SHORT-TIME FOURIER ANALYSIS

We have seen the importance of short-time Fourier analysis in speech processing:

1. Speech waveform has time-varying characteristics
2. During short epochs (10 – 30 msec.), speech may be modelled as output of an LTI system — spectral characteristics provide important clues about nature of speech.

So far, we have only considered analog systems for short-time Fourier analysis — now look at digital systems to accomplish this purpose.

A fundamental concept here is the notion of windowing the data to look at only a finite portion.

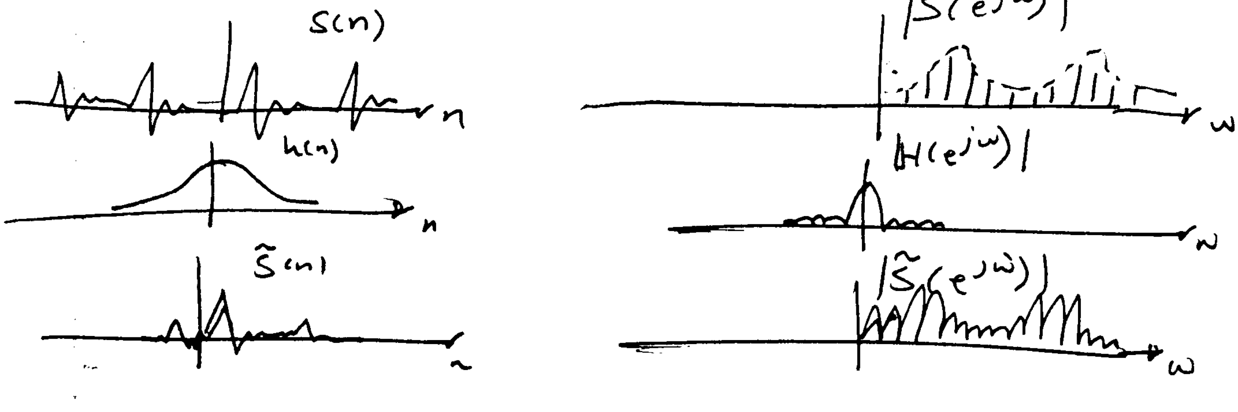
Recall discussion of <u>leakage</u> in calculation of DTFT of finite-length sinusoid. Let's look at more general case:

complete speech signal	$s(n)$	$-\infty < n < \infty$
window	$h(n)$	
windowed signal	$\tilde{s}(n) = h(n)\, s(n)$	

In frequency domain (DTFT):

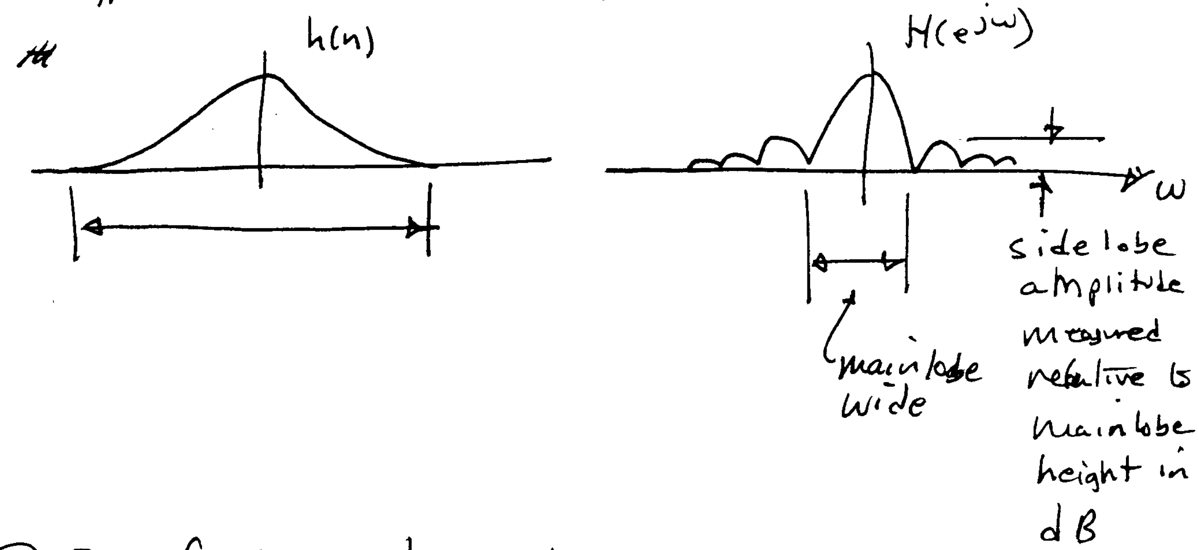
$$\tilde{S}(e^{j\omega}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} H(e^{j(\omega-\mu)})\, S(e^{j\mu})\, d\mu$$

examples



Comments on windows

1. Typical window with Spectrum



2. For fixed window type,

   main lobe width $\propto \frac{1}{N}$

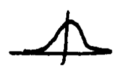
   sidelobe amplitude is fixed

3. For fixed window length N

   mainlobe width $\sim \dfrac{1}{\text{sidelobe amplitude}}$

   as we consider different window types.

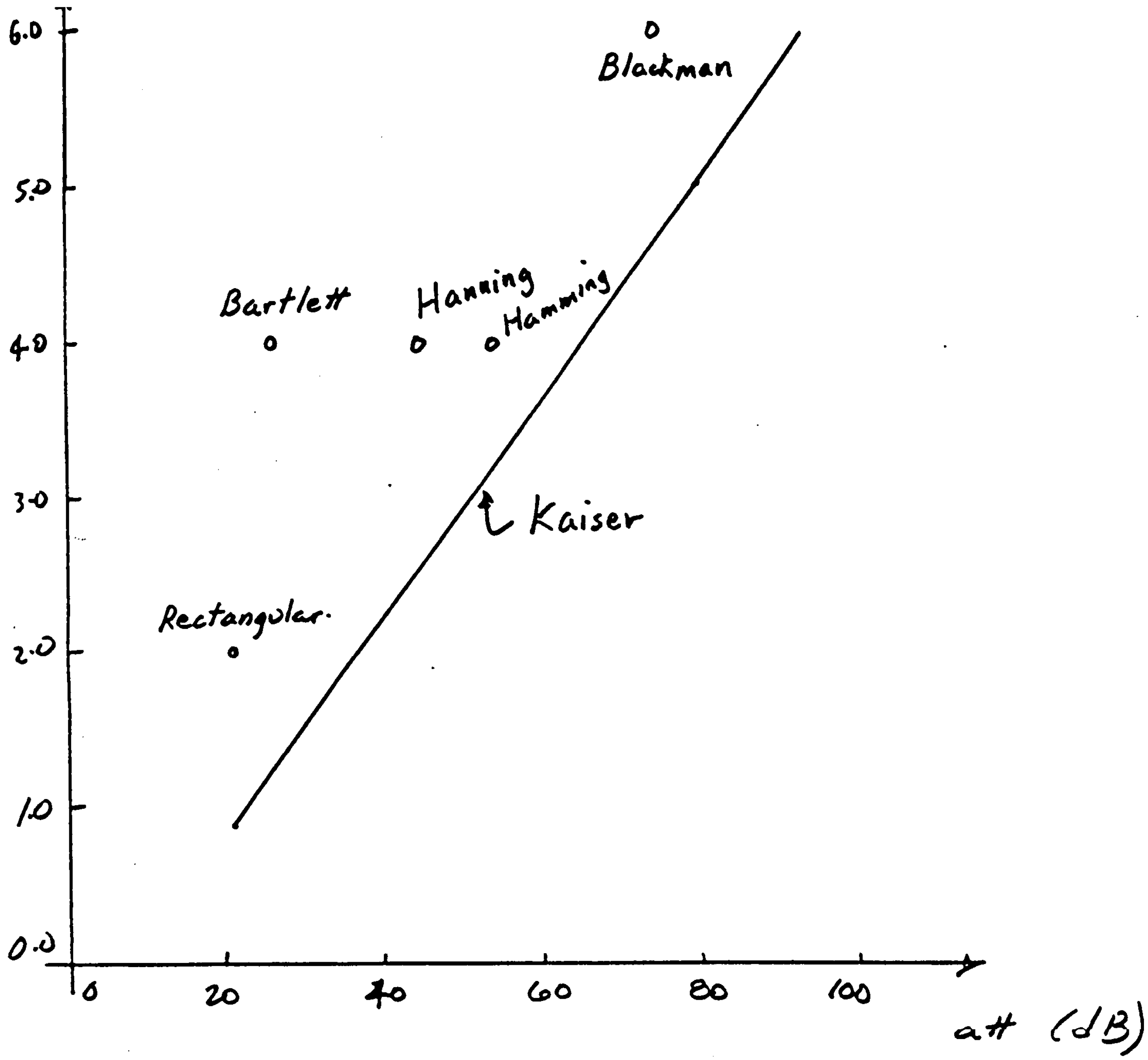
Examples

Type		Main lobe (rad./sample)	Side lobe (dB)
1. Rectangular		$\frac{4\pi}{N}$	−13
2. Bartlett		$\frac{8\pi}{N}$	−27
3. Hanning		$\frac{8\pi}{N}$	−32
4. Hamming		$\frac{8\pi}{N}$	−43
5. Blackman		$\frac{12\pi}{N}$	−58

See Cadzow, pp. 104 – 105

(4) Mainlobe width is responsible for blurring

(5) Sidelobes cause ~~[illegible]~~ noisy appearance in spectrum



N · ΔF vs. Attenuation for Several Windows

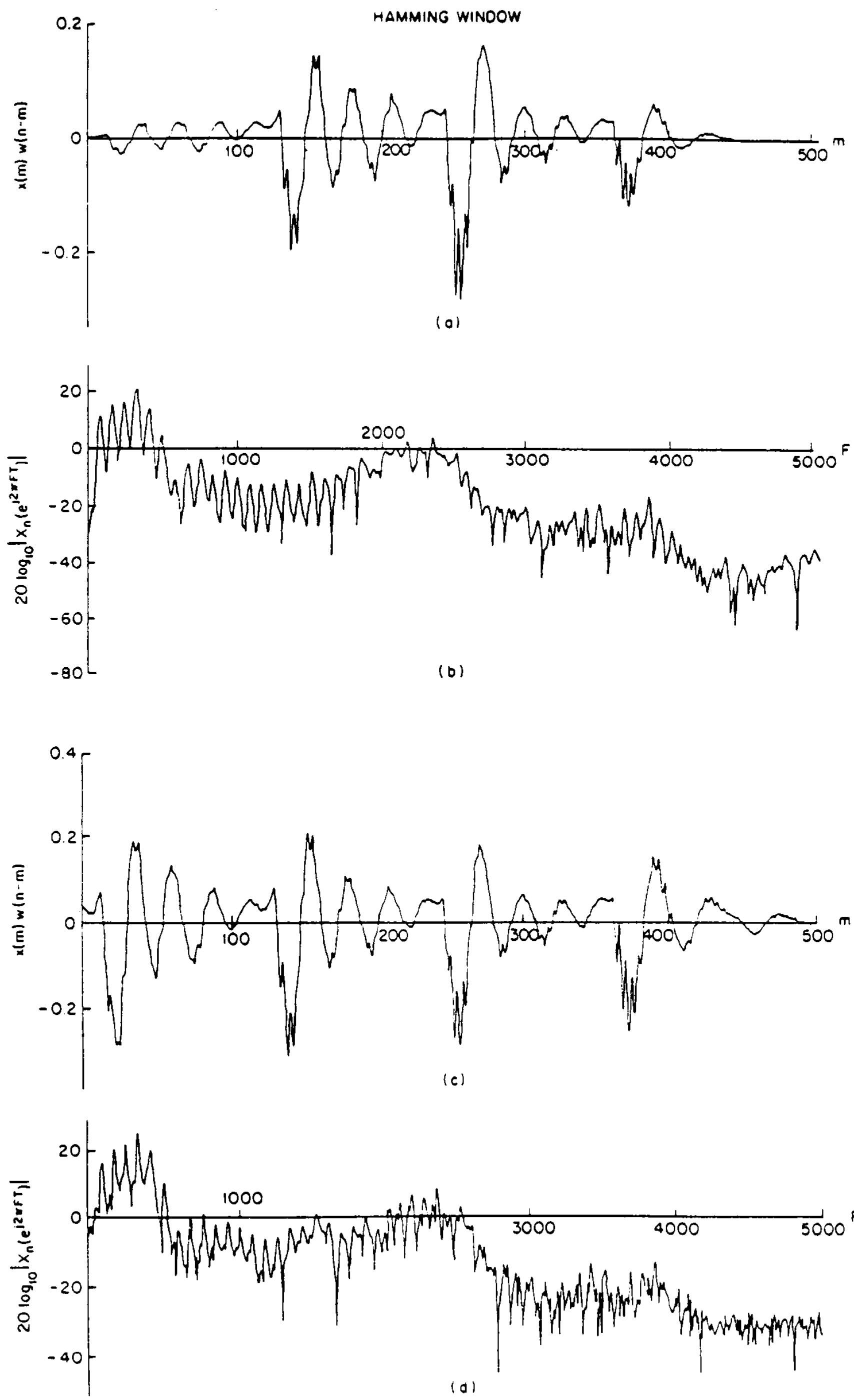


Fig. 6.2 Spectrum analysis for voiced speech using a 50 msec (a,b) Hamming window; (c,d) rectangular window. Parts (a) and (c) show time waveforms; parts (b) and (d) show corresponding spectra.

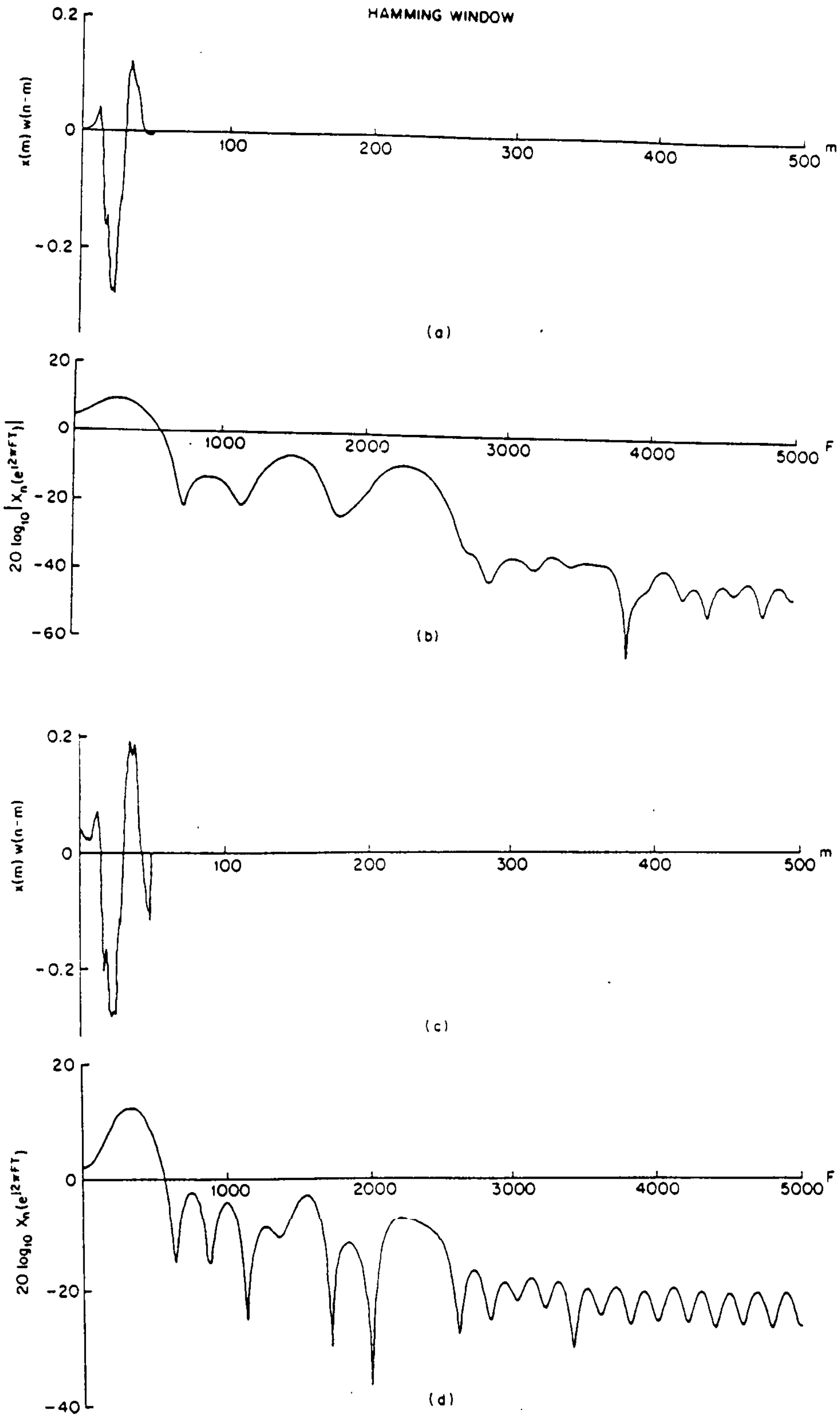


Fig. 6.3 Spectrum analysis of voiced speech using a 5 msec (a,b) Hamming window; (c,d) rectangular window. Parts (a) and (c) show time waveforms; parts (b) and (d) show corresponding spectra.

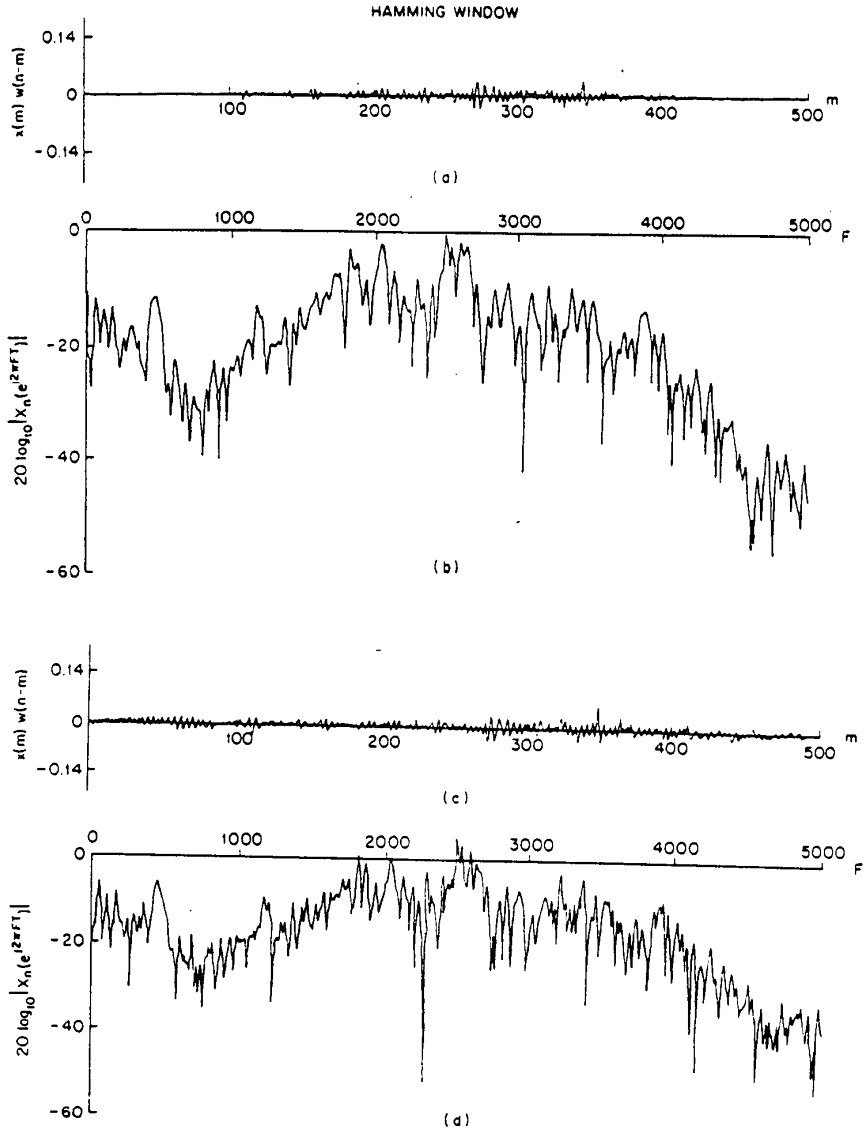


Fig. 6.4 Spectrum analysis of unvoiced speech using a 50 msec (a,b) Hamming window; (c,d) rectangular window. Parts (a) and (c) show time waveforms; parts (b) and (d) show corresponding spectra.

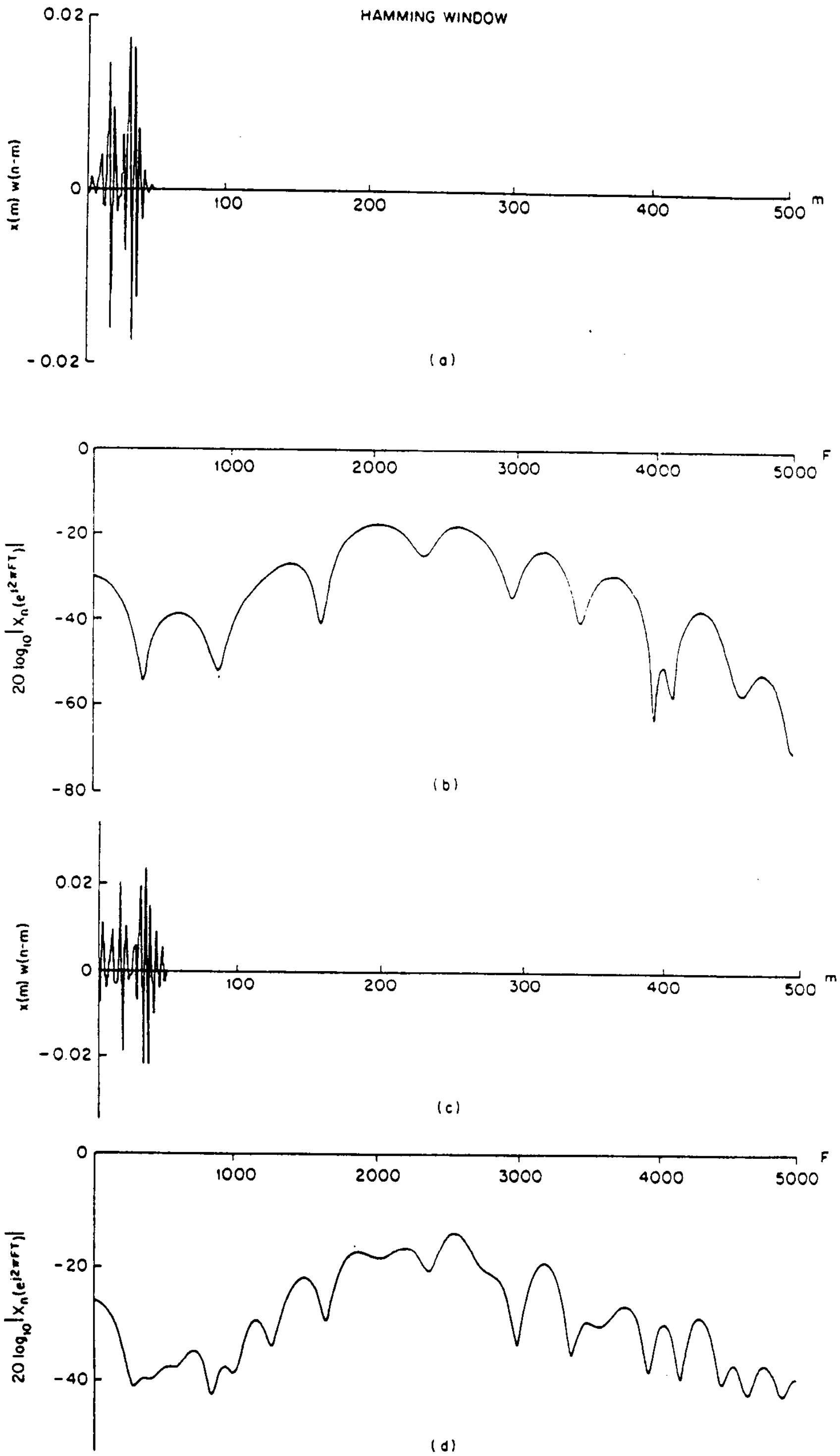


Fig. 6.5 Spectrum analysis of unvoiced speech using a 5 msec (a,b) Hamming window; (c,d) rectangular window. Parts (a) and (c) show time waveforms; parts (b) and (d) show corresponding spectra.