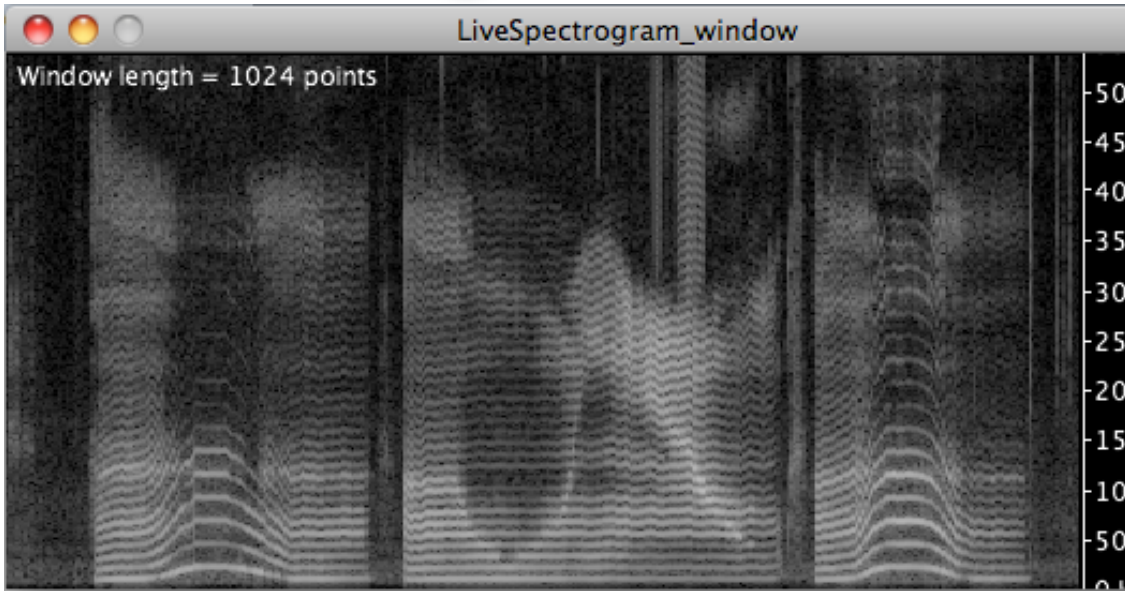
Lecture 8: Pitch Tracking

1. Pitch Tracking
2. Spectral Approaches
3. Time Domain
4. Example algorithms

Dan Ellis
Dept. Electrical Engineering, Columbia University
dpwe@ee.columbia.edu http://www.ee.columbia.edu/~dpwe/e4896/

1. Pitch Tracking

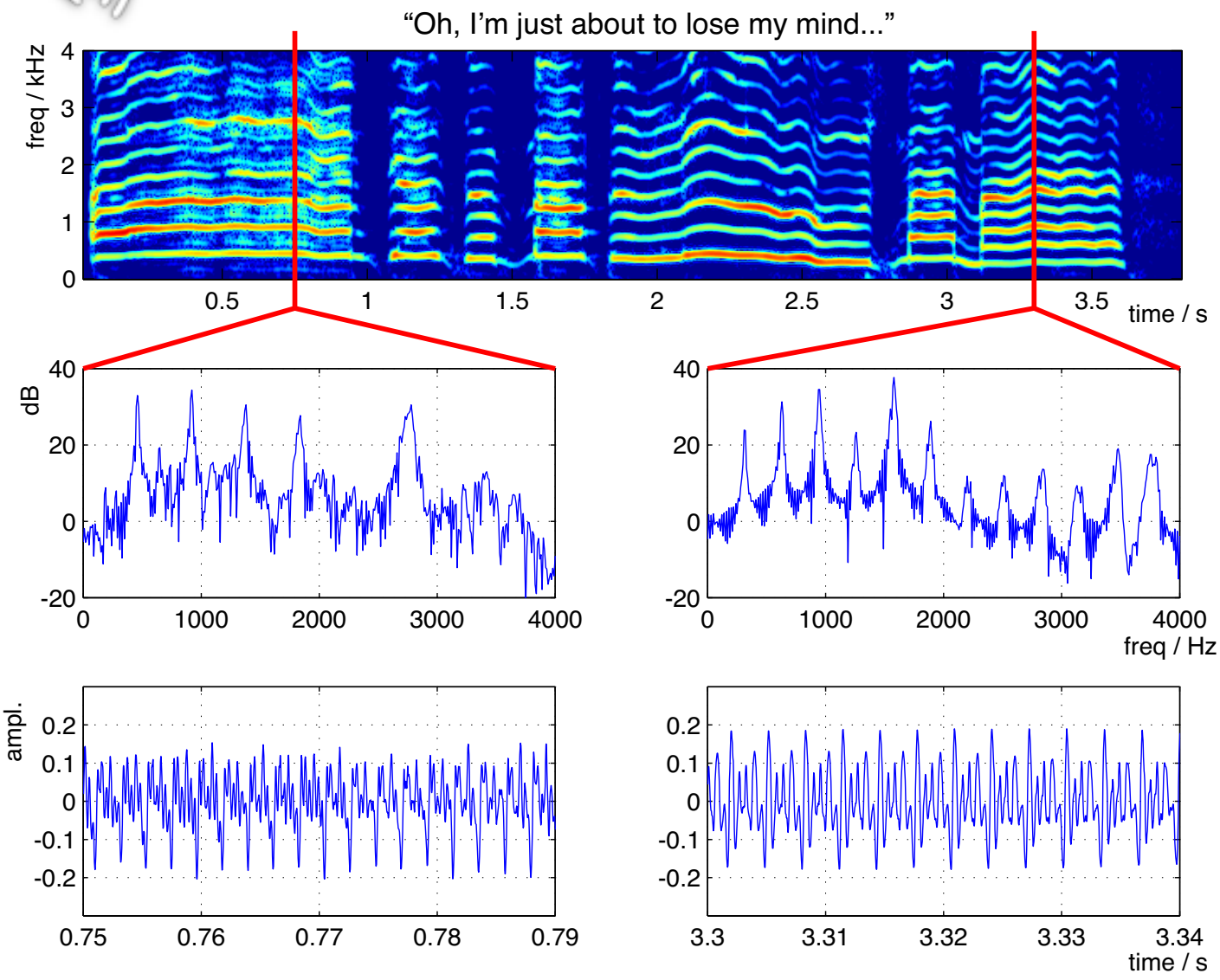


- **Pitch is a big part of hearing**
  - orthogonal to formants (vowels) in speech
  - follows fundamental frequency (f_0) of sounds
  - because of speech? because of periodicity?
- **Pitch extraction (tracking) is useful**
  - for coding / representation (telephony)
  - for extracting information
- **.. but can be tricky**
  - sounds without clear f_0

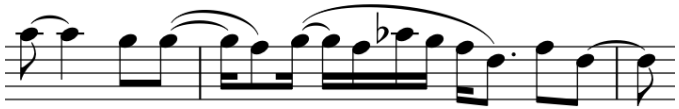
Challenges to Pitch Extraction

Talkin, 1995

- Pitch extraction challenges include...

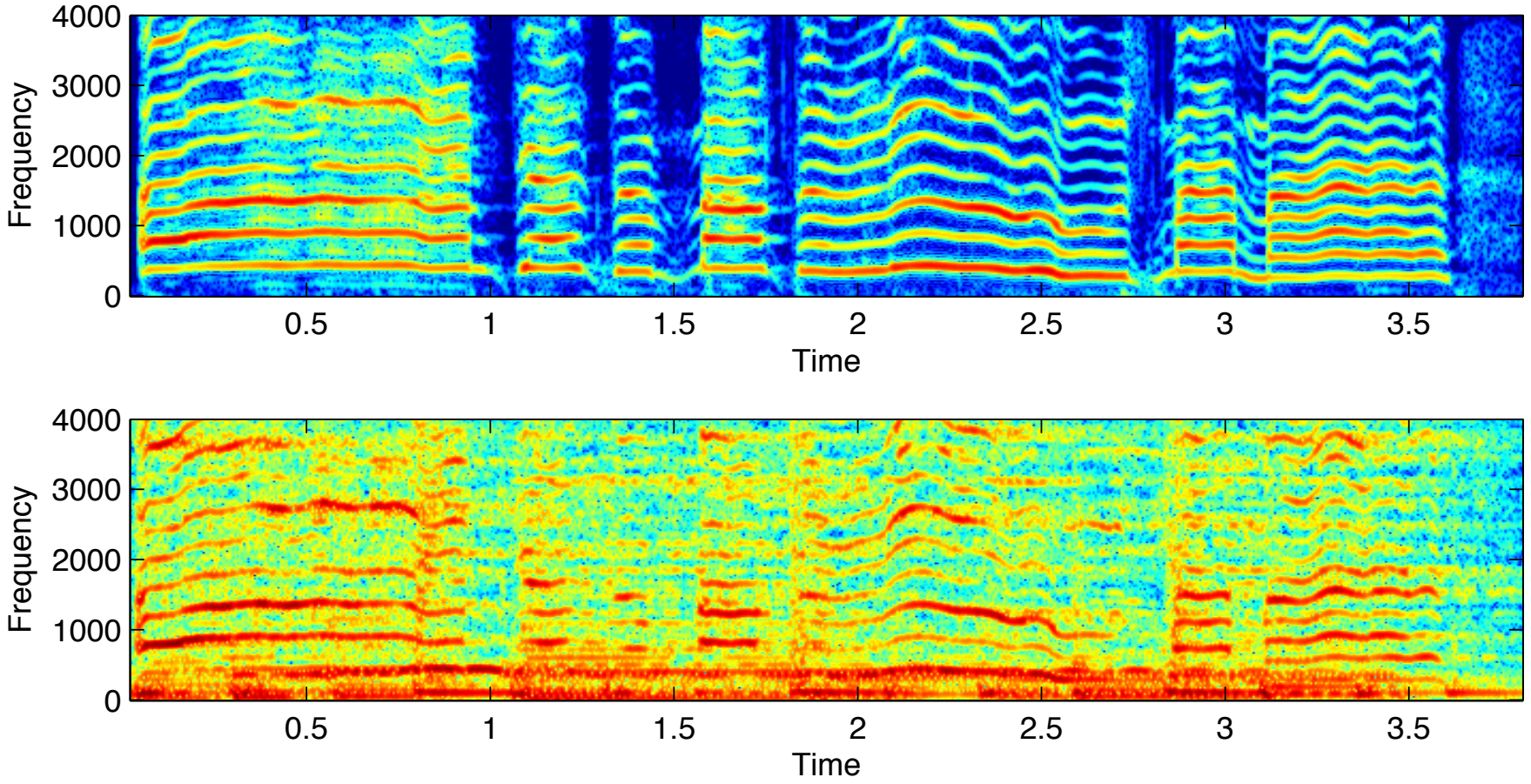


- noise / multiple f_0s
- Voice Activity Detection
- unclear f_0
- note segmentation



Multiple f_0s

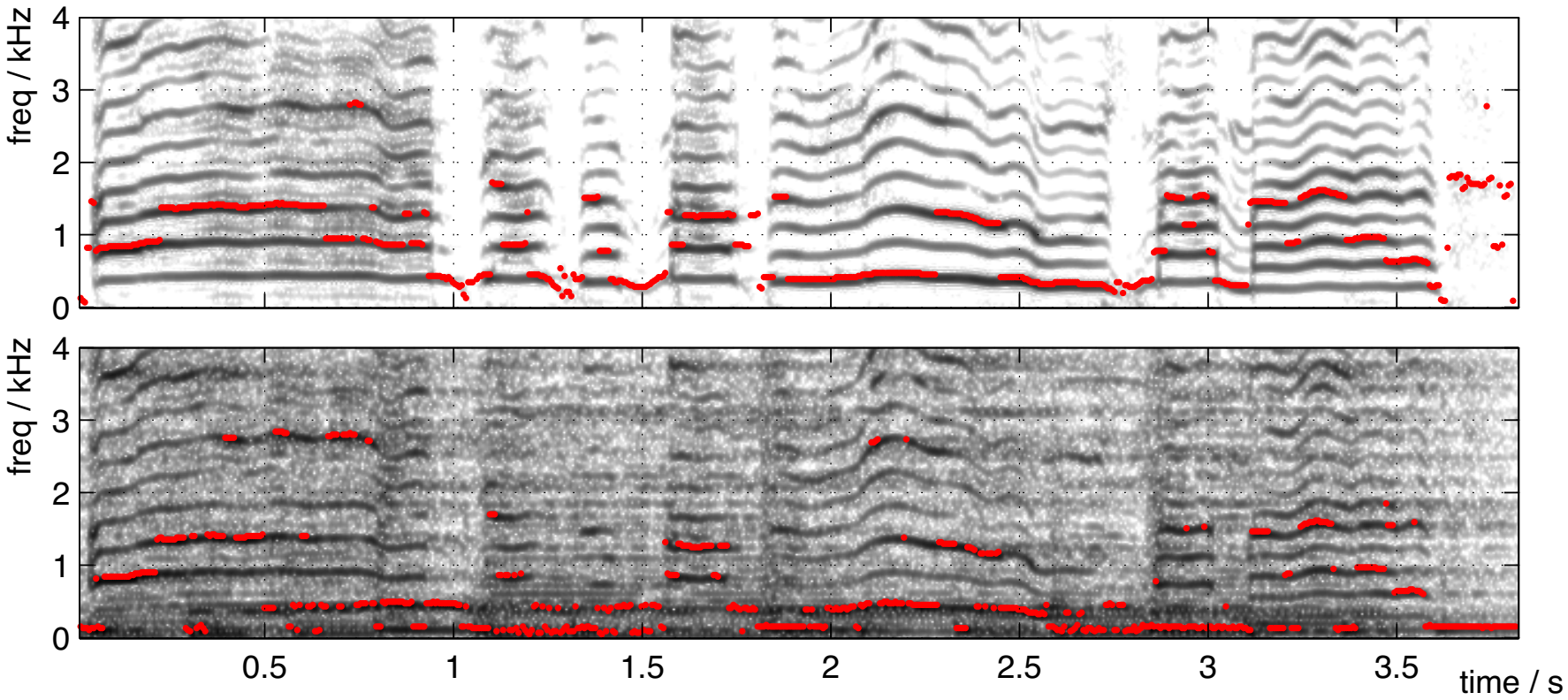
- Common in music ("polyphony")



- cross-terms...
- identify-remove-repeat?

2. Spectral Approaches

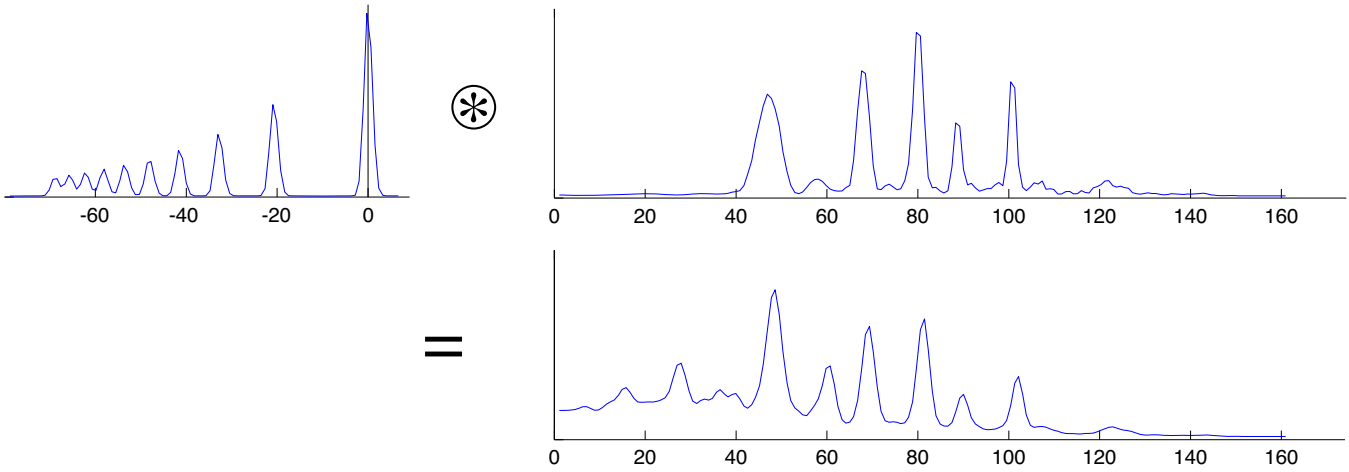
- Just pick the most intense spectral peak?

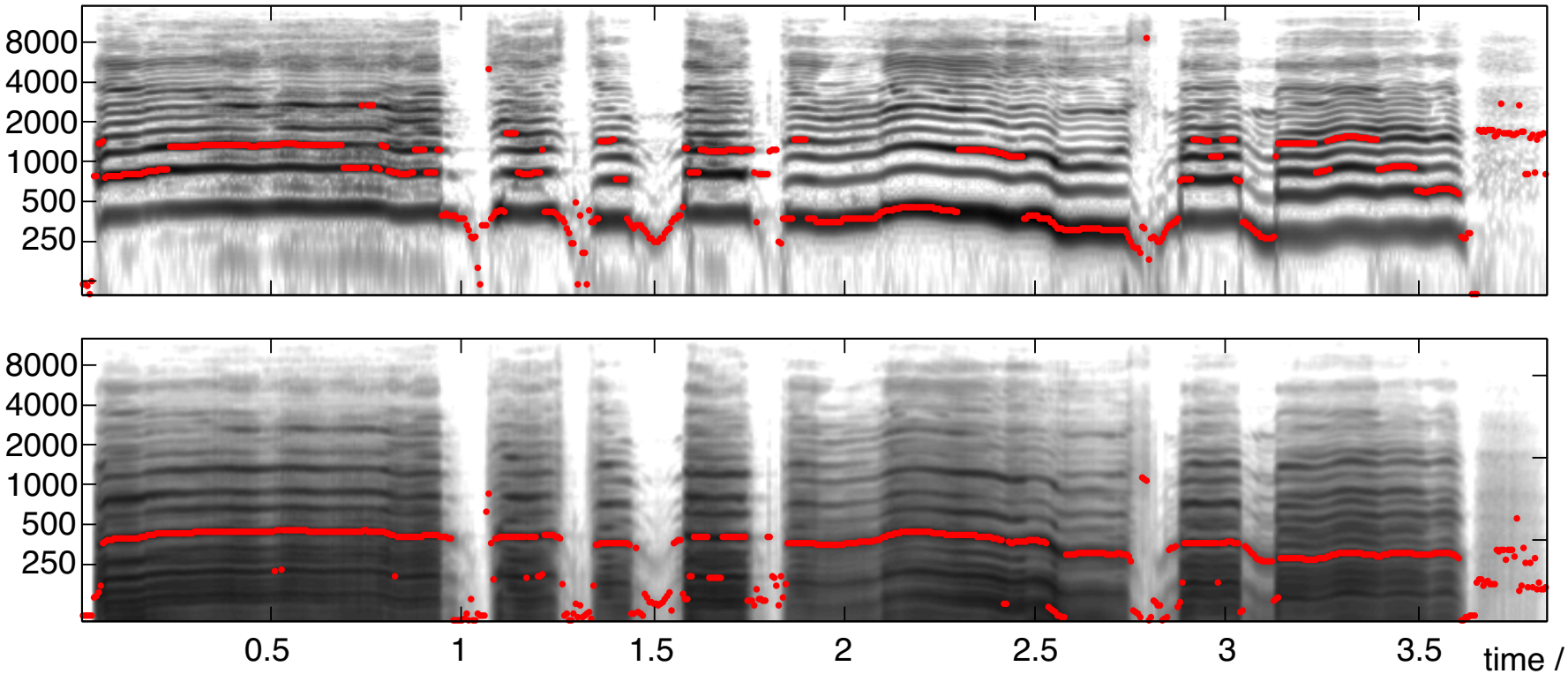


- fundamental may not be strongest...

Spectral Templates

- "Filter" log-freq spectrum with template
  - matched filter ...

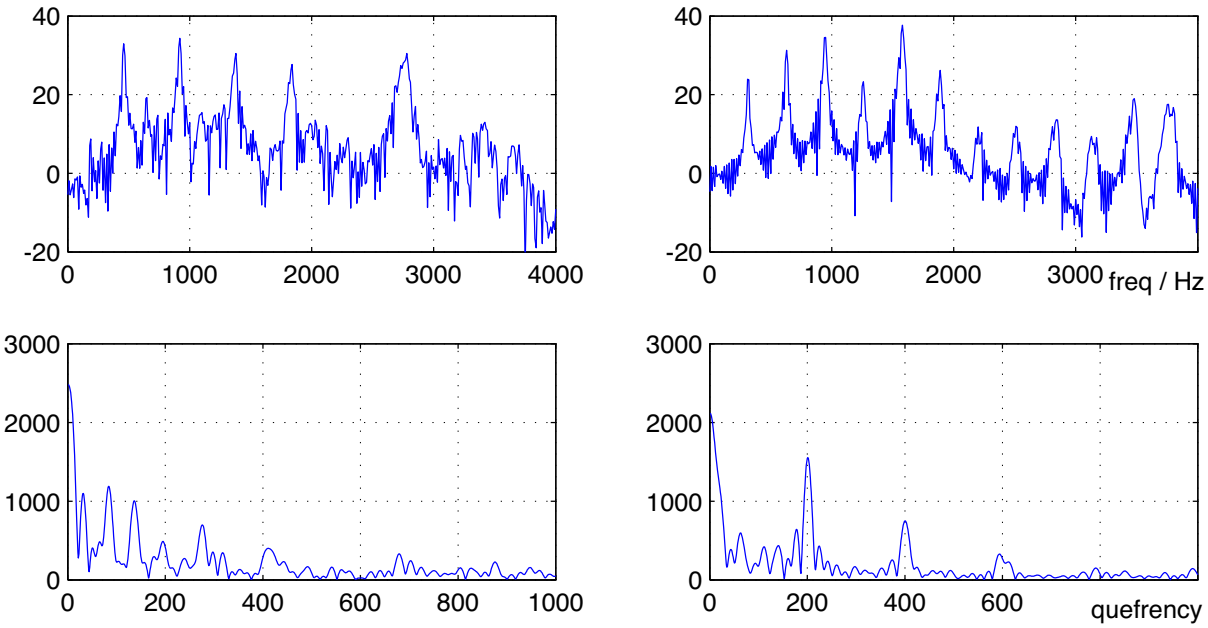


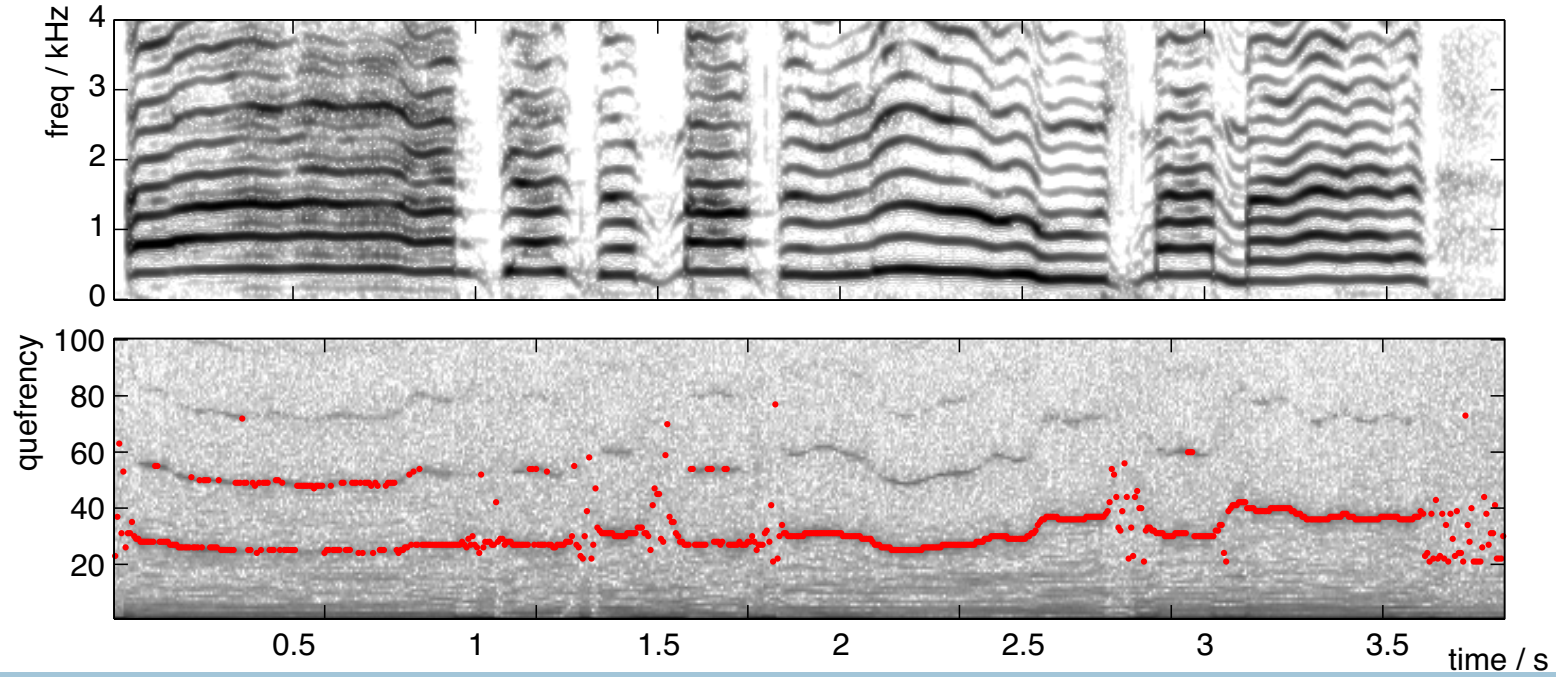


Cepstrum

Noll, 1967

- Spectrum shows periodicity at f_0
  - reveal via Fourier transform of log|Y| = cepstrum

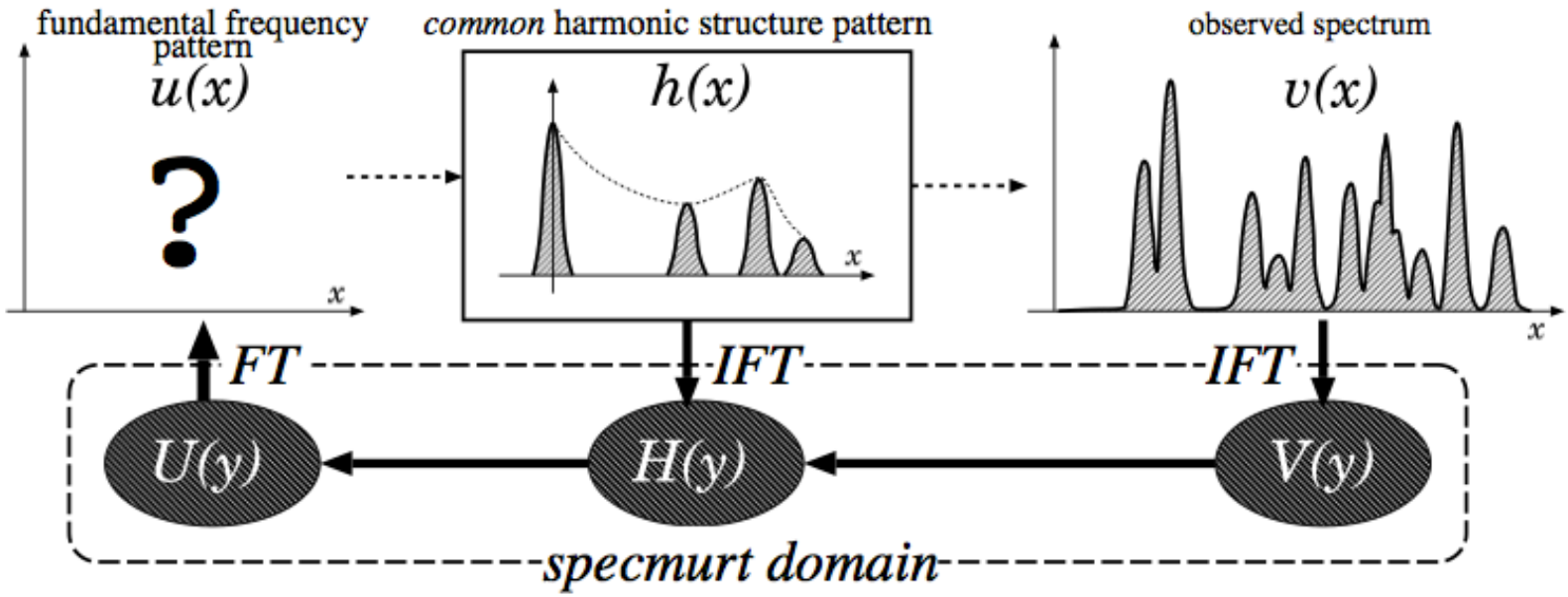




"Specmurt"

Sagayama et al., 2004

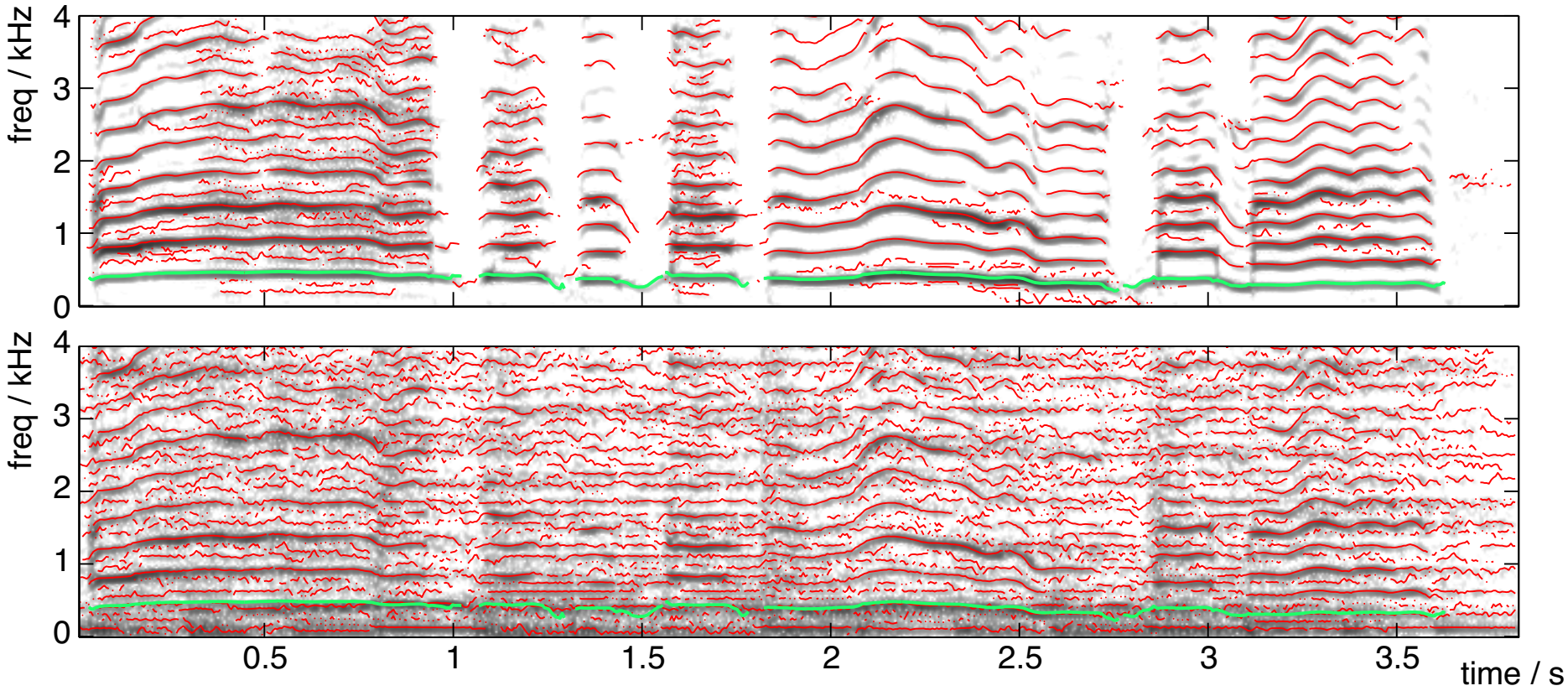
- Log-spectrum is like convolution of pitch peaks and harmonics profile
  - esp. for multiple notes by one instrument (piano)
- Separate out by deconvolution
  - inverse Fourier transform of log-f spectrum $V(y)$
  - divide out harmonic template $H(y)$



Sinusoid Tracks

Maher & Beauchamp 1994

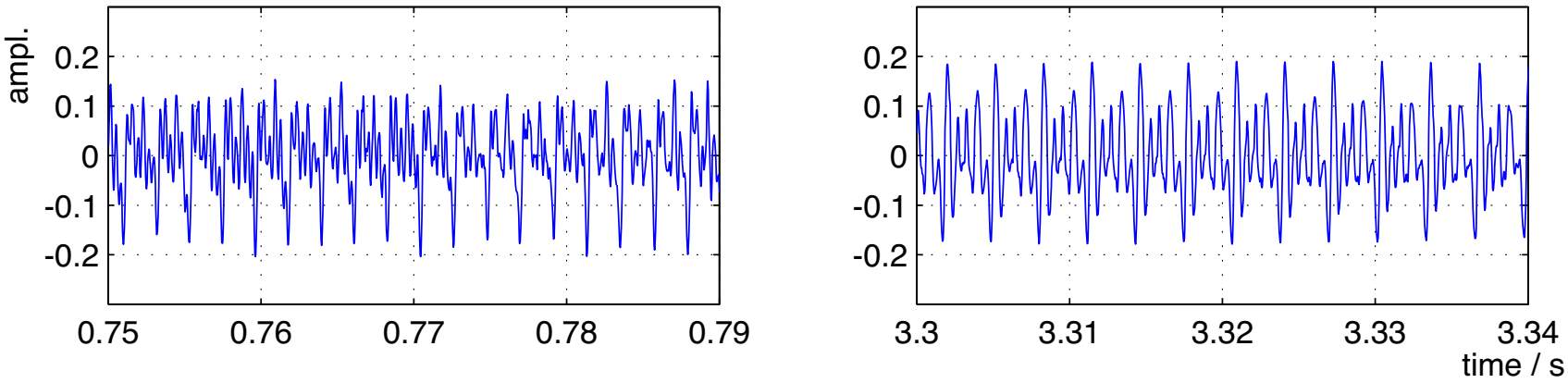
- Do sinusoid tracking before pitch tracking
  - apply templates / sieve on tracks
  - consider each strong sinusoid as a candidate
  - limits to clearly periodic energy (normalized?)



3. Time Domain Approaches

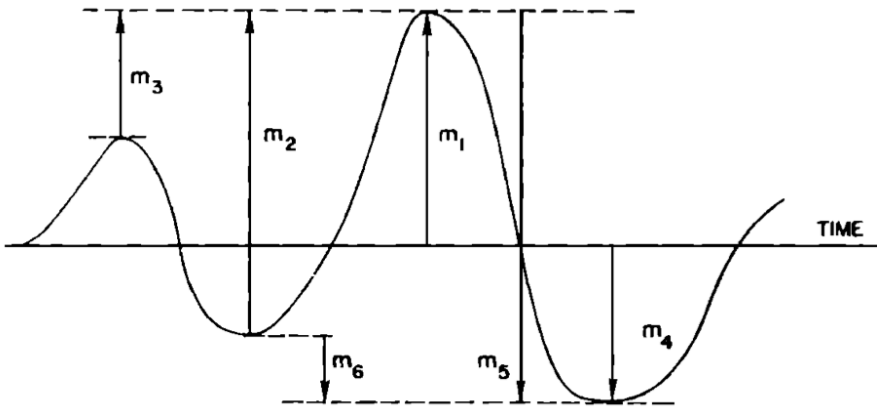
Gold & Rabiner 1969

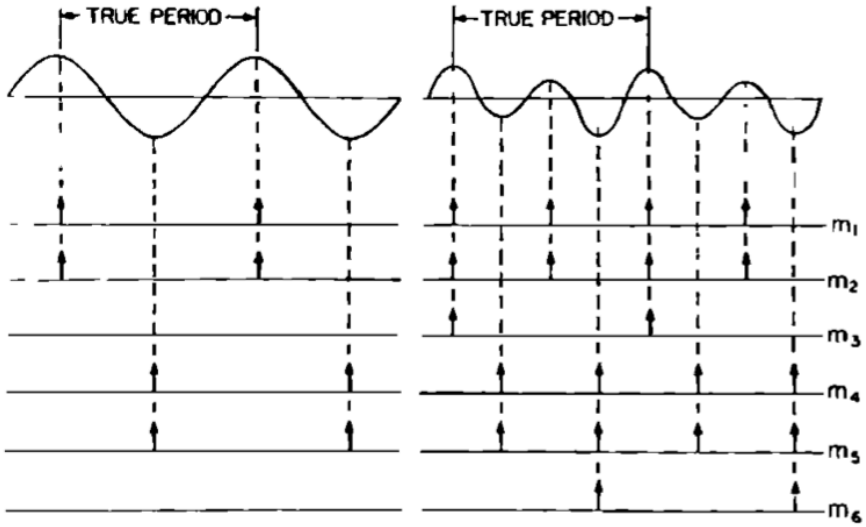
- Periodicity is primarily time-domain



  - time-domain processing avoids FFT

- Early approaches: Look for common gap between wave features

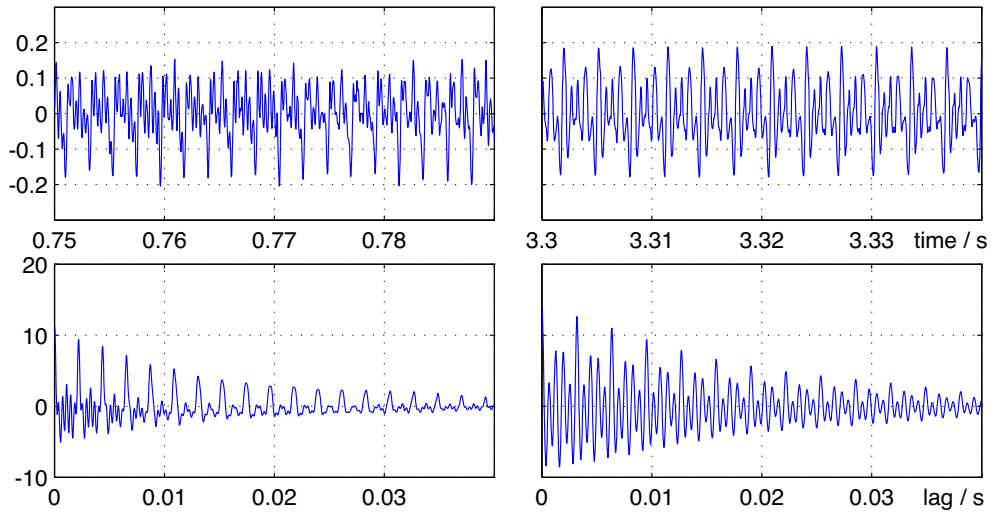


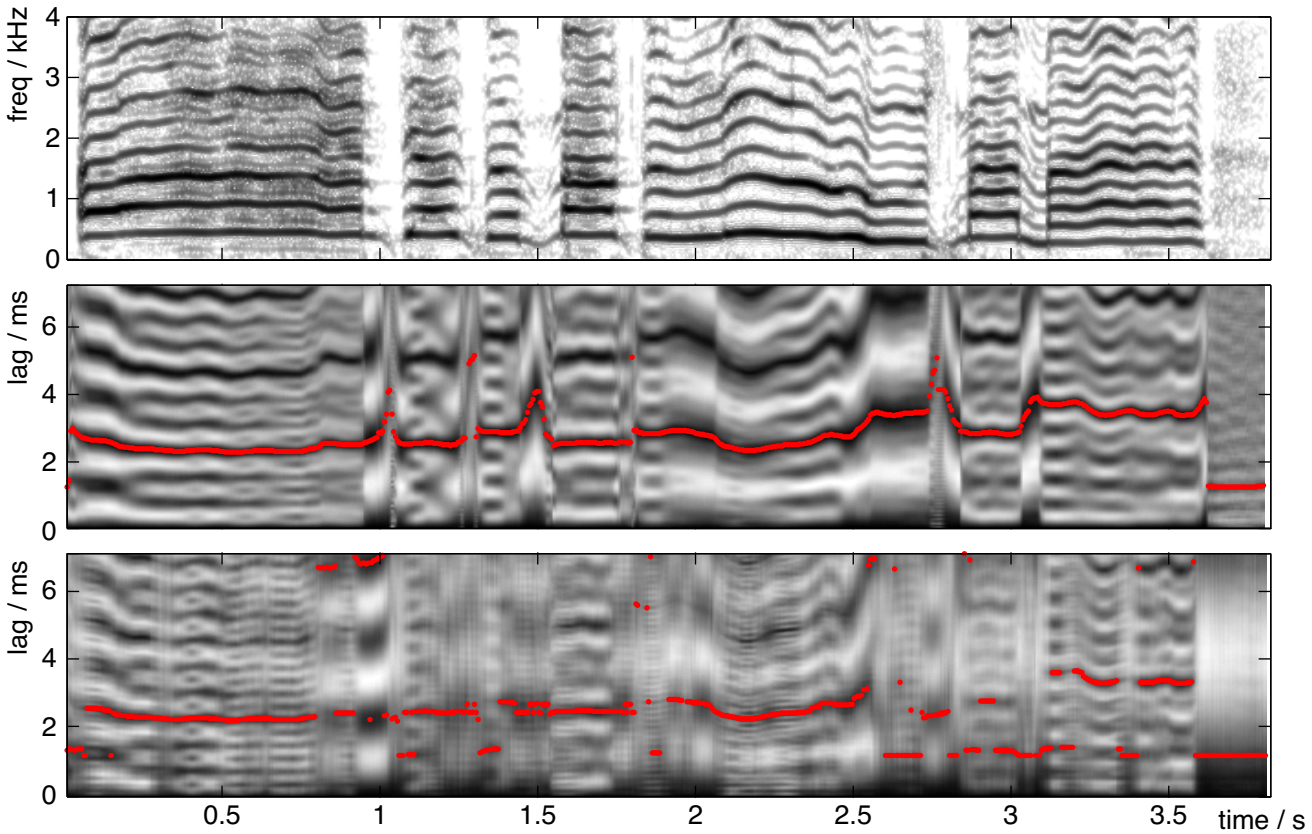


Autocorrelation

Sondhi 1967

- Directly measures period of self-similarity
  - (autoco is inverse FT of magnitude spectrum)

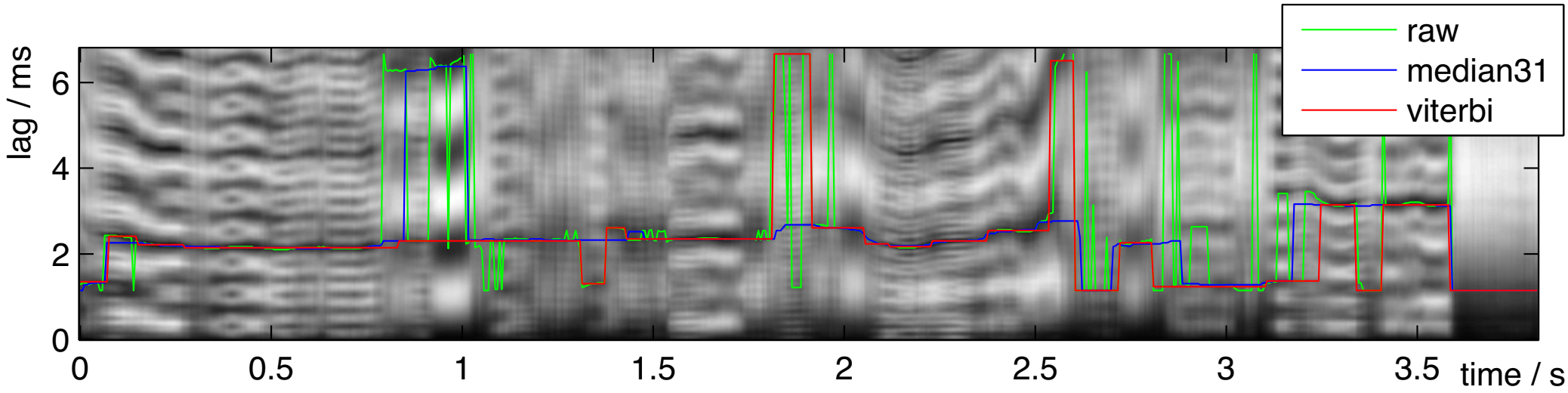




Post-processing

Talkin, 1995

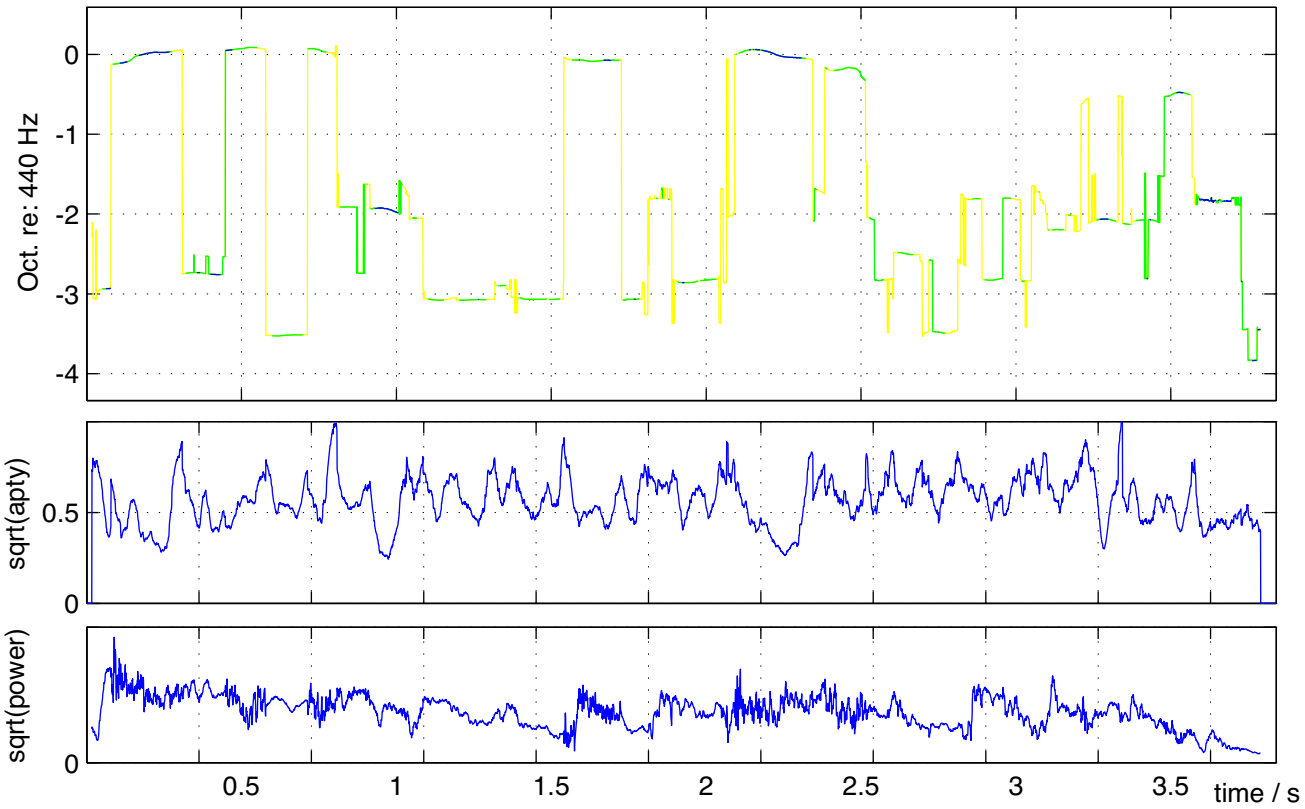
- Can assume & exploit pitch smoothness
- Median filtering
  - only knows best point
- Dynamic programming (Viterbi path)
  - can look at all underlying values



4. Examples: YIN

de Cheveigné & Kawahara, 2002

- **Like autocorrelation, but**
  - difference function - robust to changes in amplitude
  - progressive normalization by cumulative average
    - eliminates lag = 0 advantage
  - outperforms others without post processing

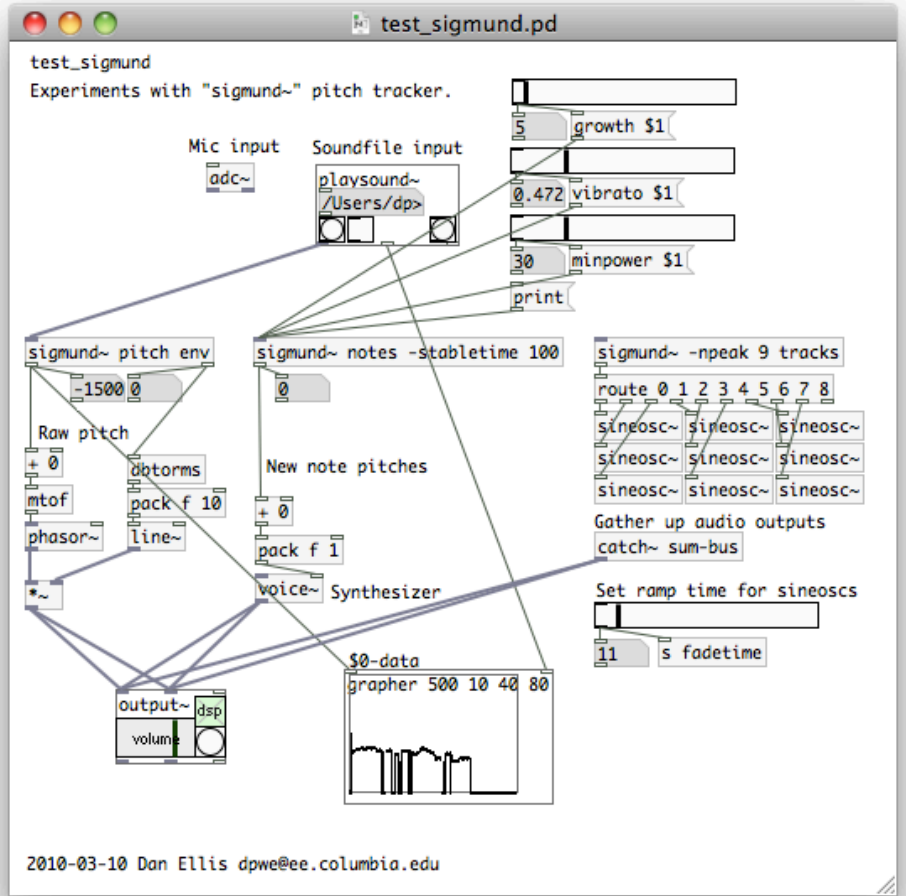


Examples: `sigmund~`

Puckette et al. 1998

- Miller Puckette's Pd pitch tracker
- Sinusoids + sieve $\mathcal{L}(f) = \sum_k \frac{\alpha}{\alpha + k} \hat{Y}(k \cdot f)$

- Also performs note segmentation
  - and polyphony



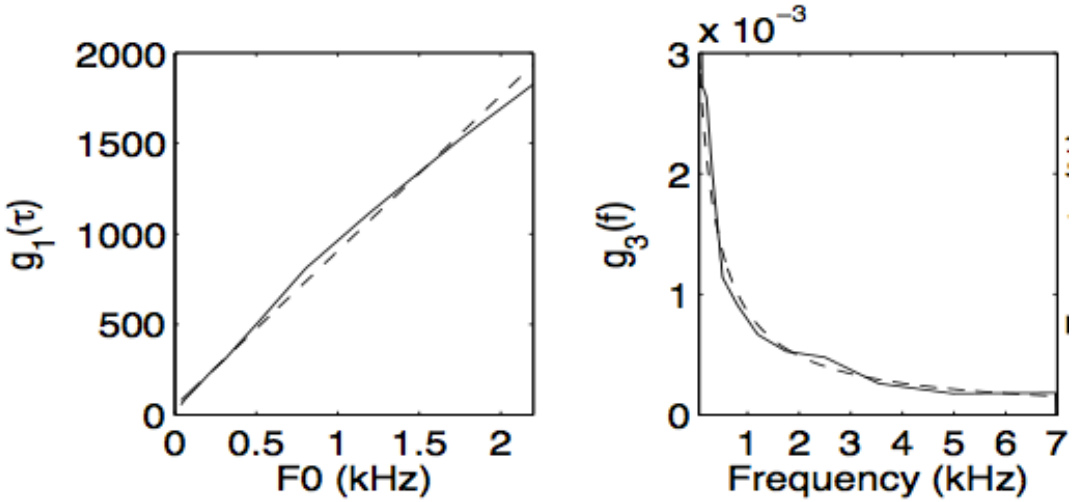
Examples: Weighted Sieve Polyphonic

Klapuri 2006

- Based on more complex multi-f_0 system
- Chooses f_0 via weighted sieve
  - spectrum is whitened before measuring peaks
  - weights are optimized for training data

$$s(\tau) = \sum_{m=1}^{M} g(\tau, m)|Y(f_{\tau,m})|$$

$$g(\tau, m) = g_1(\tau) g_3(f_{\tau,m})$$

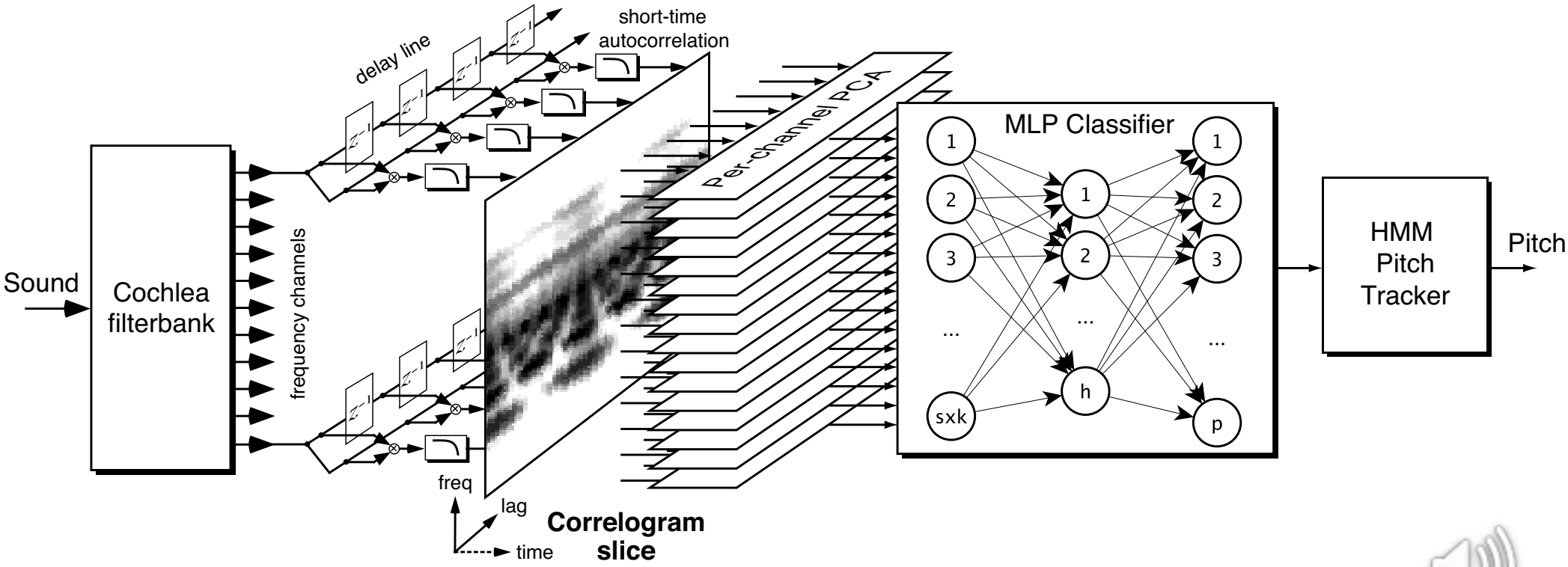


- Multiple f_0s are found via cancel+repeat, or joint optimization $G(\mathcal{I}) = \sum_{i \in \mathcal{I}} \sum_{m} g(\tau_i, m)|Y(k_{i,m})| \prod_{j \in \mathcal{I} \setminus i} (1 - Z_j(k_{i,m}))$

Example: SAcC

Lee & Ellis 2012

- Subband Autocorrelation Classification
  - multiple subbands
  - autocorrelation + PCA
  - trained MLP classifier to find pitch



Summary

- Pitch
  Important perceptually, for speech & music

- Spectrum
  Combine multiple harmonics to f_0

- Time
  Autocorrelation directly measures periodicity

References

- A. de Cheveigné & H. Kawahara, "YIN, a Fundamental Frequency Estimator for Speech and Music," *J. Acoust. Soc. Am.* 111, 1917–1930, 2002.
- B. Gold & L. Rabiner, "Parallel Processing Techniques for Estimating Pitch Periods of Speech in the Time Domain," *J. Acoust. Soc. Am.* 46(2), 442-448, 1969.
- A. Klapuri, "Multiple Fundamental Frequency Estimation by Summing Harmonic Amplitudes," *Proc. Int. Symp. Music IR*, 216-221, 2006.
- B. S. Lee & D. Ellis, "Noise Robust Pitch Tracking by Subband Autocorrelation Classification," *Proc. Interspeech-12*, P3b.05, 2012.
- R. Maher & J. Beauchamp, "Fundamental frequency estimation of musical signals using a two-way mismatch procedure," *J. Acoust. Soc. Am.* 95(4), 2254-2263, 1994.
- A. Noll, "Cepstrum Pitch Determination," *J. Acoust. Soc. Am.* 41, 293-309, 1967.
- M. Puckette, T. Apel, D. Zicarelli, "Real-time audio analysis tools for Pd and MSP," *Proc. Int. Comp. Music Conf.*, 109-112, 1998.
- S. Sagayama, K. Takahashi, H. Kameoka and T. Nishimoto, "Specmurt Anasylis: A Piano-Roll-Visualization of Polyphonic Music Signal by Deconvolution of Log-Frequency Spectrum," *Proc. SAPA-2004*, 2004.
- M. Sondhi, "New Methods of Pitch Extraction," *IEEE Trans. Audio and Electroacoustics* AU-16(2), 262-266, 1968.
- D. Talkin, "A robust algorithm for pitch tracking (RAPT)," in *Speech Coding and Synthesis* (W. B. Kleijn and K. K. Paliwal, eds.), 495-518, Elsevier, 1995.