Speech Analysis:
What are we looking for?

So far:
- STFT (spectrogram)
- LP Analysis

Next, we consider a method to separate convolved components of a signal.

The "cepstrum" of a signal $x[n]$ is:

$$c[n] = \hat{x}[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |X(e^{j\omega})| e^{j\omega n} d\omega$$

Now,

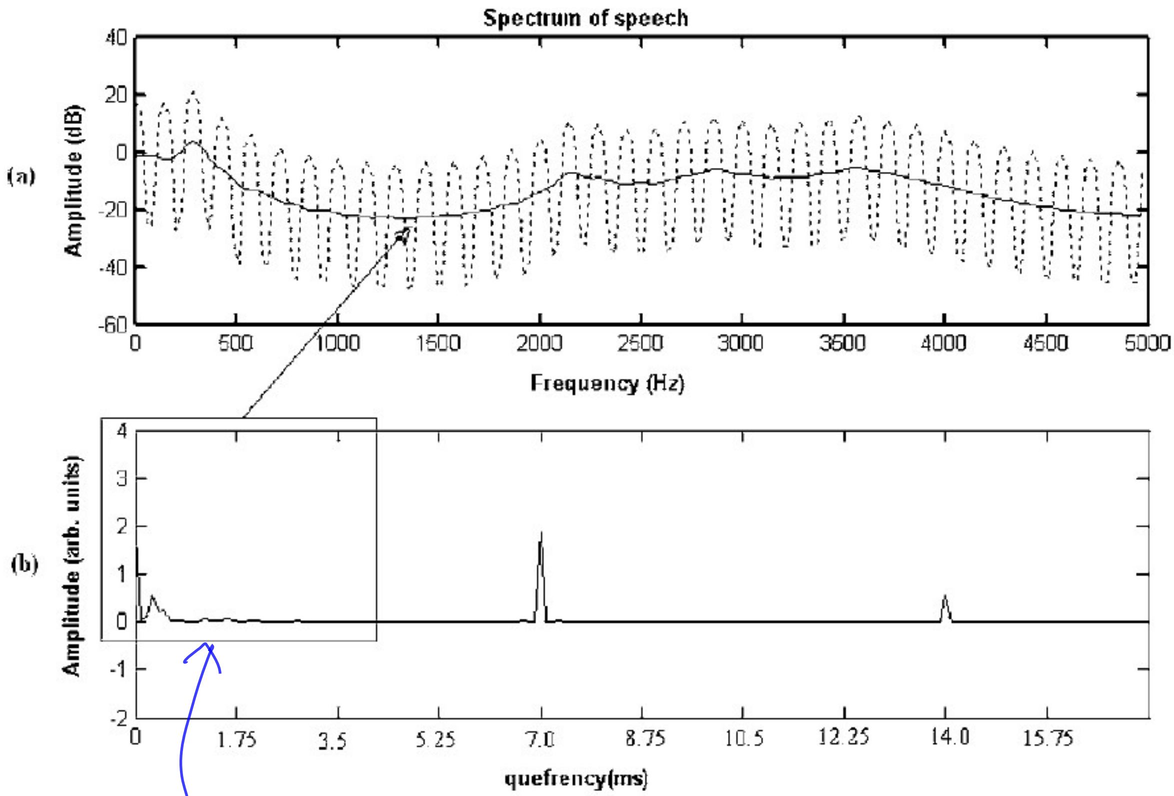
$$s[n] = u[n] * h[n]$$

$$\Rightarrow \quad S(e^{j\omega}) = U(e^{j\omega}) \cdot H(e^{j\omega})$$

$$\Rightarrow \quad \log|S(e^{j\omega})| = \log|U(e^{j\omega})| + \log|H(e^{j\omega})|$$

$$\Rightarrow \quad \hat{s}[n] = \hat{u}[n] + \hat{h}[n]$$

Figure 2



cepstrum window

Figure 3

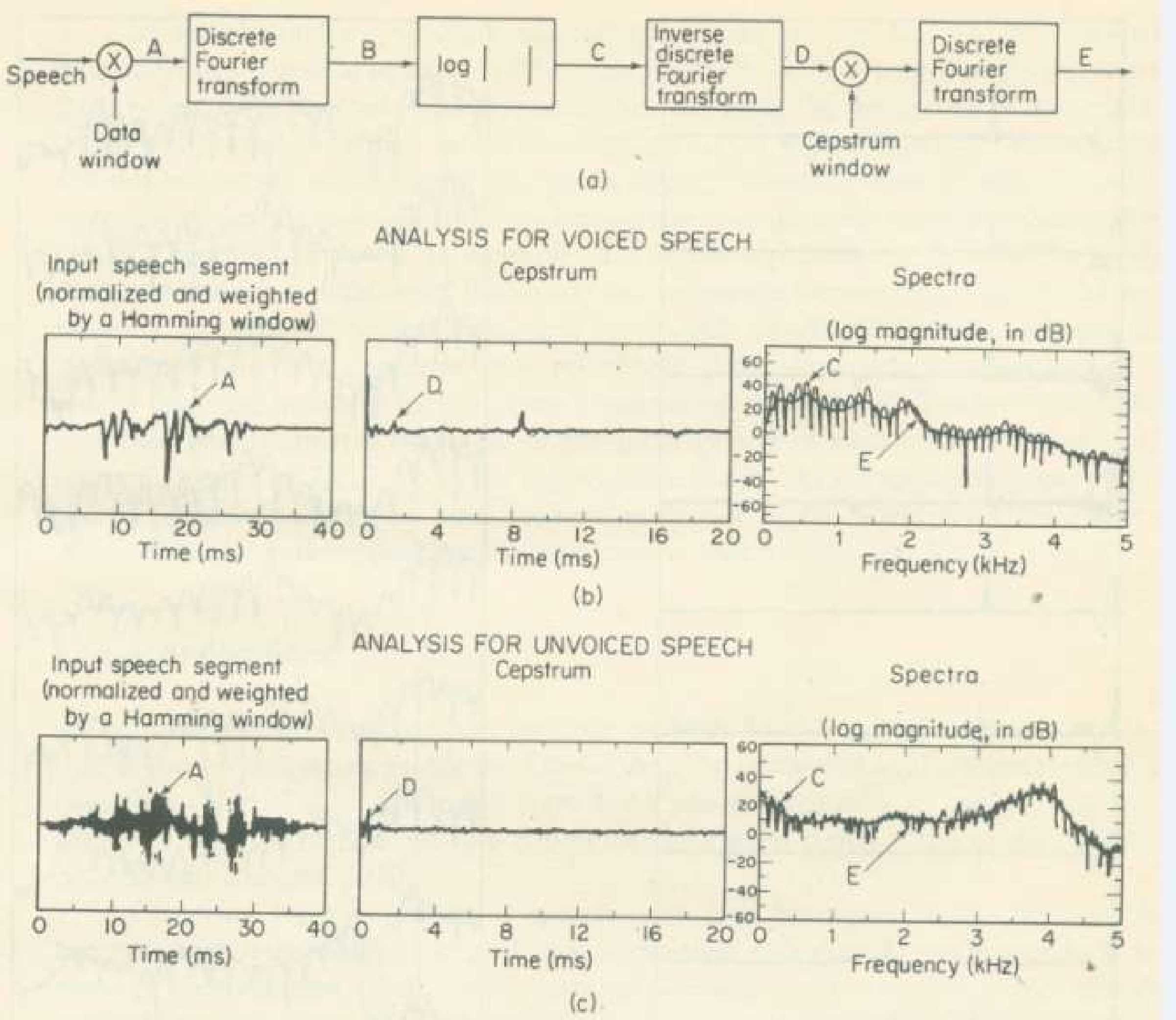


Figure 12.27 (a) System for cepstrum analysis of speech signals. (b) Analysis for voice speech. (c) Analysis for unvoiced speech.

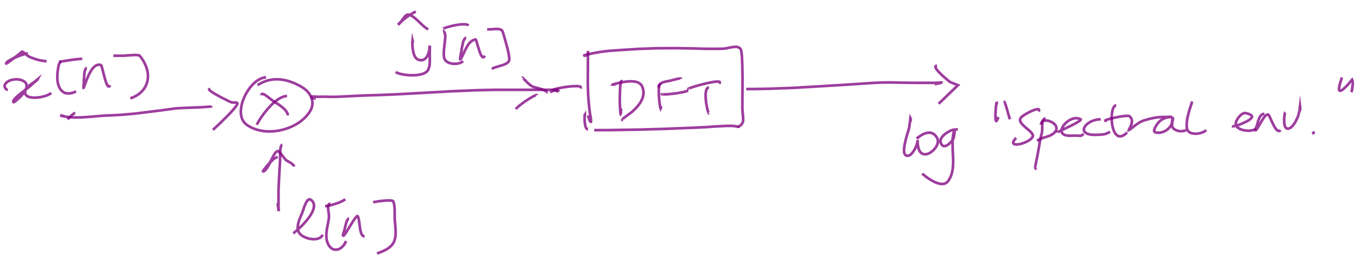
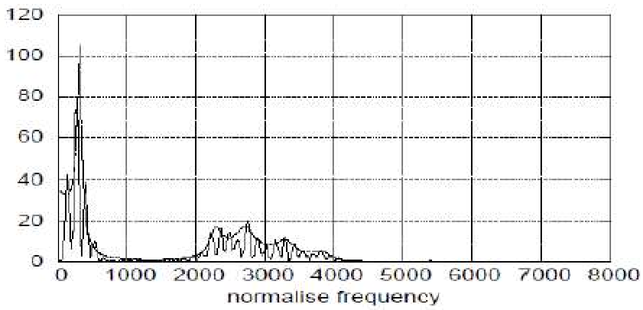
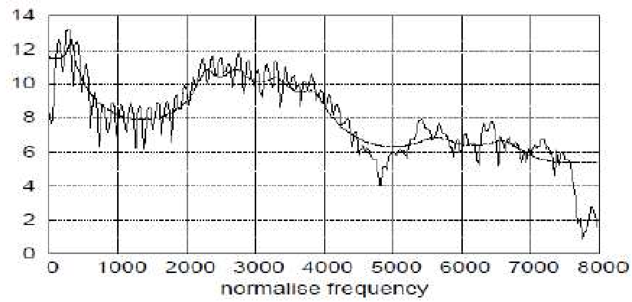


Figure 1

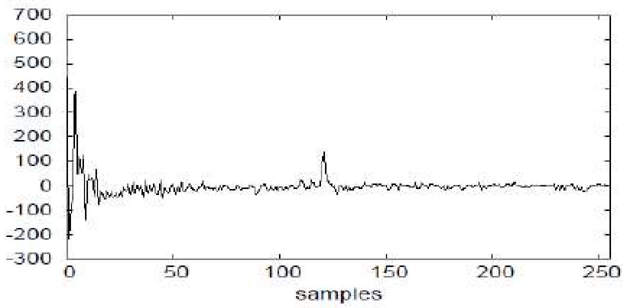
$\mathcal{F}\{x[n]\}$



$\log|\mathcal{F}\{x[n]\}|$



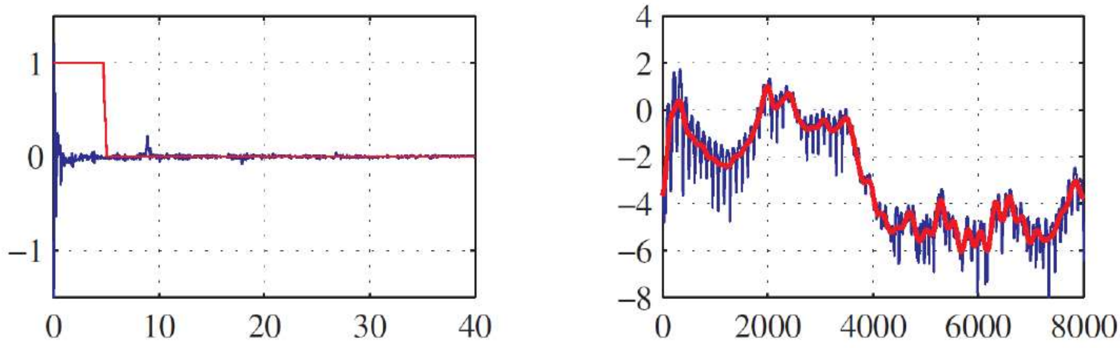
$\mathcal{F}^{-1}\{\log|\mathcal{F}\{x[n]\}|\}$



[Taylor, 2009]

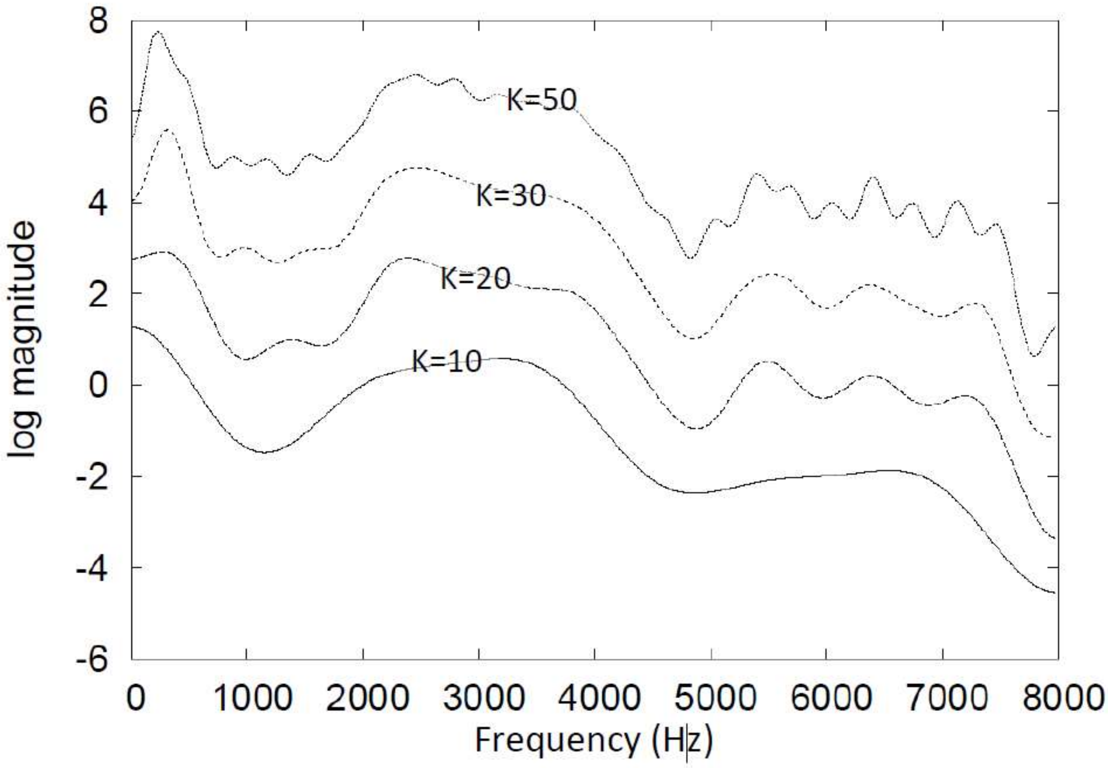
Figure 4

Liftering in the cepstral domain



[Rabiner & Schafer, 2007]

Changing cepstrum window length



[Taylor, 2009]

Applications

"Spectral Distance" Measure:

L2 spectral norm is widely used in speech.

$$d_2^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\log |S_1(e^{j\omega})| - \log |S_2(e^{j\omega})| \right]^2 d\omega$$

By Parseval's thm, we have:

$$d_2^2 = \sum_{m=-\infty}^{\infty} (c_1[m] - c_2[m])^2$$

can restrict to 'spec env.'

Formant tracking (& pitch detection)

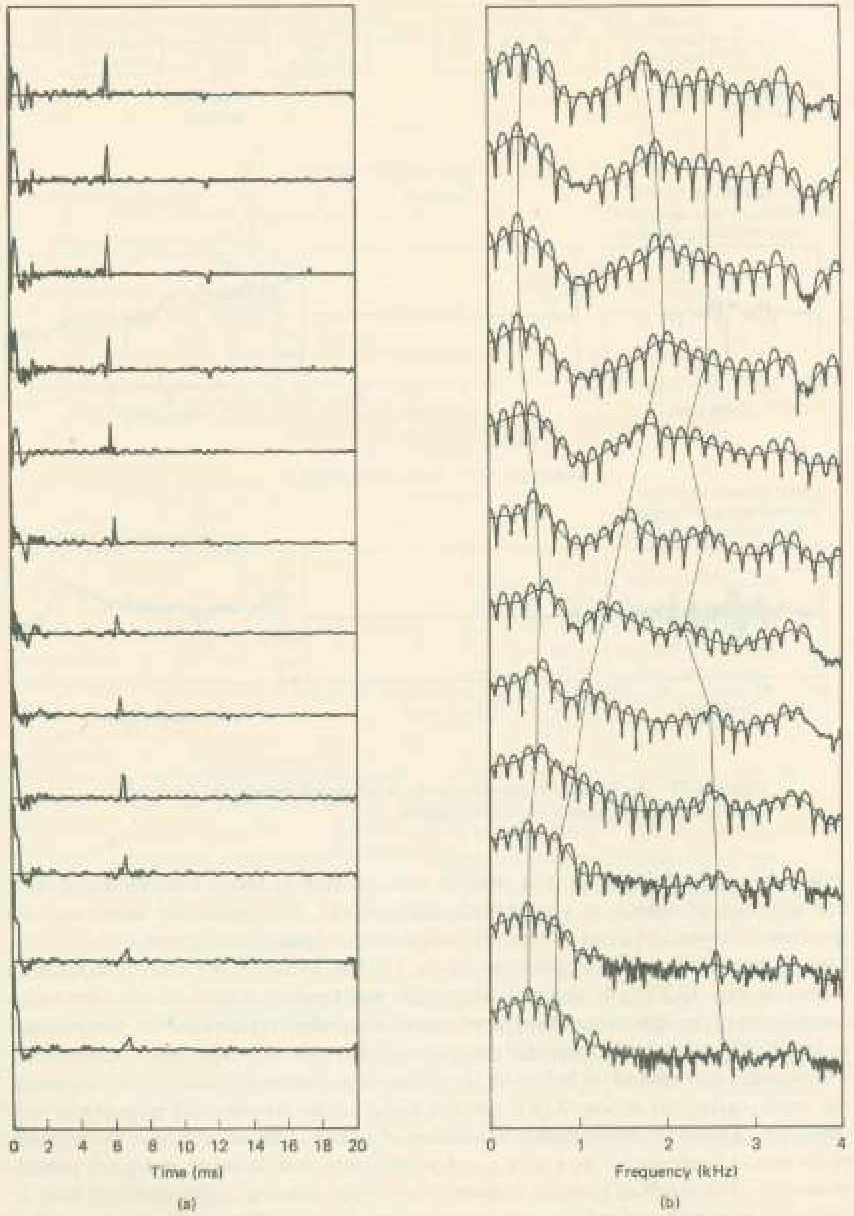


Figure 12.28 (a) Cepstra and (b) log spectra for sequential segments of voiced speech.