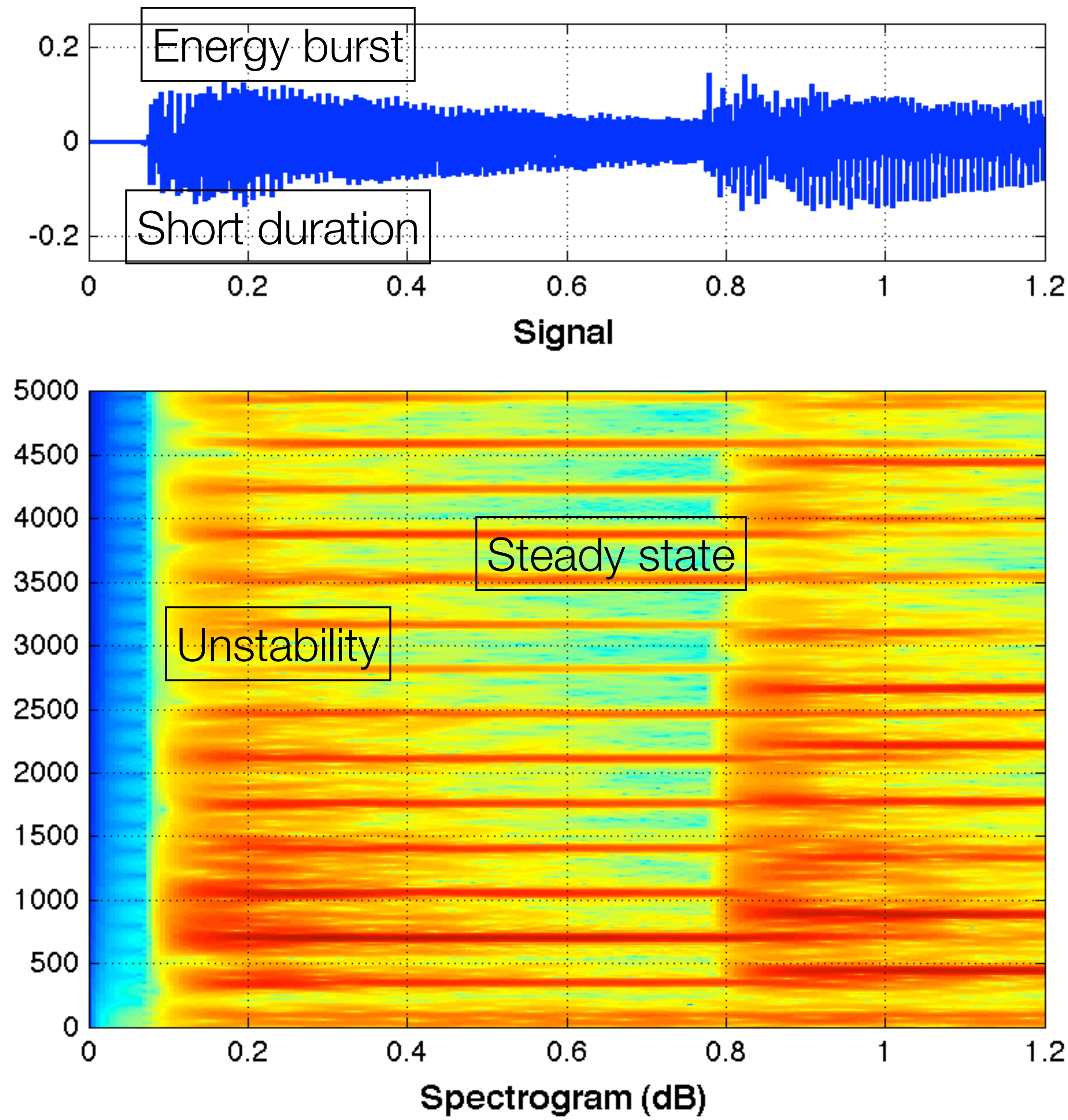
Novelty detection

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

Novelty detection

- Find the start time (onset) of new events in the audio signal.

- Onset: single instant chosen to mark the start of the (attack) transient.

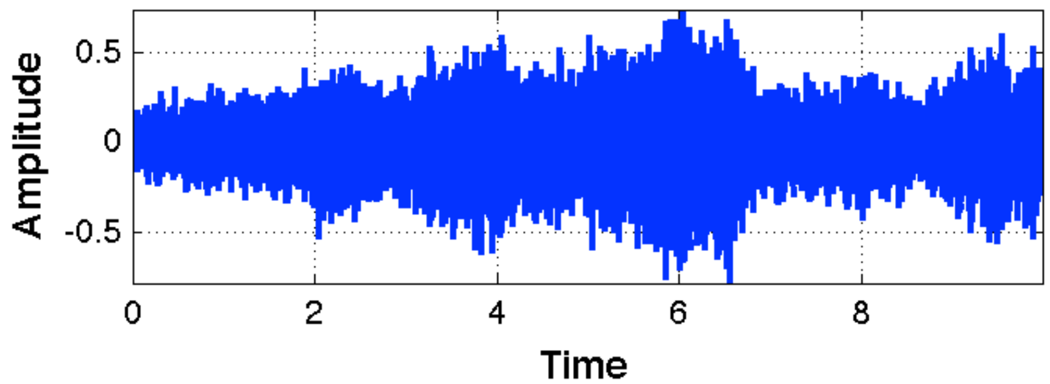


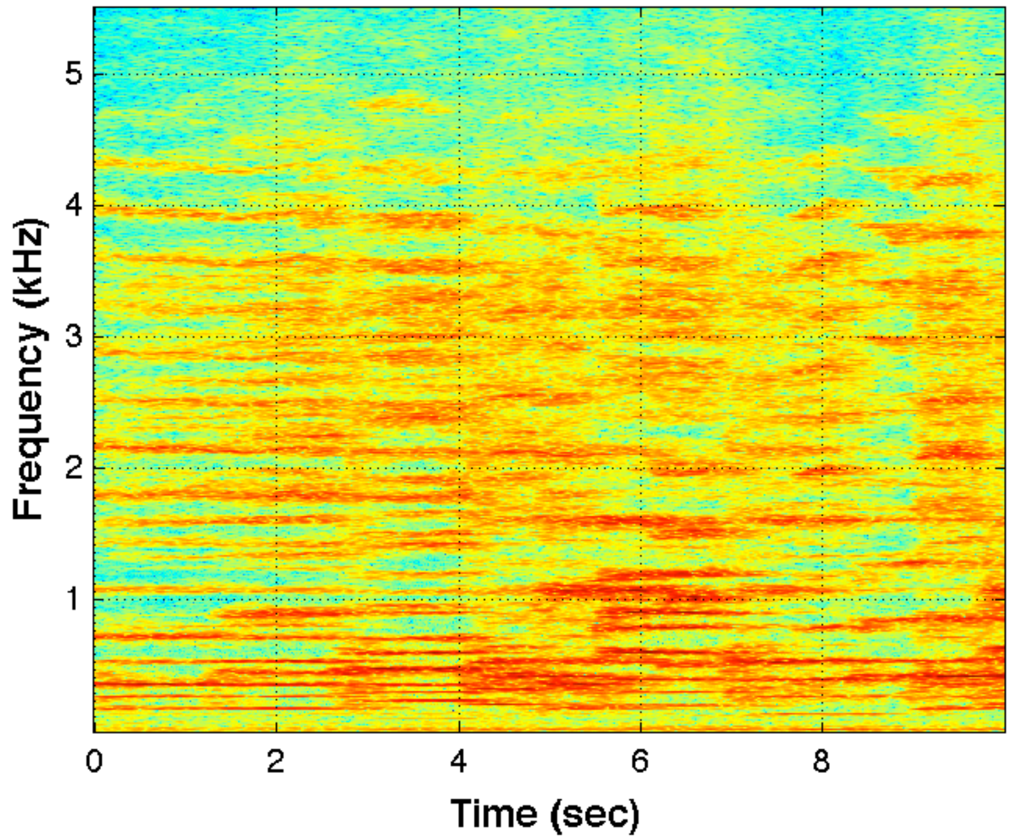
Applications

- Identifying regions of interest in environmental recordings
- Segmentation of word/phonemes in speech, notes in music
- First layer of rhythm analysis
- Sound manipulation and synthesis: http://www.music.mcgill.ca/~hockman/projects/ARTMA/index.html
- Computational biology?!! http://isophonics.net/content/calcium-signal-analyser

Difficulties

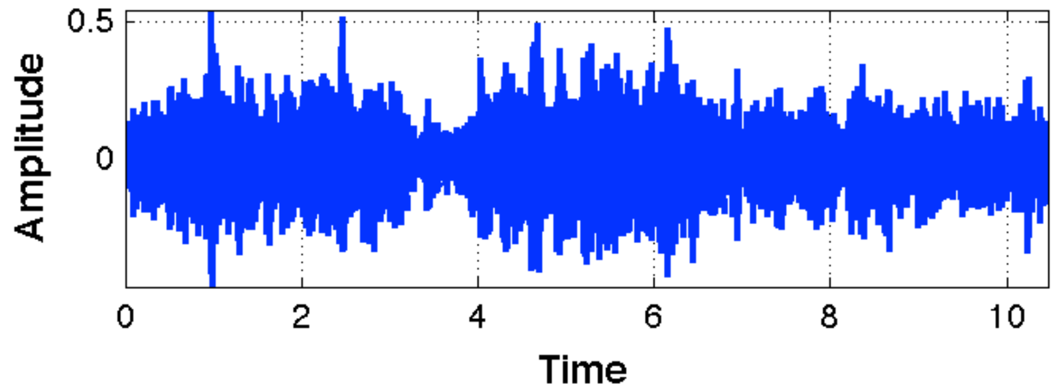
- transient extended in time
- multiple voices -> (a)synchronous onsets
- ambiguous events (vibrato, tremolo, glissandi)
- perceptual vs physical

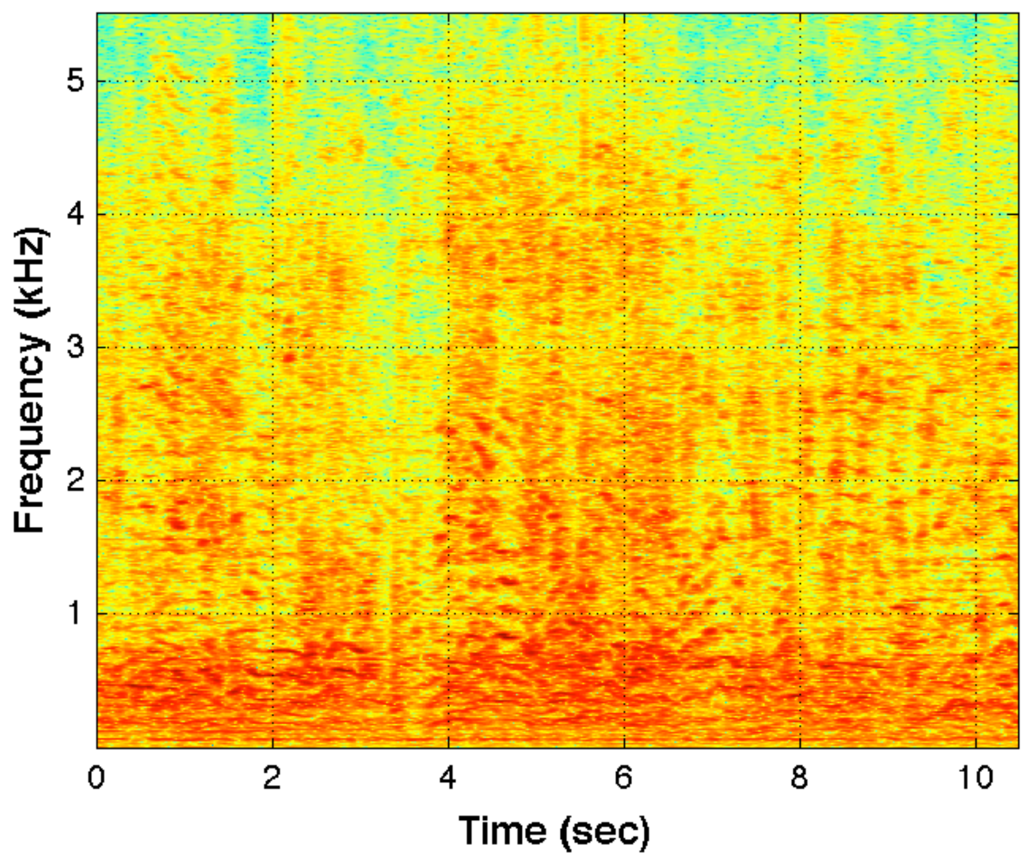




Difficulties

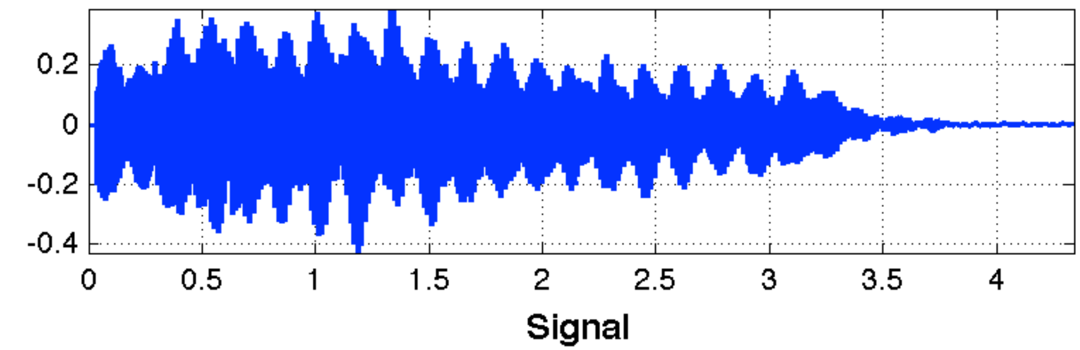
- transient extended in time
- multiple voices -> (a)synchronous onsets
- ambiguous events (vibrato, tremolo, glissandi)
- perceptual vs physical

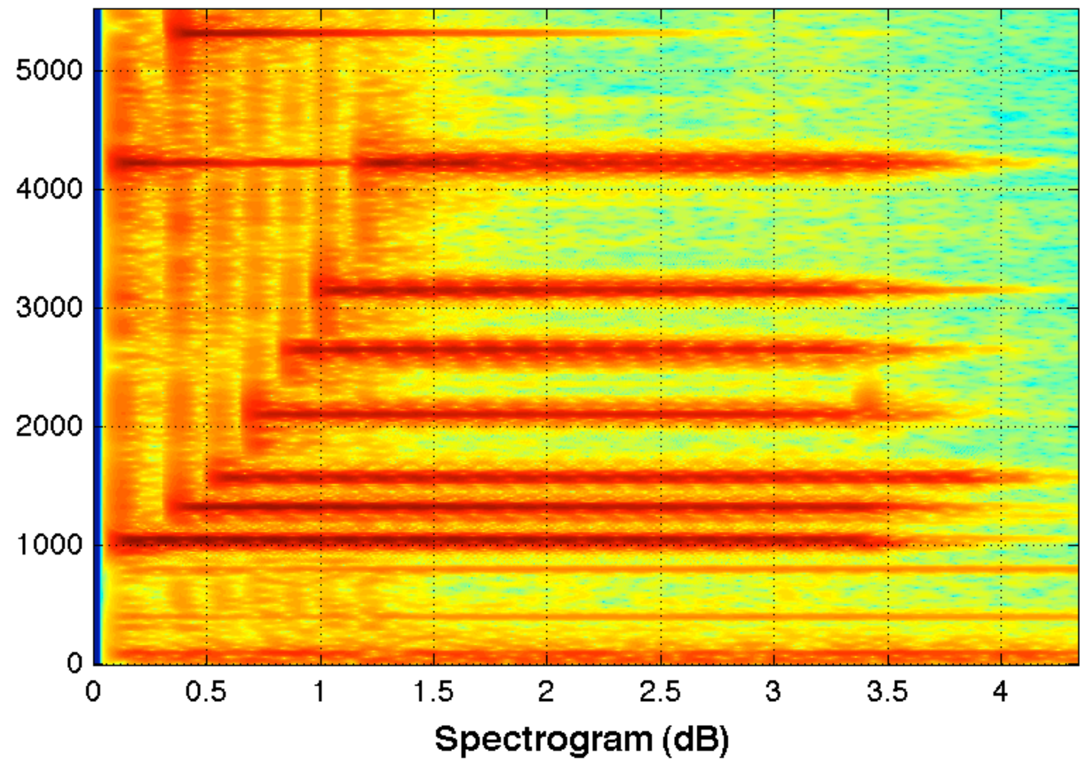




Difficulties

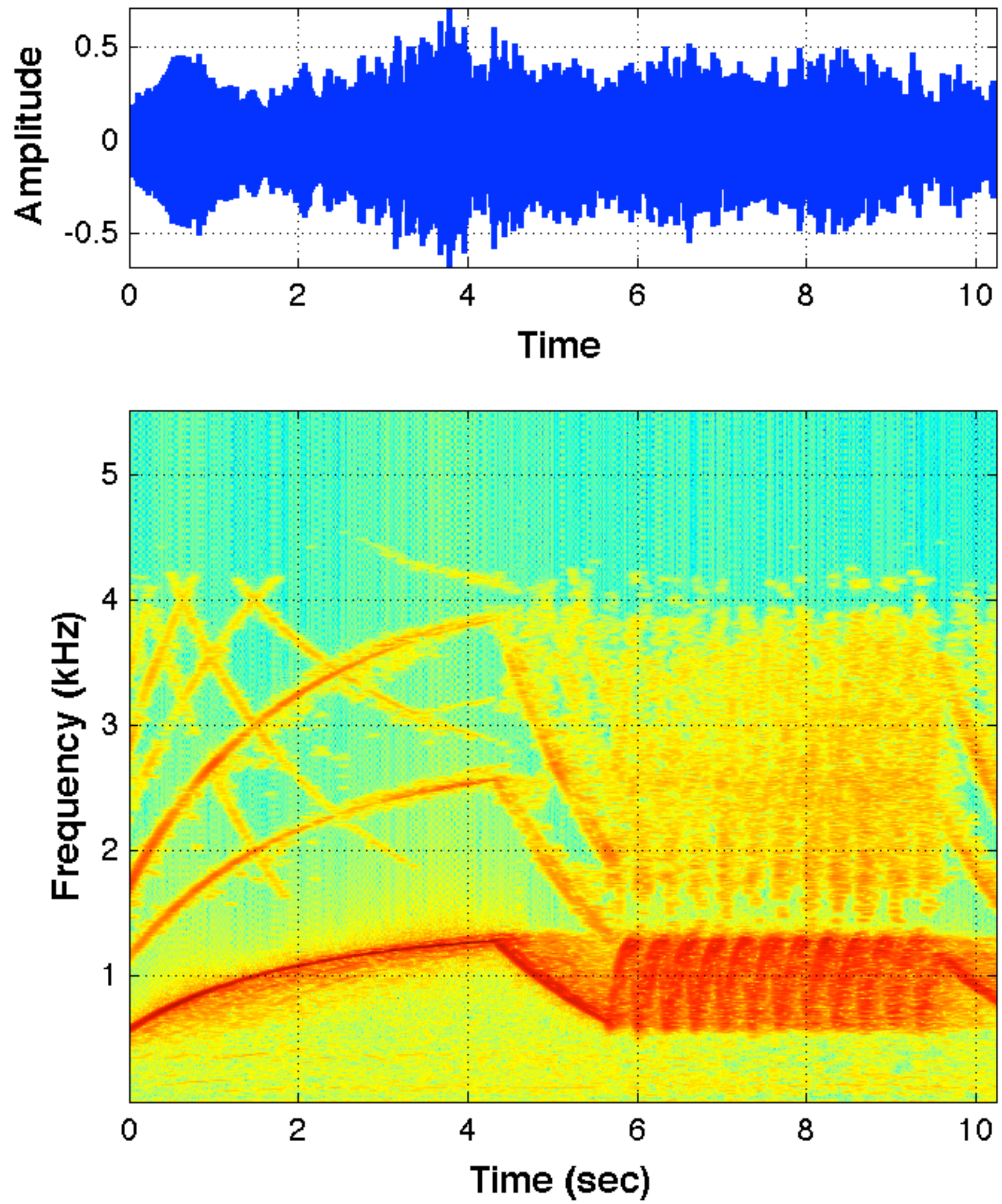
- transient extended in time
- multiple voices -> (a)synchronous onsets
- ambiguous events (vibrato, tremolo, glissandi)
- perceptual vs physical



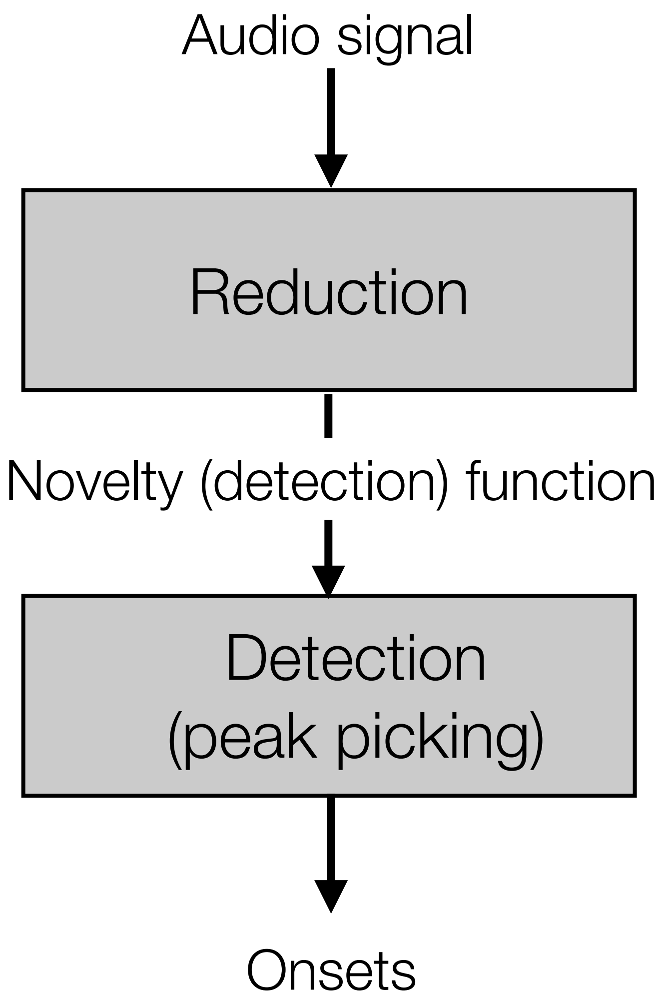


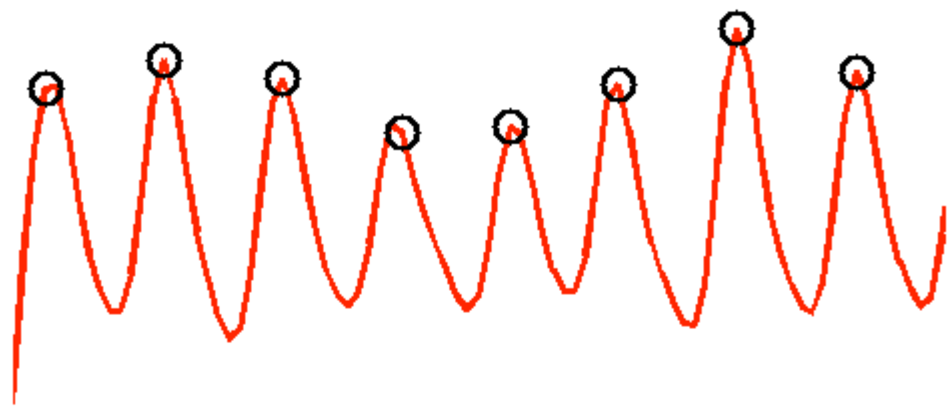
Difficulties

- transient extended in time
- multiple voices -> (a)synchronous onsets
- ambiguous events (vibrato, tremolo, glissandi)
- perceptual vs physical



Architecture





Time-domain

- Onsets: often characterized by an amplitude increase

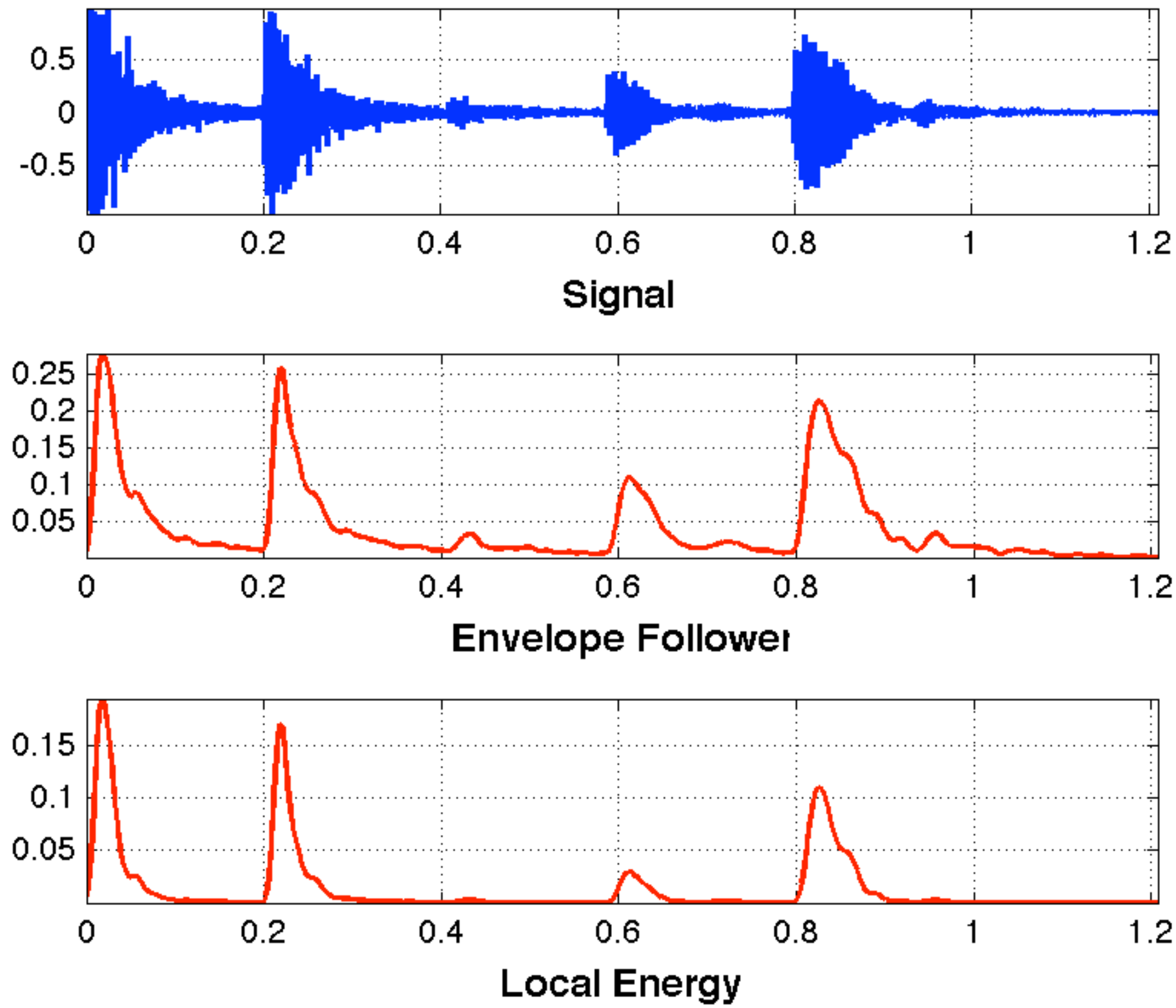
- Envelope following (full-wave rectification + smoothing):

$$E_0(m) = \frac{1}{N} \sum_{n=-N/2}^{N/2} |x(n+mh)| w(n)$$

- Squaring instead of rectifying, results in the local energy:

$$E(m) = \frac{1}{N} \sum_{n=-N/2}^{N/2} (x(n+mh))^2 \, w(n)$$

Time-domain



0.5
0
-0.5
0
0.2
0.4
0.6
0.8
1
1.2
Signal
0.25
0.2
0.15
0.1
0.05
0
0.2
0.4
0.6
0.8
1
1.2
Envelope Follower
0.15
0.1
0.05
0
0.2
0.4
0.6
0.8
1
1.2
Local Energy

Time-domain

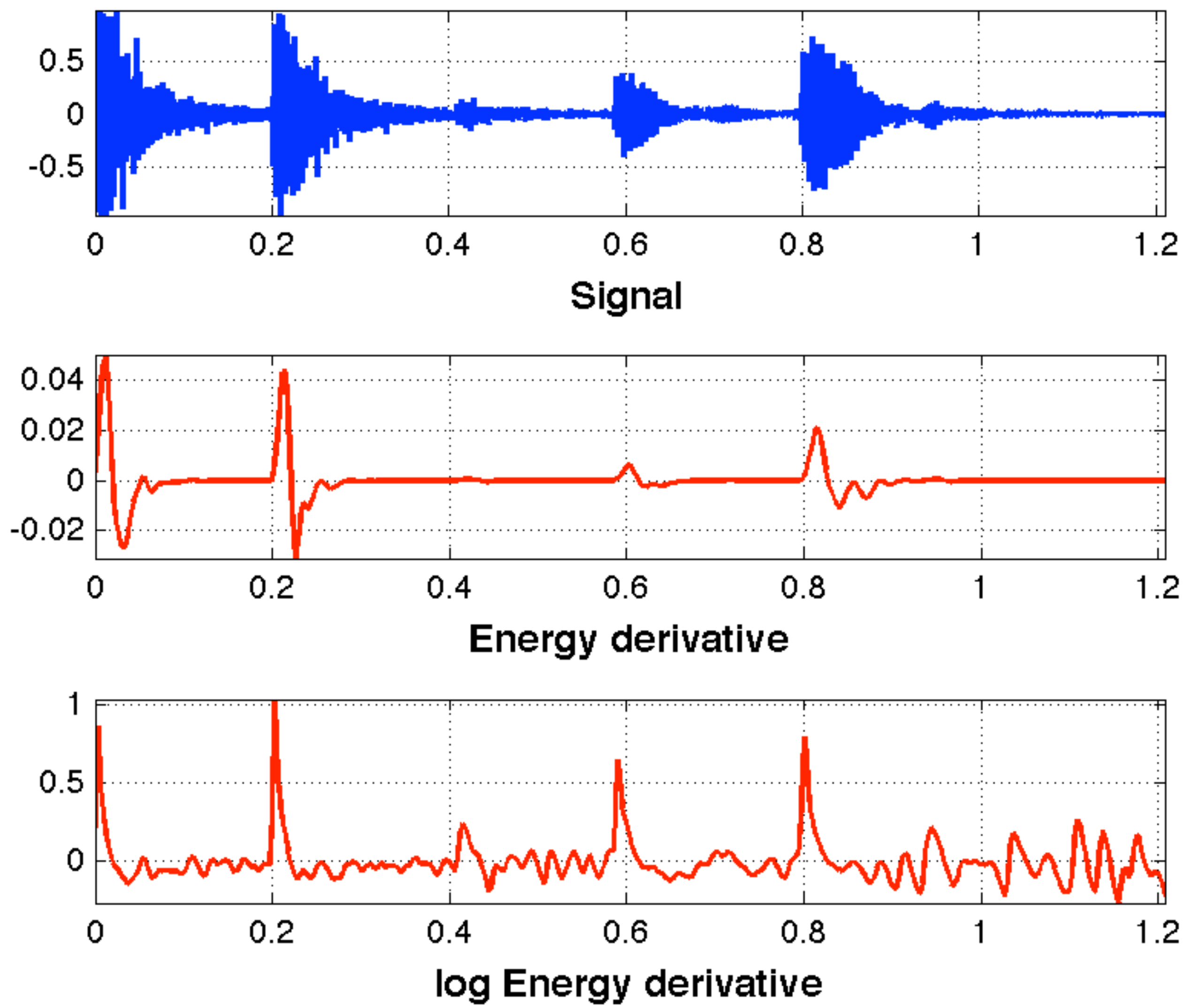
- We can use the derivative of energy w.r.t. time -> sharp peaks during energy rise

- Detectable changes in loudness are proportional to the overall loudness of the sound.

$$\frac{\partial E(m)/\partial m}{E(m)} = \frac{\partial log(E(m))}{\partial m}$$

- Simulates the ear's perception of loudness (Klapuri, 1999)

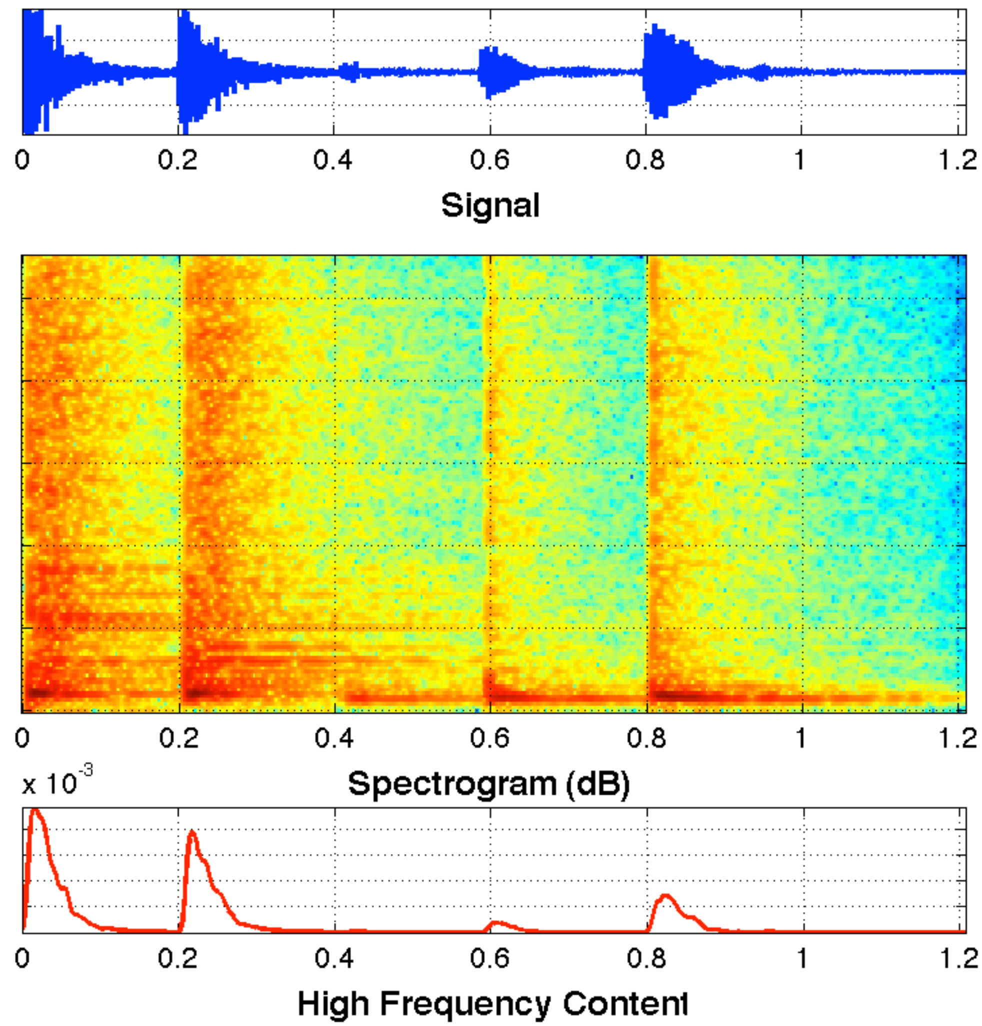
Time-domain


0.5
0
-0.5
0
0.2
0.4
0.6
0.8
1
1.2
Signal
0.04
0.02
0
-0.02
0
0.2
0.4
0.6
0.8
1
1.2
Energy derivative
1
0.5
0
0
0.2
0.4
0.6
0.8
1
1.2
log Energy derivative

Frequency-domain

- Impulsive noise in time -> wide band noise in frequency
- More noticeable in high frequencies.
- Linear weighting of Energy

$$HFC(m) = \frac{2}{N} \sum_{k=0}^{N/2} |X_k(m)|^2 k$$



Frequency-domain

- We can also measure change (flux) in spectral content (Duxbury, 02).

$$SF(m) = \frac{2}{N} \sum_{k=0}^{N/2} \left(|X_k(m)| - |X_k(m-1)| \right)$$

- Use half-wave rectification to only take energy increases into account

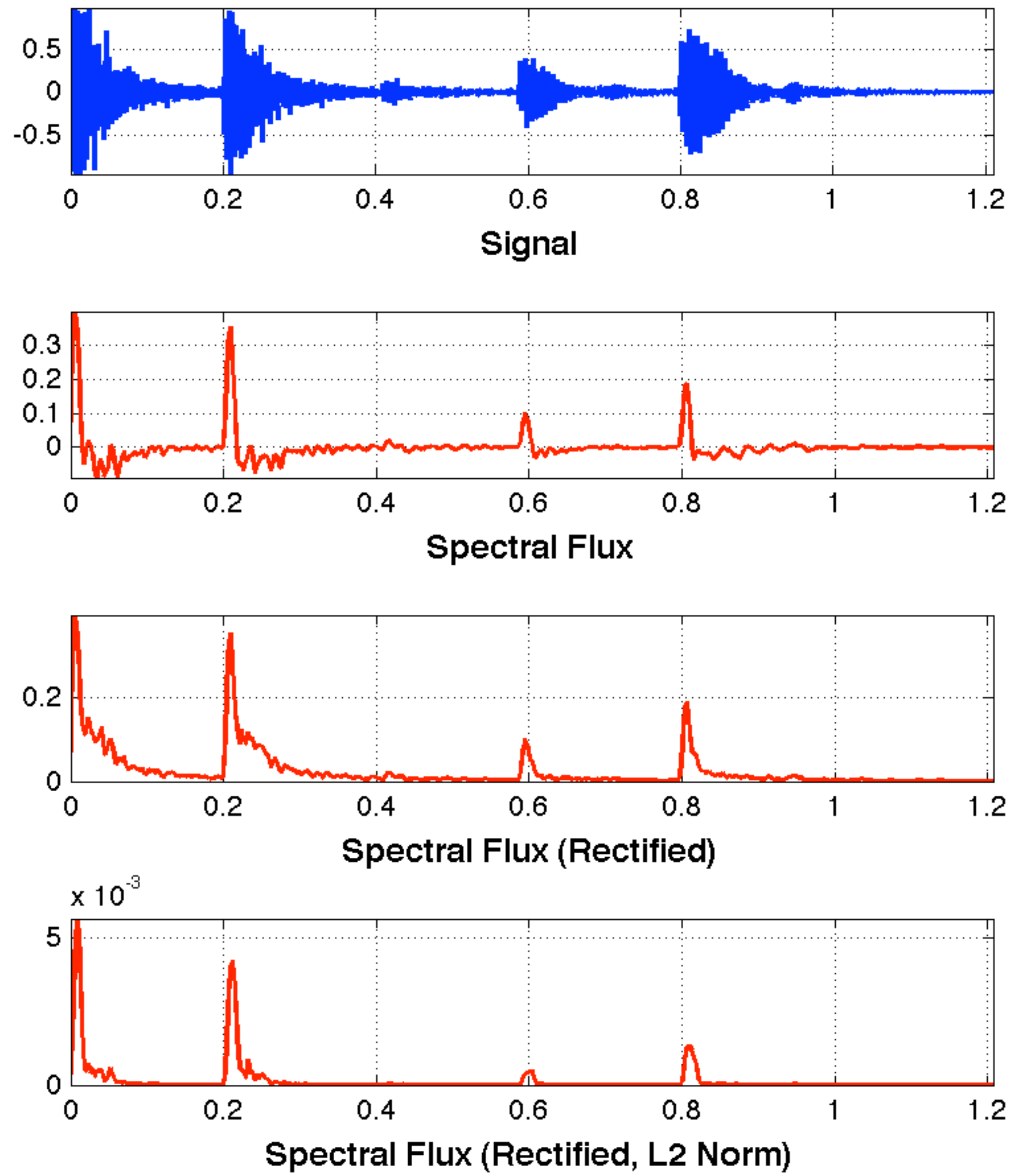
$$SF_R(m) = \tfrac{2}{N} \textstyle\sum_{k=0}^{N/2} H\left(|X_k(m)| - |X_k(m-1)| \right)$$

$$H(x) = (x + |x|)/2$$

- Use the (squared) L_2 norm of the flux vector

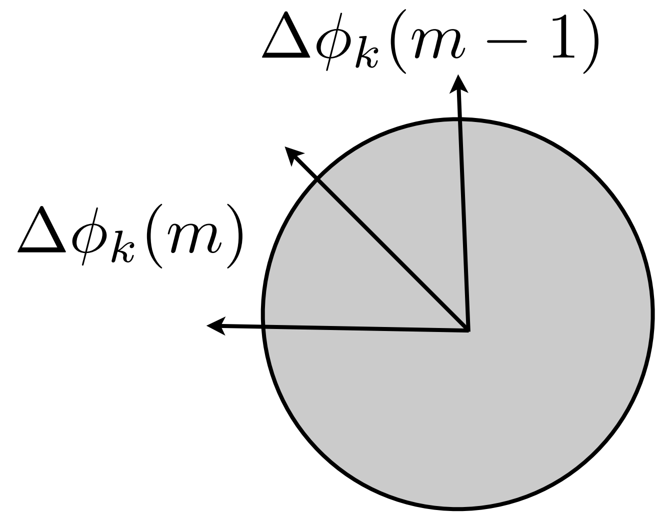
$$SF_{RL_2}(m) = \frac{2}{N} \sum_{k=0}^{N/2} \left[H\left(|X_k(m)| - |X_k(m-1)| \right) \right]^2$$

Spectral Flux



Phase deviation

- The change in instantaneous frequency in bin k can be used to detect onsets (Bello, 03)



$$PD(m) = \frac{2}{N}\sum_{k=0}^{N/2}|\phi_k''(m)| = \frac{2}{N}\sum_{k=0}^{N/2}|\Delta\phi_k(m) - \Delta\phi_k(m-1)|$$

$$= \boxed{\frac{2}{N}\sum_{k=0}^{N/2}|princarg\left(\phi_k(m) - 2\phi_k(m-1) + \phi_k(m-2)\right)|}$$

Phase deviation

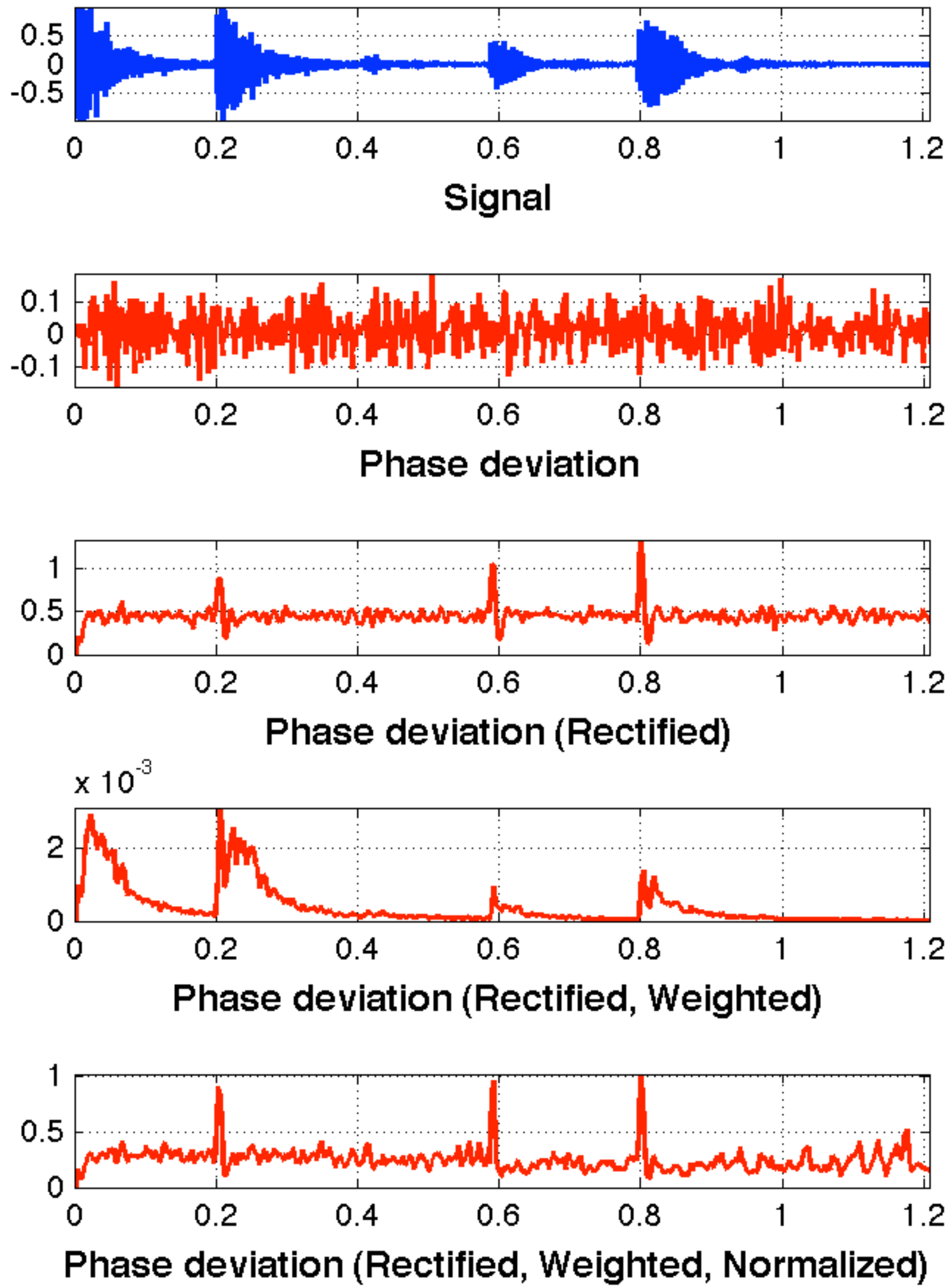
- This function can be improved by weighting frequency bins by their magnitude (Dixon, 06):

$$PD_W(m) = \frac{2}{N} \sum_{k=0}^{N/2} |X_k(m) \phi_k''(m)|$$

- and normalized:

$$PD_{WN}(m) = \frac{\sum_{k=0}^{N/2} |X_k(m) \phi_k''(m)|}{\sum_{k=0}^{N/2} |X_k(m)|}$$

Phase deviation



Complex domain

- We can combine the spectral flux and phase deviation strategies, such that:

$$\hat{X}_k(m) = |\hat{X}_k(m)| e^{j\hat{\phi}_k(m)}$$

where,

$$|\hat{X}_k(m)| = |X_k(m-1)|$$

$$\hat{\phi}_k(m) = princarg\left(2\phi_k(m-1) - \phi_k(m-2)\right)$$

such that,

$$\boxed{CD(m) = \frac{2}{N} \sum_{k=0}^{N/2} |X_k(m) - \hat{X}_k(m)|}$$

Complex domain

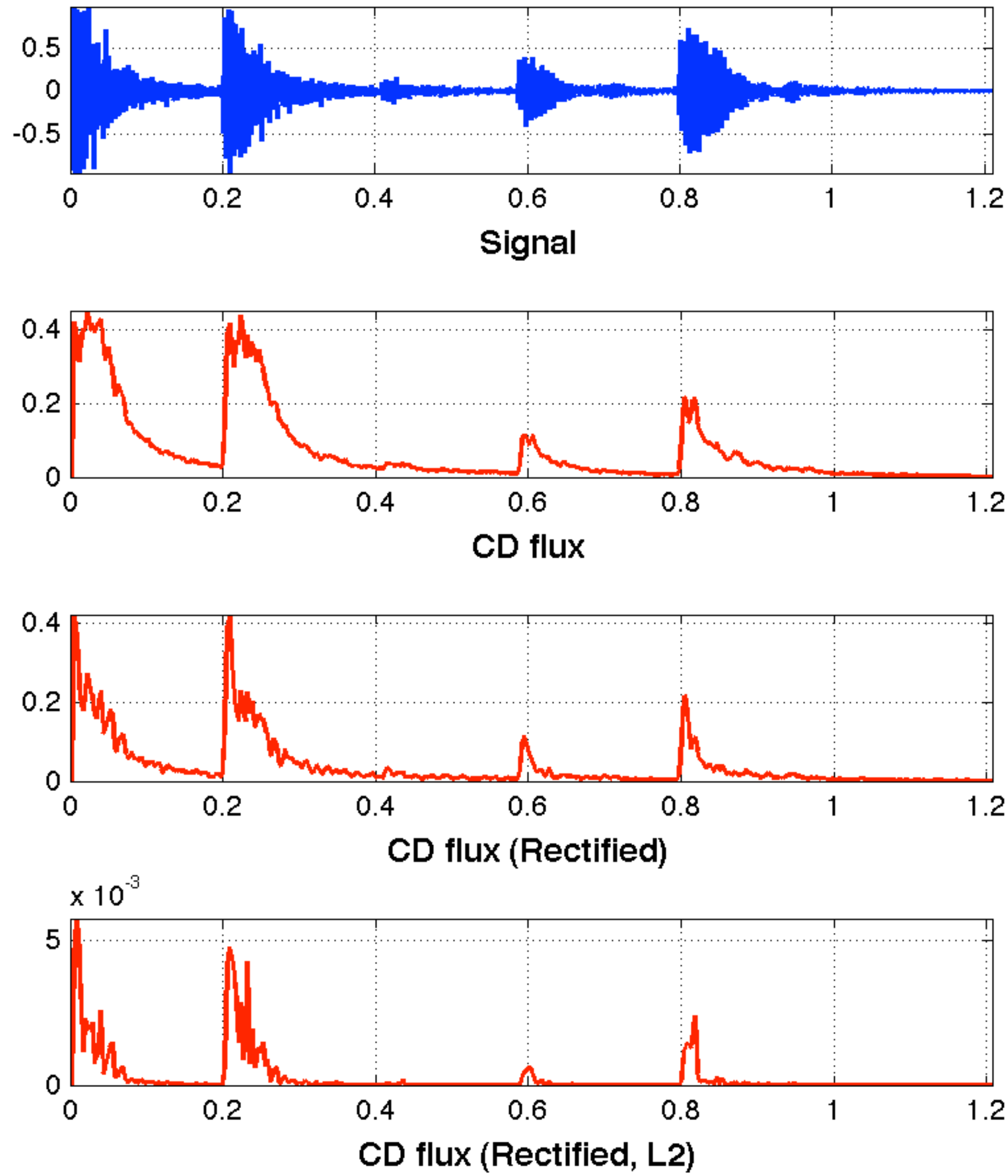
- As before, we can use half-wave rectification to improve the function (Dixon, 06):

$$CD(m) = \tfrac{2}{N} \textstyle\sum_{k=0}^{N/2} RCD_k(m)$$

where,

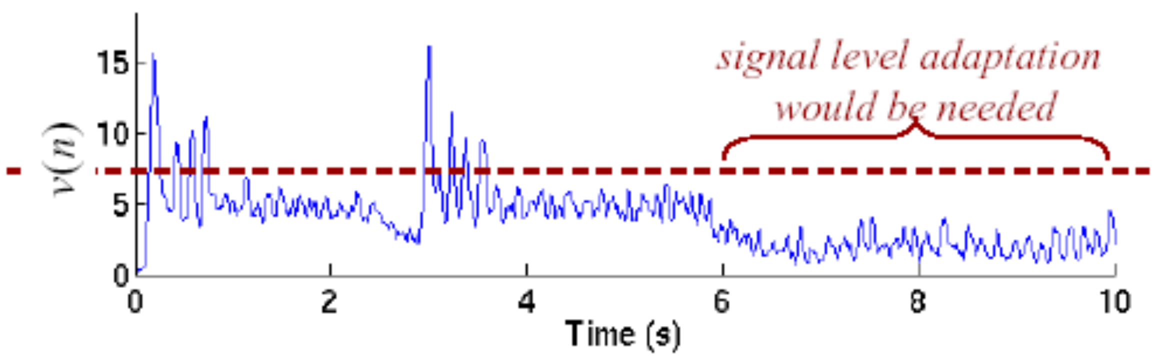
$$RCD_k(m) = \begin{cases} |X_k(m) - \hat{X}_k(m)| & \text{if } |X_k(m)| \geq |X_k(m-1)| \\ 0 & \text{otherwise} \end{cases}$$

Complex domain



Peak picking

- The function is post-processed to facilitate peak picking:
  - Smoothing -> decrease jaggedness
  - Normalization -> generalization of threshold values
  - Thresholding -> eliminate spurious peaks



Peak picking

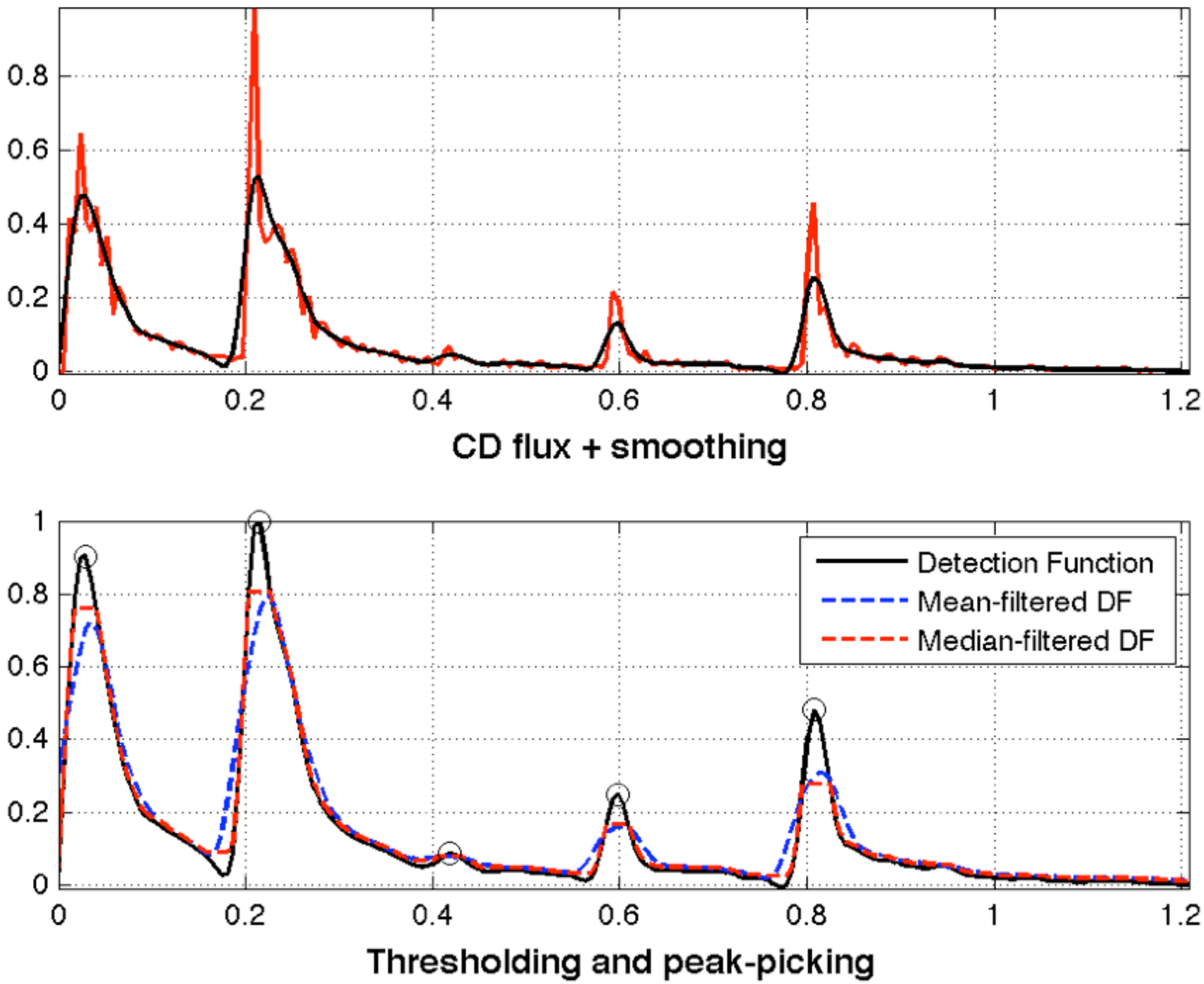
- Adaptive thresholding is a more robust choice, typically defined as:

$$\delta(m) = \alpha + f(\bar{m}), \quad m - \beta L \leq \bar{m} \leq m + L$$

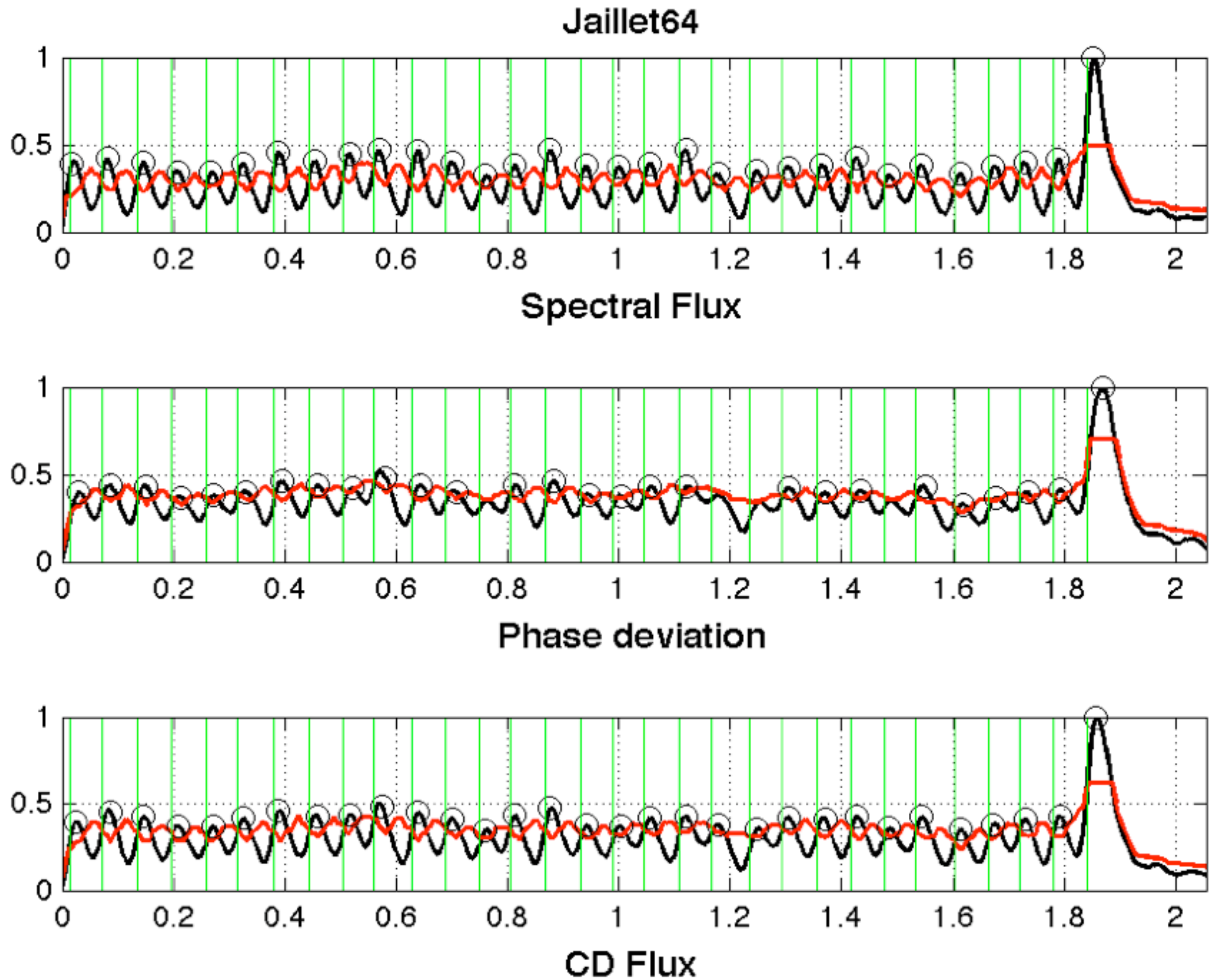
- where f is a function, e.g. the local mean or median, of the detection function; $\boldsymbol{\beta}$ increases the window length before the peak; and $\boldsymbol{\alpha}$ is an offset value

- Peak picking reduces to selecting local maxima above the threshold

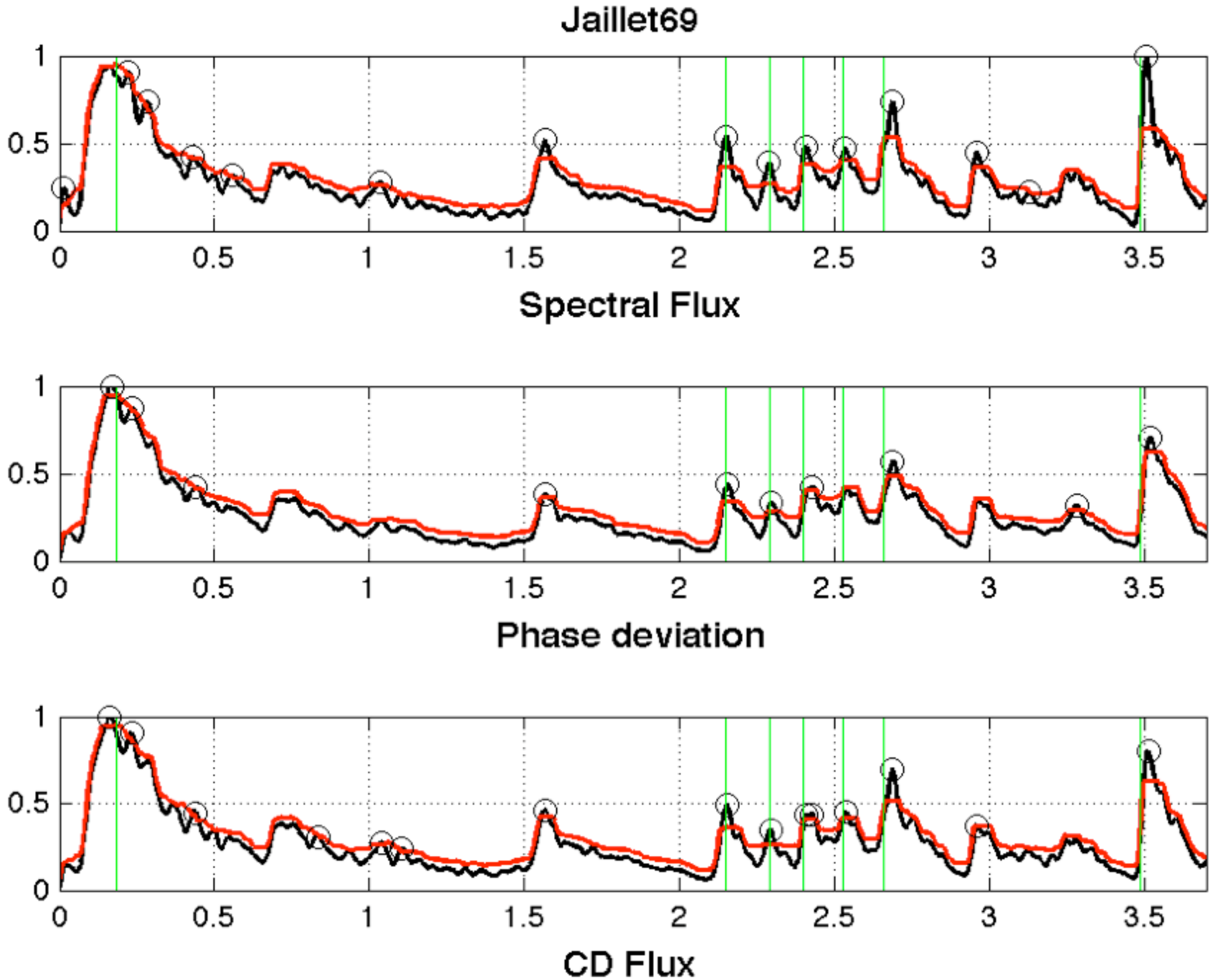
Peak picking



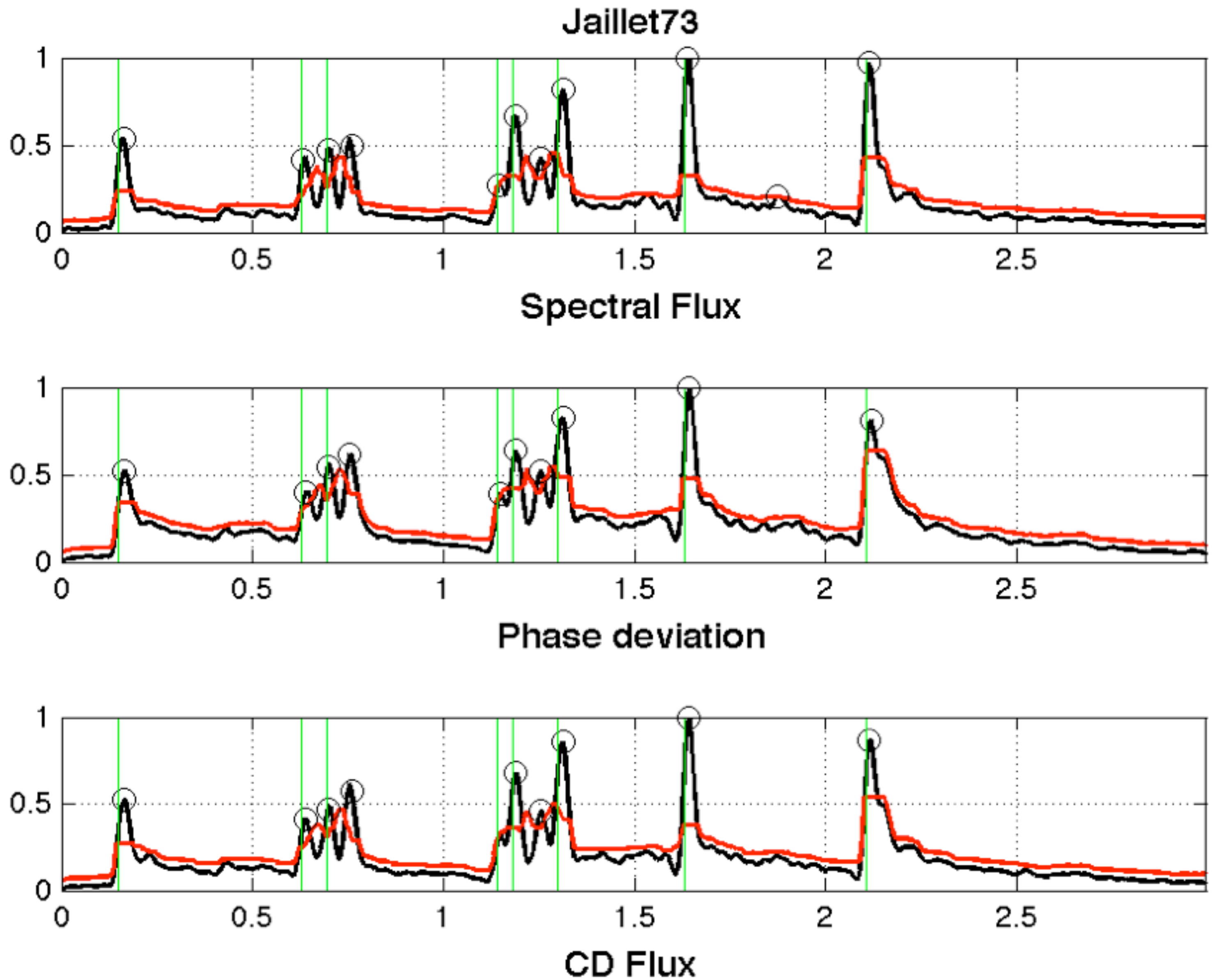
Comparing detection functions



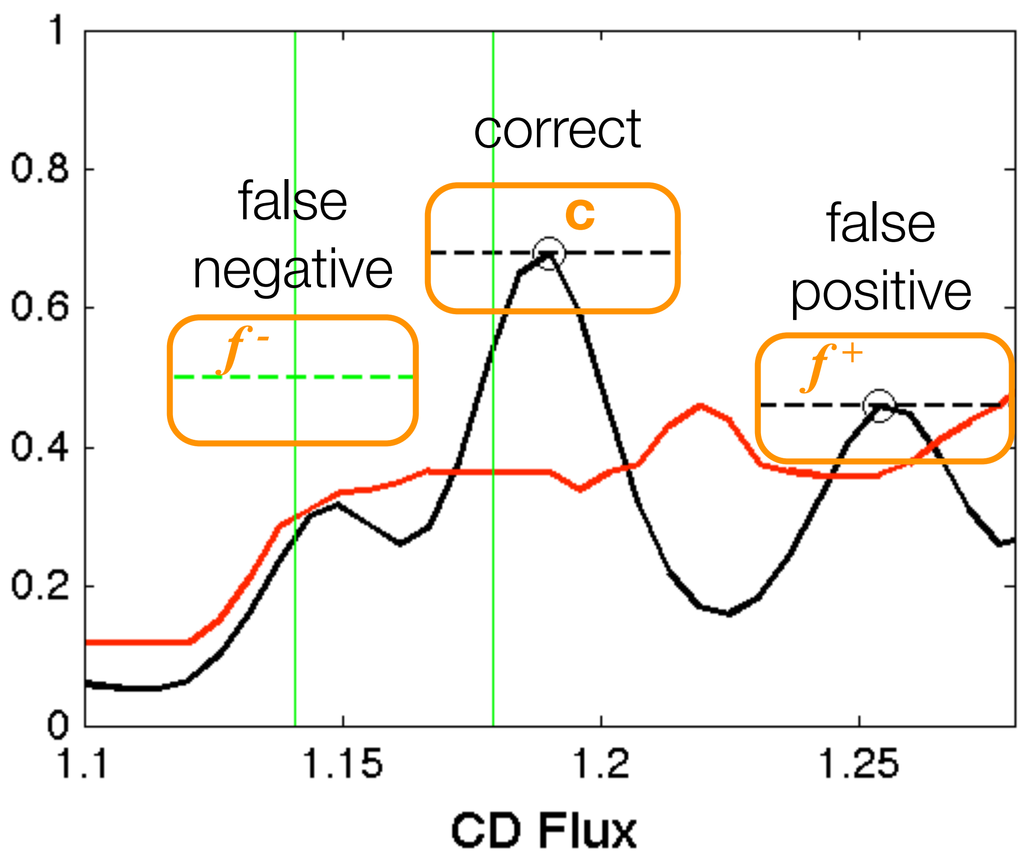
Comparing detection functions



Comparing detection functions



Benchmarking



$$P = \frac{c}{c + f^+}$$

$$R = \frac{c}{c + f^-}$$

$$F = \frac{2PR}{P + R}$$

References

- Dixon, S. “Onset Detection Revisited”. Proceedings of the 9th International Conference on Digital Audio Effects (DAFx06), Montreal, Canada, 2006.
- Collins, N. “A Comparison of Sound Onset Detection Algorithms with Emphasis on Psycho-Acoustically Motivated Detection Functions”. Journal of the Audio Engineering Society, 2005.
- Bello , J.P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M. and Sandler, M.B. “A tutorial on onset detection in music signals”. IEEE Transactions on Speech and Audio Processing. 13(5), Part 2, pages 1035-1047, September, 2005.
- Bello , J.P., Duxbury, C., Davies, M. and Sandler, M. On the use of phase and energy for musical onset detection in the complex domain. IEEE Signal Processing Letters. 11(6), pages 553-556, June, 2004.
- Klapuri, A. “Sound Onset Detection by Applying Psychoacoustic Knowledge”. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Phoenix, Arizona, 1999.