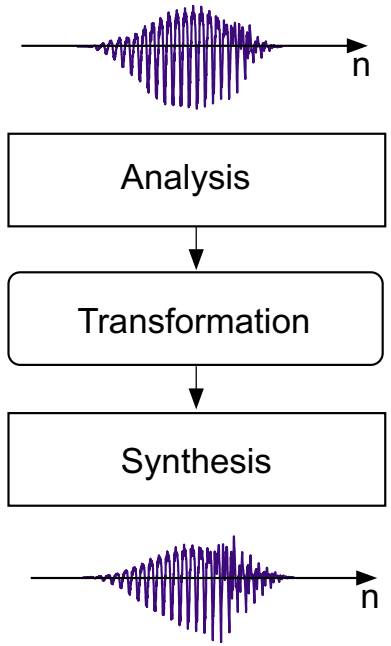
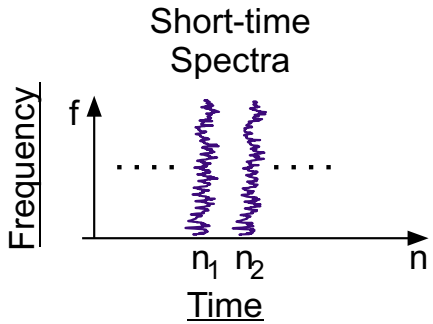
E85.2607: Lecture 7 – Phase Vocoder

Analysis/Resynthesis

General audio processing framework:

1. Analyze input signal
   - e.g. transform into different domain
2. Modify signal in new domain
3. Invert analysis to synthesize time-domain signal

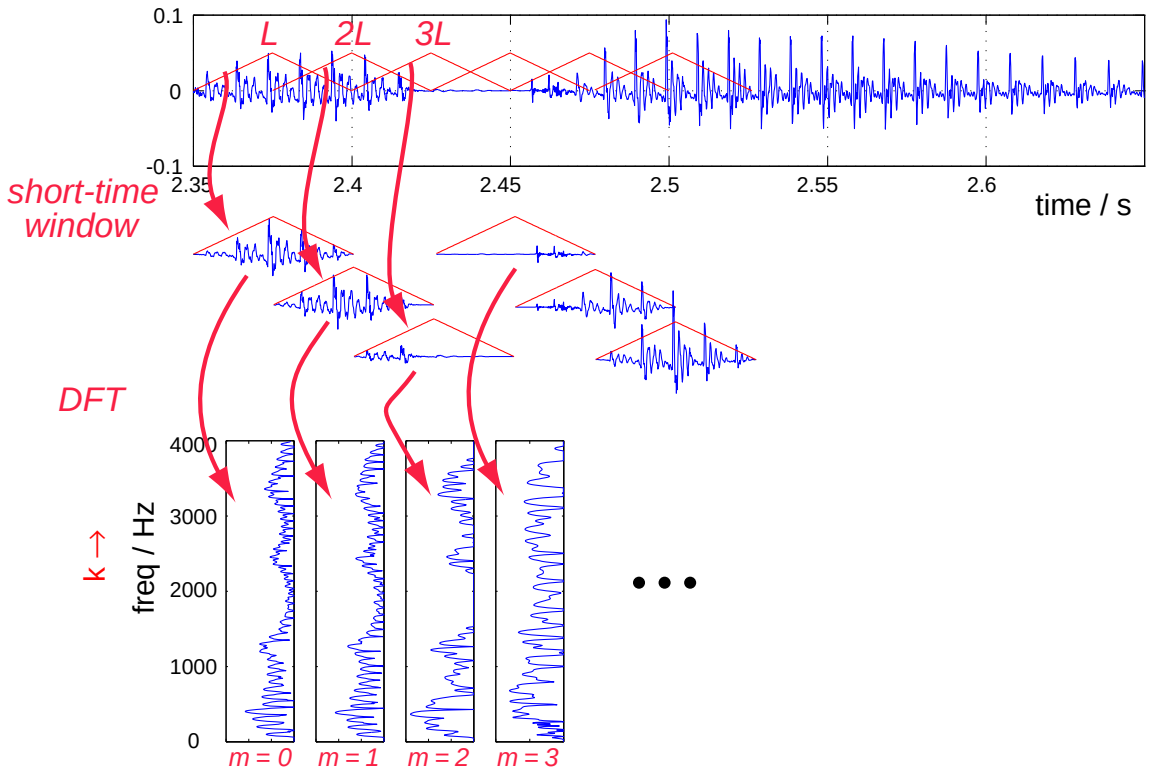




Review: Short-time Fourier Transform (analysis)

Want to localize energy in time and frequency

- break sound into short-time pieces
- calculate DFT of each one



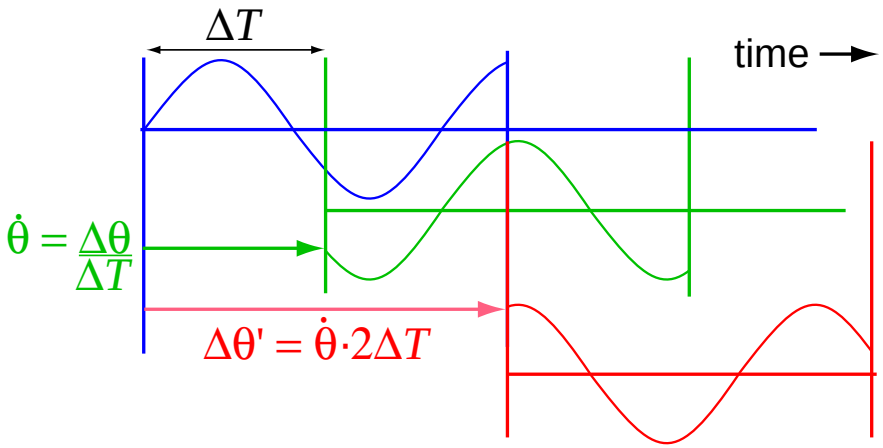
$$X[n,k] = \sum_{m=-\infty}^{\infty} x[m]\, w[n-m] e^{-j\frac{2\pi}{N}mk} = |X[n,k]|\, e^{-j\phi[n,k]}$$

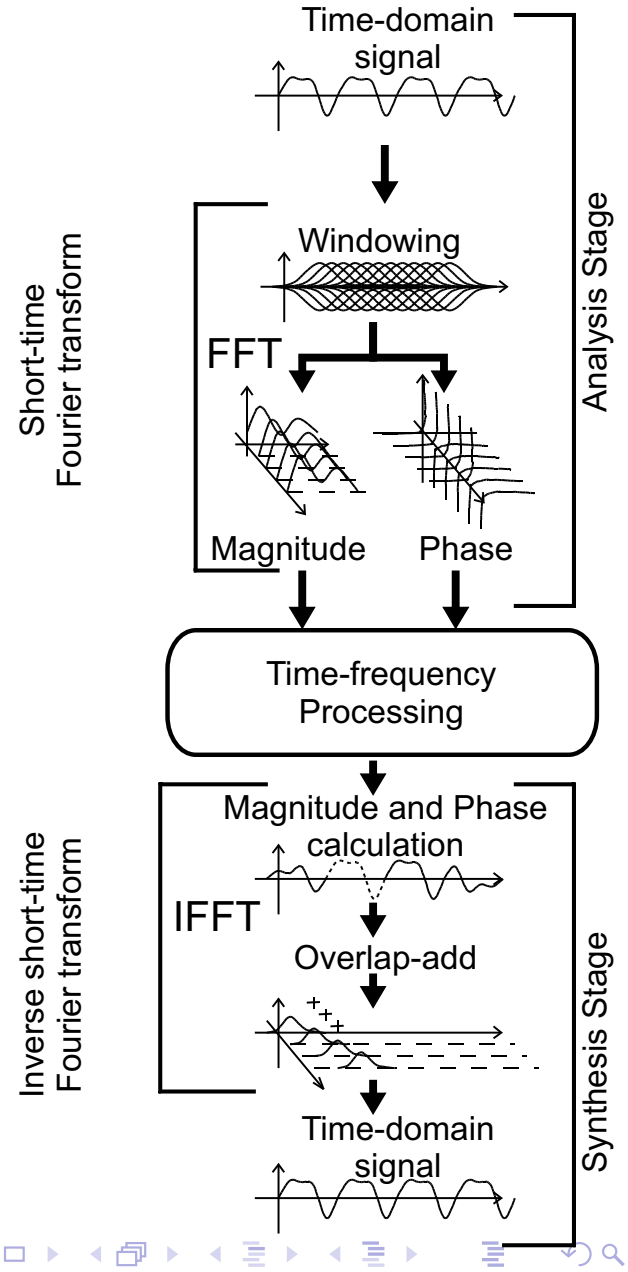
Phase vocoder

- Dates back to the 1960s [Flanagan, 1966]
- Like STFT, but track rate of phase change between frames

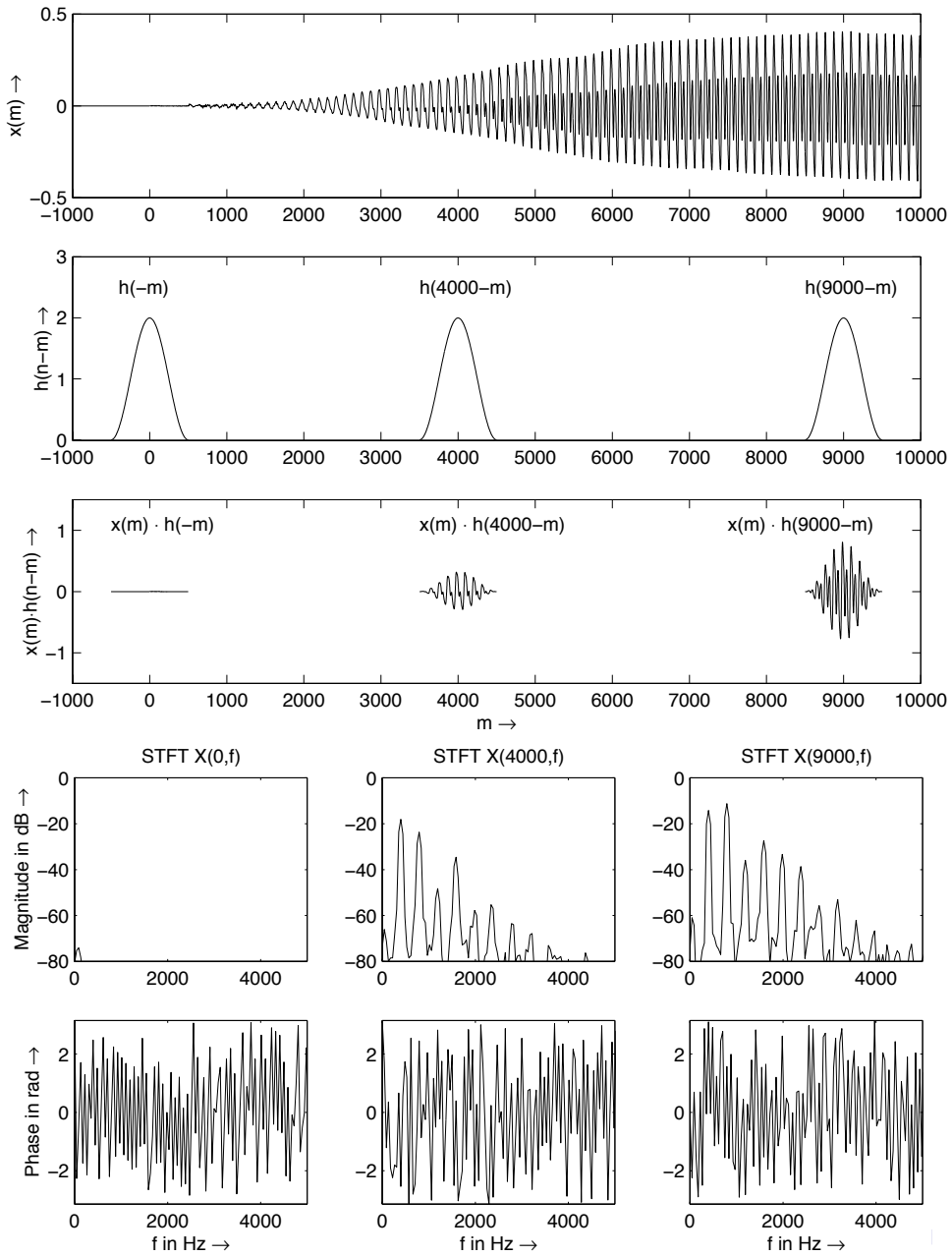
  Instantaneous frequency:

$$\Delta\phi[n,k] = \frac{d}{dt}\,\phi[n,k]$$

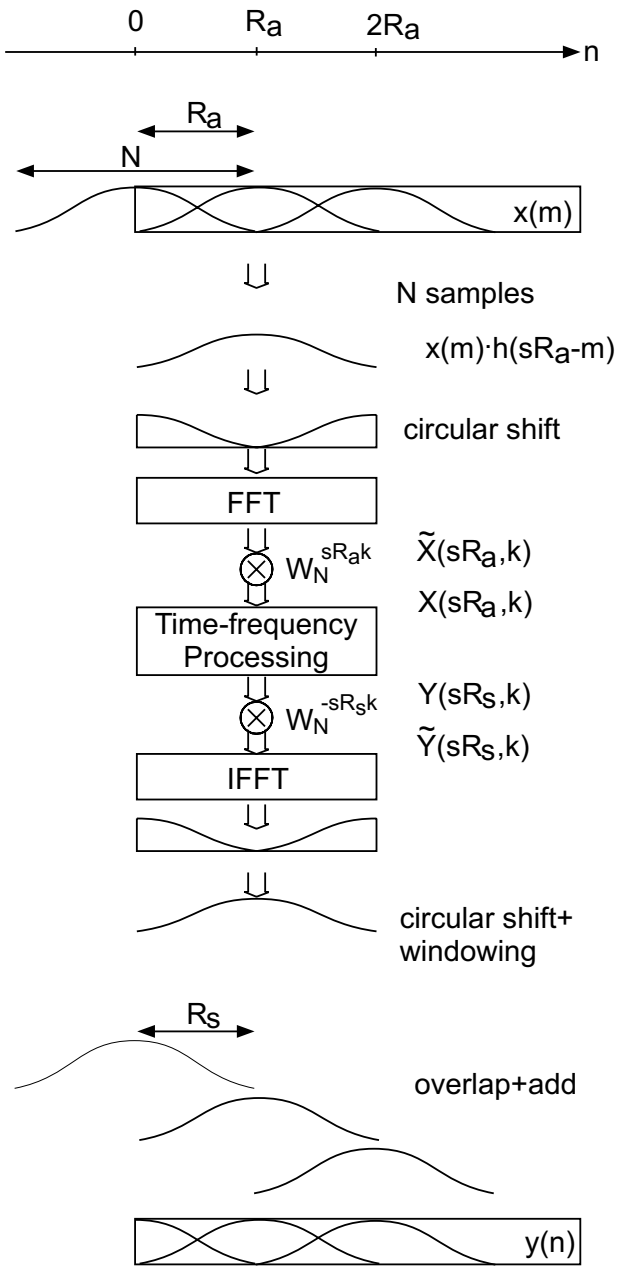


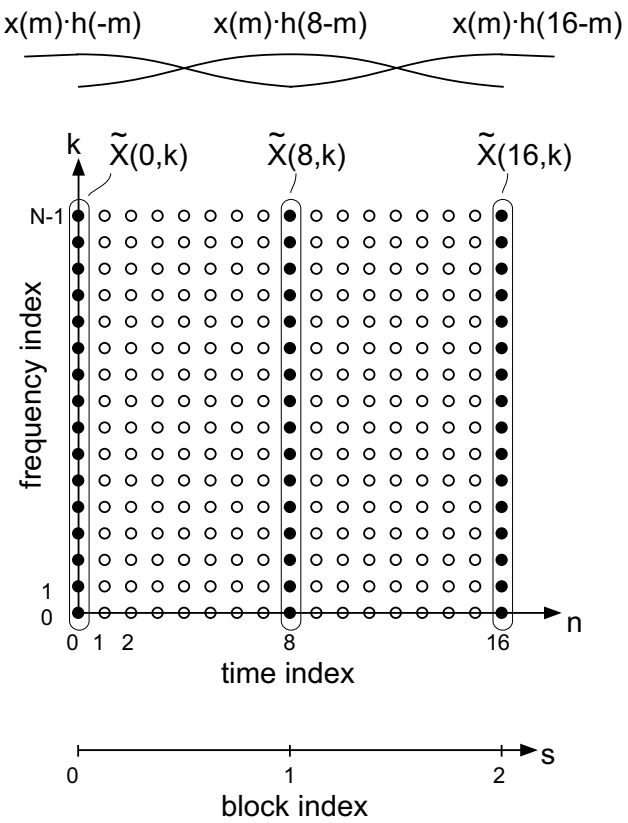


Sliding window



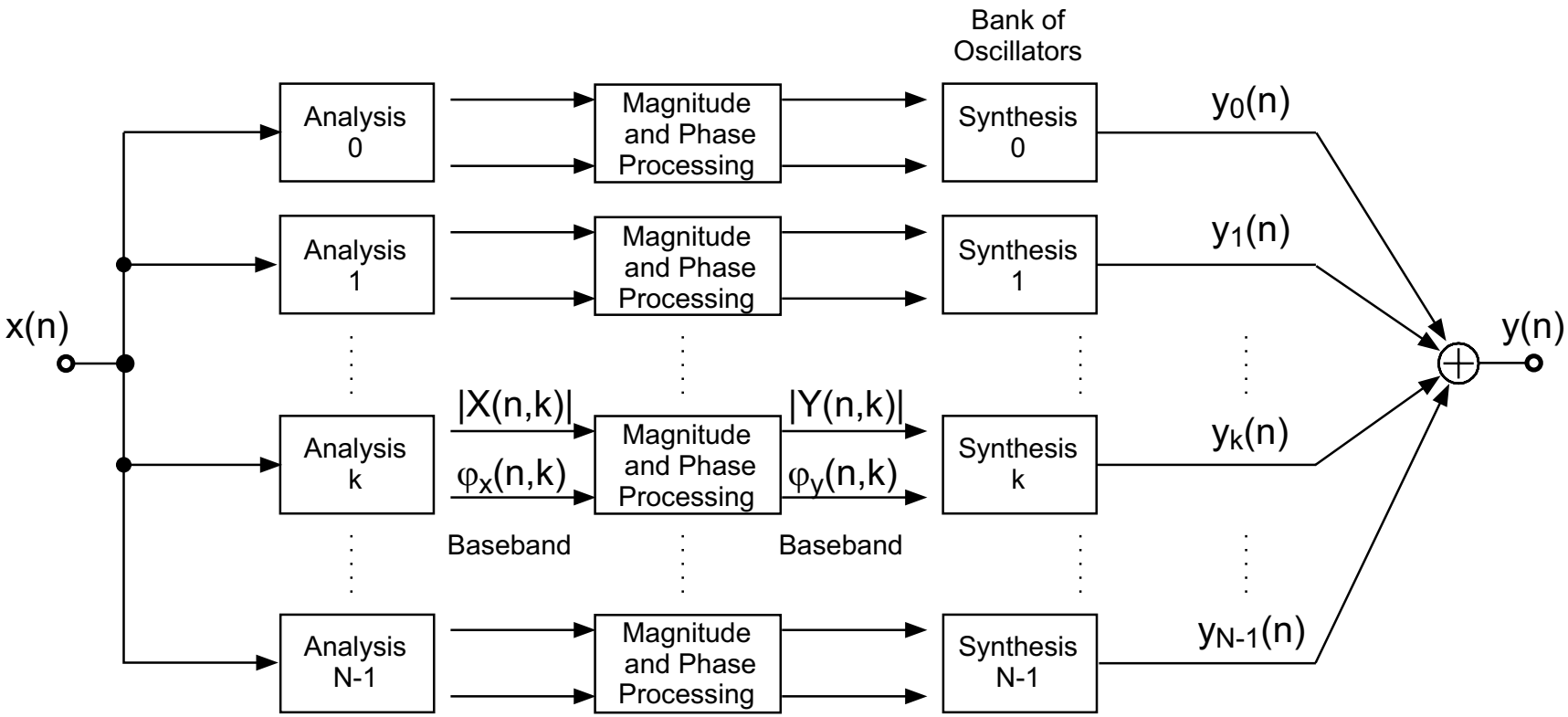
FFT/IFFT interpretation



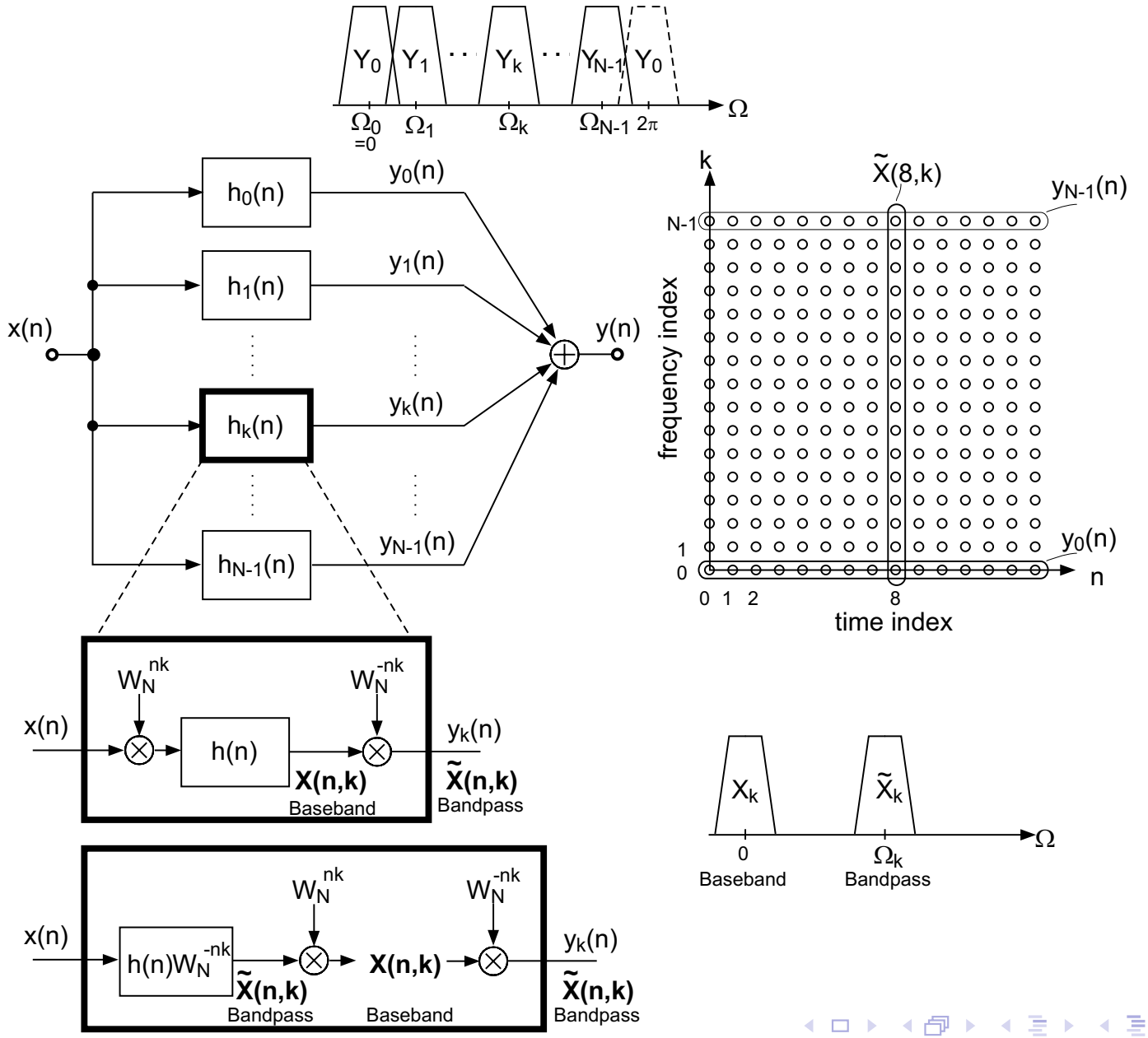


Phase vocoder as a filter bank

Each row of STFT is output of a bandpass filter

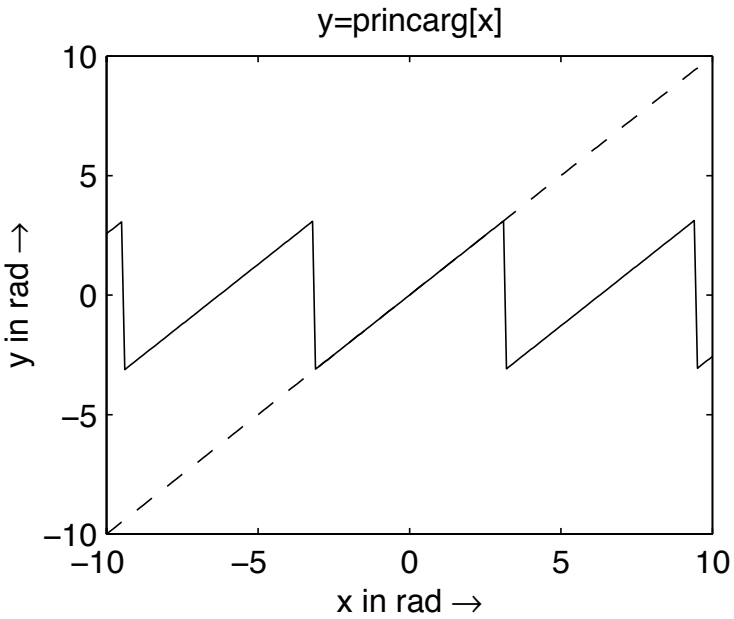


Filterbank interpretation



Phase unwrapping

- Phase $\phi[n, k]$ is only measured modulo 2π
  - Looks like noise...



- Can't differentiate directly to compute instantaneous frequency
- Unwrap phase by keeping track of cumulative phase variation:

$$\tilde{\phi}[n, k] = \frac{2\pi k}{N} n + \phi[n, k]$$

Predicting phase

Given stable sinusoid with frequency Ω_k, can predict the phase of the next frame:

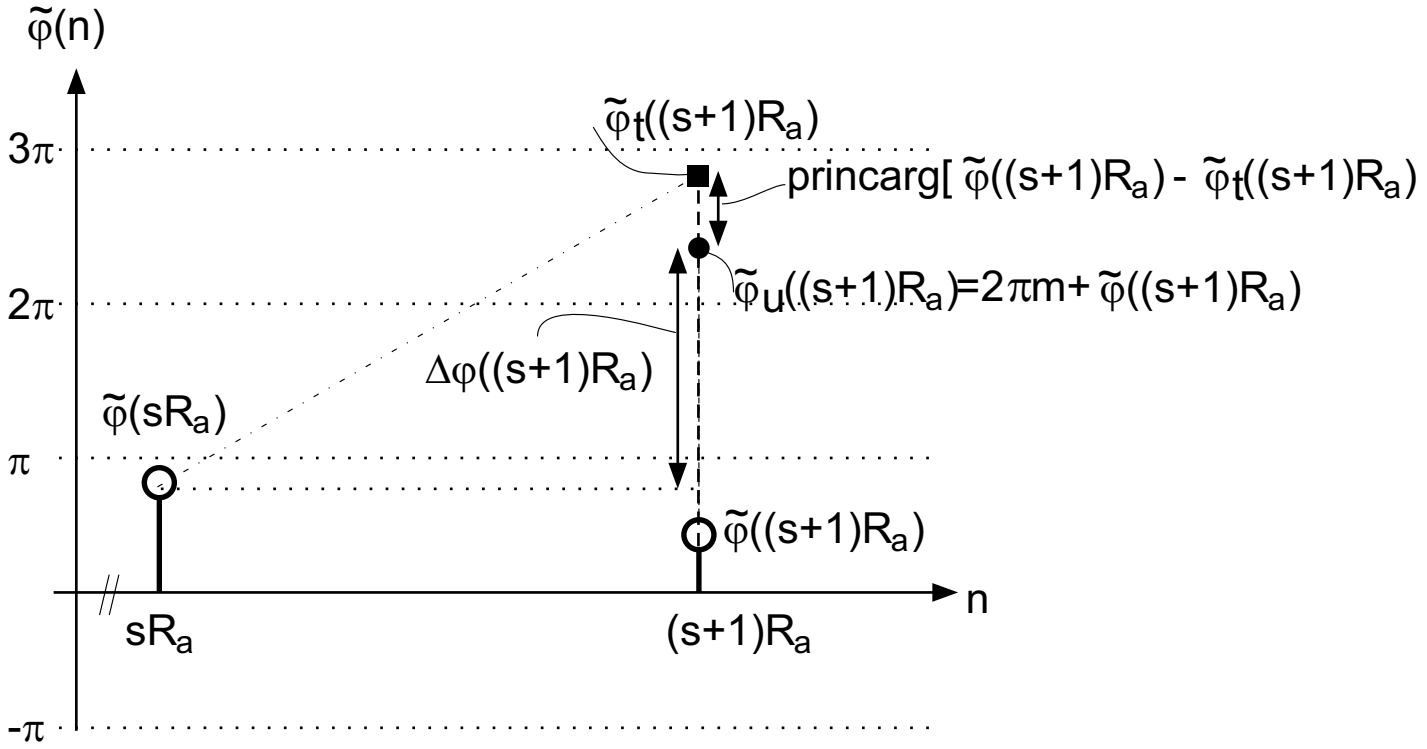
Observed phase for current frame

$$\underbrace{\tilde{\varphi}_t\big((s+1)R\big)}_{\text{Target phase for next frame}} = \overbrace{\tilde{\varphi}(sR)} + \underbrace{\Omega_k R}_{\text{phase increment}}$$

$$R = \text{hop size}$$

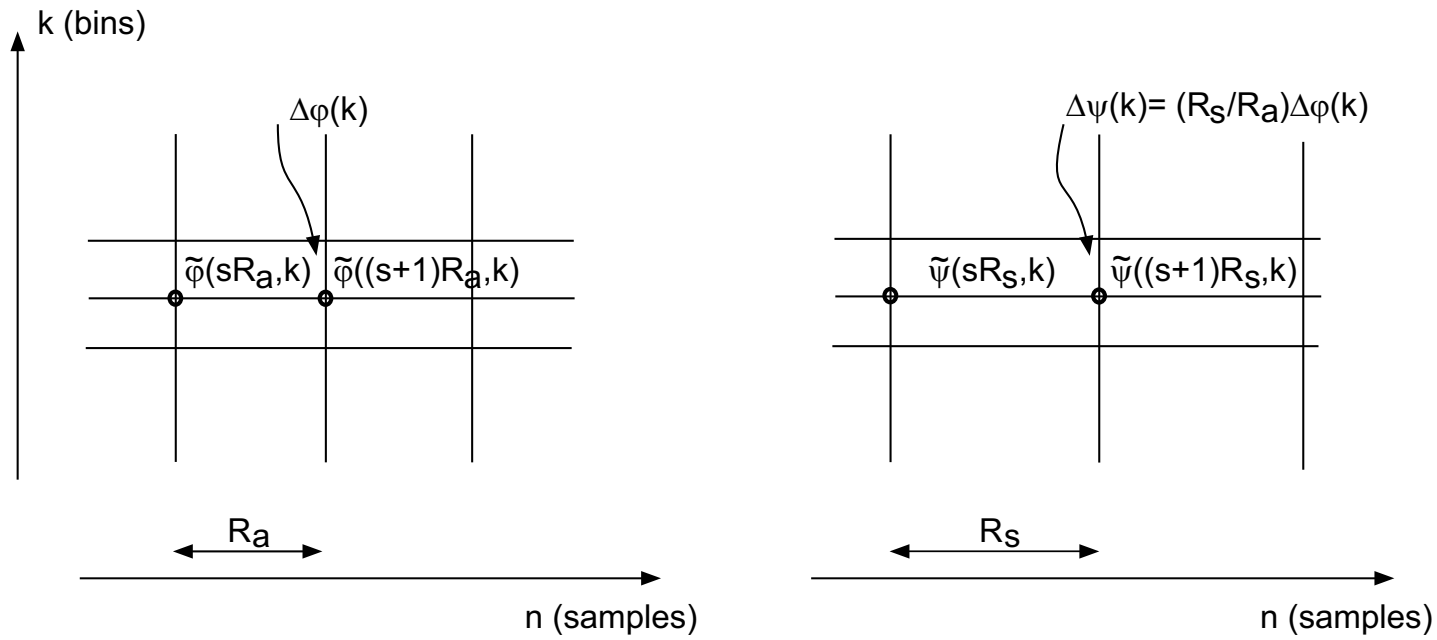
$$\Omega_k = \frac{2\pi k}{N}$$

Instantaneous frequency



$$f_i\left((s+1)R\right) = \frac{\Delta\phi\left((s+1)R\right)}{2\pi R} f_s$$

Applications: timescale modification



- Magnitude from 'stretched' spectrogram:

$$|Y[n,k]| = |X[vn,k]|$$

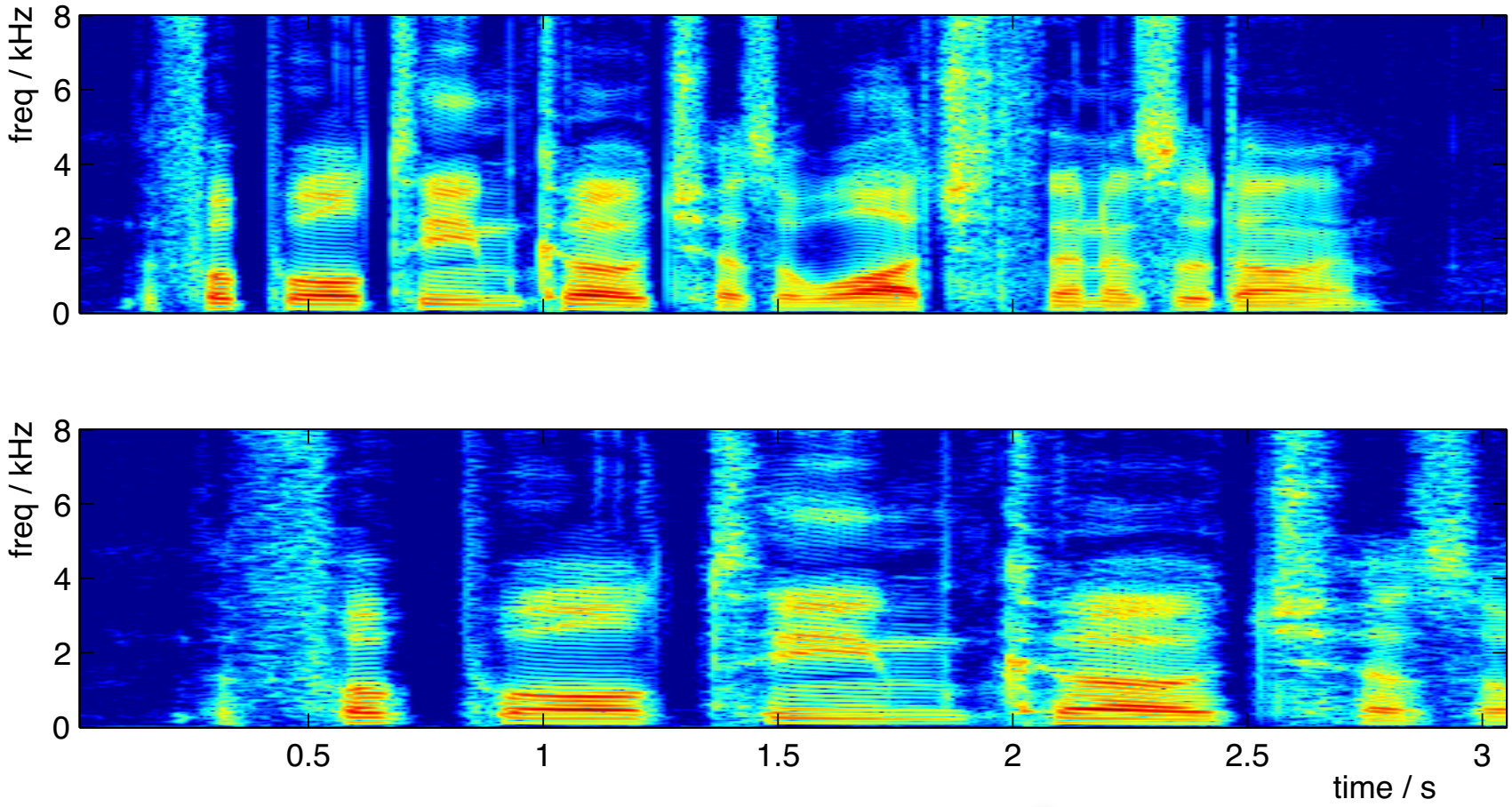
- But preserve phase increment (f_i) between slices:

$$\Delta\phi_Y[n,k] = \Delta\phi[vn,k]$$

Guitar (20%)

Pop (15%)

Phase vocoder timescale modification example



Tends to introduce "metallic" noise

Other PV/STFT effects

- Mutation between sounds
  - Reconstruct using magnitude from one sound and phase from another
- Whisperization
  - Small window, replace phase with noise

Reading

DAFX Chapter 8 - Time-frequency processing