CTP 431 Music and Audio Computing

Short-Time Fourier Transform And Feature Extraction

Graduate School of Culture Technology (GSCT)
Juhan Nam

Outlines

- Short-Time Fourier Transform
  - Spectrogram

- Feature Extraction
  - Amplitude envelope
  - Pitch analysis
  - Timbre features

Short-Time Fourier Transform (STFT)

- DFT assumes that the signal is stationary
  - It is not a good idea to apply DFT to a very long and dynamically changing signal like music
  - Instead, we segment the signal and apply DFT separately

- Short-Time Fourier Transform

$$X(l,k) = \sum_{n=0}^{N-1} w(n)x(n + l \cdot H)e^{-j2\pi kn/N}$$

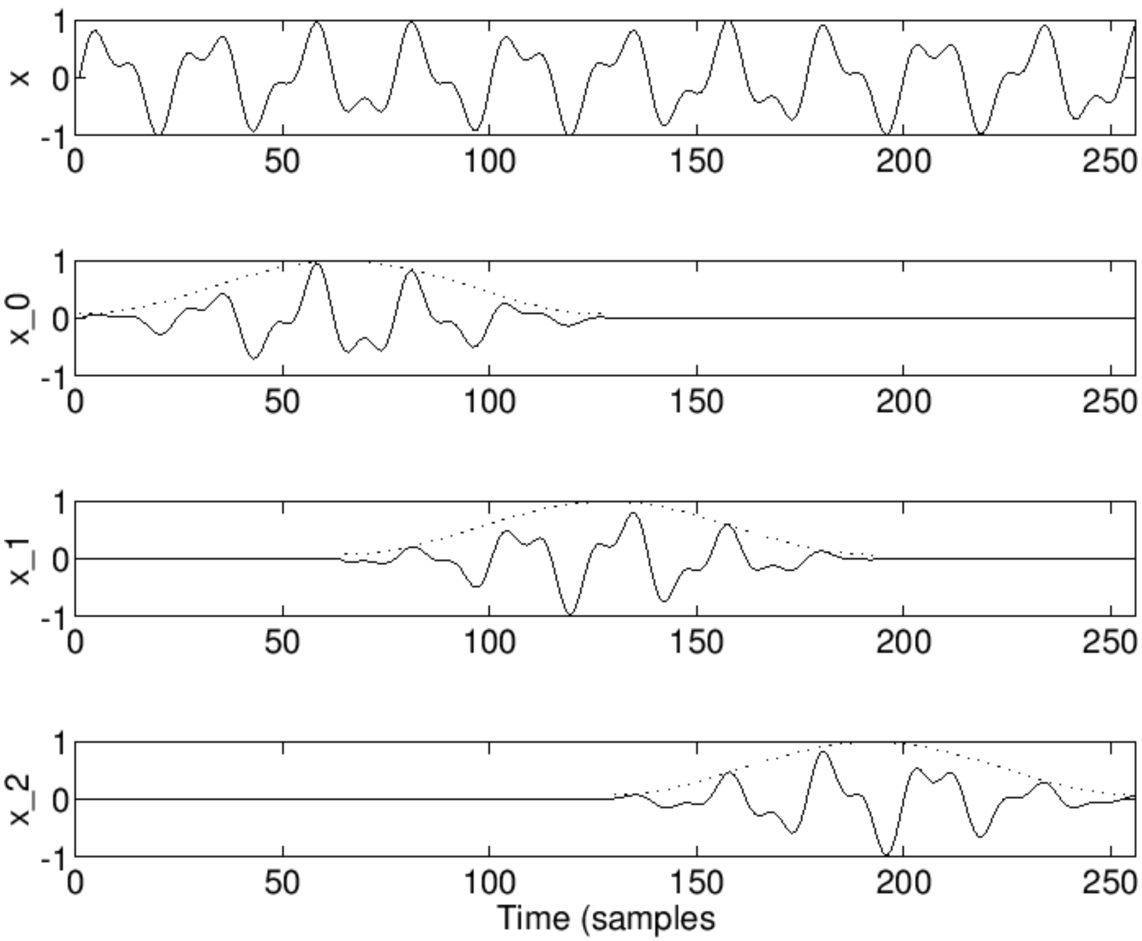
H : hop size

$w(n)$: window

N : FFT size

- This produces 2-D time-frequency representations
  - Get "spectrogram" from the magnitude
  - Parameters: window size, window type, FFT size, hop size

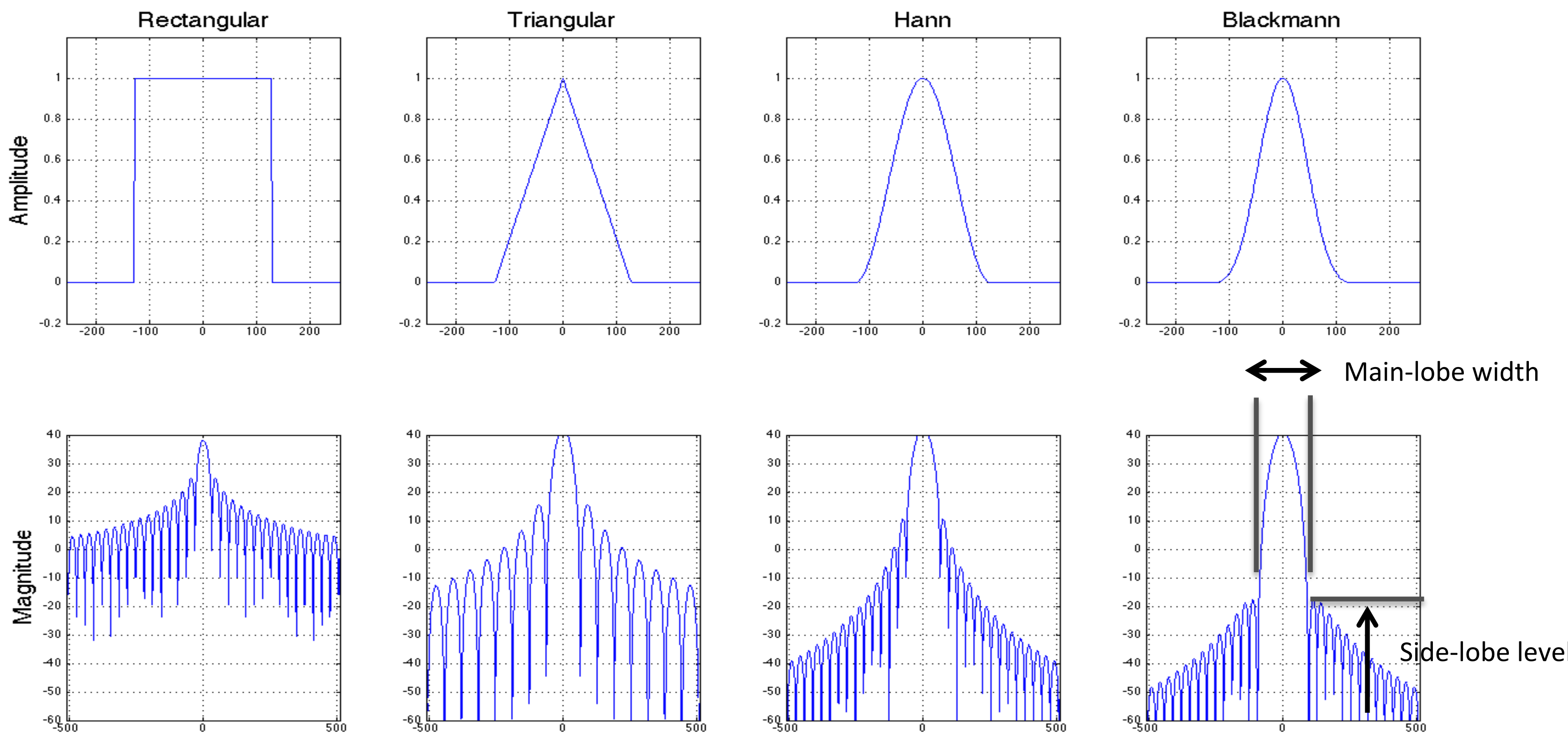
Short-Time Fourier Transform (STFT)



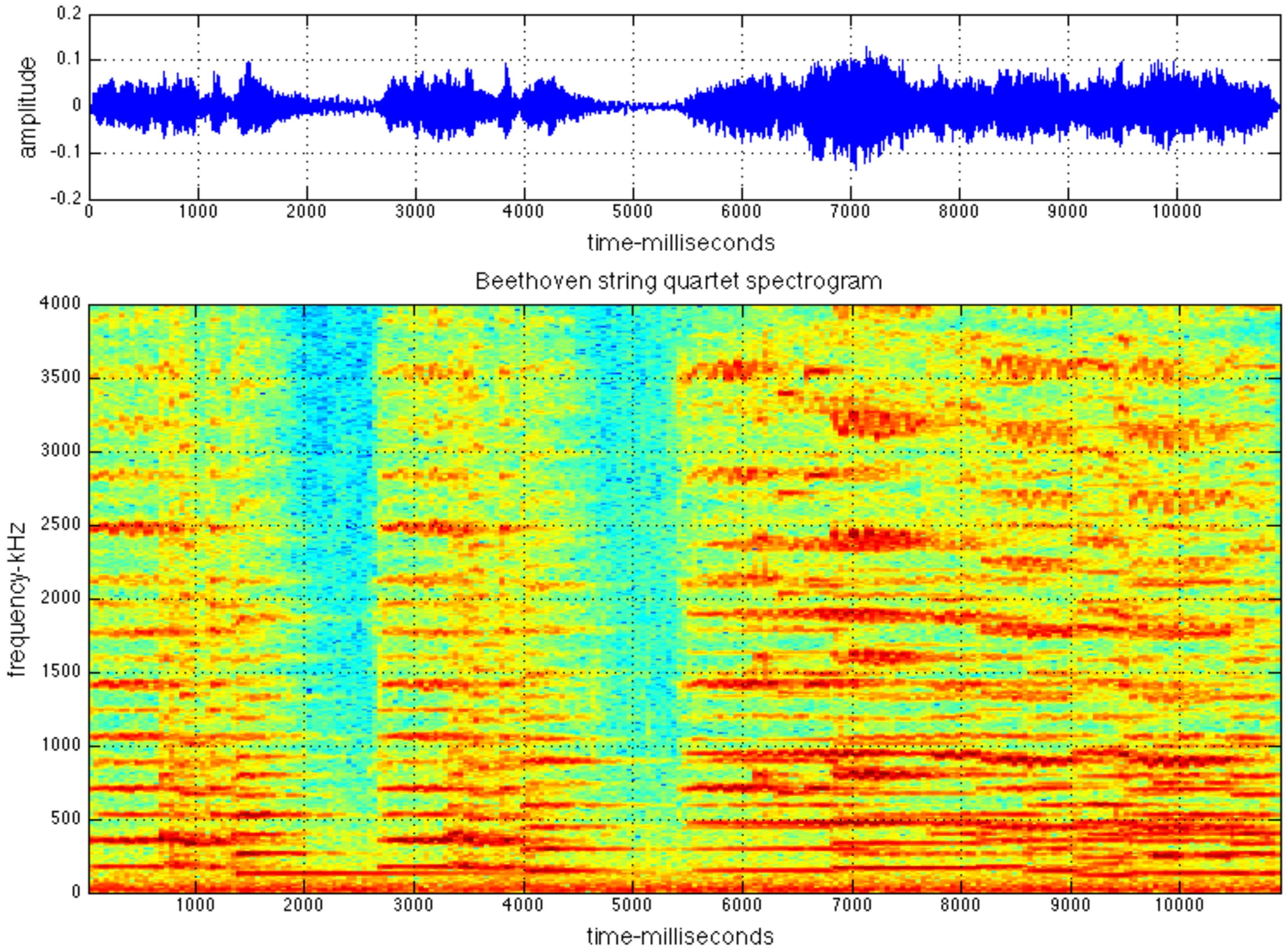
50% overlap

Windowing

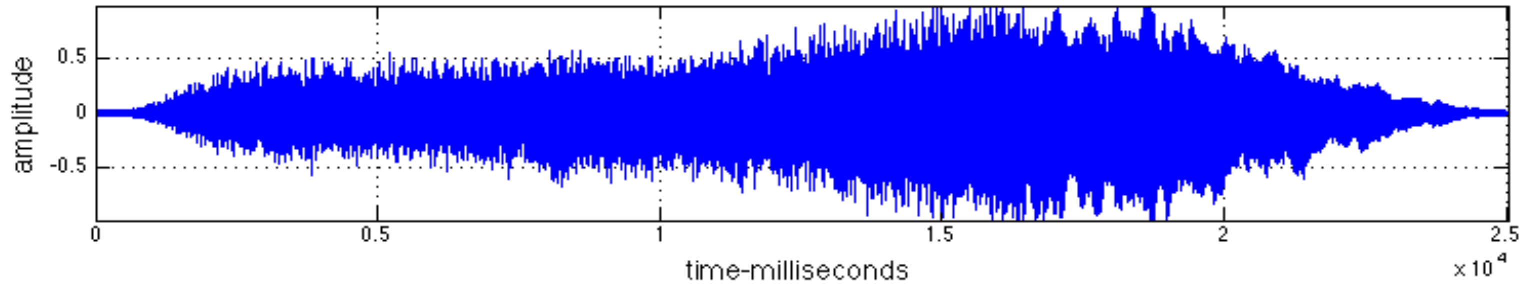
- Types of window functions
  - Trade-off between the width of main-lobe and the level of side-lobe

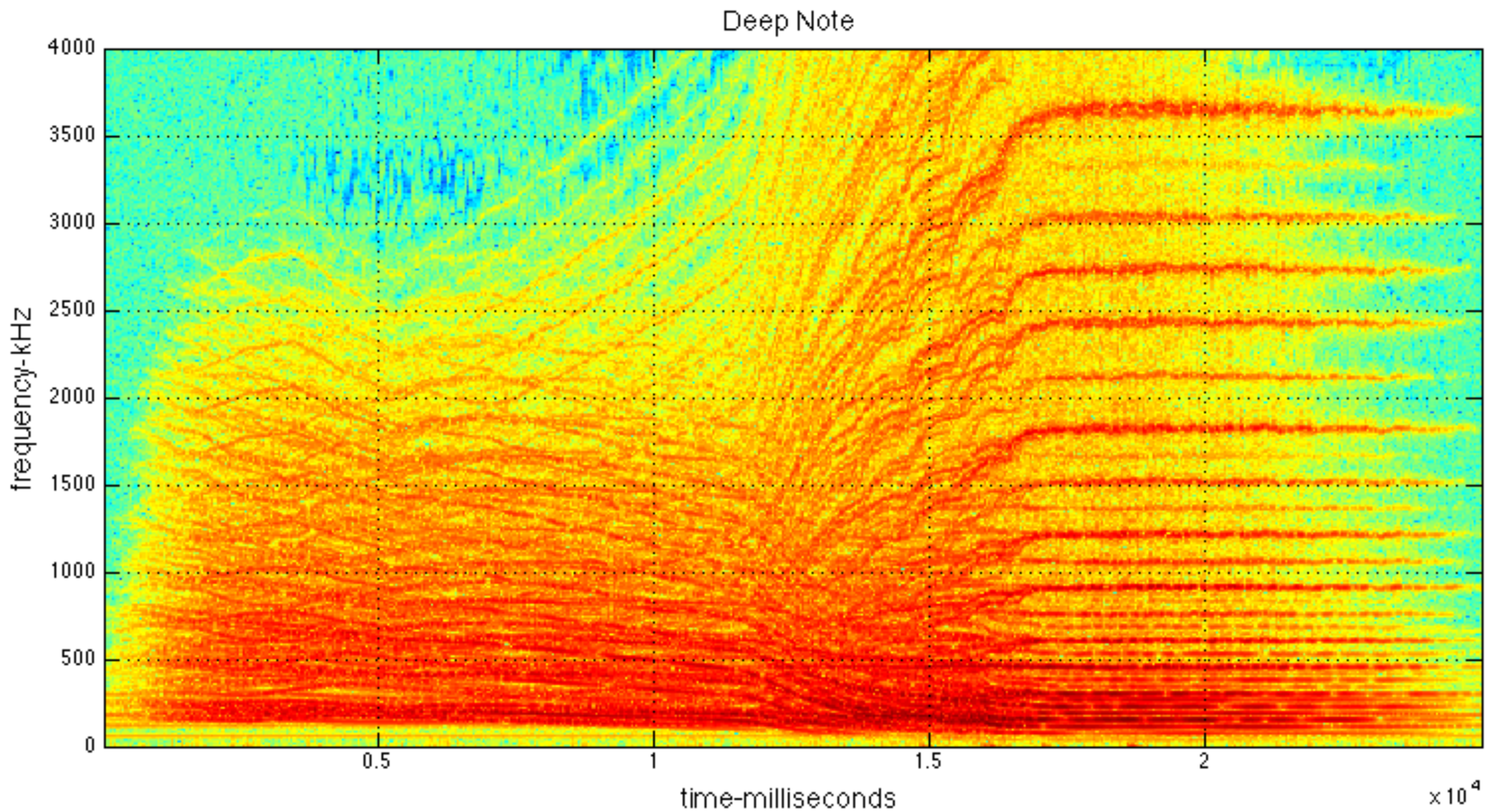


Example: Music



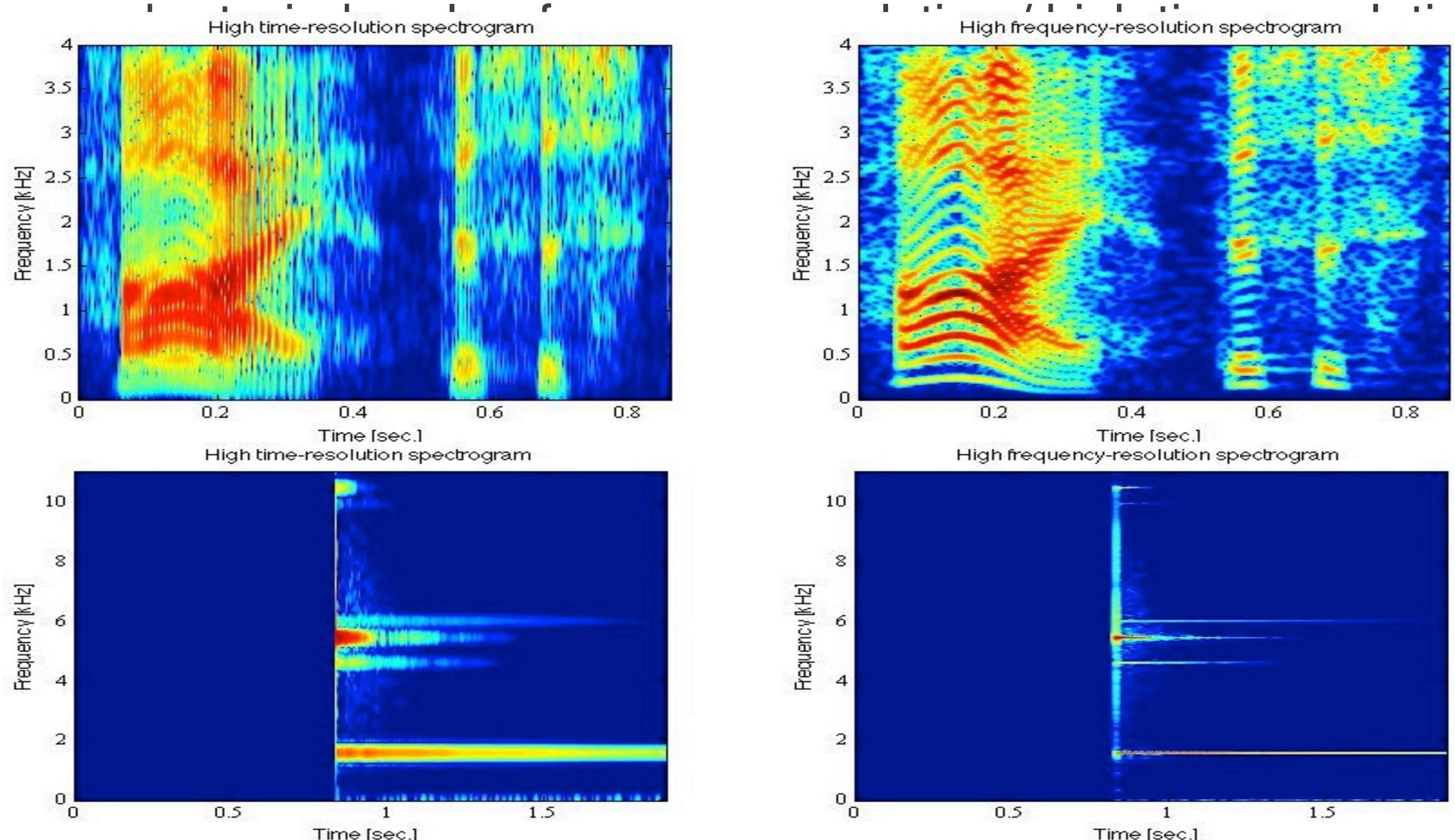
Example: Deep Note





Time-Frequency Resolutions in STFT

- Trade-off between time-resolution and frequency-resolution
  - Long window: high frequency-resolution / low time-resolution



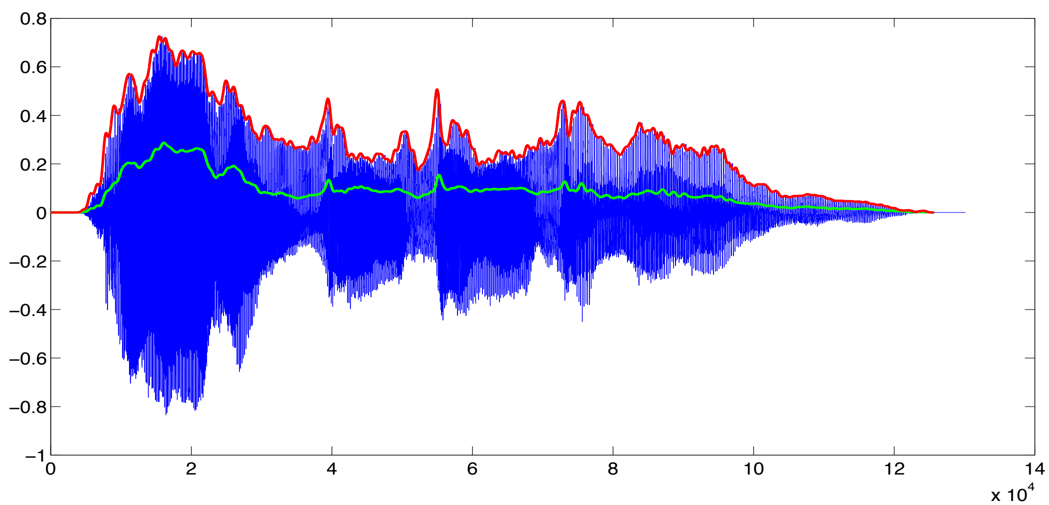
Amplitude Envelope

- Root Mean Square:

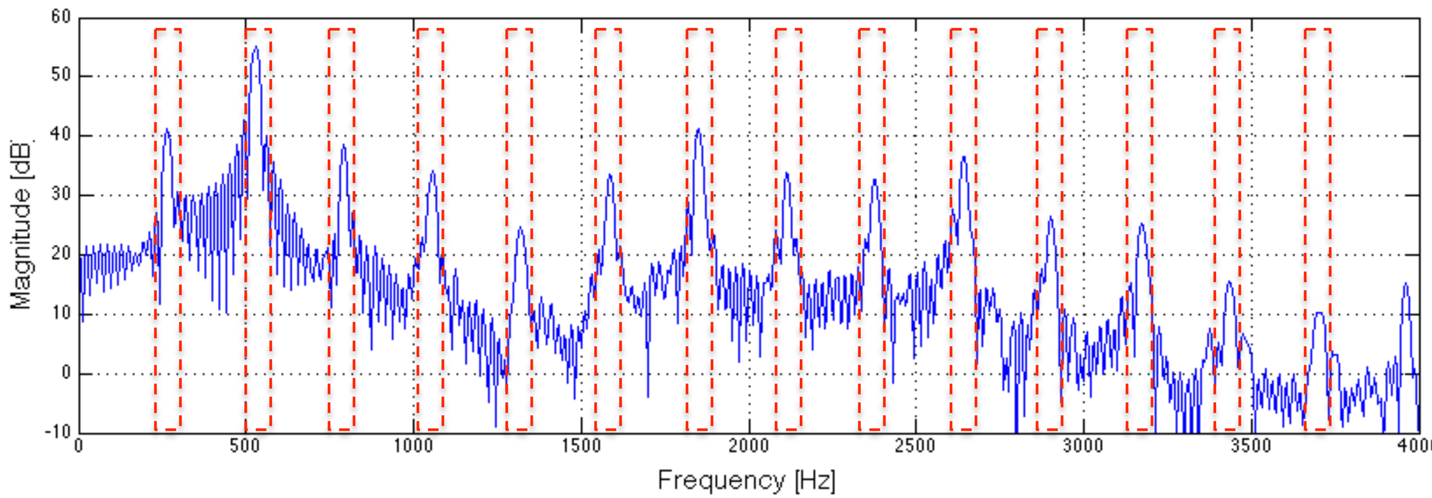
$$RMS(l) = \sqrt{\sum_{n=0}^{N-1} (w(n)x(n + l \cdot H))^2} \qquad l = 0,1,2,\cdots N-1$$

- Max-Peak:

$$MAX\_PEAK(l) = \max(x(n + l \cdot H))$$

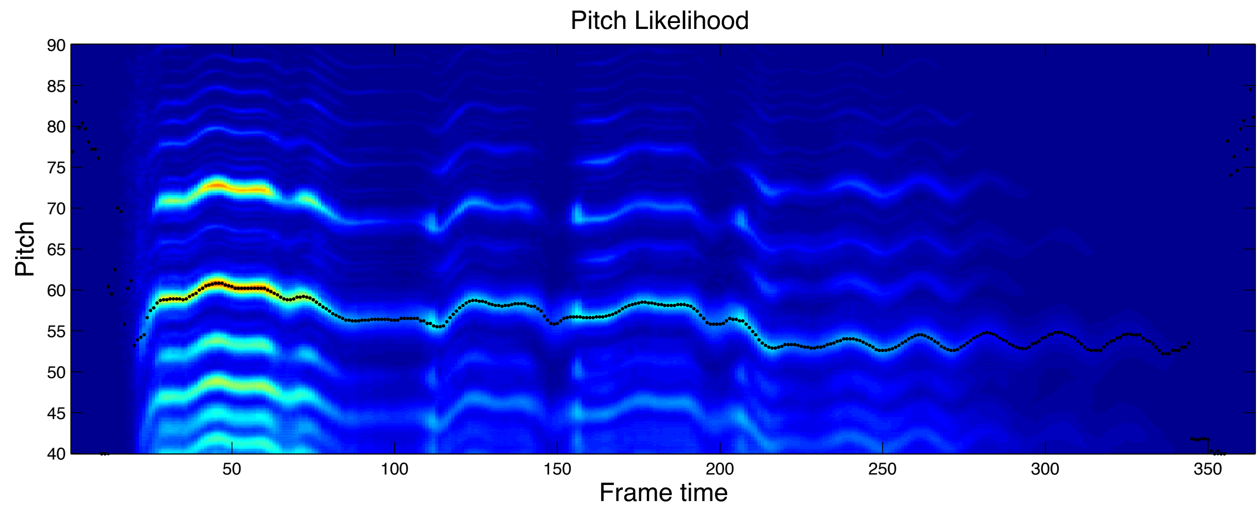


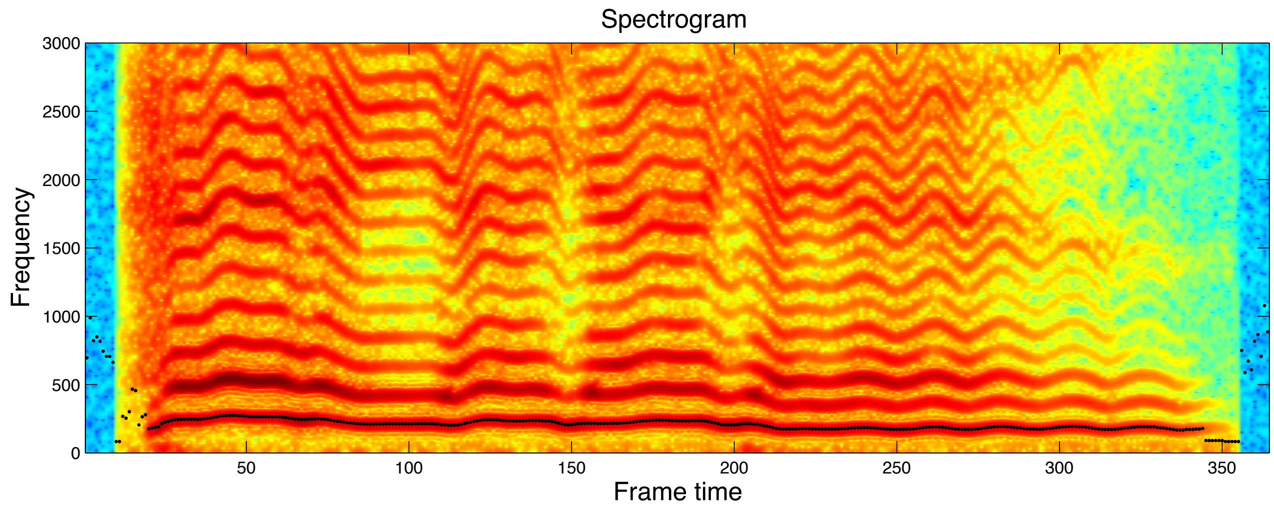
Pitch Detection



- Frequency domain approach
  - Define harmonic sieves or comb filter for each of pitch candidates
  - Sum all spectral energy at the sieve locations
  - Associated with the auto-correlation method (frequency domain version)

Pitch Detection





Timbre Features: Spectral Summary

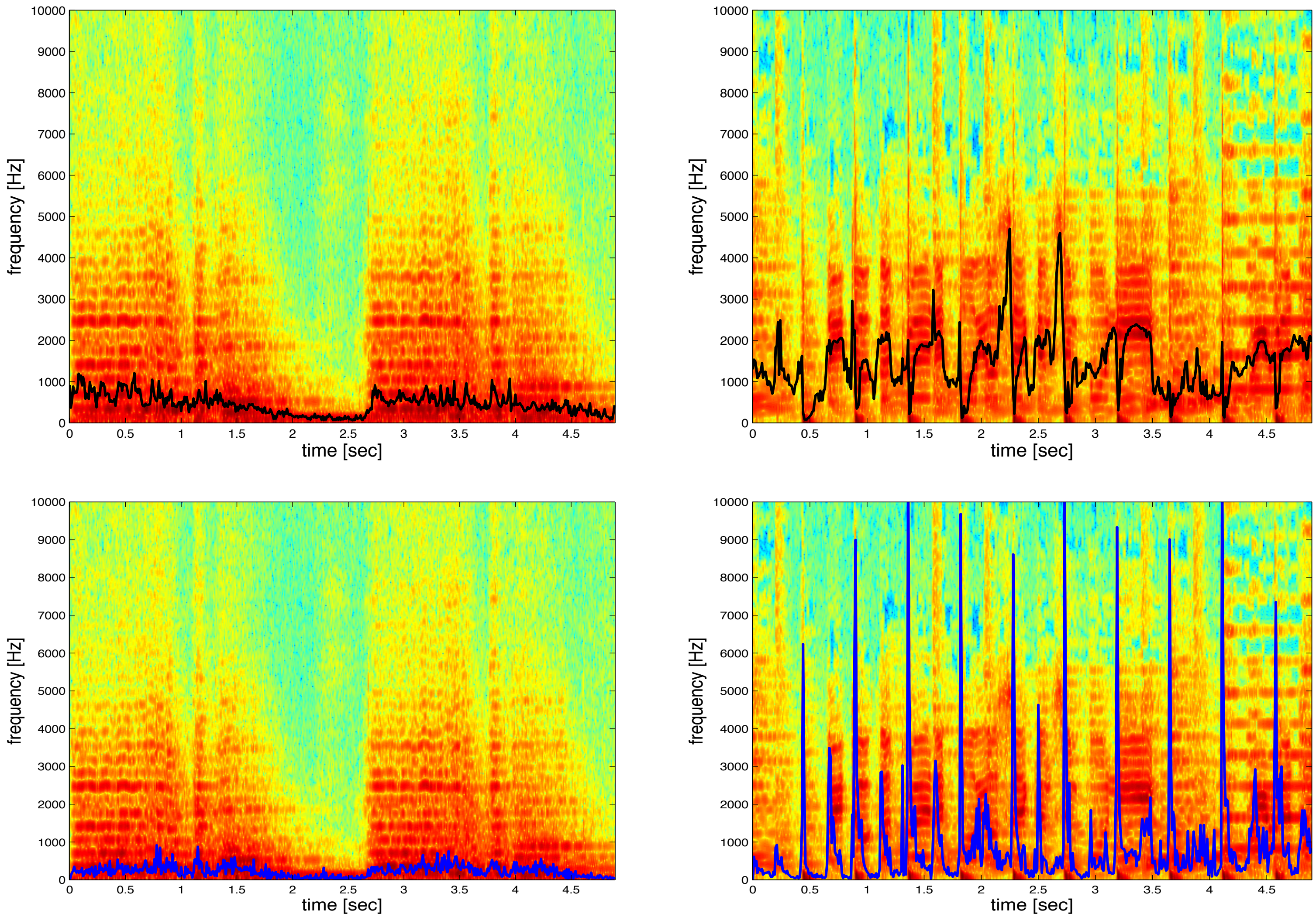
- Spectral centroid
  - "Center of gravity" of the spectrum
  - Associated with the brightness of sounds

$$SC(t) = \frac{\sum_k f_k \left| X_t(k) \right|^2}{\sum_k \left| X_t(k) \right|^2}$$

- Spectral flux
  - Positive changes between successive spectra
  - Used for onset detection

$$X_t = \sum_k \max(\left| X_t(k) \right| - \left| X_{t-1}(k) \right|, 0) \quad \text{(Dixon, 2006)}$$

Examples of Spectral Centroid and Flux



Spectral Centroid

Spectral Flux

Classical: “Beethoven String Quartet”

Pop: “Video killed the radio star”