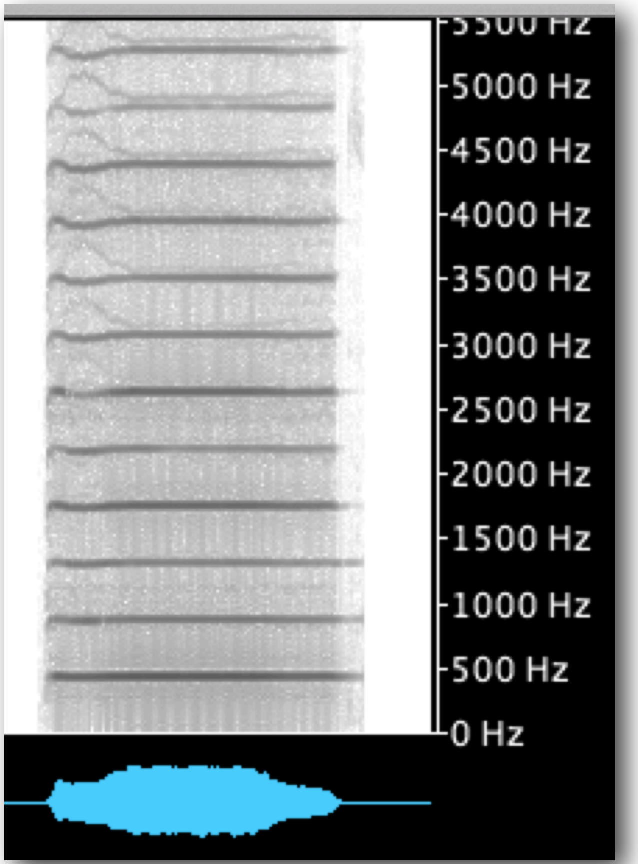
Lecture 5: Sinusoidal Modeling

1. Sinusoidal Modeling
2. Sinusoidal Analysis
3. Sinusoidal Synthesis & Modification
4. Noise Residual

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

1. Sinusoidal Modeling

- Periodic sounds
  → ridges in spectrogram
  - each ridge is a sinusoidal harmonic
  - .. with smoothly-varying parameters
  - .. an efficient & flexible description?



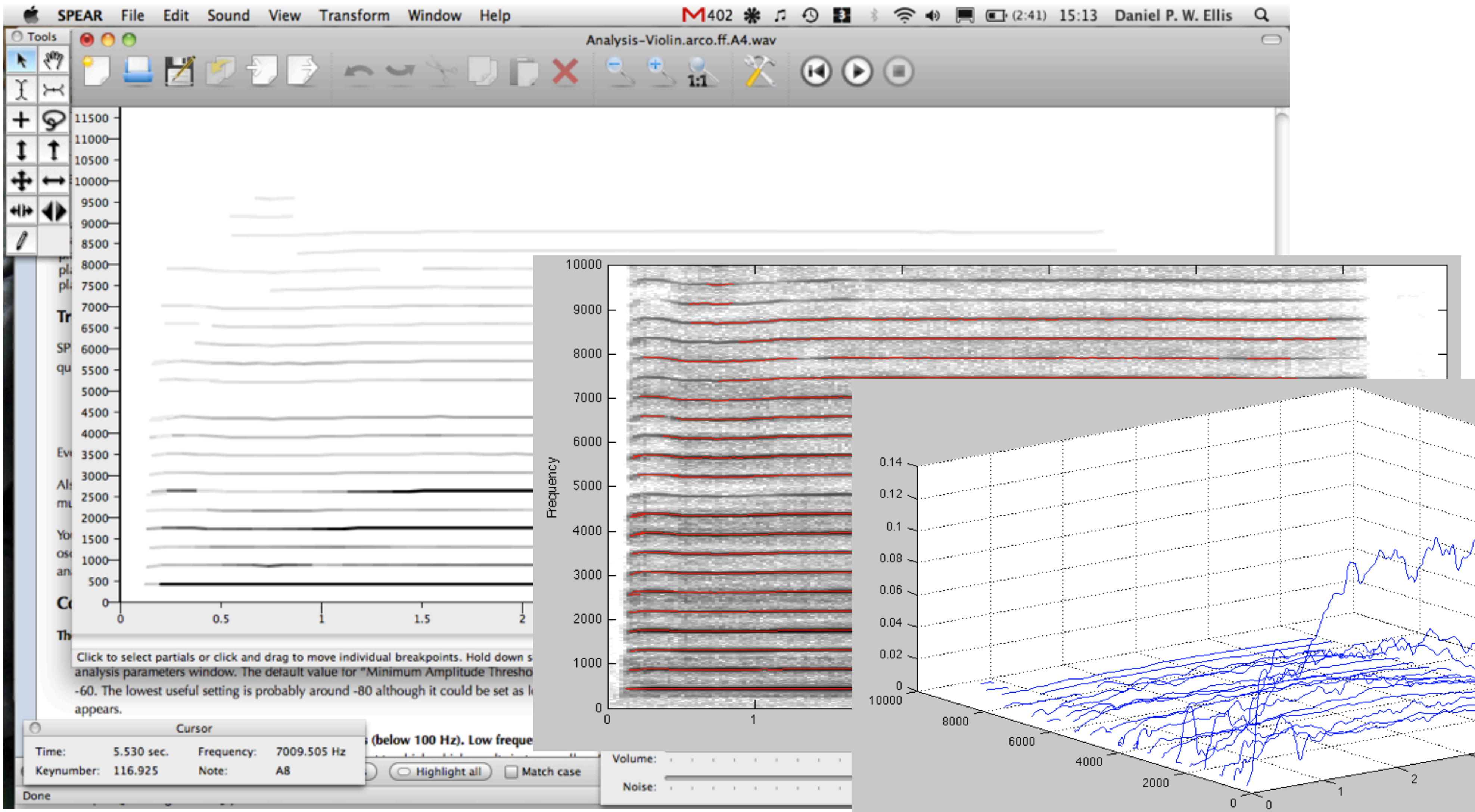
Violin.arco.ff.A4

Sinusoid Modeling

- Analogous to Fourier series
  - model harmonics explicitly? e.g.
    $$x[n] = \sum_k a_k[n] cos(\theta_k[n])$$
  - ... for pitched signal with fundamental $\omega_0[n]$
    $$\theta_k[n] = k \cdot \omega_0[n] \cdot n$$
- Additional constraints
  - harmonicity
  - smoothness of $a_k[n]$
- Arbitrarily accurate given enough sinusoids

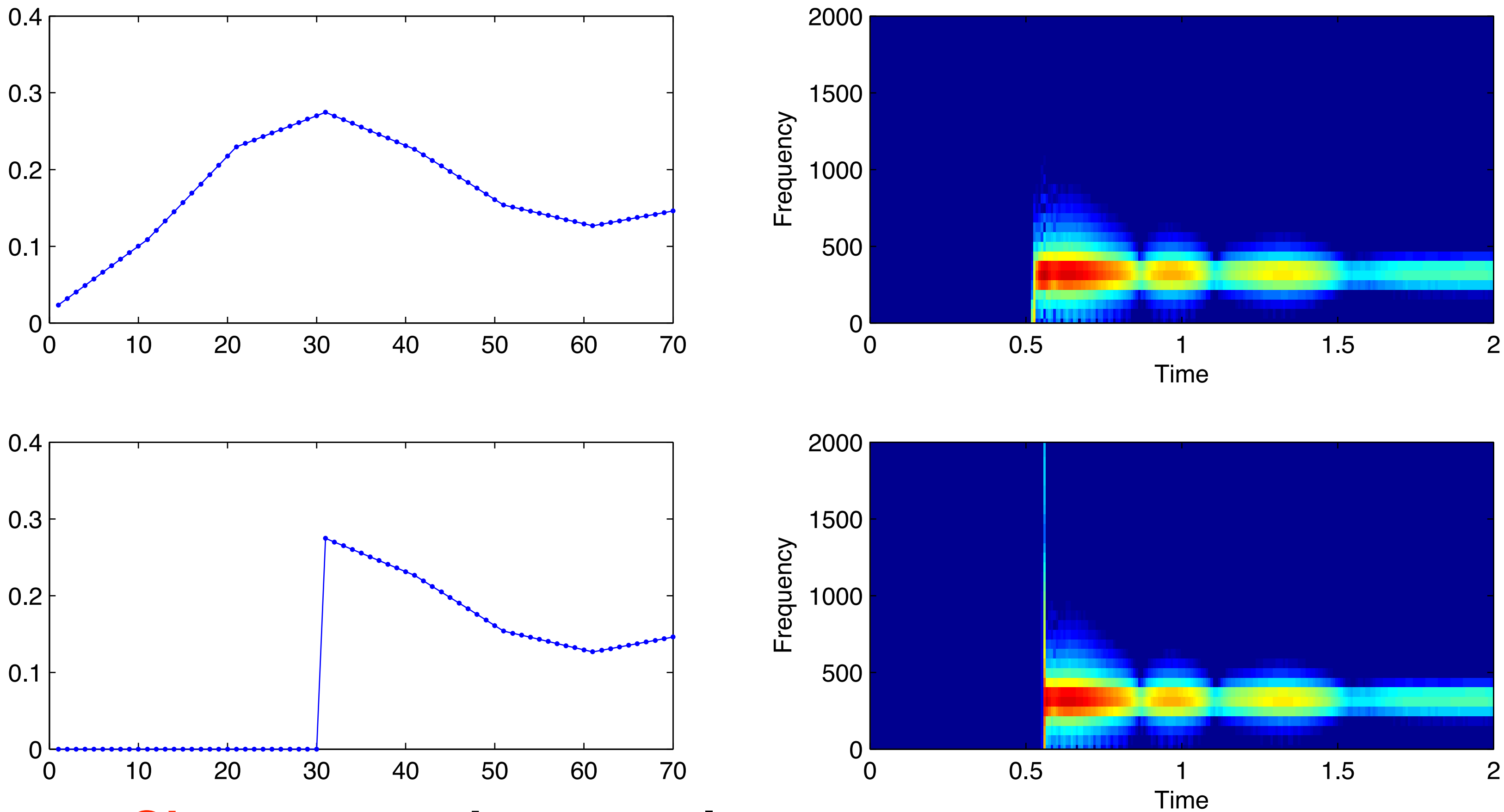
Examples

- Using Michael Klingbeil's SPEAR
  - http://www.klingbeil.com/spear/



Envelope Limitations

- Extracted envelope reflects analysis window



- Sharp window violates assumptions

2. Sinusoidal Analysis

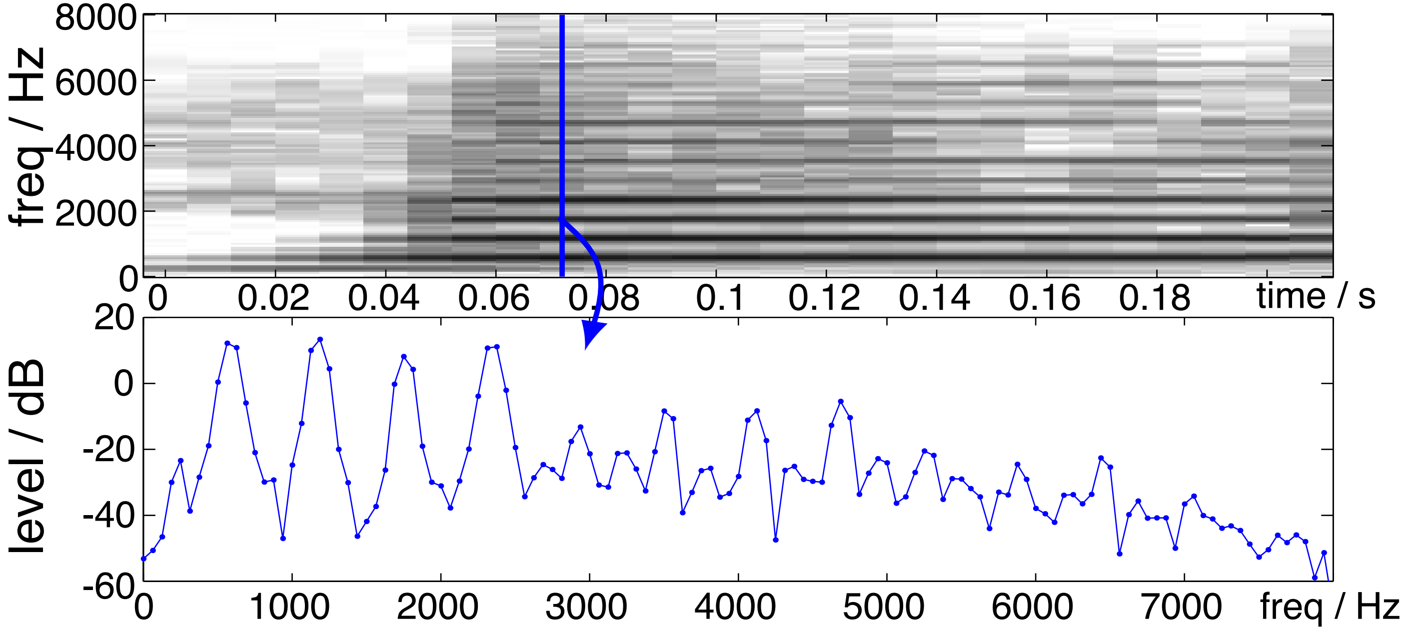
- Sinusoids = peaks in spectrogram slices

  = DFT frames $X[k,m] = \sum_{n=0}^{N-1} x[n+mL]\,w[n]e^{-j\frac{2\pi kn}{N}}$

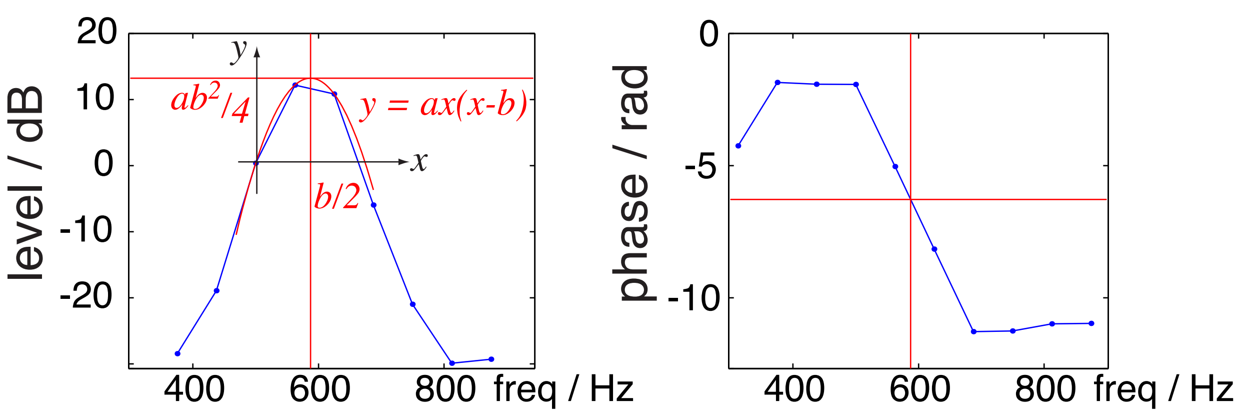
- DFT length N
  - window determines frequency resolution: $X(e^{j\omega}) \circledast W(e^{j\omega})$
  - long enough to "see harmonics"
    e.g. 2-3x longest pitch cycle typically 50-100 ms
  - but: too long → blurs amplitude envelope $a_k[n]$
- Hop advance L
  - choose N/2 or N/4
  - .. denser for simpler interpolation along time

Sinusoidal Peak Picking

- Local maxima in DFT frames

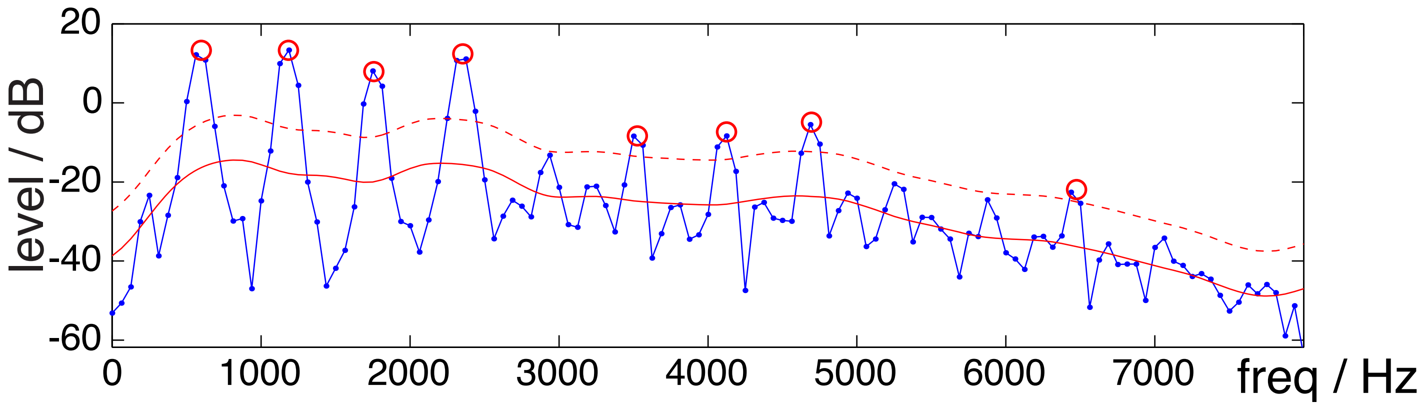


- Quadratic fit for sub-bin resolution



Peak Selection

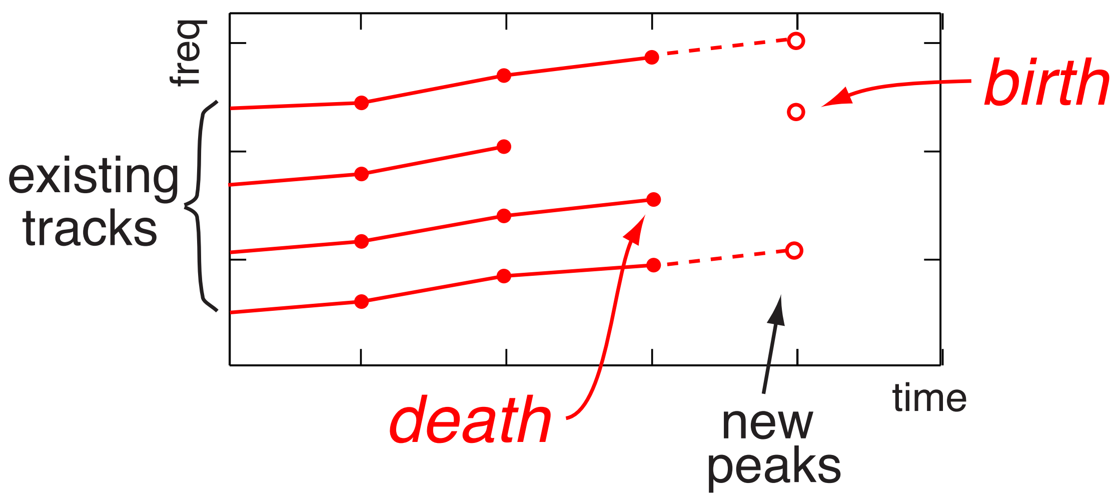
- Don't want every peak
  - just "true" sinusoids
  - threshold?



  - local shape - fits $\delta(\omega - \omega_0) \circledast W(e^{j\omega})$
- Look for stability
  - of frequency & amplitude in successive time frames
  - phase derivative in time/freq

Track Formation

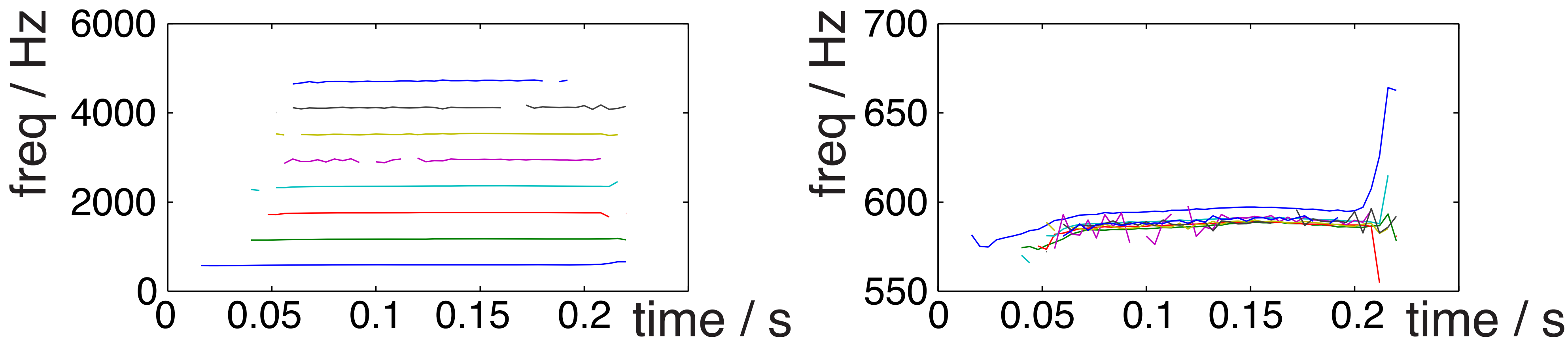
- Connect peaks in adjacent frames to form sinusoids
  - can be ambiguous if large frequency changes



- Unclaimed peak → create new track
- No continuation of track → termination
  - hysteresis

Pitch Tracking

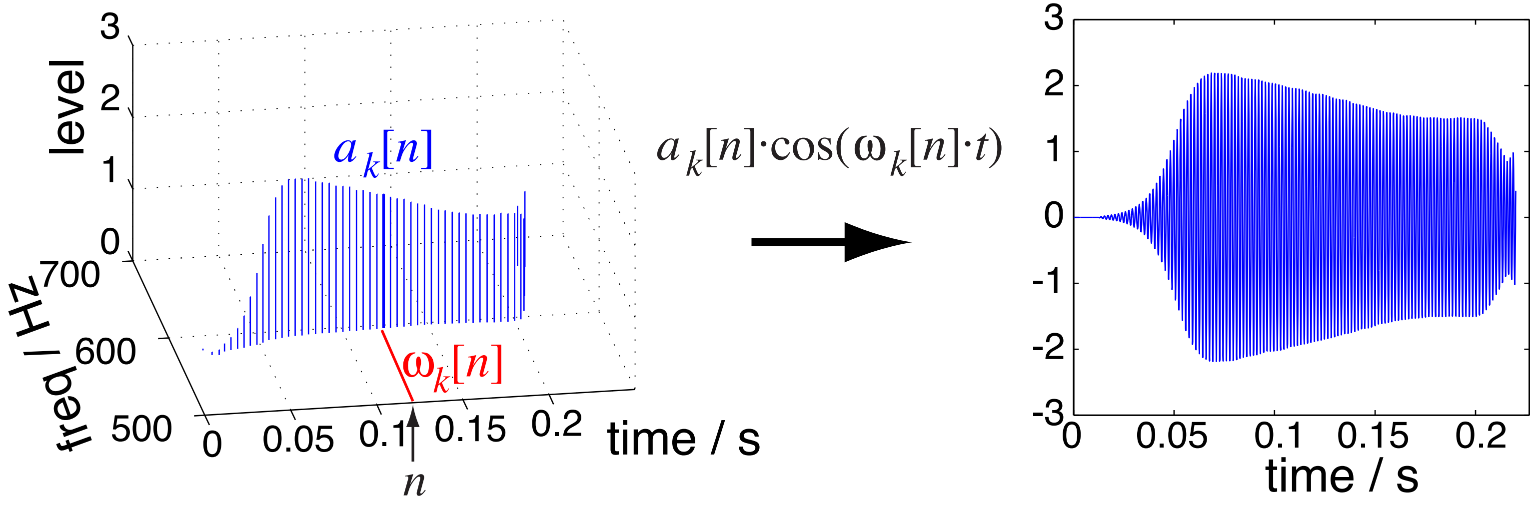
- Extracted sinusoids could be anywhere
  - but often expect them to be in harmonic series



- Find pitch by searching for common factor
  - can then "regularize" pitch $\omega_k[n] = k \cdot \omega_0[n]$

3. Sinusoidal Synthesis

- Each sinusoid track $\{a_k[n], \omega_k[n]\}$ drives an oscillator

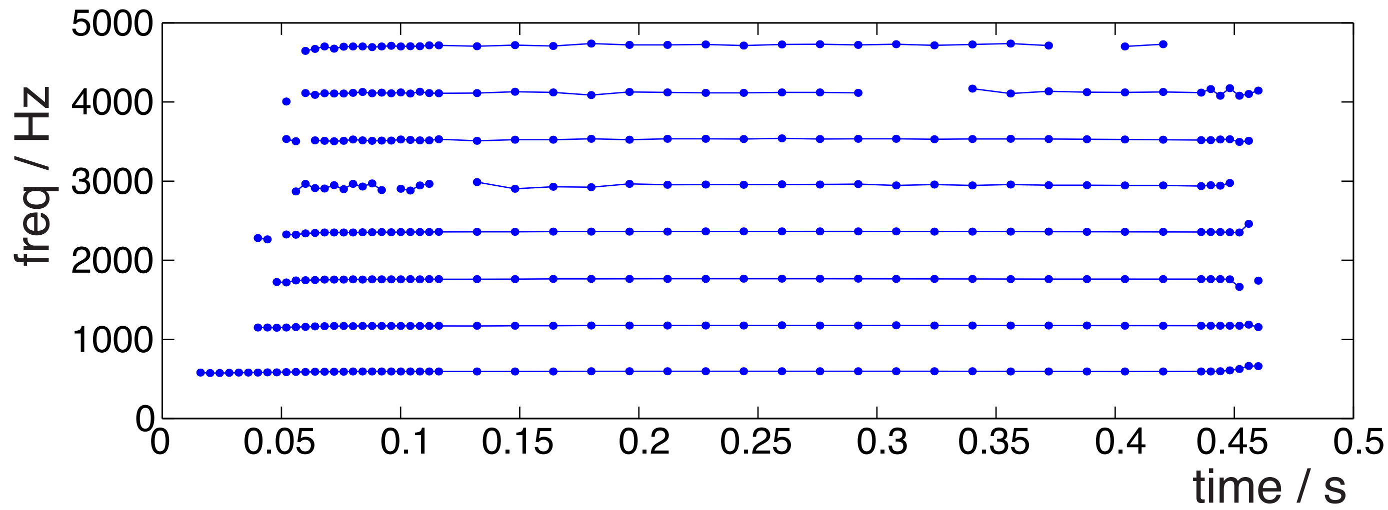


  - can interpolate amplitude, frequency samples

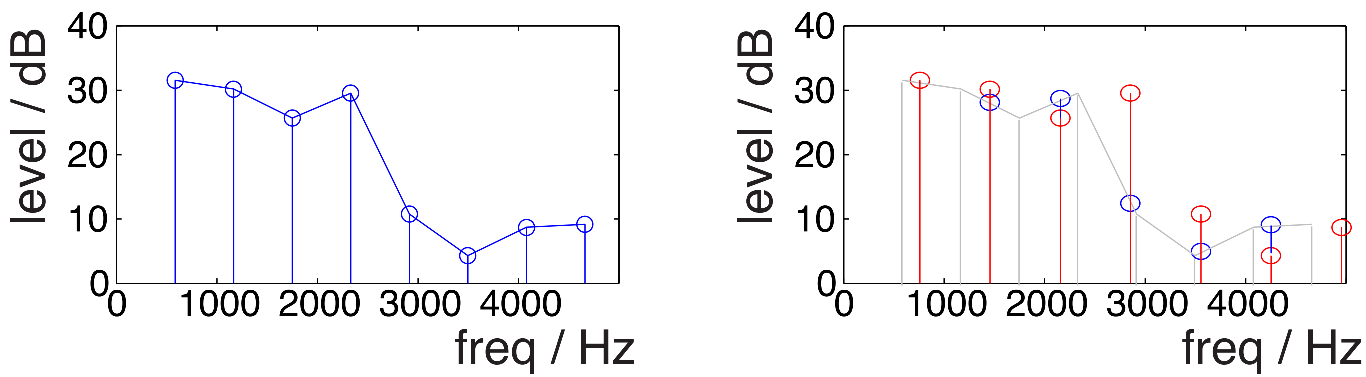
- Faster method synthesizes DFT frames
  - then overlap-add
  - trickier to achieve frequency modulation

Sinusoidal Modification

- Sinusoidal description very easy to modify
  - e.g. changing time base of sample points



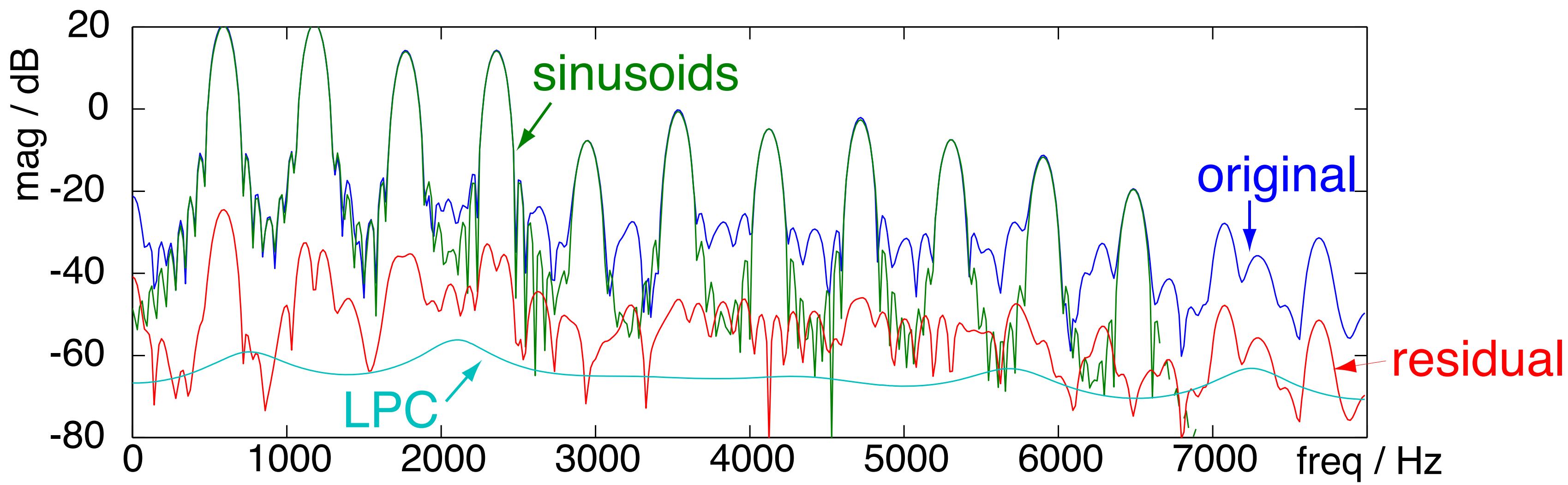
- Frequency stretch
  - preserve formant envelope?



4. Noise Residual

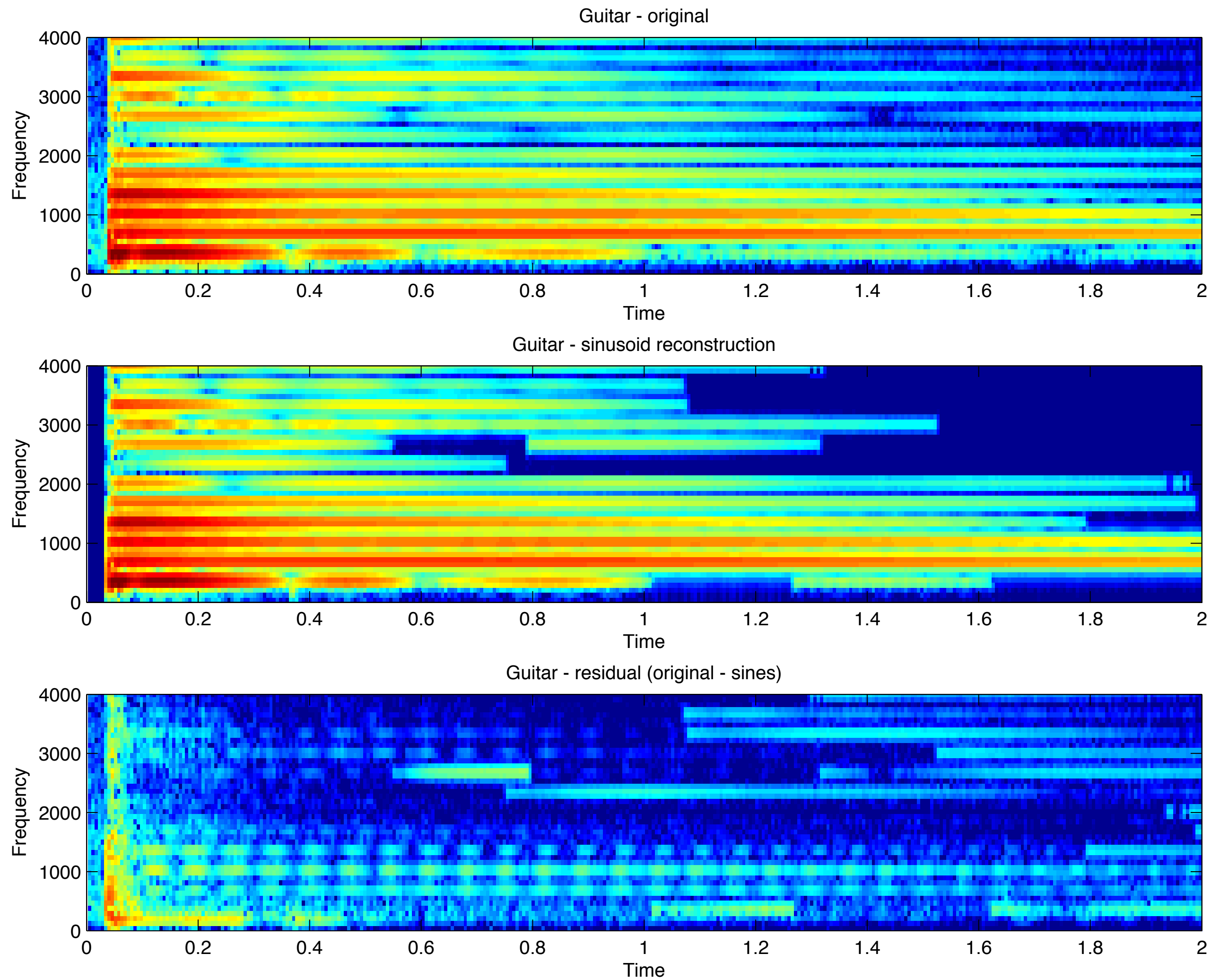
- Some energy is not well fit with sinusoids
  - e.g. noisy energy
- Can just keep it as residual
  - or model it some other way
- Leads to "sinusoidal + noise" model

$$x[n] = \sum_k a_k[n] cos(\omega_k[n]n) + e[n]$$



Sinusoids + Noise Decomposition

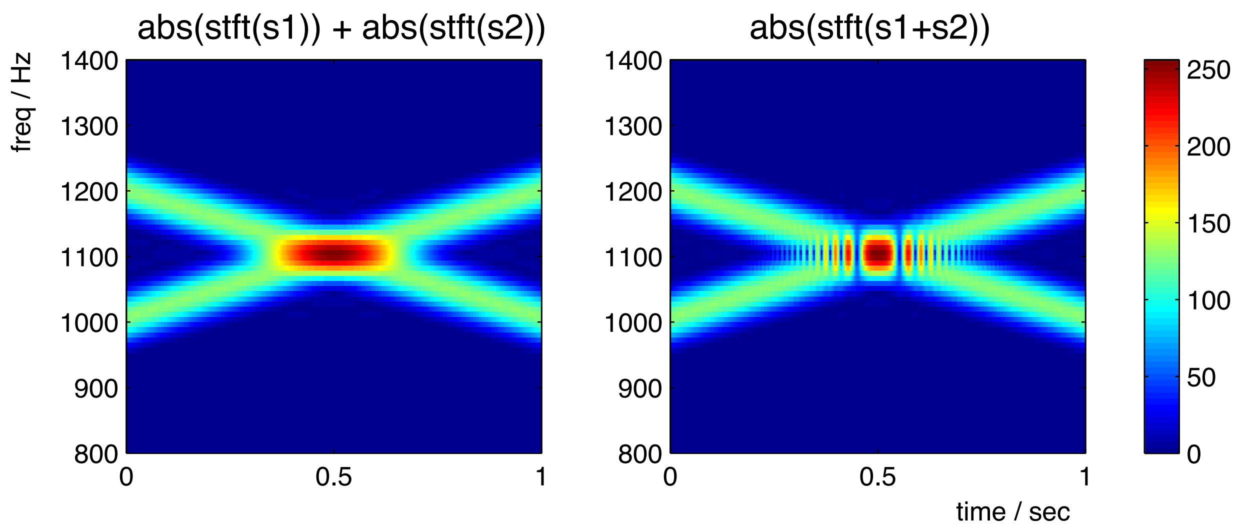
- Removing sines reveals noise & transients



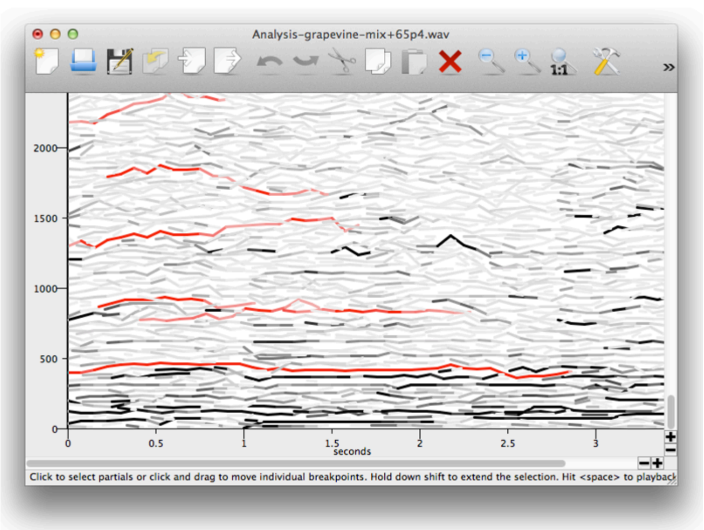
- Different representation approaches...

5. Limitations

- The spectrogram (mag STFT) is not linear
  - superpositions suffer from phase effects



- Separating sources is generally hard...
  - parameters
  - tracking



Summary

- Spectrogram shows sinusoid harmonics in many sounds
- Peak picking in spectrogram can effectively extract them
- Sinusoidal domain extremely flexible for modification
- Noise residual can add even more realism