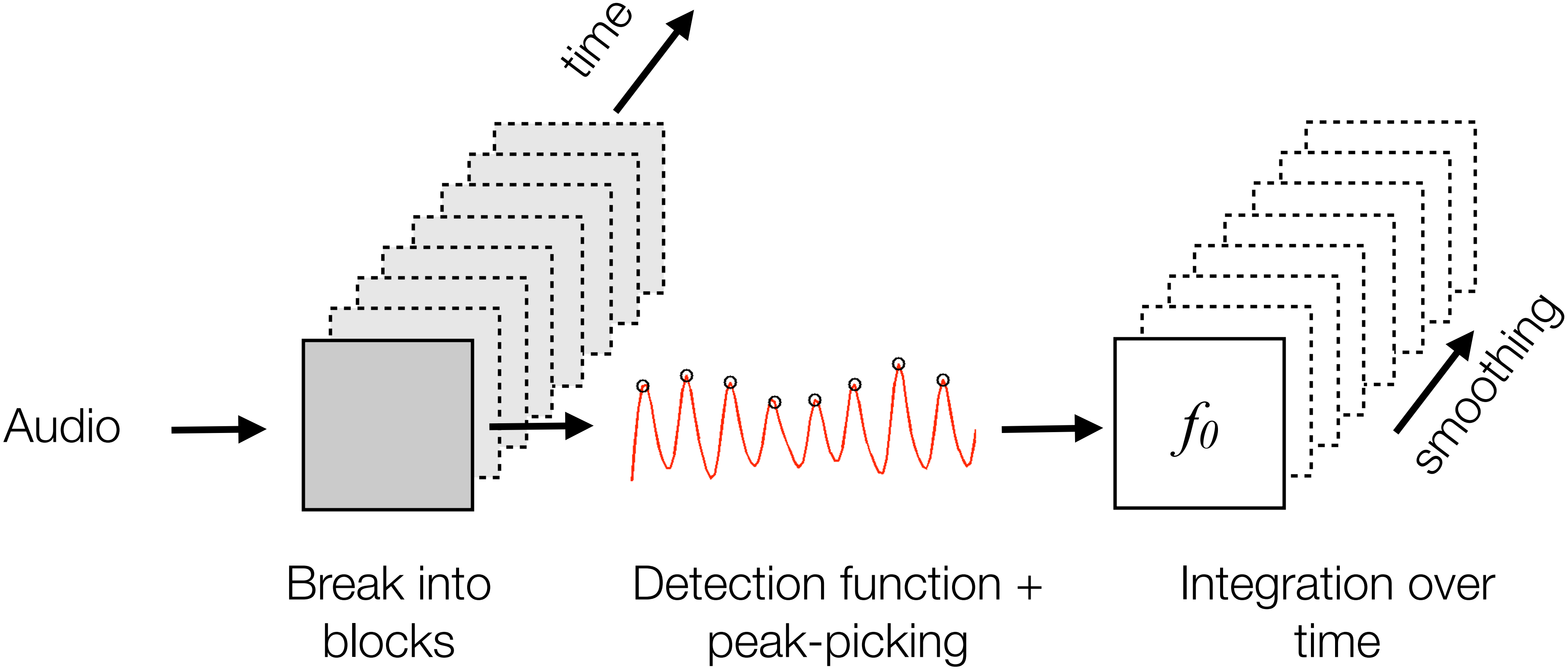


Overlap



50
0
-50
dB
C4
C4+E4
E4
0
500
1000
1500
2000
2500
3000
(a) C4 and E4 - a third
C4
C4+G4
G4
(b) C4 and G4 - a fifth
C4
C4+C5
C5
(c) C4 and C5 - an octave

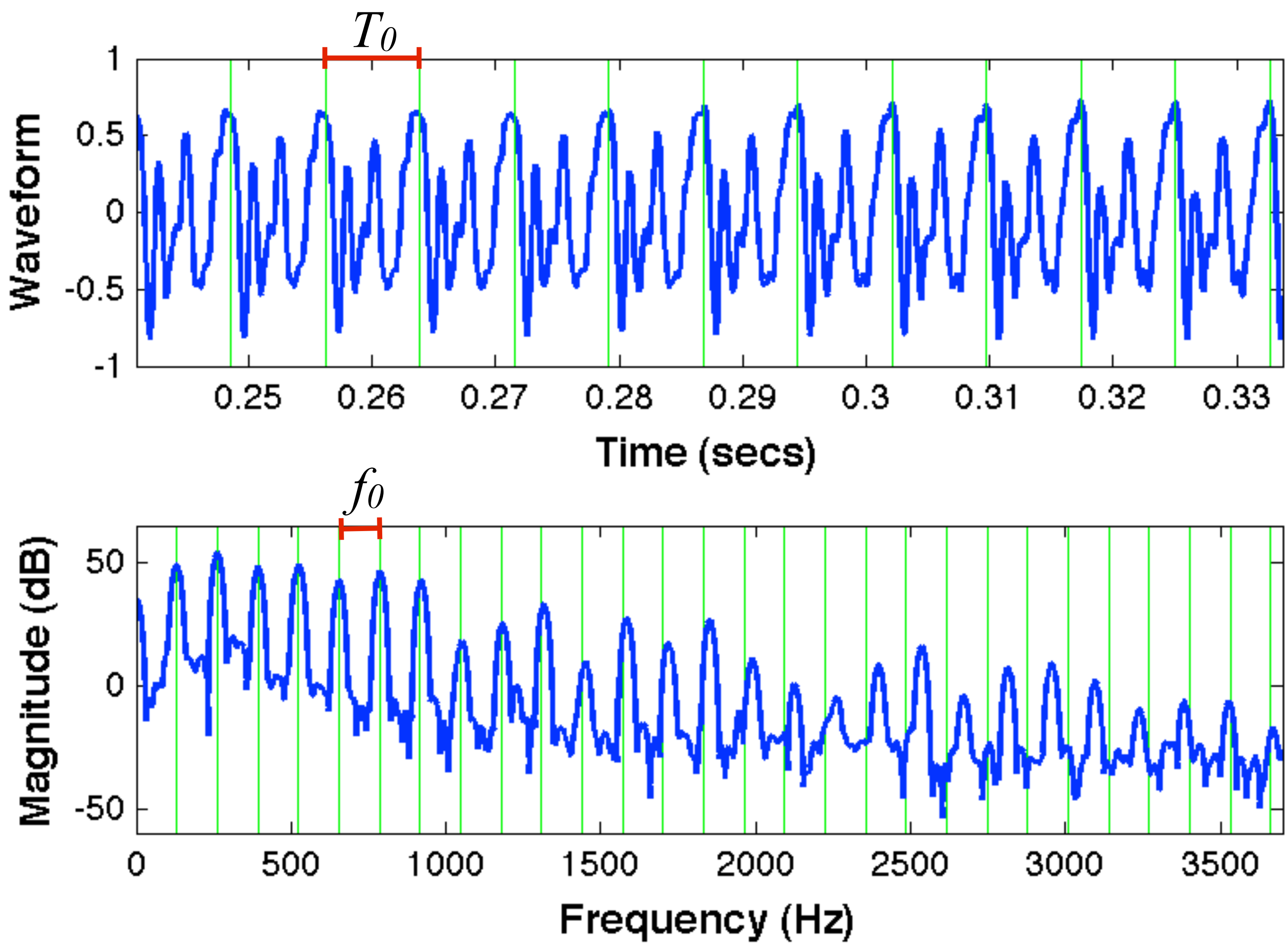
Architecture



Overview of Methods

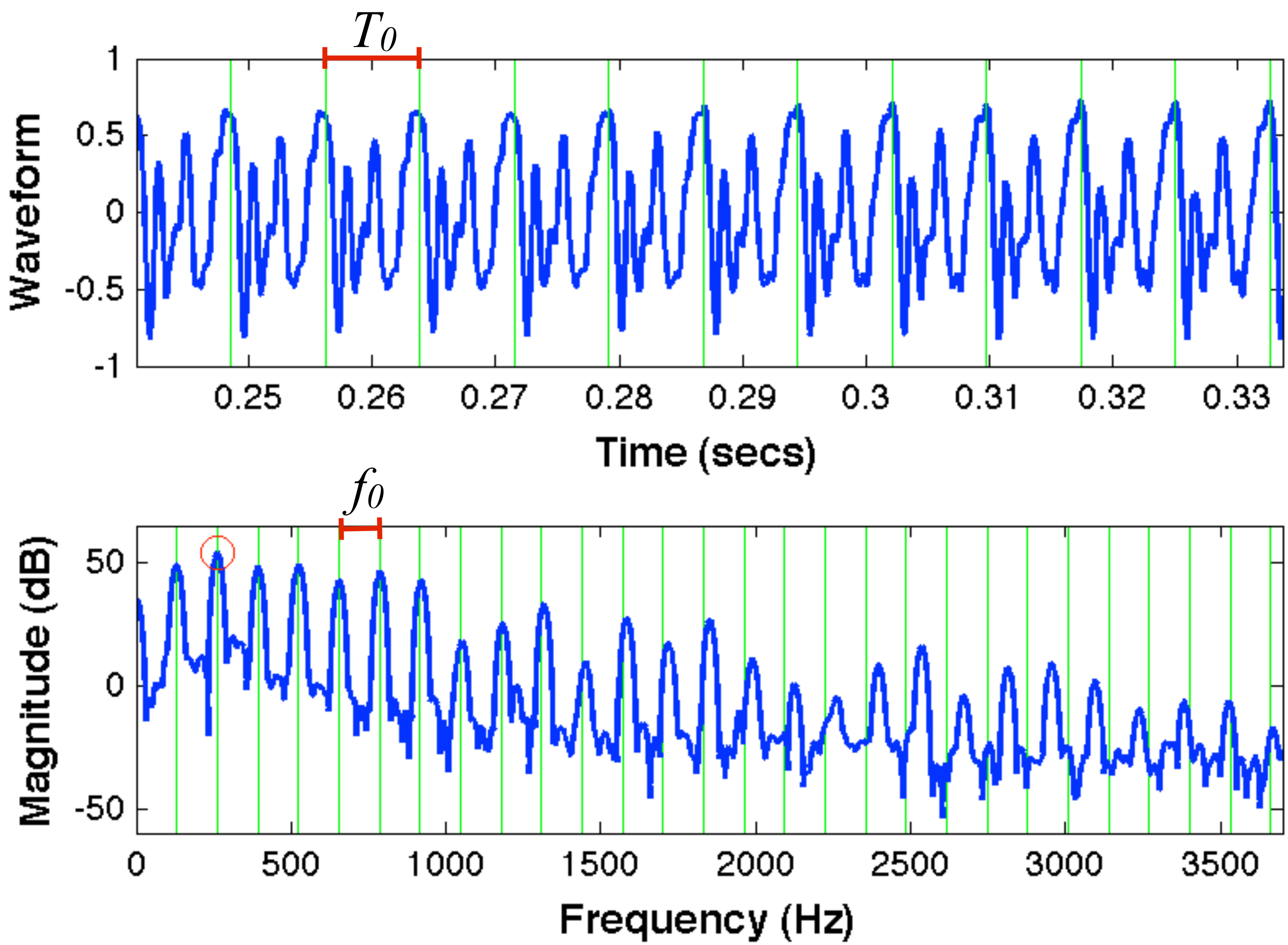
- DFT
- Autocorrelation
- Spectral Pattern Matching
- Cepstrum
- Spectral Autocorrelation
- YIN
- Auditory model

DFT

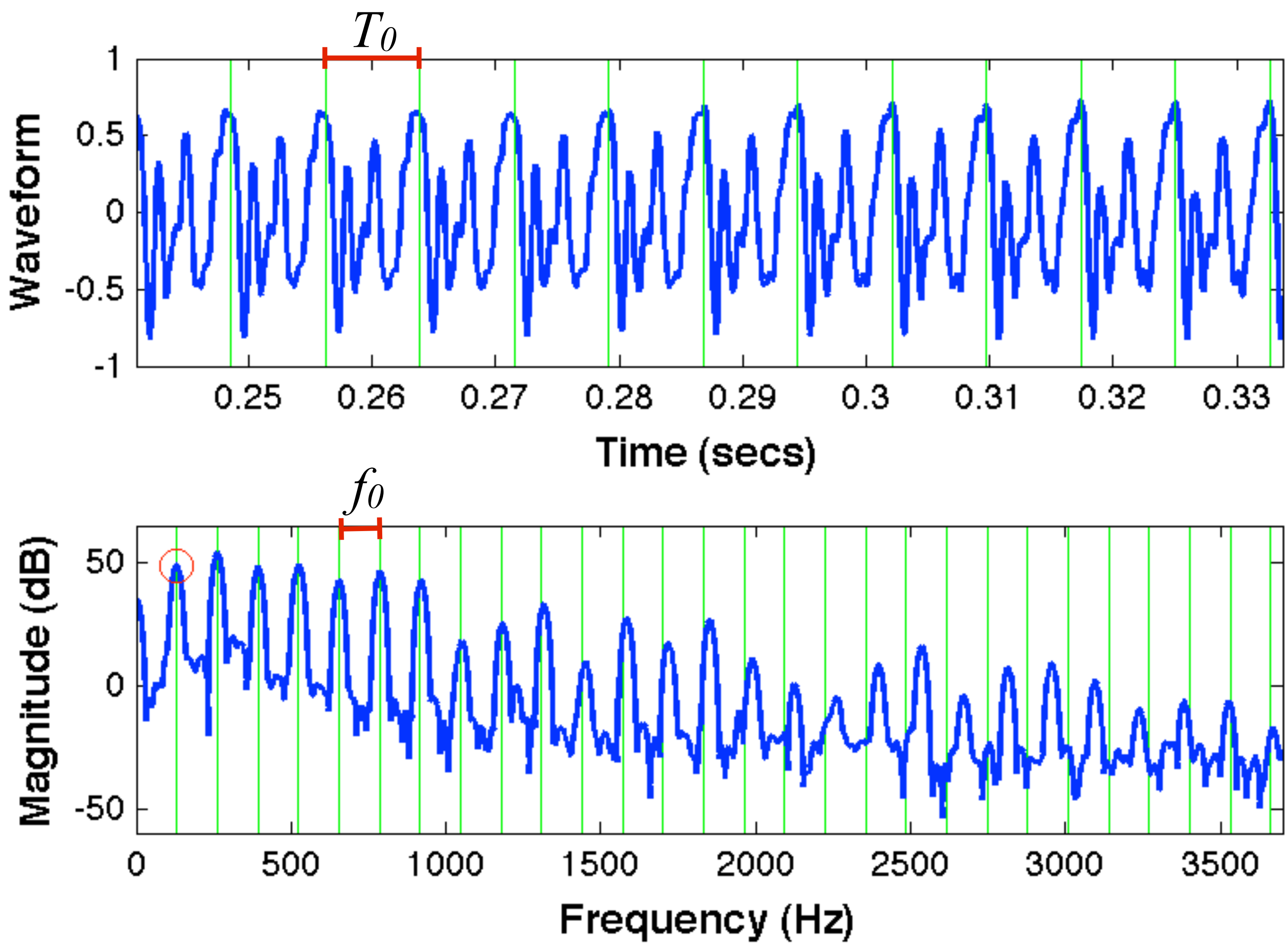


T_0
Waveform
1
0.5
0
-0.5
-1
0.25
0.26
0.27
0.28
0.29
0.3
0.31
0.32
0.33
Time (secs)
f_0
Magnitude (dB)
50
0
-50
0
500
1000
1500
2000
2500
3000
3500
Frequency (Hz)

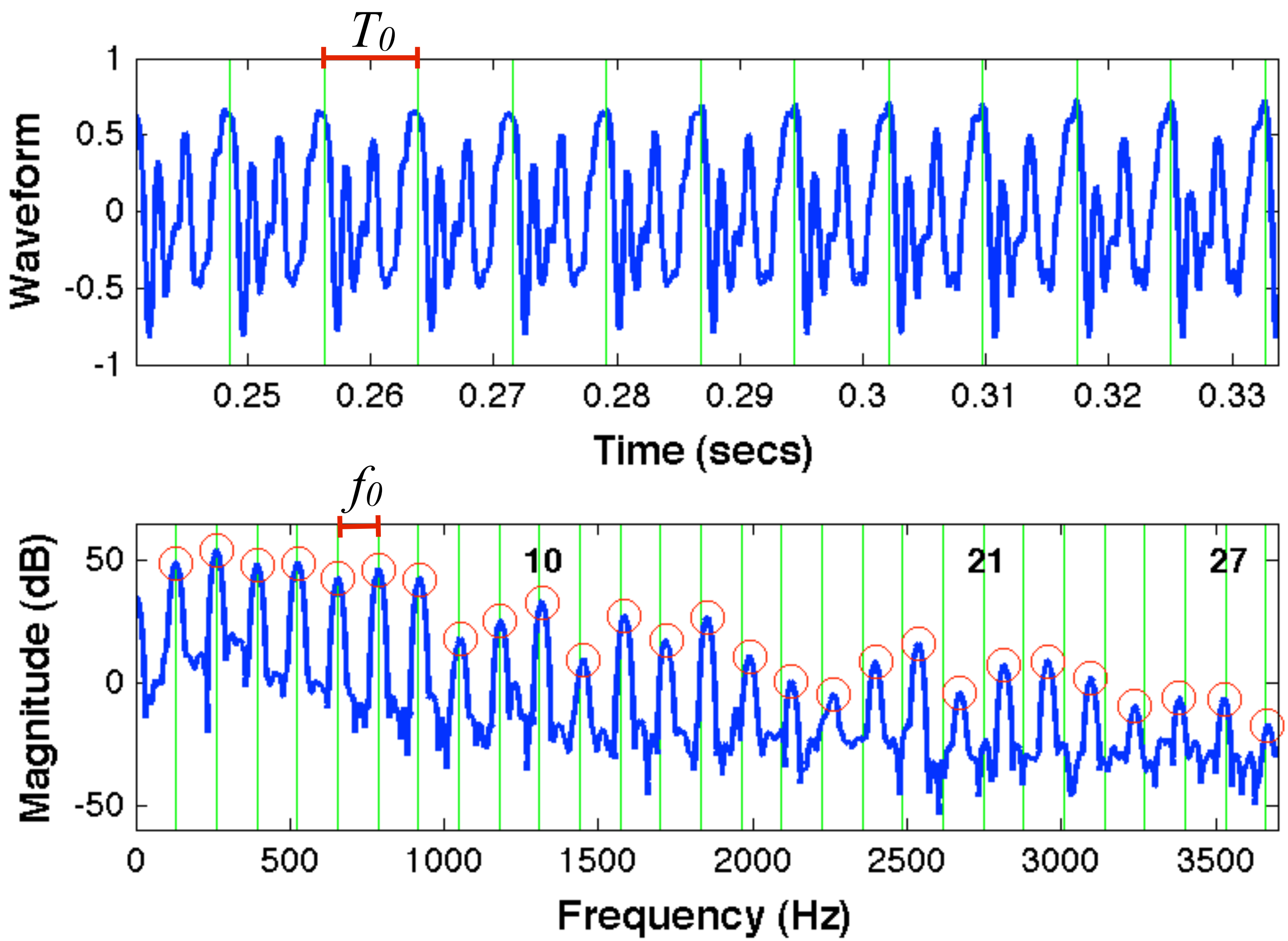
DFT



DFT


T0
1
0.5
0
-0.5
-1
Waveform
0.25
0.26
0.27
0.28
0.29
0.3
0.31
0.32
0.33
Time (secs)
f0
50
0
-50
Magnitude (dB)
0
500
1000
1500
2000
2500
3000
3500
Frequency (Hz)

DFT



T_0
Waveform
1
0.5
0
-0.5
-1
0.25
0.26
0.27
0.28
0.29
0.3
0.31
0.32
0.33
Time (secs)
f_0
Magnitude (dB)
50
0
-50
10
21
27
0
500
1000
1500
2000
2500
3000
3500
Frequency (Hz)

Autocorrelation

- Cross-product measures similarity across time
- Cross-correlation of two real-valued signals x and y:

$$r_{xy}(l) = \frac{1}{N} \sum_{n=0}^{N-1} x(n)y(n+l) \quad \leftarrow \text{modulo } N$$

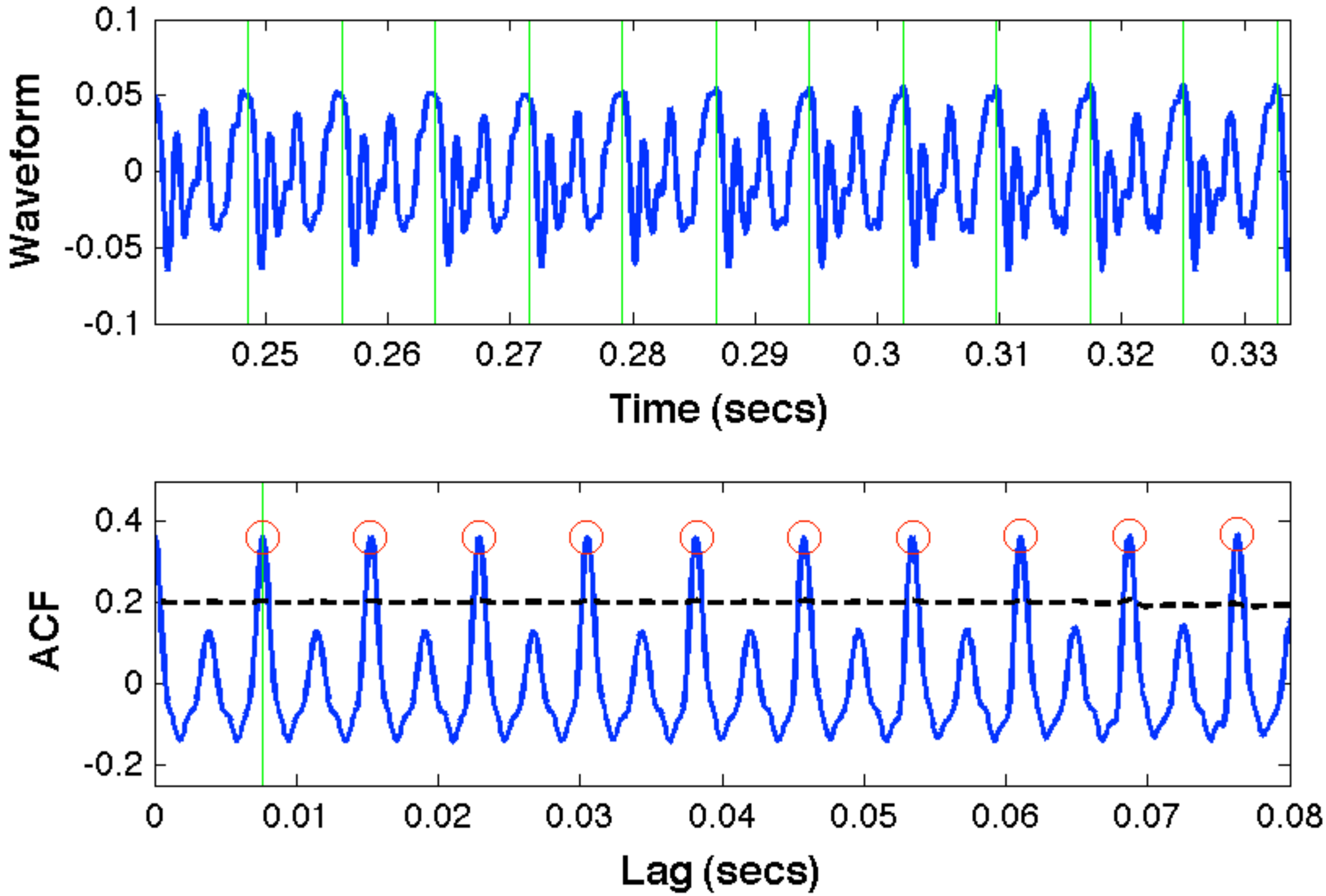
$$l = 0, 1, 2, \cdots, N-1$$

- Unbiased (short-term) Autocorrelation Function (ACF):

$$r_x(l) = \frac{1}{N-l} \sum_{n=0}^{N-1-l} x(n)x(n+l)$$

$$l = 0, 1, 2, \cdots, L-1$$

Autocorrelation



Autocorrelation

- The short-term ACF can also be computed as:

$$r_x(l) = \left(\frac{1}{N-l}\right) real(IFFT(|X|^2))$$

$$X \rightarrow FFT(x)$$

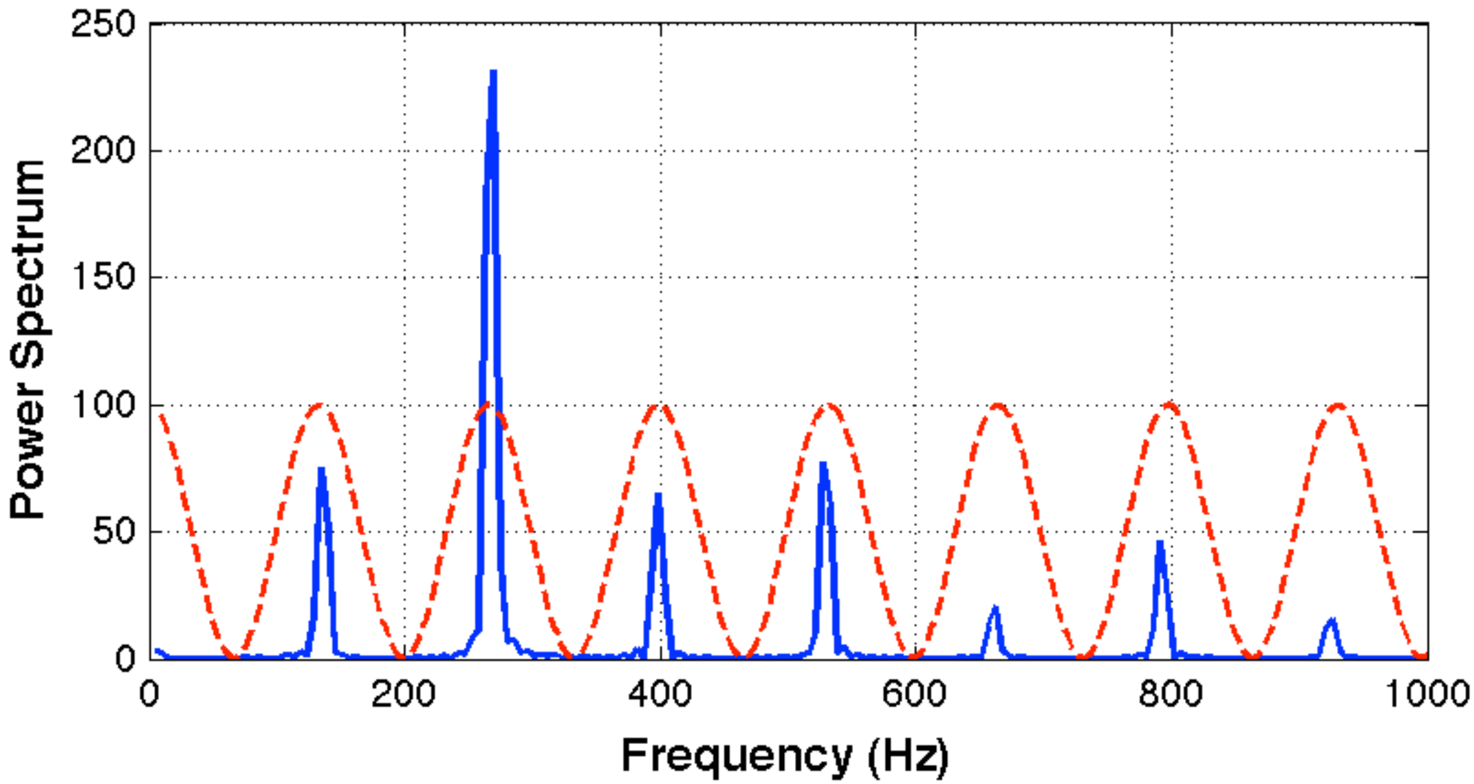
x zero-padded to next power of 2

after $(N+L)-1$

Autocorrelation

- This is equivalent to the following correlation:

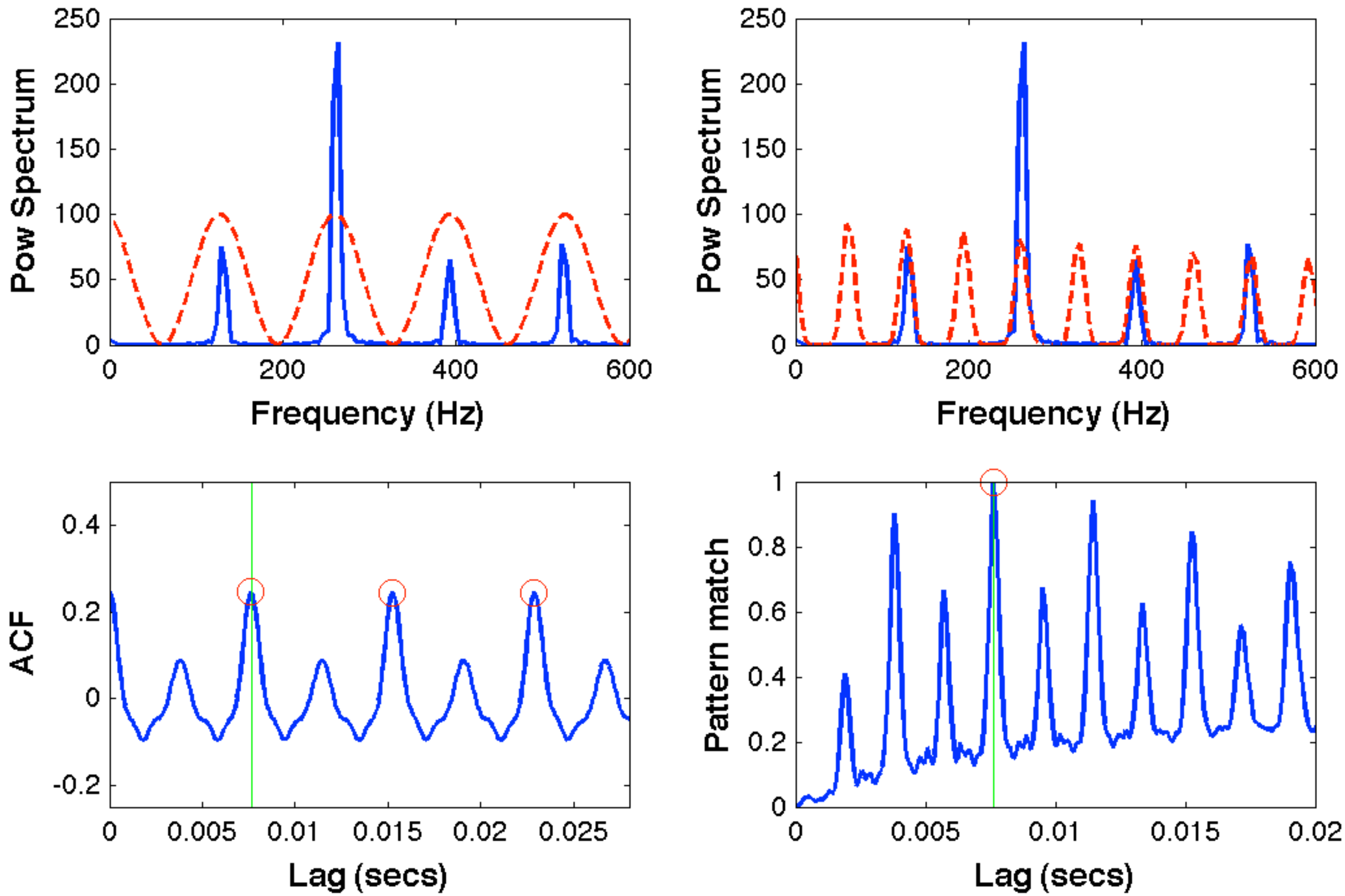
$$r_x(l) = \frac{1}{K-l} \sum_{k=0}^{K-1} cos(2\pi lk/K)|X(k)|^2$$



Pattern Matching

- Comb filtering is a common strategy
- Any other template that realistically fits the magnitude spectrum
- Templates can be specific to sound sources
- Matching strategies vary: correlation, likelihood, distance, etc.

Pattern Matching

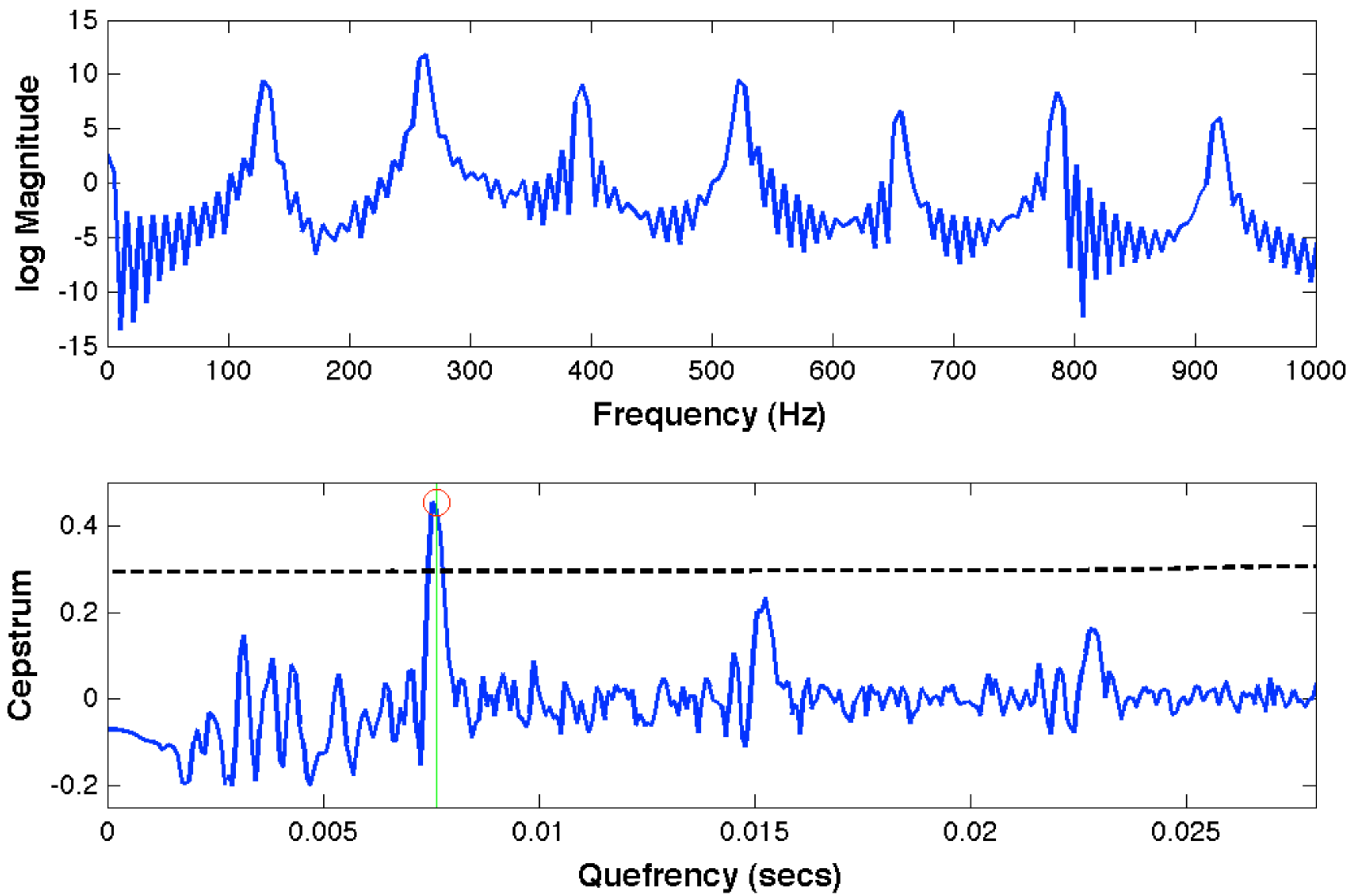


Cepstrum

- Treat the log magnitude spectrum as if it were a signal -> take its (I)DFT
- Measures rate of change across frequency bands (Bogert et al., 1963)
- Cepstrum -> Anagram of Spectrum (same for quefrency, liftering, etc)
- For a real-valued signal is defined as:

$$c_x(l) = real(IFFT(log(|FFT(x)|)))$$

Cepstrum


15
10
5
0
-5
-10
-15
log Magnitude
0
100
200
300
400
500
600
700
800
900
1000
Frequency (Hz)
0.4
0.2
0
-0.2
Cepstrum
0
0.005
0.01
0.015
0.02
0.025
Quefrency (secs)

Spectral ACF

- Spectral location -> sensitive to quasi-periodicities

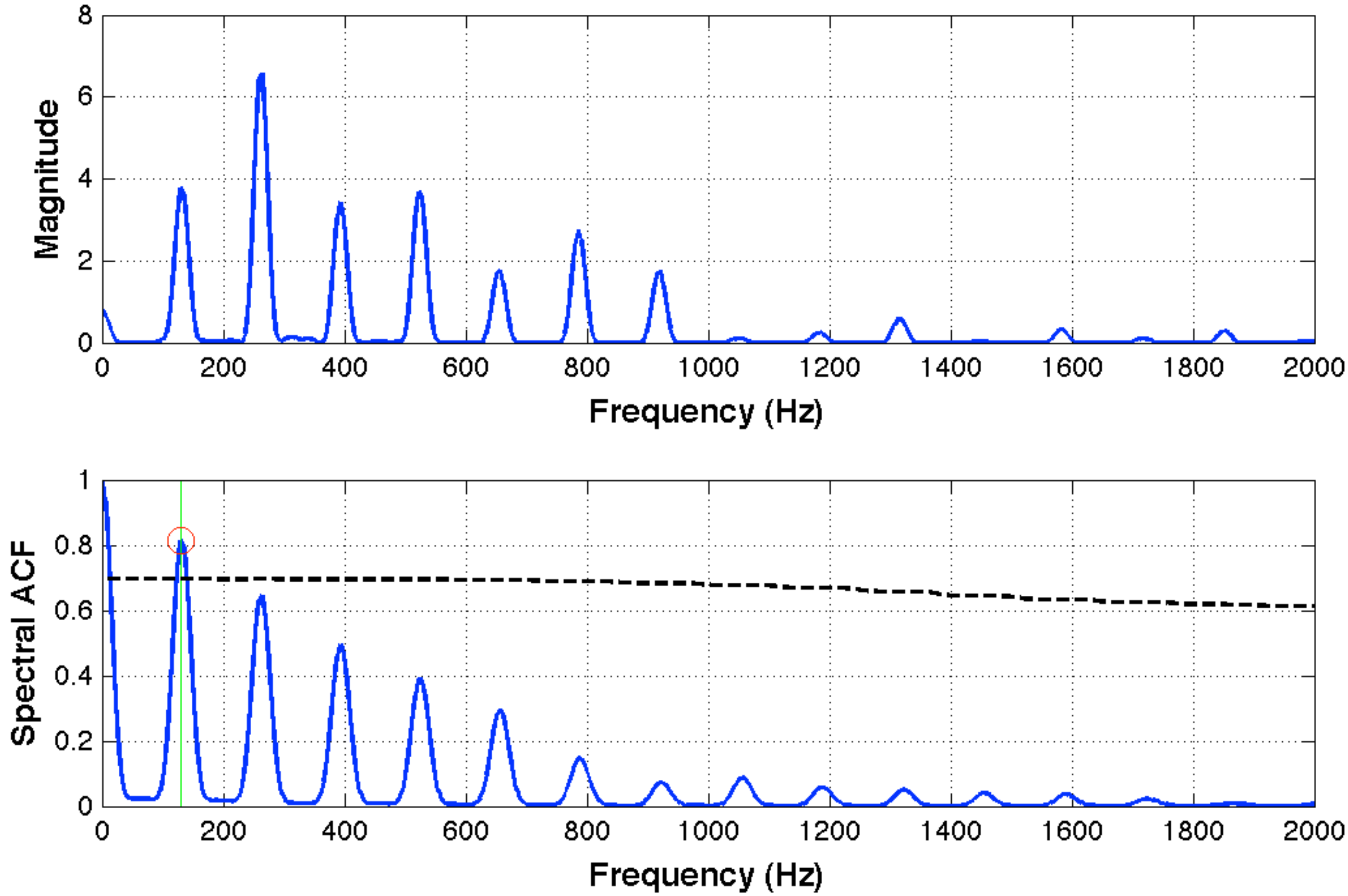
- (Quasi-)Periodic Spectrum, Spectral ACF.

$$r_X(l_f) = \frac{1}{N - l_f} \sum_{k=0}^{N-1-l_f} |X(k)||X(k+l_f)|$$

$$l_f = 0, 1, 2, \cdots, L-1$$

- Exploits intervalic information (more stable than locations of partials), while adding shift-invariance.

Spectral ACF



YIN

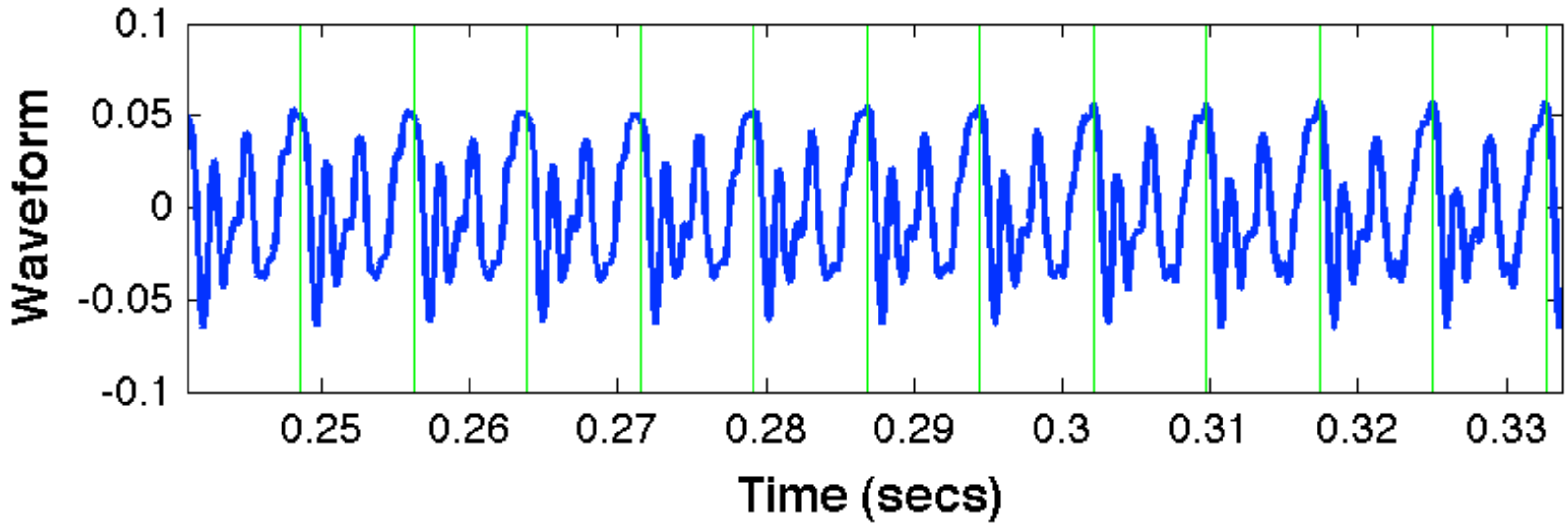
- Alternative to the ACF that uses the squared difference function (deCheveigne, 02):

$$d(l) = \sum_{n=0}^{N-1-l} \left(x(n) - x(n+l)\right)^2$$

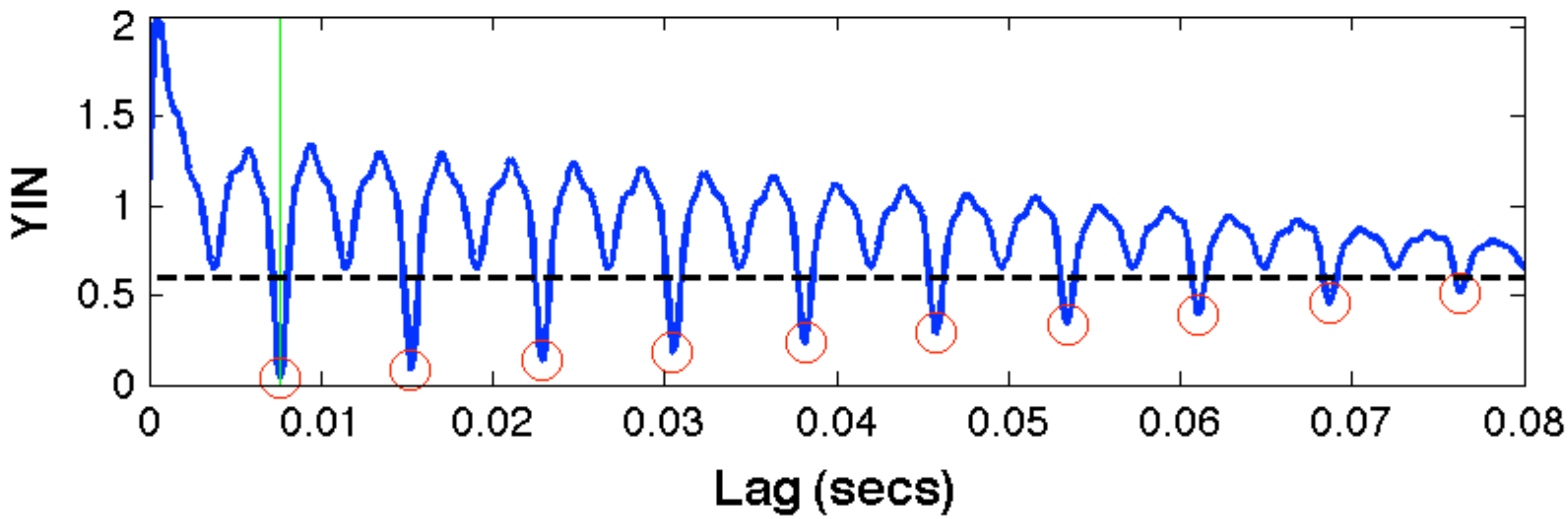
- For (quasi-)periodic signals, this functions cancel itself at $l = 0, l_0$ and its multiples. Zero-lag bias is avoided by normalizing as:

$$\hat{d}(l) = \begin{cases} 1 & l = 0 \\ d(l)/[(1/l)\sum_{u=1}^{l} d(u)] & \text{otherwise} \end{cases}$$

YIN

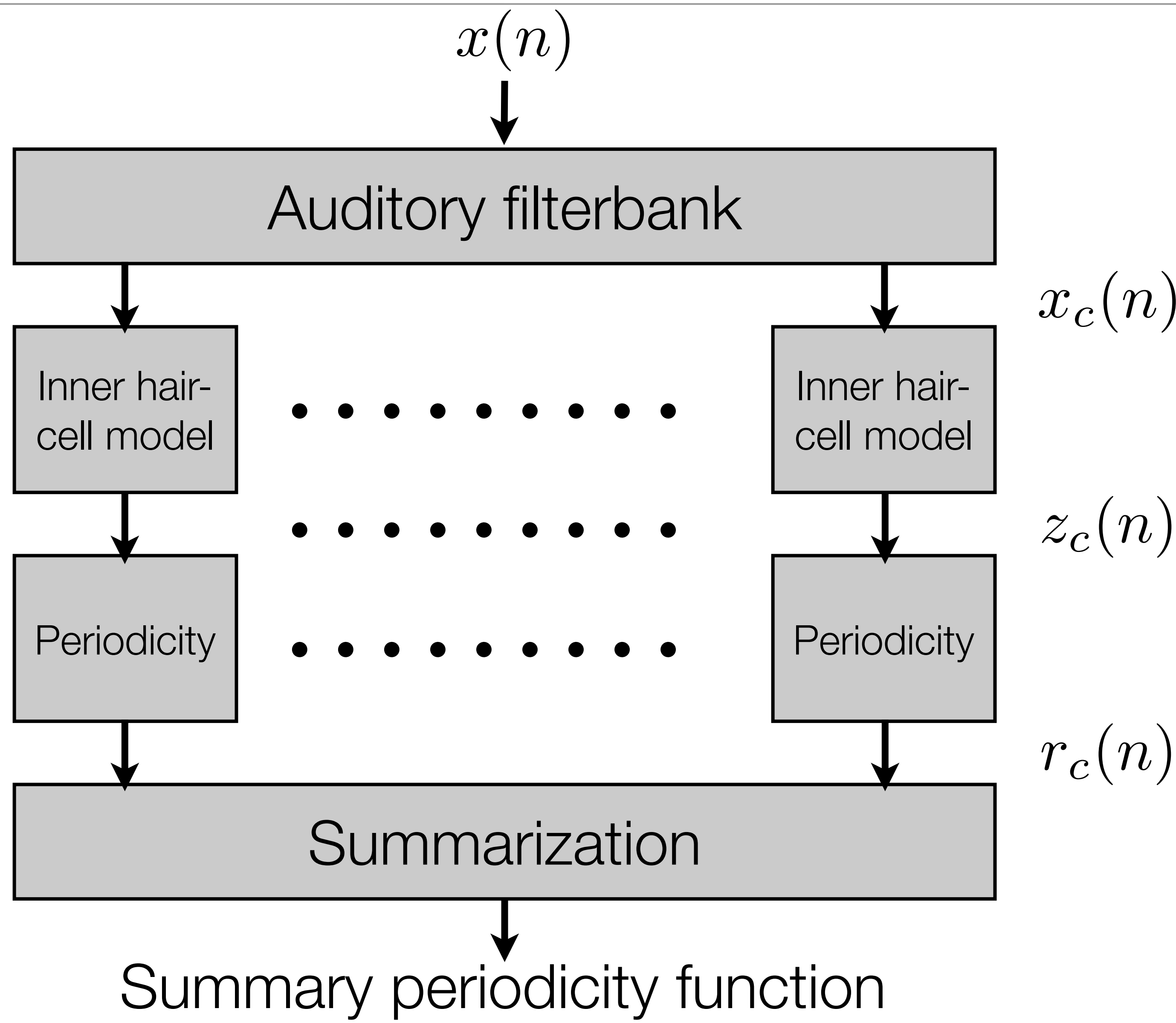


0.1
0.05
0
-0.05
-0.1
Waveform
0.25
0.26
0.27
0.28
0.29
0.3
0.31
0.32
0.33
Time (secs)



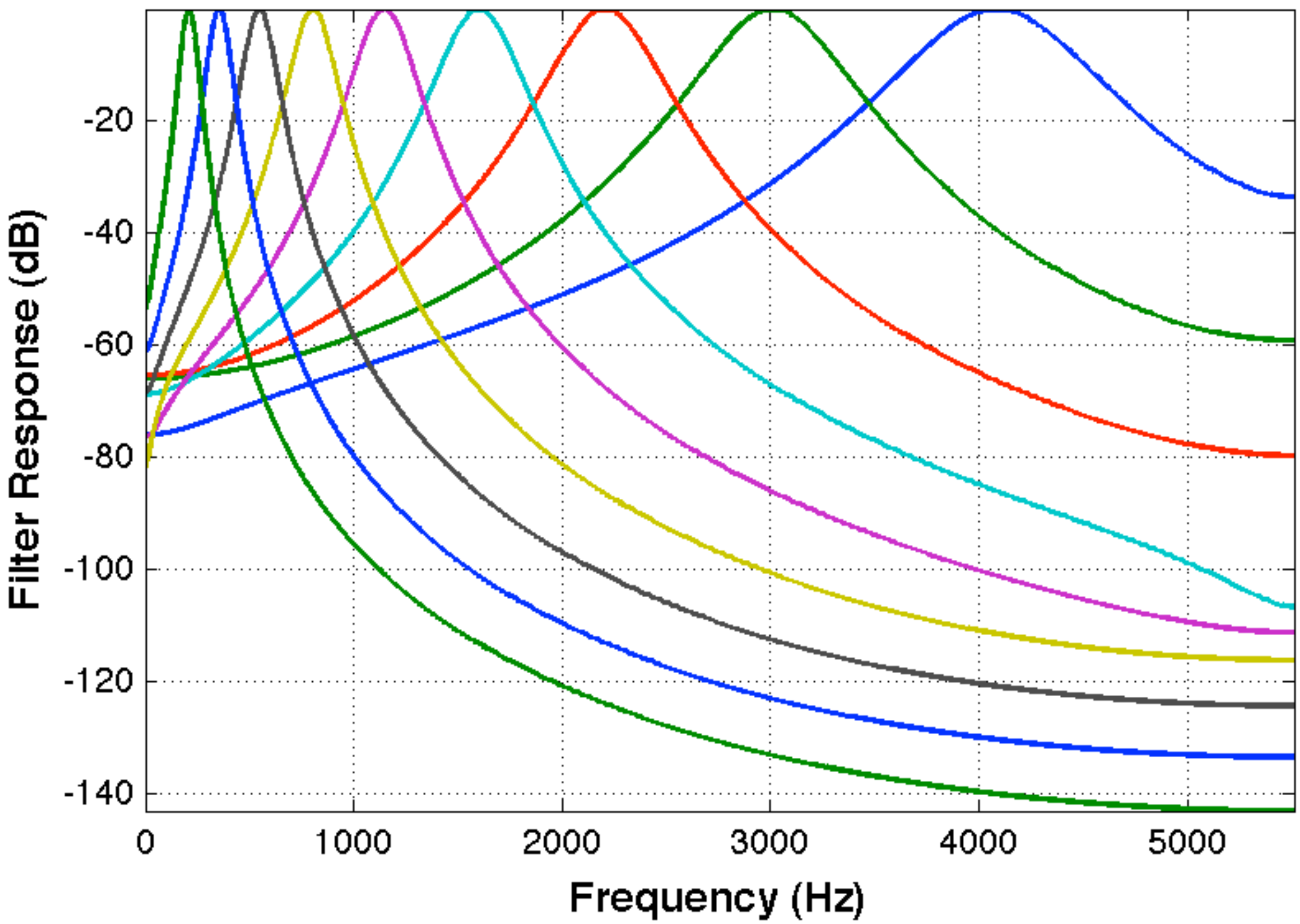
2
1.5
1
0.5
0
YIN
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
Lag (secs)

Auditory model



Auditory model

- Auditory filterbank: gammatone filters (Slaney, 93; Klapuri, 06):



Auditory model

The Equivalent Rectangular Bandwidths (ERB) of the filters:

$$b_c = 0.108 f_c + 24.7$$

$$f_c = 229 \times \left(10^{\psi/21.4} - 1\right)$$

$$\psi = \psi_{min} : (\psi_{max} - \psi_{min})/F : \psi_{max}$$

$$\psi_{min/max} = 21.4 \times \log_{10}(0.00437 f_{min/max} + 1)$$

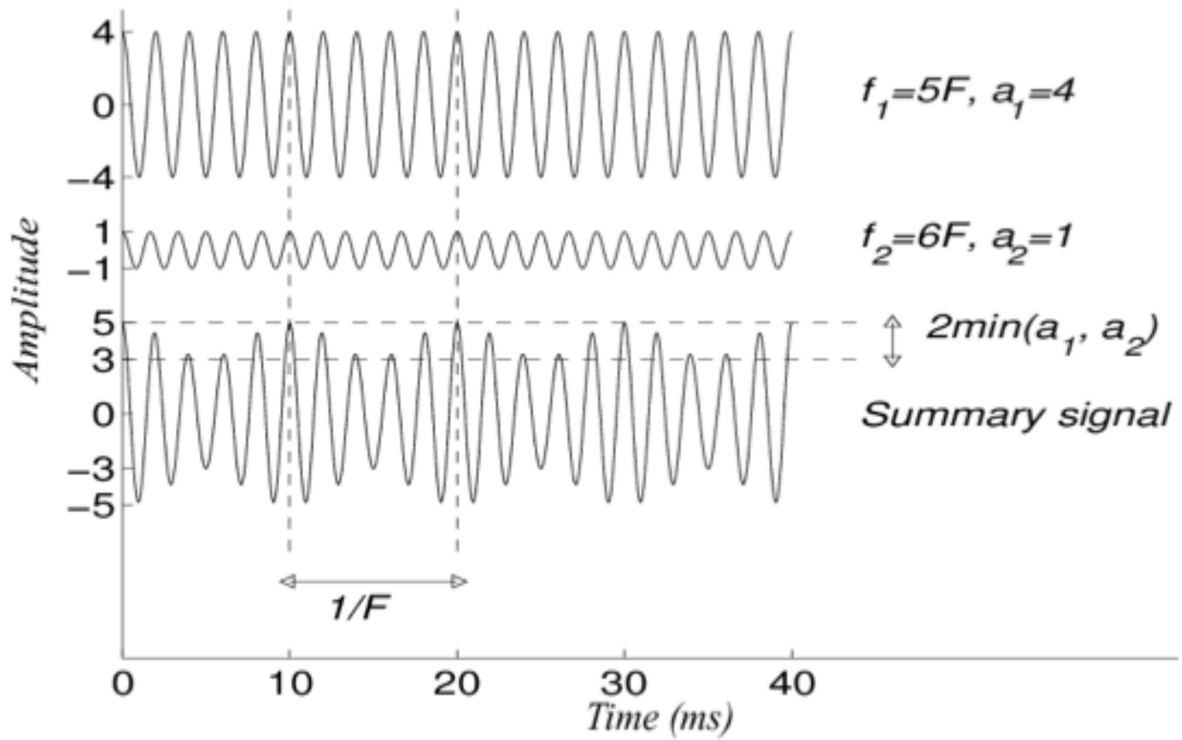
F = number of filters.

Auditory model

- Beating: interference between sounds of frequencies f_1 and f_2

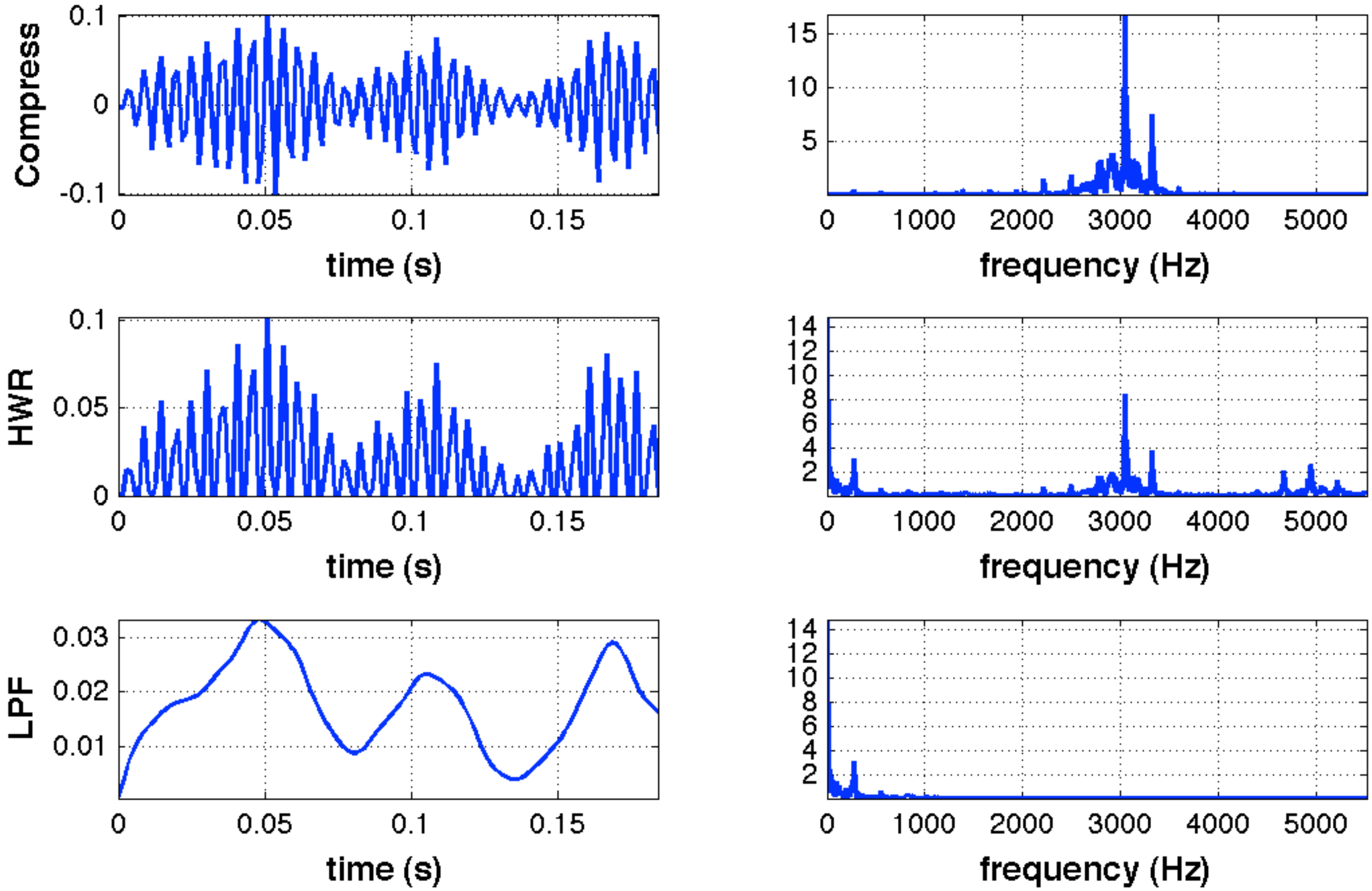
- Fluctuation of amplitude envelope of frequency $|f_2 - f_1|$

- The magnitude of the beating is determined by the smaller of the two amplitudes



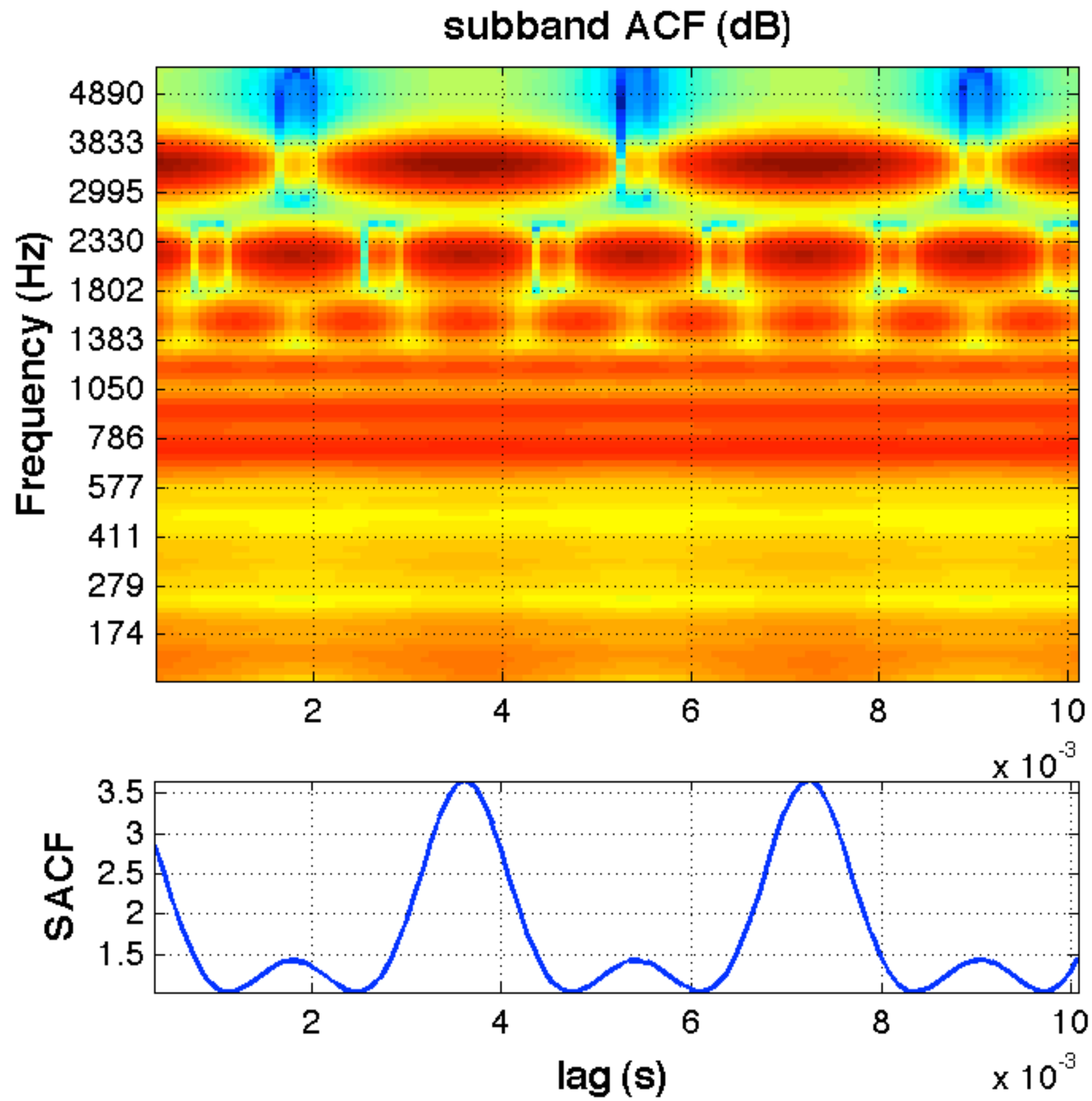
Auditory model

- Inner hair-cell (IHC) model:

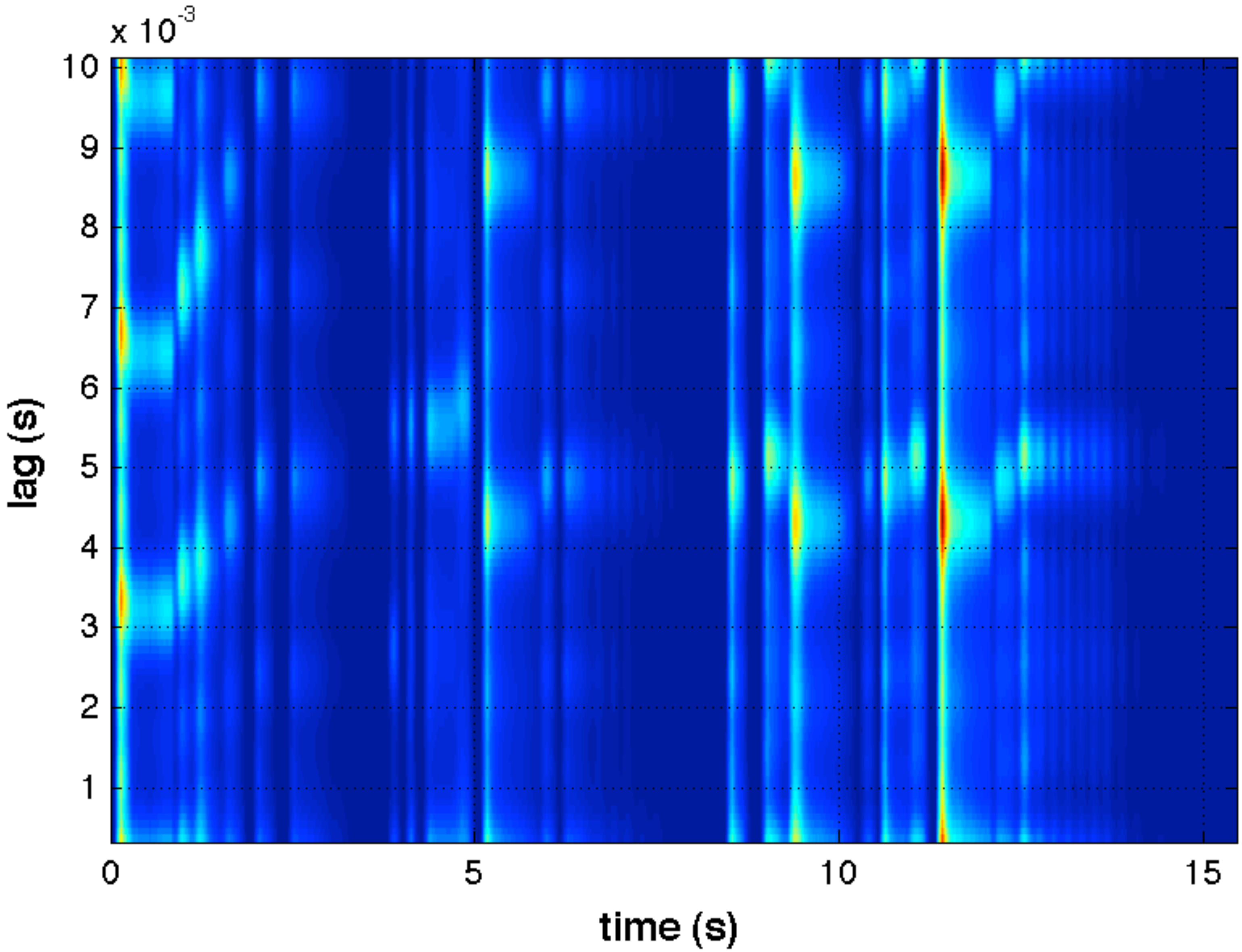


Auditory model

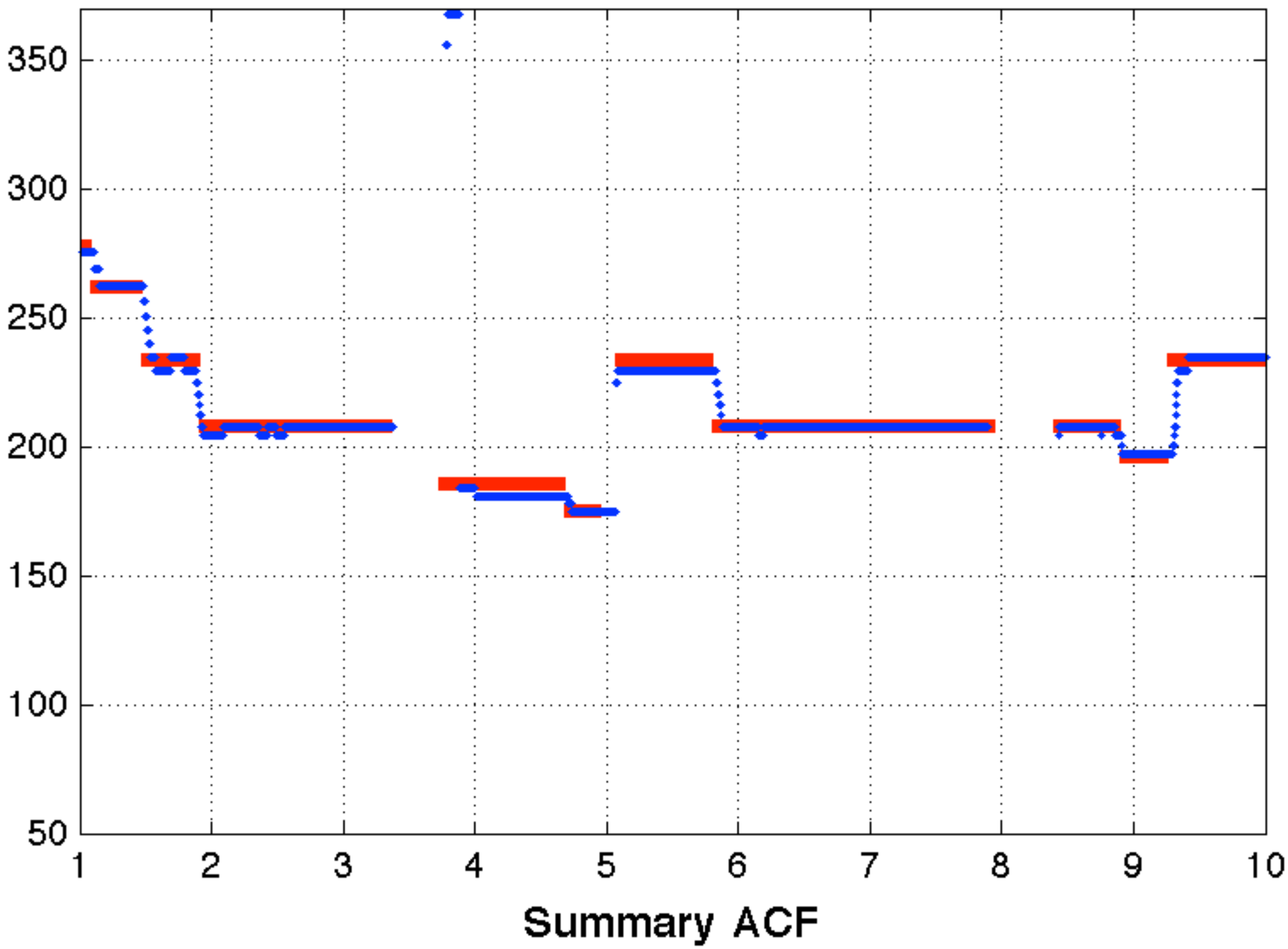
- Sub-band periodicity analysis using ACF
- Summing across channels (Summary ACF)
- Weighting of the channels changes the topology of the SACF



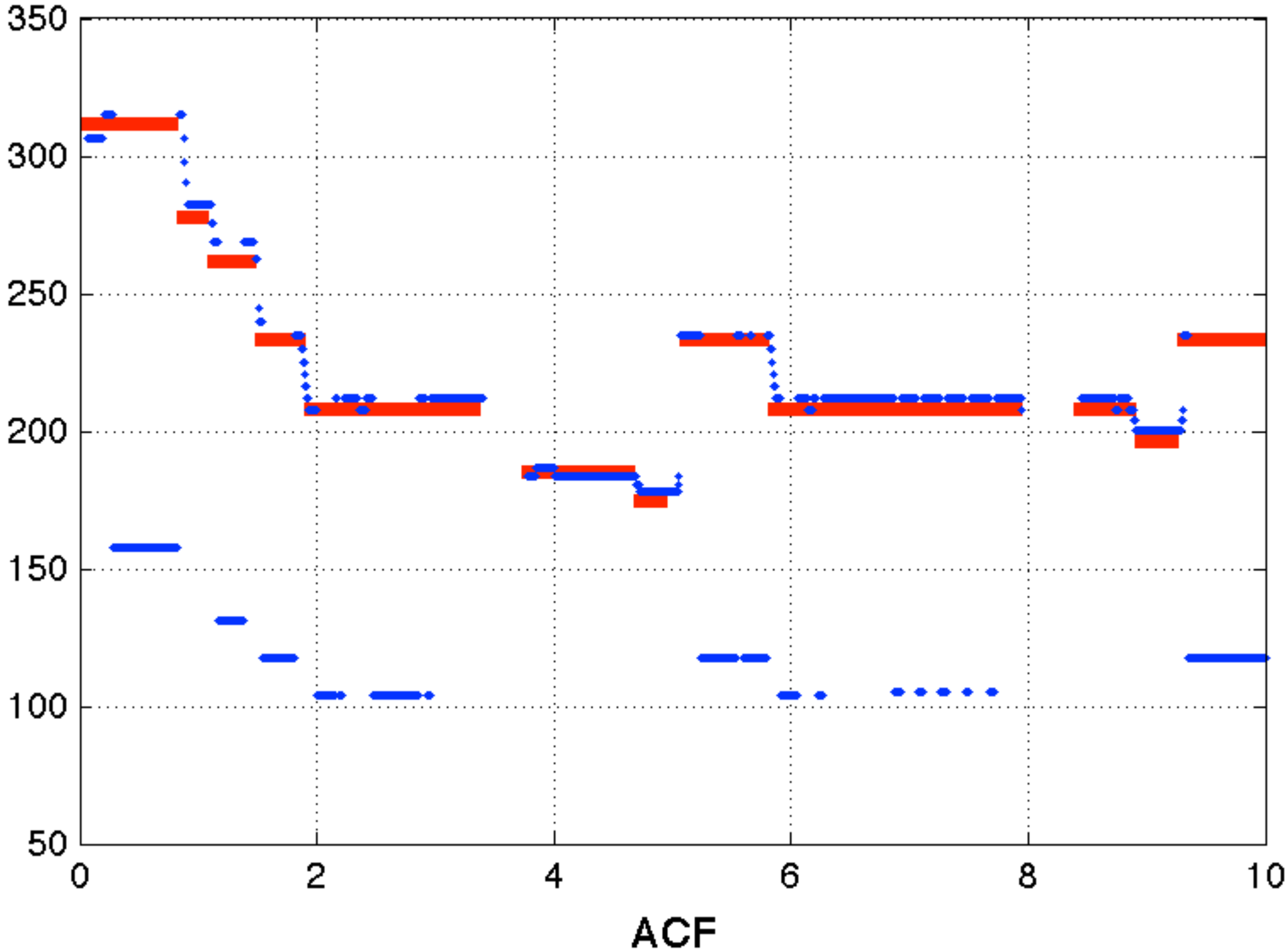
Auditory model



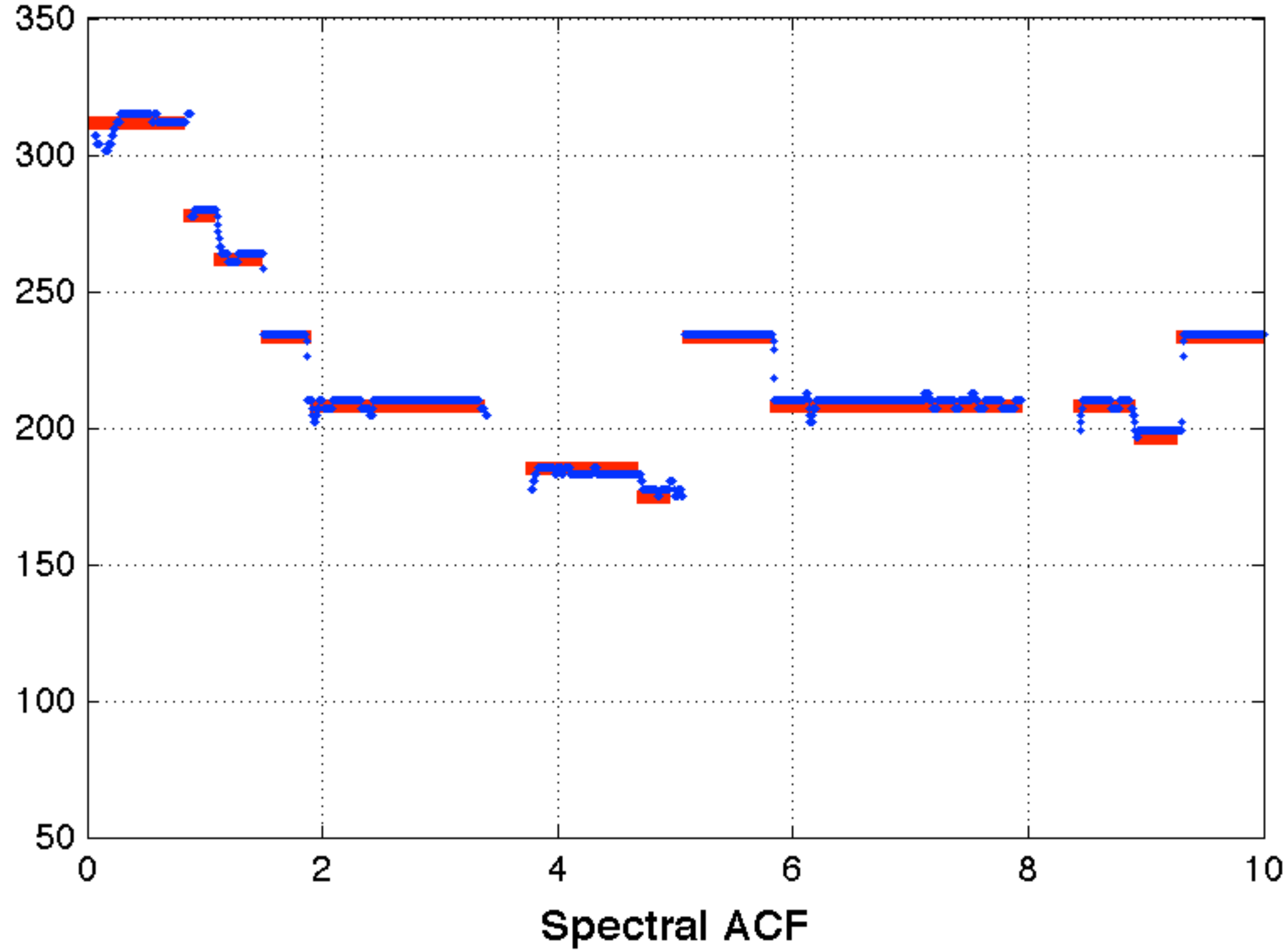
Comparing detection functions



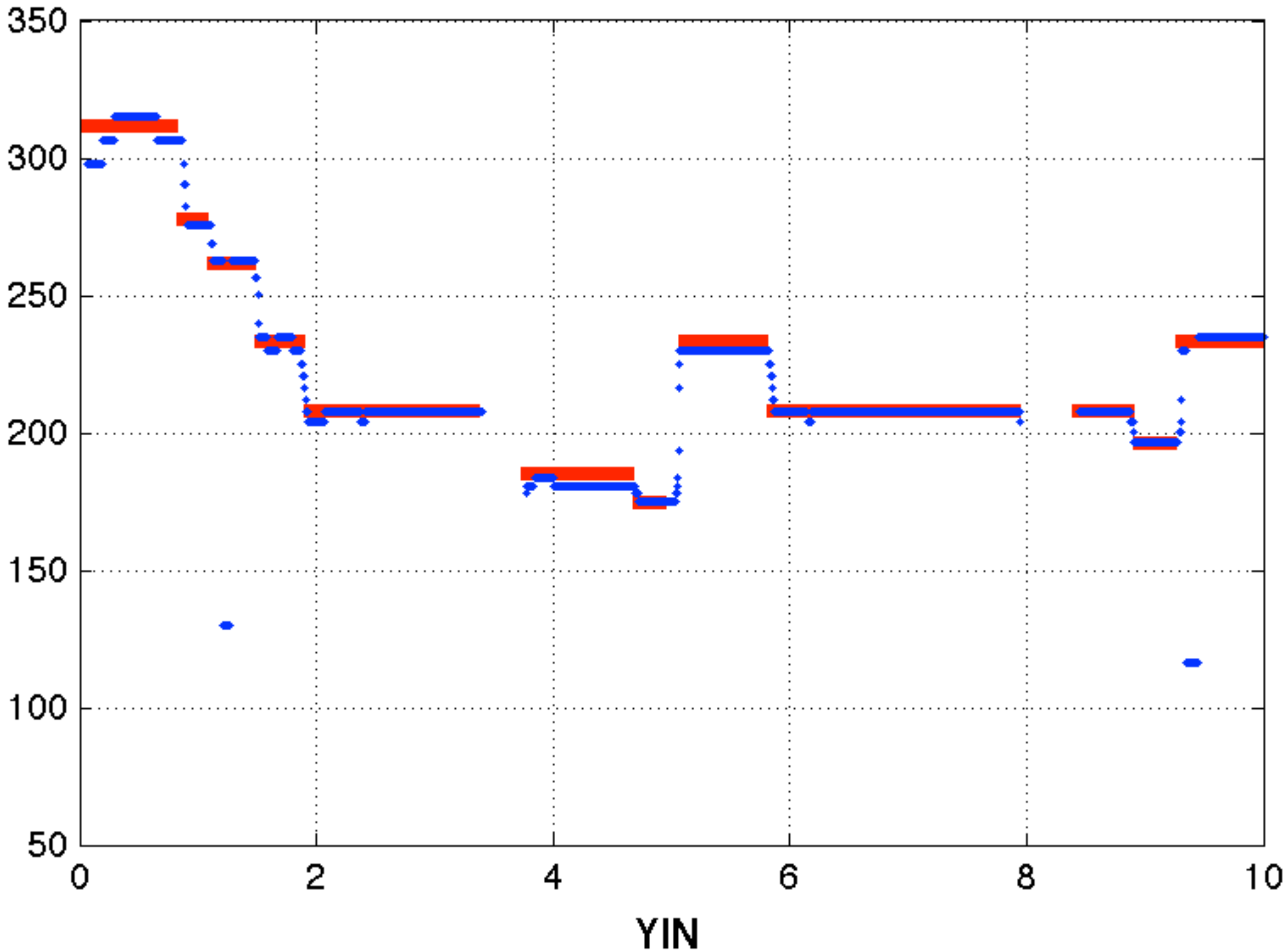
Comparing detection functions



Comparing detection functions



Comparing detection functions

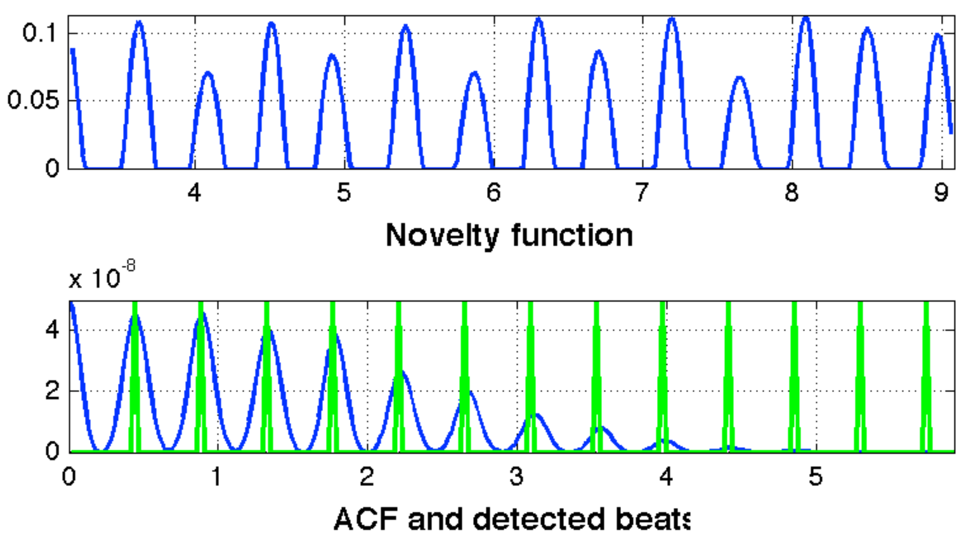


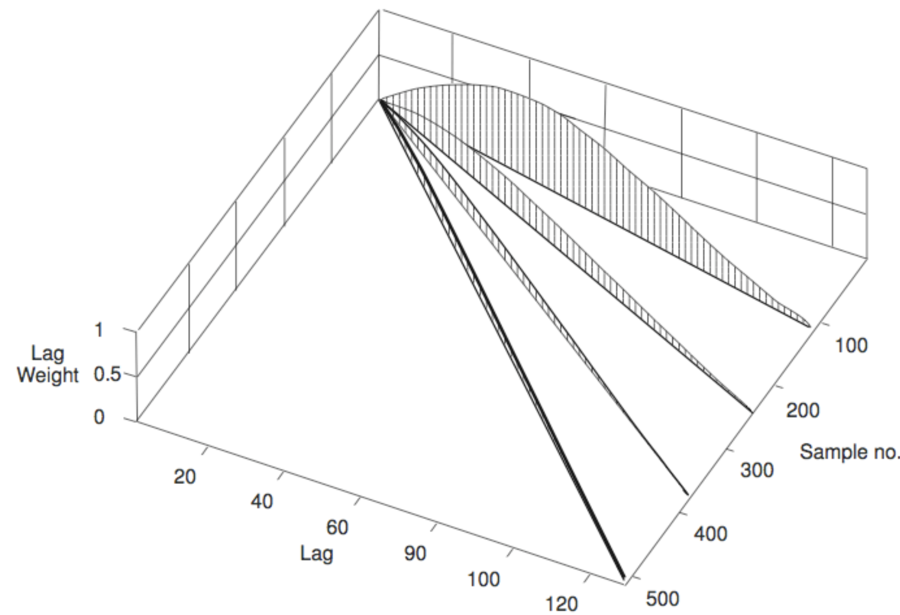
Tempo

- Tempo refers to the pace of a piece of music and is usually given in beats per minutes (BPM).
- Global quality vs time-varying local characteristic.
- Thus, in computational terms we differentiate between tempo estimation and tempo (beat) tracking.
- In tracking, beats are described by both their rate and phase.
- Vast literature: see, e.g. Hainsworth, 06; or Goto, 06 for reviews.

Tempo estimation and tracking (Davies, 05)

- Novelty function (NF): remove local mean + half-wave rectify
- Periodicity: dot multiply ACF of NF with a weighted comb filterbank



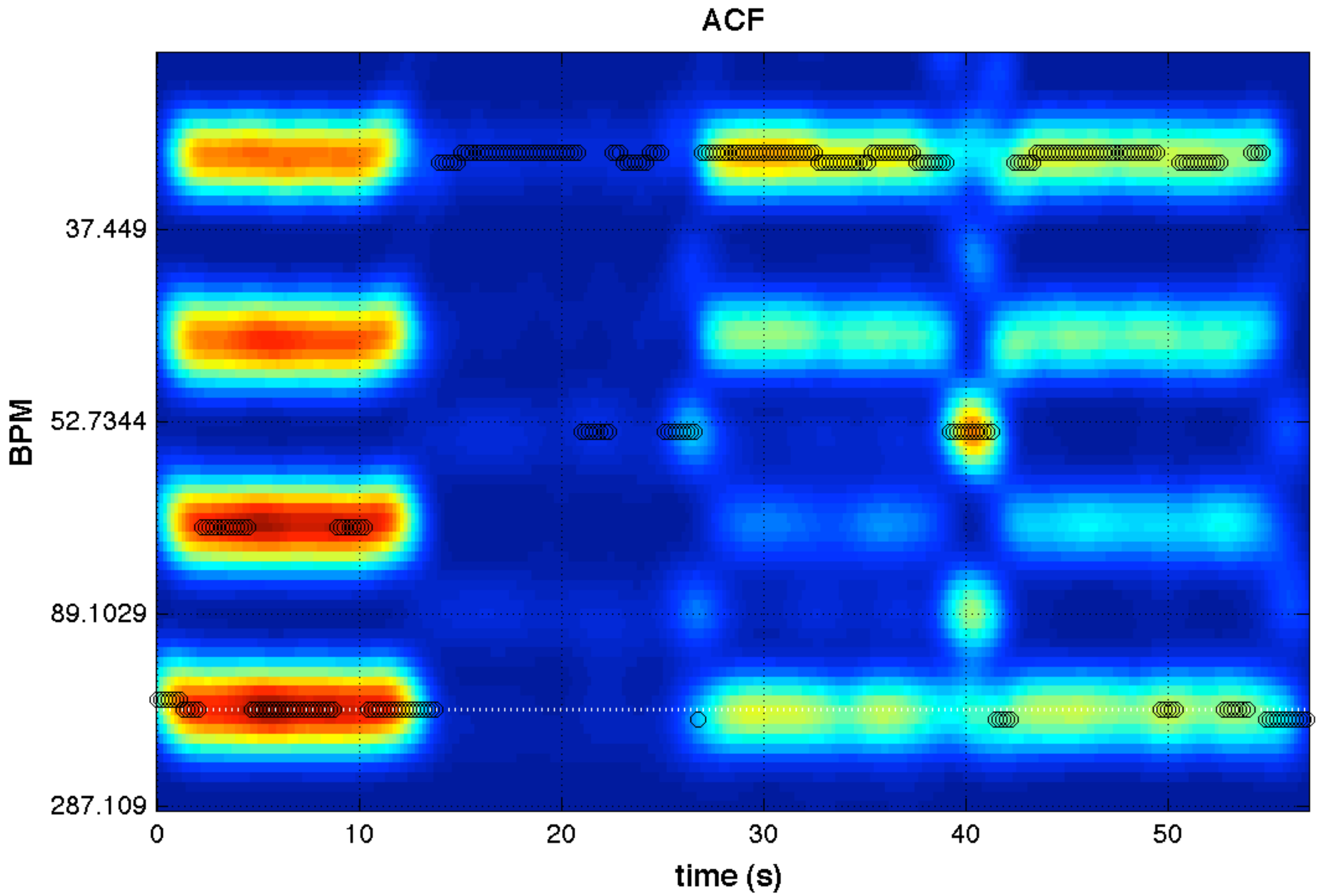


$$R_w(l) = (l/b^2)e^{\frac{-l^2}{2b^2}}$$

*From Davies and Plumbley, ICASSP 2005

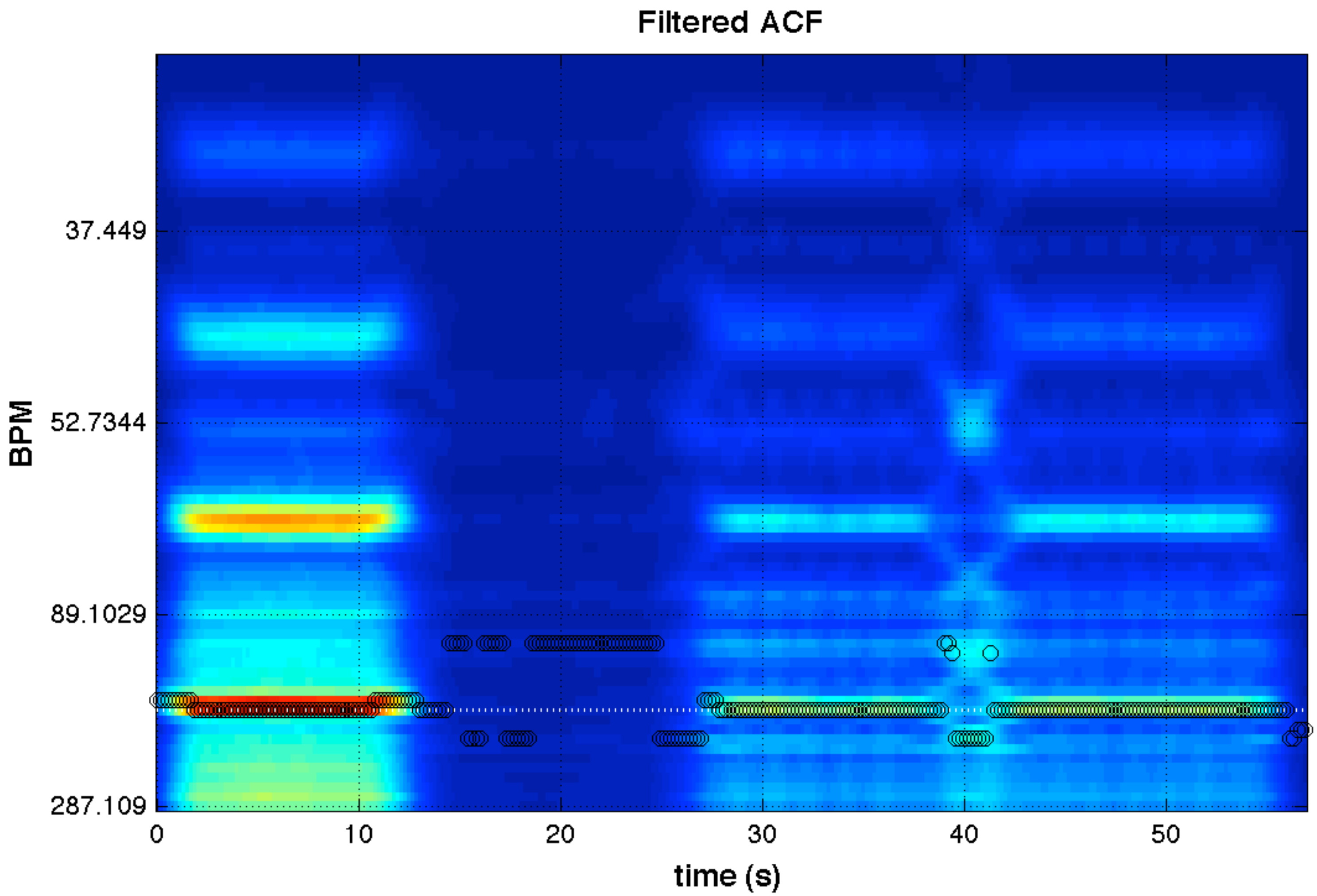
Tempo estimation and tracking (Davies, 05)

- Choose lag that maximizes the ACF

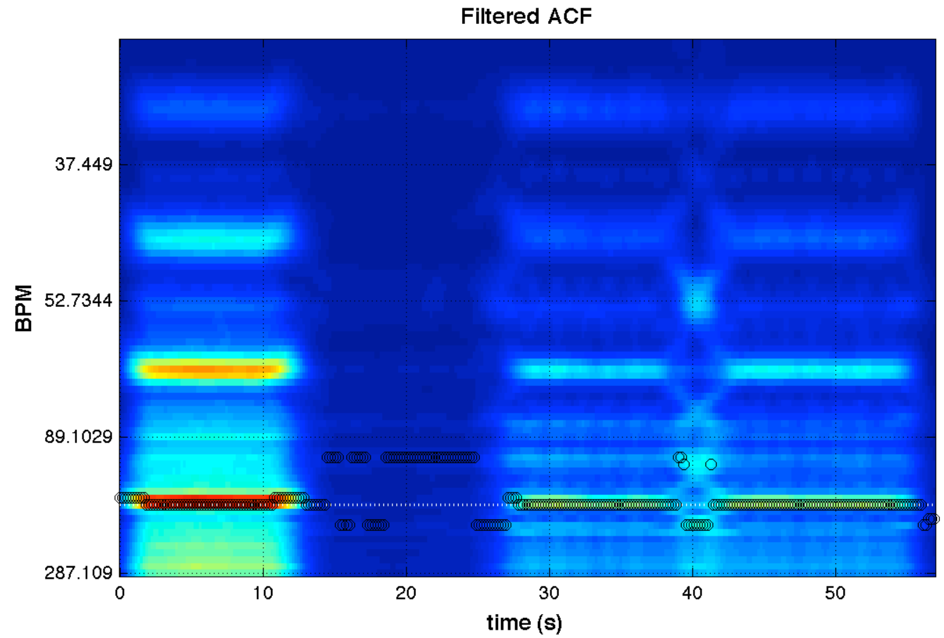


Tempo estimation and tracking (Davies, 05)

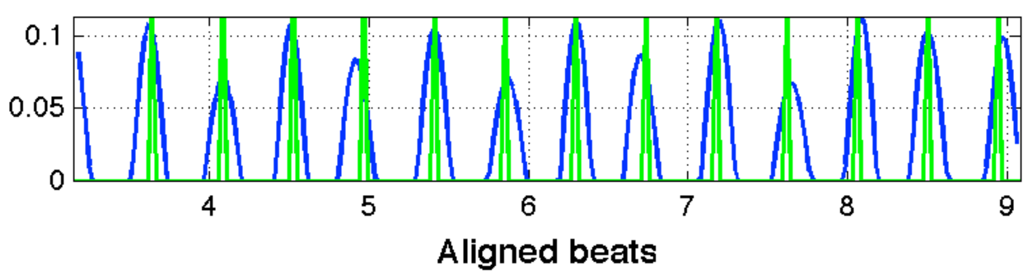
- Choose filter that maximizes the dot product

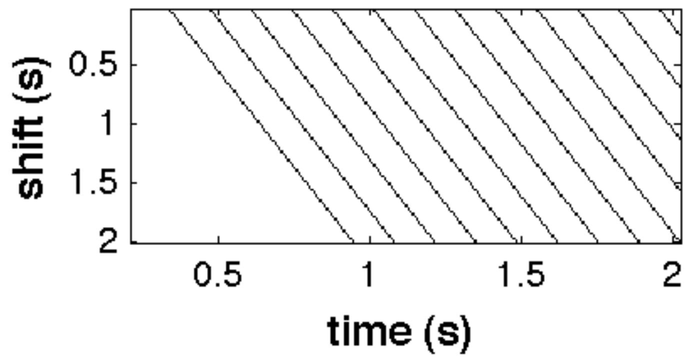


Tempo estimation and tracking (Davies, 05)



- Phase: dot multiply DF with shifted versions of selected comb filter





Tempo estimation and tracking (Grosche, 09)

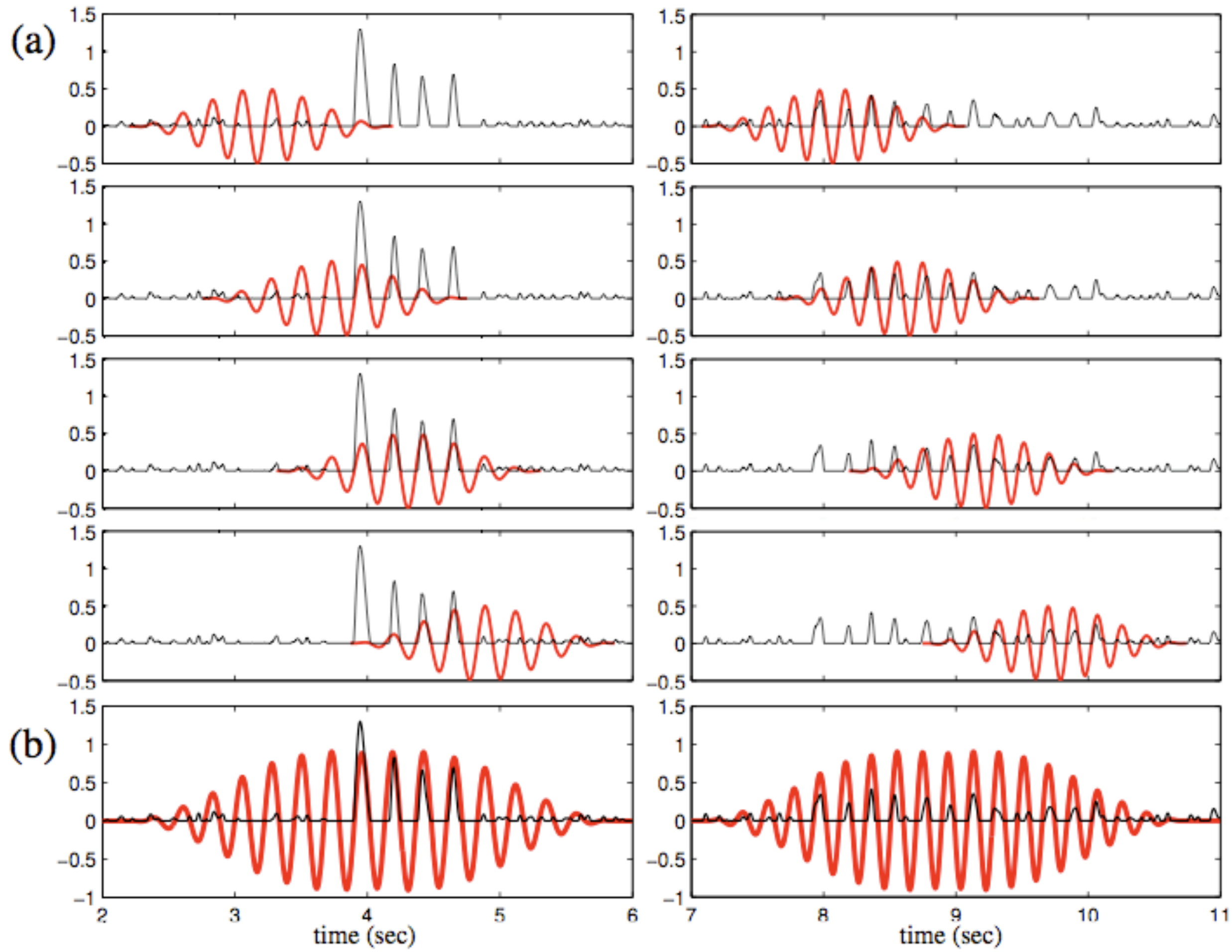
- DFT of novelty function γ(n) for frequencies: $\omega \in [30:600]/(60 \times \mathrm{fs}_\gamma)$
- Choose frequency that maximizes the magnitude spectrum at each frame
- Construct a sinusoidal kernel: $\kappa(m) = w(m-n)cos(2\pi(\hat{\omega}m - \hat{\varphi}))$
- In Grosche, 09 phase is computed as:

$$\hat{\varphi} = \frac{1}{2\pi}\arccos\left(\frac{Re(F(\hat{\omega},n))}{|F(\hat{\omega},n)|}\right)$$

- Alternatively, we can find the phase that maximizes the dot product of γ(n) with shifted versions of the kernel, as before.

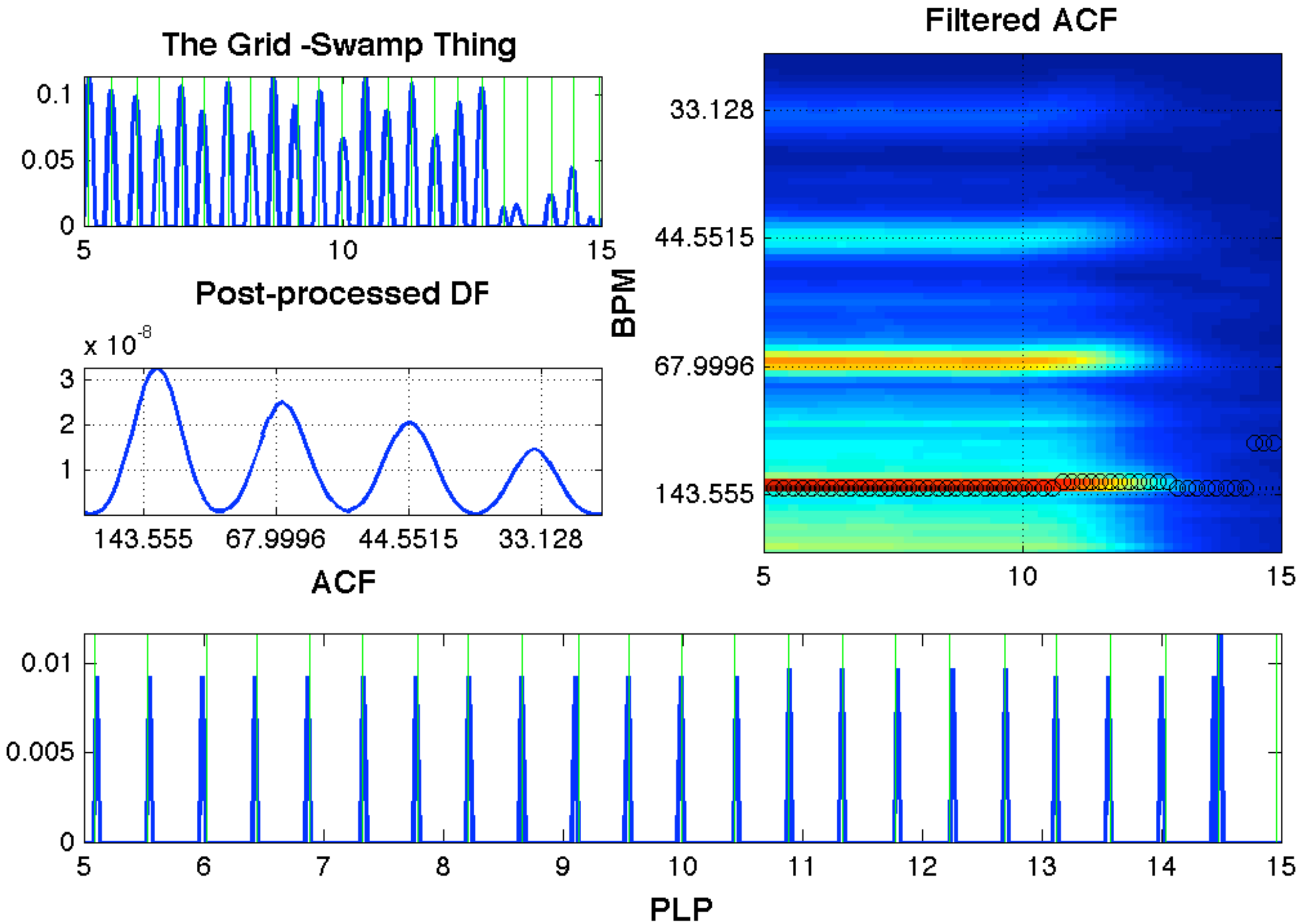
Tempo estimation and tracking (Grosche, 09)

- tracking function: Overlap-add of optimal local kernels + half-wave rectify

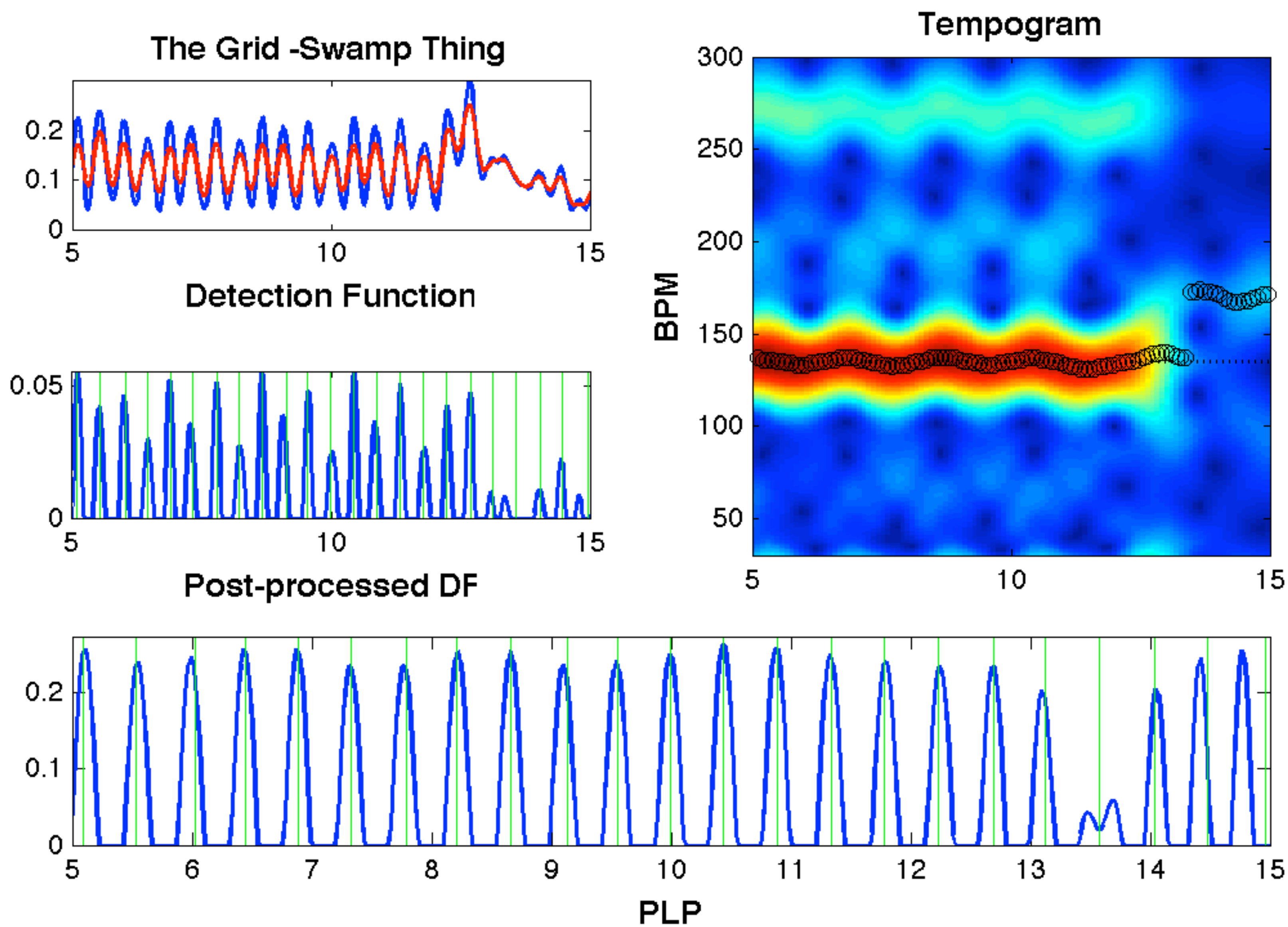


*From Grosche and Mueller, WASPAA 2009

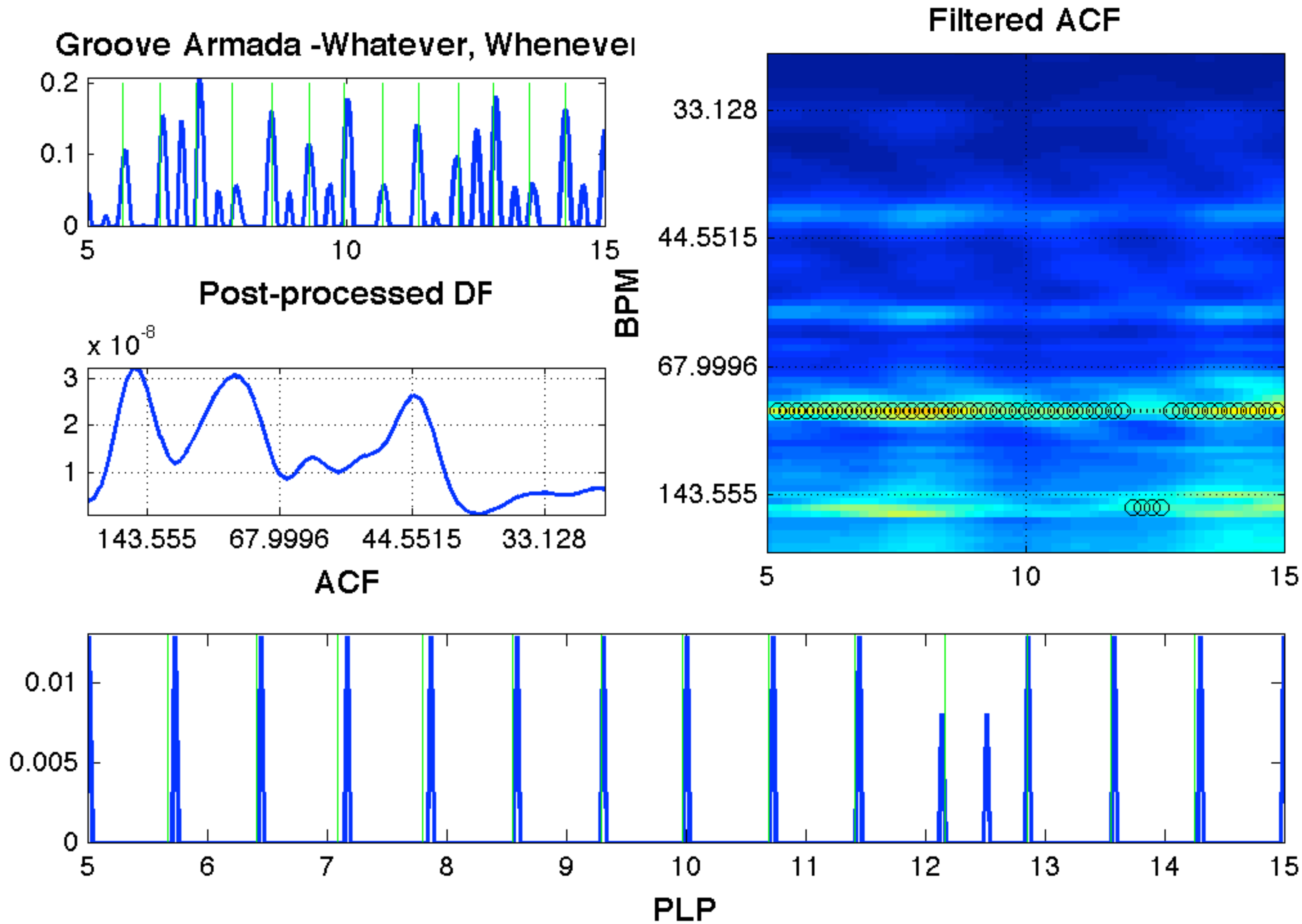
Tempo estimation and tracking (Davies, 05)



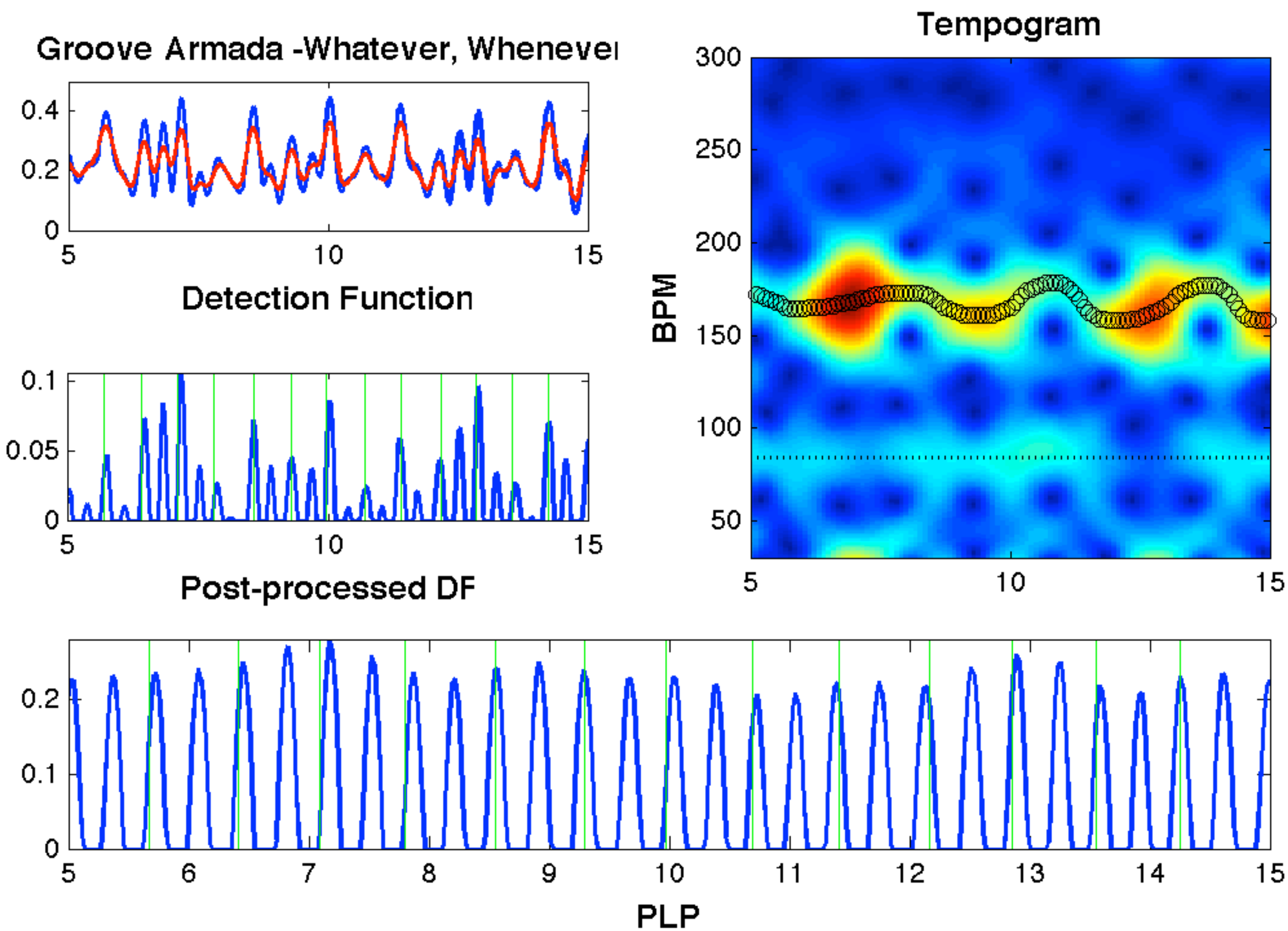
Tempo estimation and tracking (Grosche, 09)



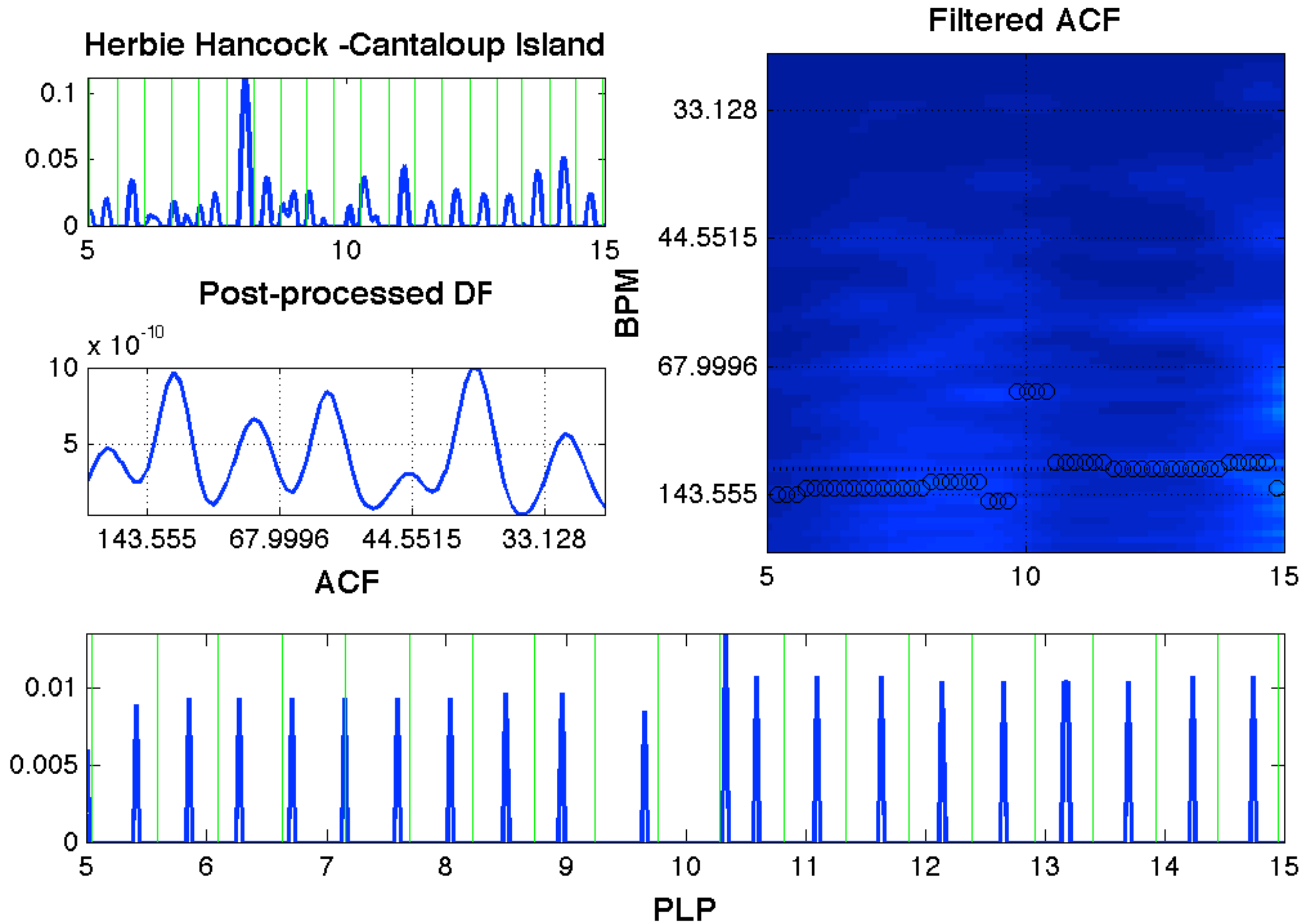
Tempo estimation and tracking (Davies, 05)



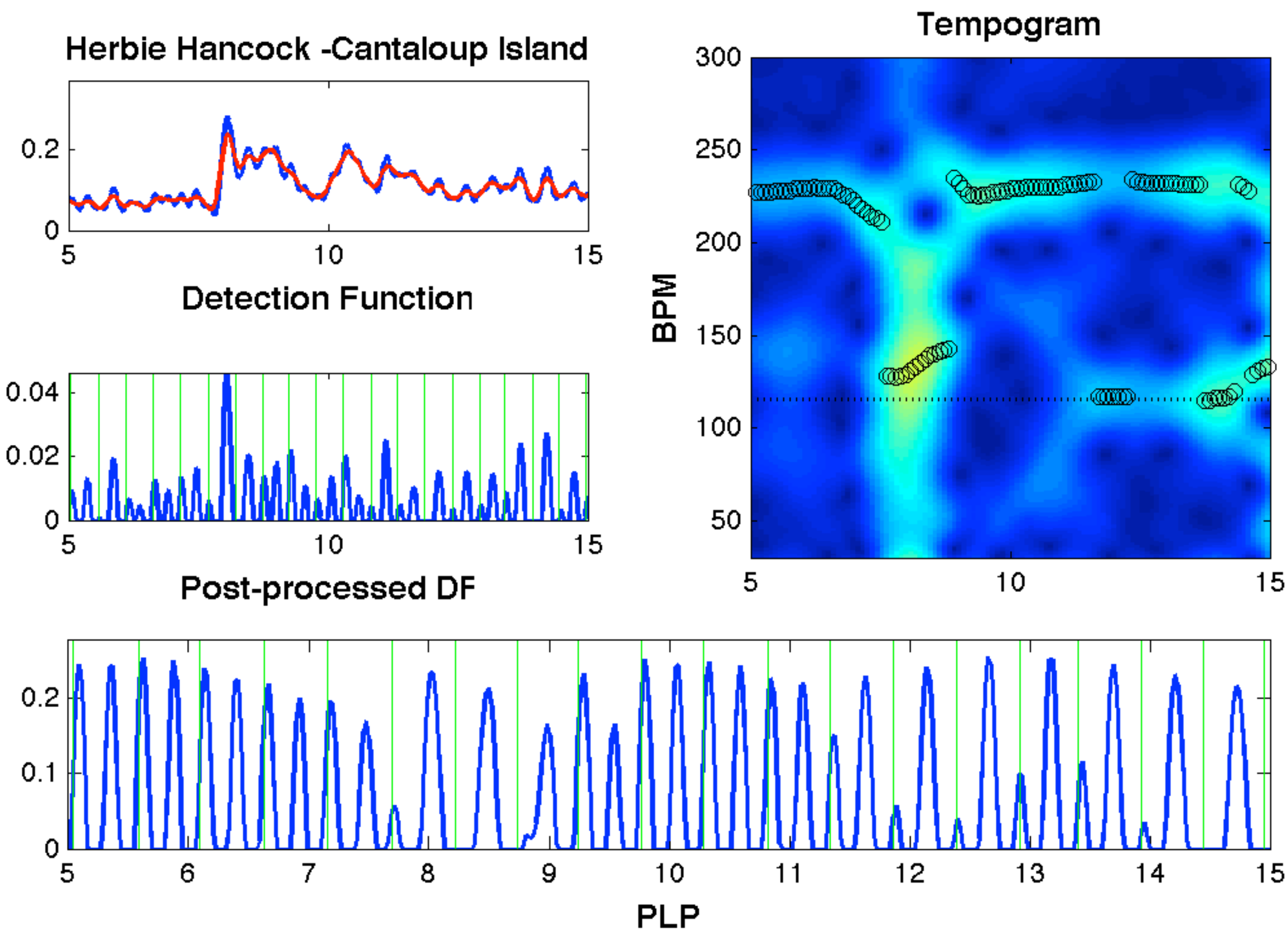
Tempo estimation and tracking (Grosche, 09)



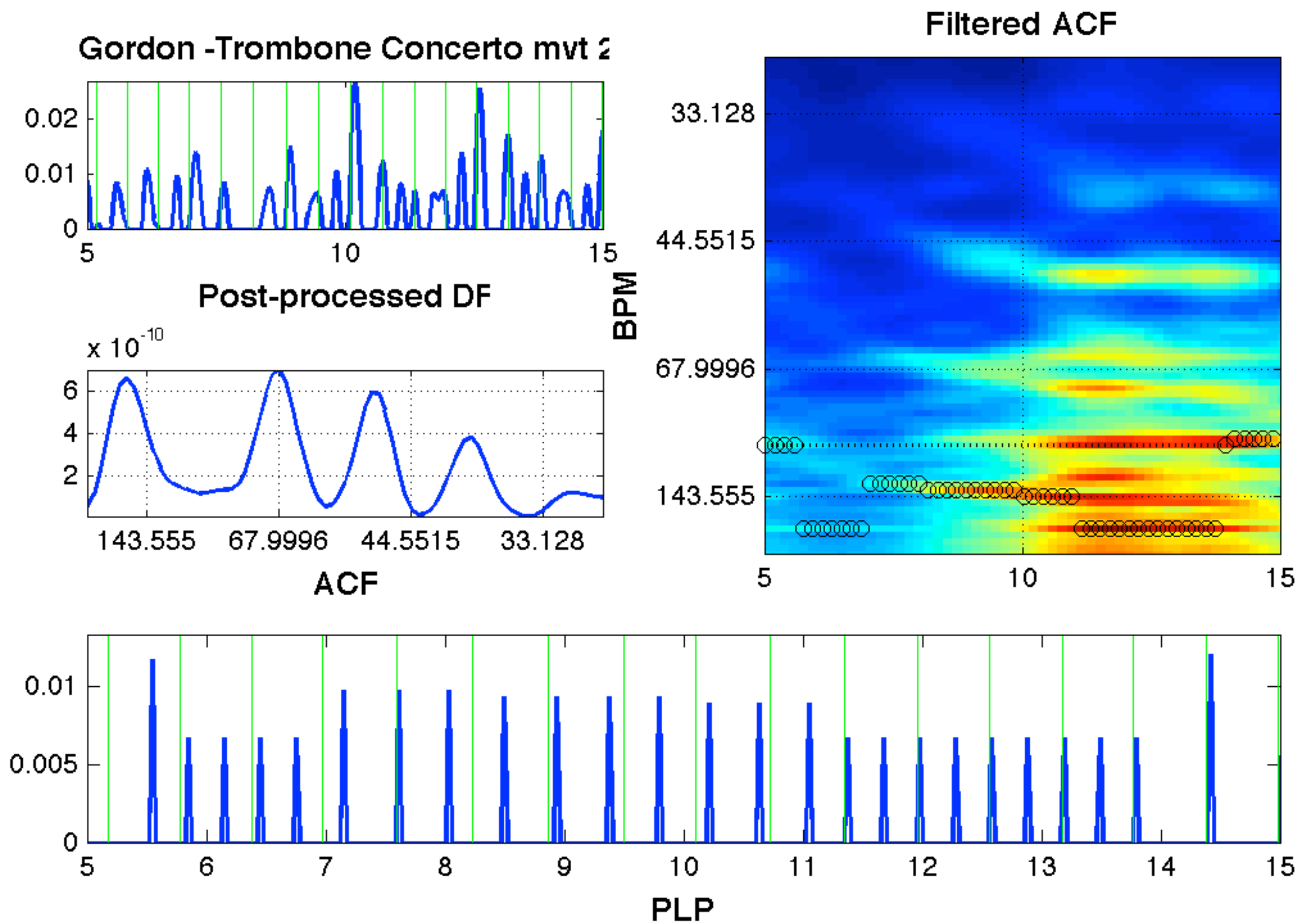
Tempo estimation and tracking (Davies, 05)



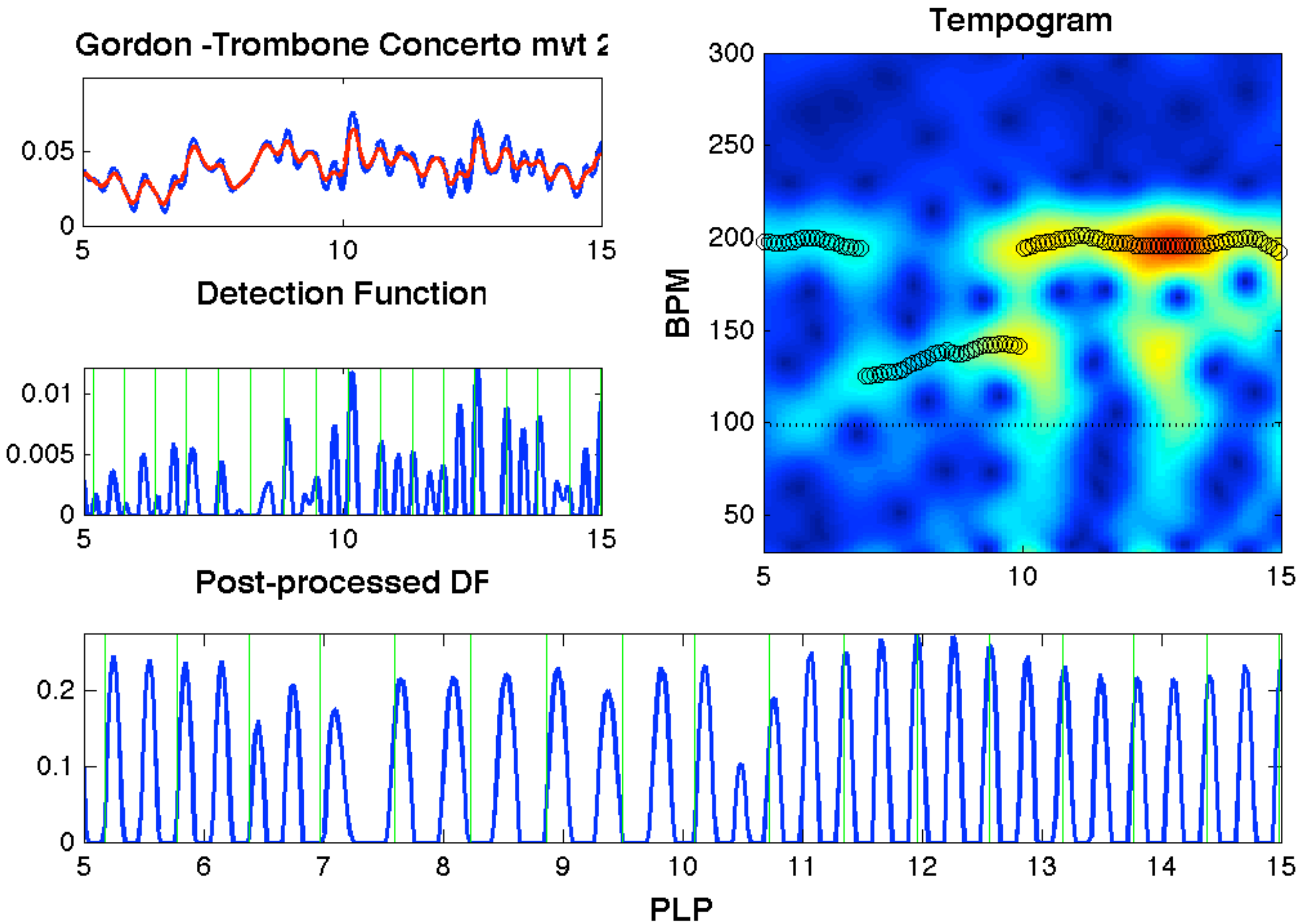
Tempo estimation and tracking (Grosche, 09)



Tempo estimation and tracking (Davies, 05)



Tempo estimation and tracking (Grosche, 09)



References

- Wang, D. and Brown, G. (Eds) “Computational Auditory Scene Analysis”. Wiley-Interscience (2006): chapter 2, de Cheveigné, A. “Multiple F0 Estimation”; and chapter 8, Goto, M. “Analysis of Musical Audio Signals”.
- Klapuri, A. and Davy, M. (Eds) “Signal Processing Methods for Music Transcription”. Springer (2006): chapter 1, Klapuri, A. “Introduction to Music Transcription”; chapter 4, Hainsworth, S. “Beat Tracking and Musical Metre Analysis”; and chapter 8, Klapuri, A. “Auditory Model-based methods for Multiple Fundamental Frequency Estimation”
- Slaney, M. “An efficient implementation of the Patterson Holdsworth auditory filter bank”. Technical Report 35, Perception Group, Advanced Technology Group, Apple Computer, 1993.
- Smith, J.O. “Mathematics of the Discrete Fourier Transform (DFT)”. 2nd Edition, W3K Publishing (2007): chapter 8, “DFT Applications”.

References

- Grosche, P. and Müller, M. “Computing Predominant Local Periodicity Information in Music Recordings”. Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, 2009.
- Davies, M.E.P. and Plumbley, M.D. “Beat Tracking With A Two State Model”. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vol. 3, pp 241-244 Philadelphia, USA, March 19-23, 2005.
- This lecture borrows heavily from: Emmanuel Vincent’s lecture notes on pitch estimation (QMUL - Music Analysis and Synthesis); and from Anssi Klapuri’s lecture notes on F0 estimation and automatic music transcription (ISMIR 2004 Graduate School: http://ismir2004.ismir.net/graduate.html)