University of California
Berkeley

College of Engineering
Department of Electrical Engineering
and Computer Sciences

Professors : N.Morgan / B.Gold
EE225D Spring,1999

LPC Analysis

Lecture 22

Hearing and Speech Engineering

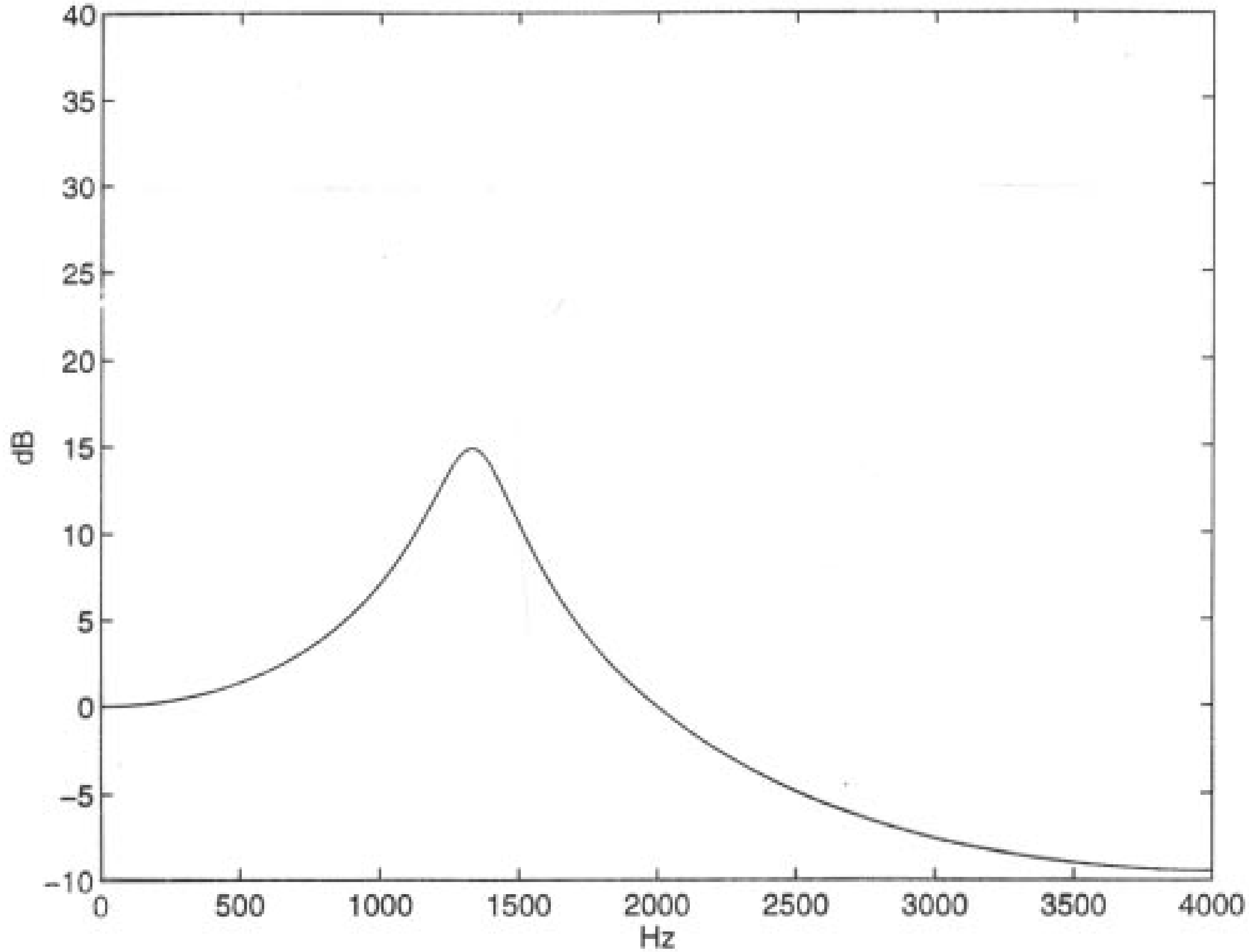
- Focus on power spectrum (not phase)
- Spectral envelope for phonetic discrimination
- Less accuracy required at high frequencies
- Emphasis on spectral peaks

Spectral Envelope Estimation

- Filter banks
- Cepstral Analysis
- Linear Predictive Coding (LPC)

Incorporate Production

- Assume simple excitation /vocal tract model
- Assume vocal tract like series resonators
- Find best spectrum based on resonators


40
35
30
25
20
15
10
5
0
−5
−10
dB
0
500
1000
1500
2000
2500
3000
3500
4000
Hz

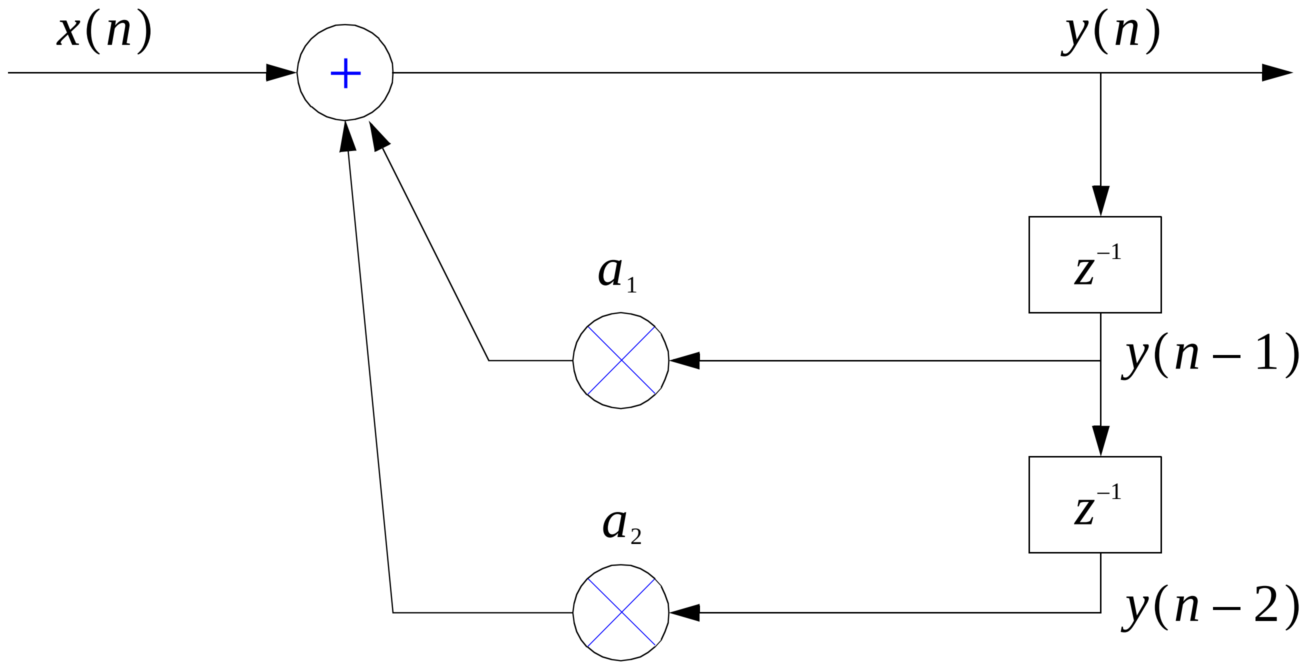
Pole-only resonator

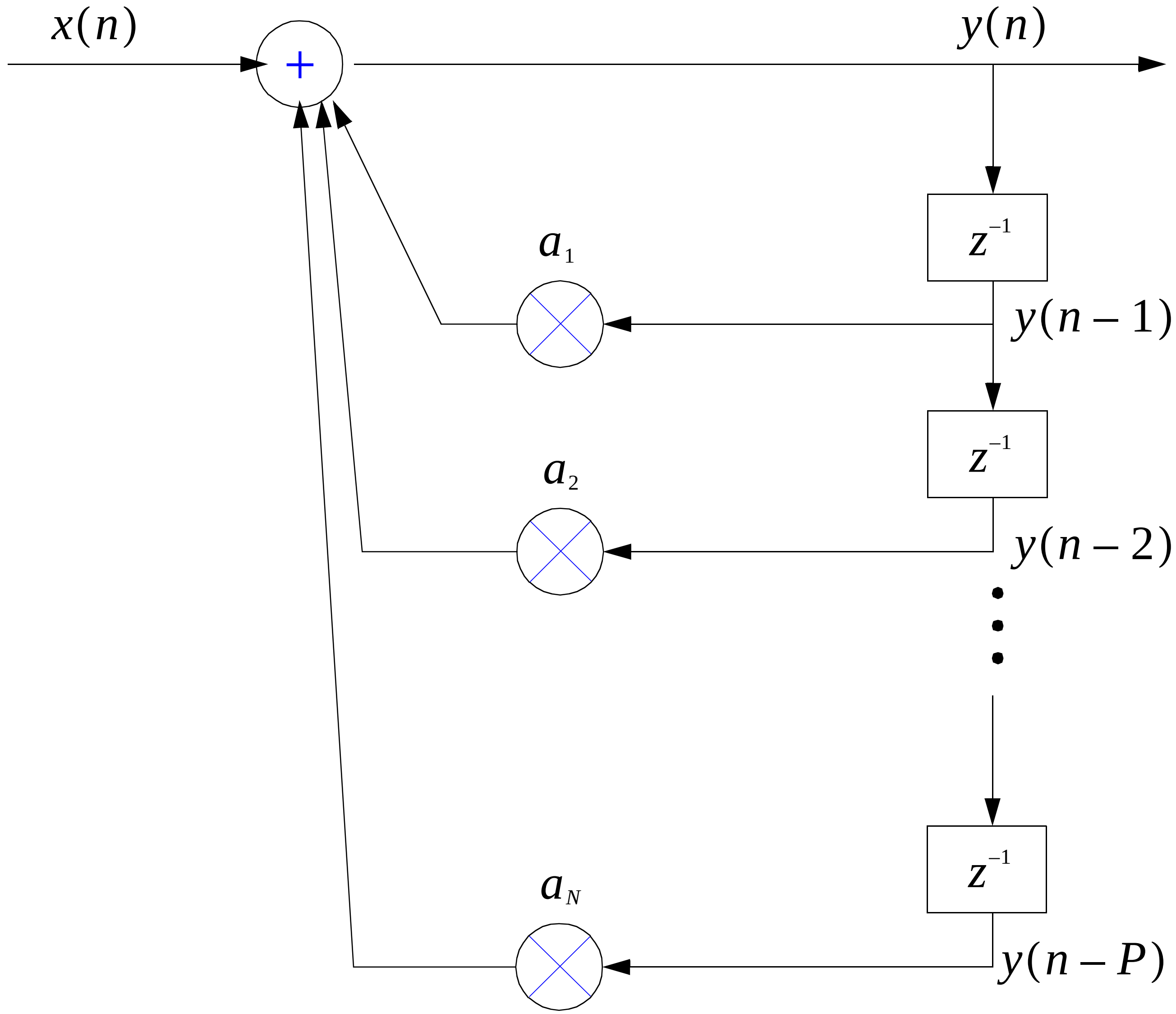
- $H_i(z) = \dfrac{1}{1 - b_i z^{-1} - c_i z^{-2}}$

For complex pole pair,

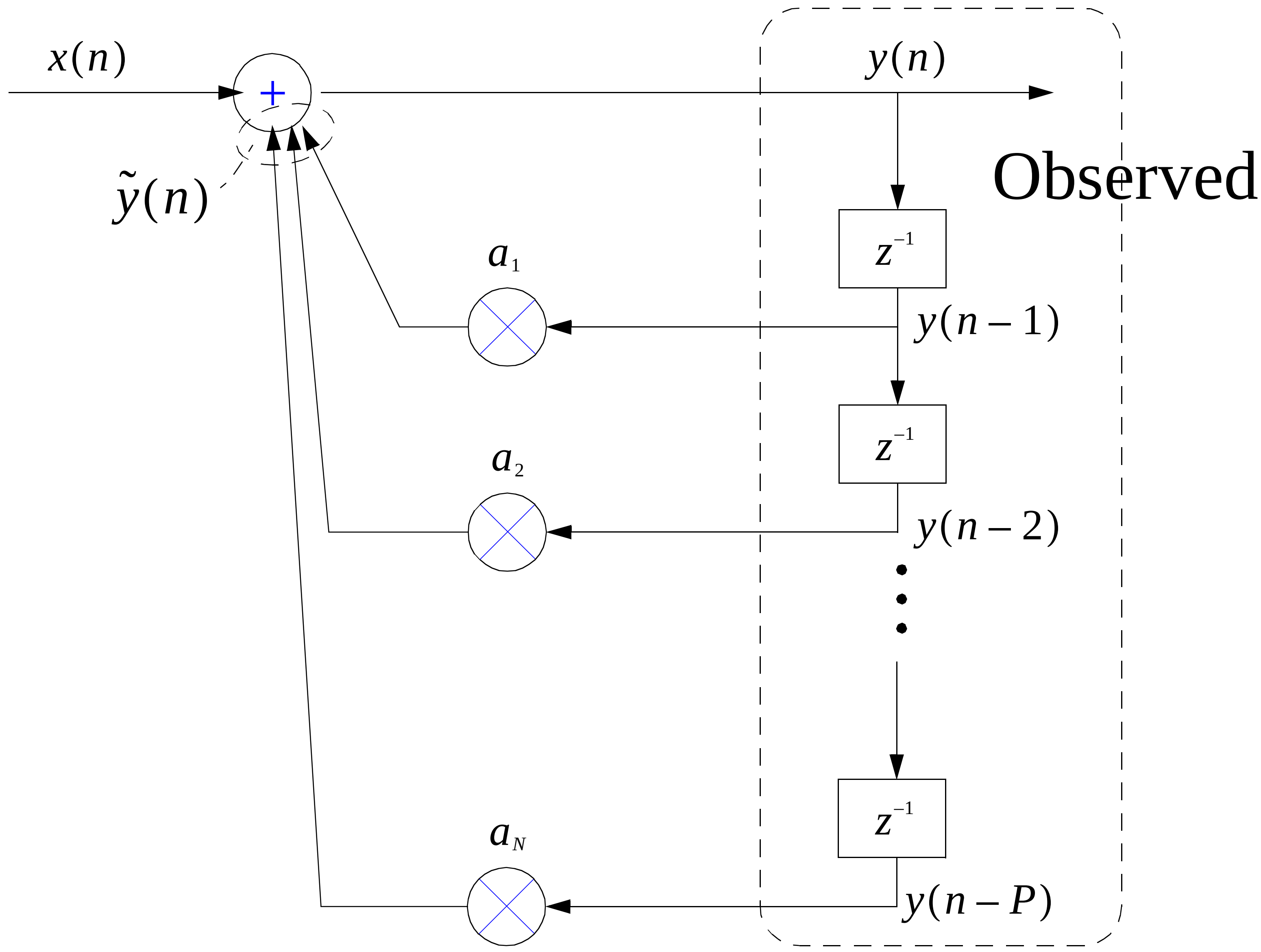
- $b_i = 2r\cos\Theta, c_i = r^2$

where r is pole magnitude and Θ is pole angle.


$x(n)$
$y(n)$
z^{-1}
a_1
$y(n-1)$
z^{-1}
a_2
$y(n-2)$



$x(n)$
$+$
$y(n)$
z^{-1}
a_1
$y(n-1)$
z^{-1}
a_2
$y(n-2)$
z^{-1}
a_N
$y(n-P)$



$x(n)$
$y(n)$
Observed
$\tilde{y}(n)$
a_1
z^{-1}
$y(n-1)$
a_2
z^{-1}
$y(n-2)$
a_N
z^{-1}
$y(n-P)$

Error Signal

$$e(n) = y(n) - \tilde{y}(n) = y(n) - \sum_{j=1}^{P} a_j y(n-j)$$

$$E(z) = Y(Z) - Y(z) = Y(z) - \sum_{j=1}^{P} a_j z^{-j} Y(z)$$

$$= Y(z)\underbrace{\left(1 - \sum_{j=1}^{P} a_j z^{-j}\right)}_{\frac{1}{H(z)}}$$

or

$$E(z) = \frac{Y(Z)}{H(z)}$$

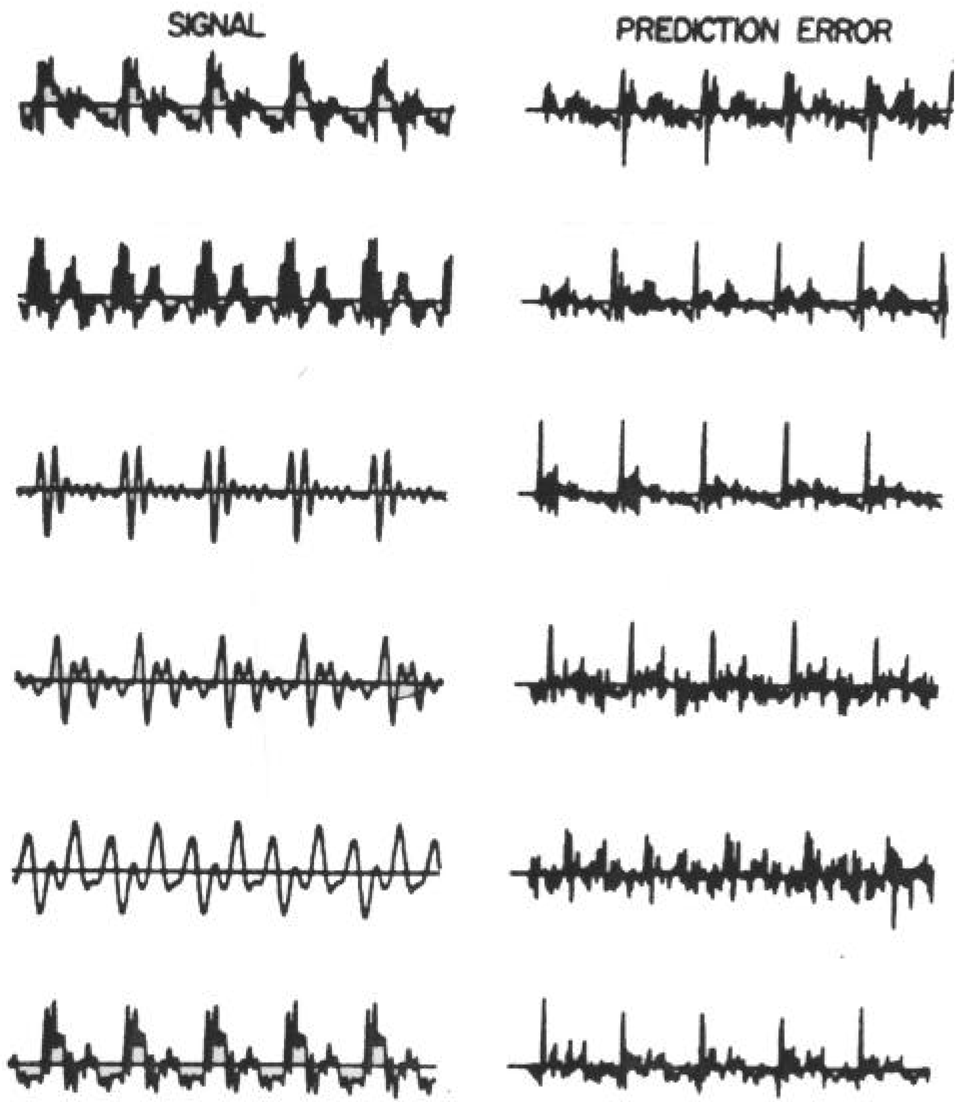


Figure 21.3 : Residual error waveforms for several vowels.

Some LPC Issues

- Error criterion for minimization
- Model order

Error Criterion

$$D = \sum_{n=0}^{N-1} e^2(n) = \int_{-\pi}^{\pi} |E(\omega)|^2 \frac{d\omega}{2\pi}$$

SO

$$D = \int_{-\pi}^{\pi} \frac{|Y(\omega)|^2}{|H(\omega)|^2} \frac{d\omega}{2\pi}$$

LPC peak modeling

- Total error constrained to be (at best) gain factor squared
- Error where model spectrum is larger contributes less
- Tends to “hug” peaks

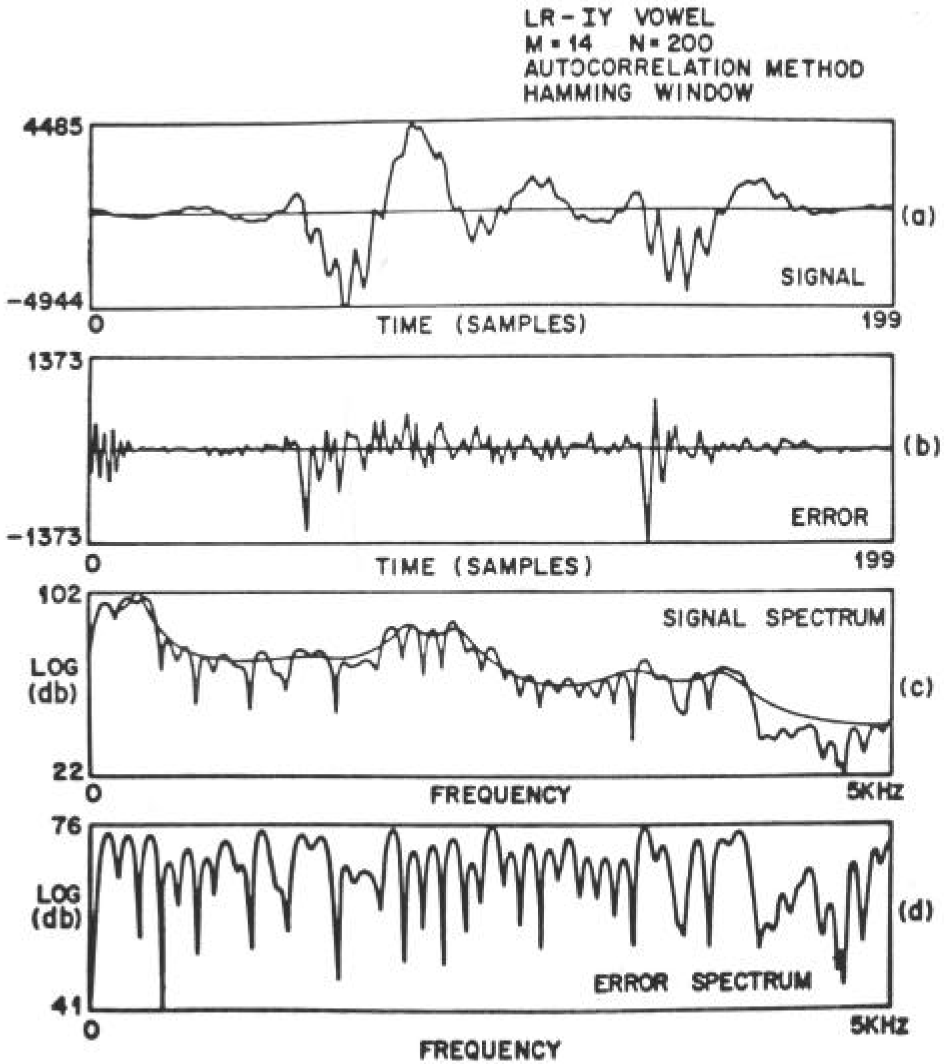


Figure 3.32 : Typical signals and spectra for LPC autocorrelation method for a segment of speech spoken by a male speaker (after Rabiner et al.)

More Effects of Error Criterion

- Globally tracks, but worse match in log spectrum for low values
- "Attempts" to model anti-aliasing filter
- Ill conditioned for wide range of values

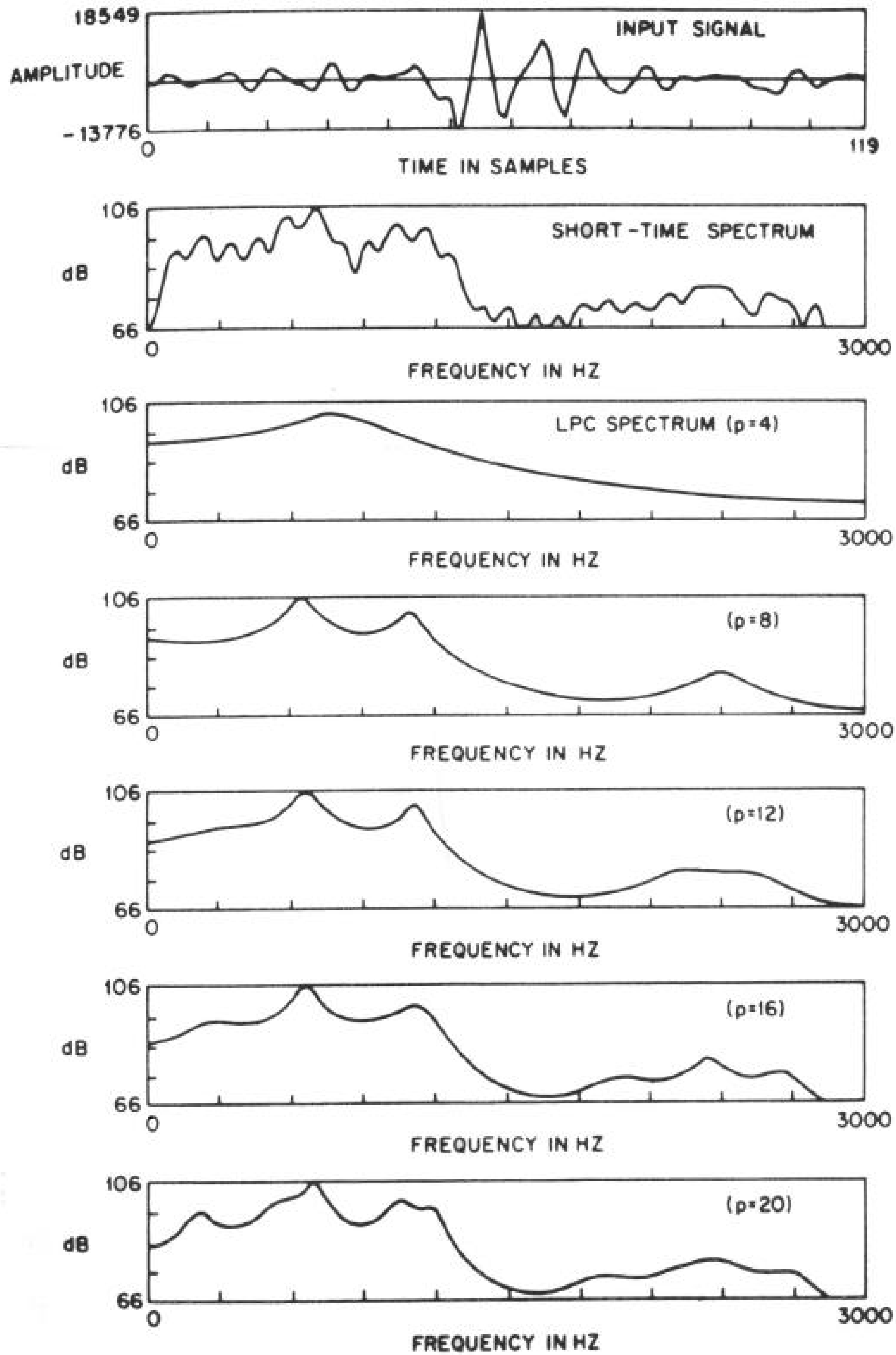
Other LPC properties

- Behavior in noise
- Sharpness of peaks
- Speaker dependence

Model Order

- Too few, can’t represent formants
- Too many, model detail, e.g., harmonics
- Too many, low error, ill-conditioned matrices

Figure 3.36 : Spectra for a vowel sound for several values of predictor order, p.



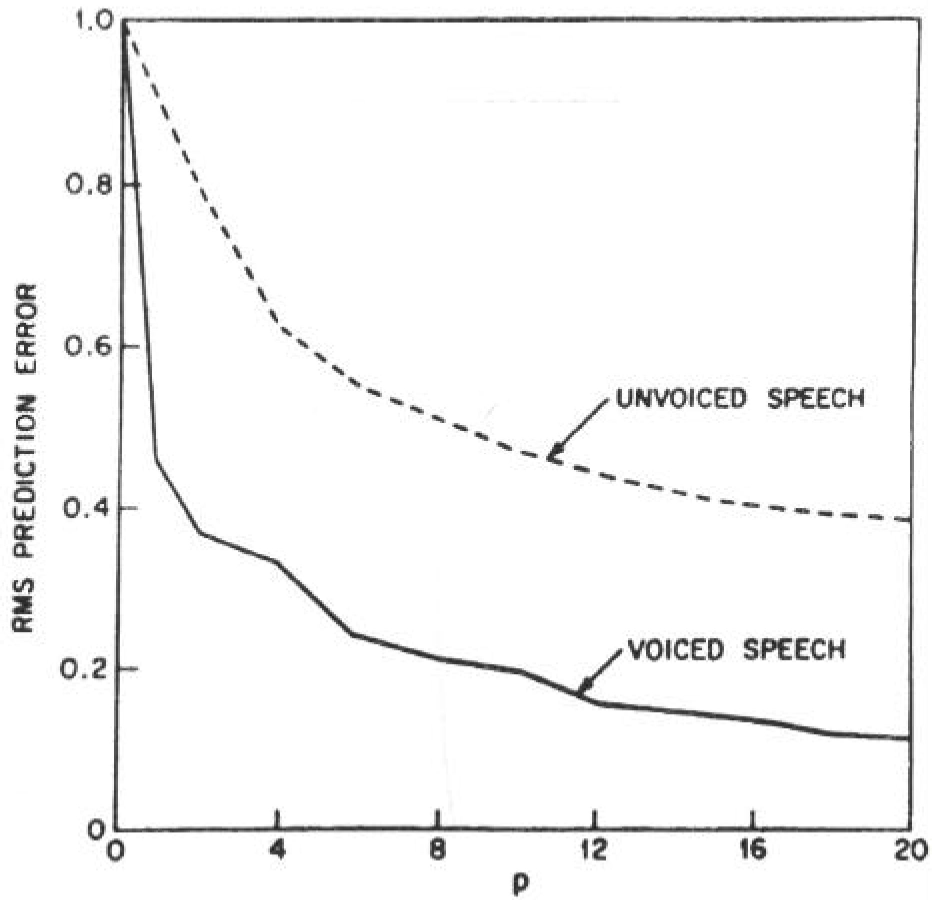


Figure 21.6 : RMS prediction error for different model orders.

Optimal Model Order

- Akaike Information Critetion (AR time series)
- Cross-validation (trial and error)

Coefficient Estimation

- Minimize squared error - set derivatives to zero
- Compute in blocks or on-line
- For blocks, use autocorrelation or covariance methods (windowing)

Minimizing the error

- $D = \sum_{n=0}^{N-1} e^2(n) = \sum_{n=0}^{N-1}\left(y(n) - \sum_{j=1}^{P} a_j y(n-j)\right)^2$

If we take partial derivatives with respect to each a, we get P equations of the form.

- $\sum_{j=1}^{P} a_j \phi(i, j) = \phi(i, 0) for i = 1, 2, \ldots, P$

Where $\phi(i, j)$ is a correlation sum between varsions of the speech signal delayed by i and j points.

Solving the equations

- Autocorrelation method : Levinson or Durbin recursions ($O(P^2)$ operations) ; uses Toeplitz property, guaranteed stable
- Covariance method : Cholesky decomposition ($O(P^3)$ operations) - just uses symmetic property

LPC-based representations

- Predictor Polynomial
- Root pairs
- Reflection coefficients
- Log area ratios
- Cepstrum

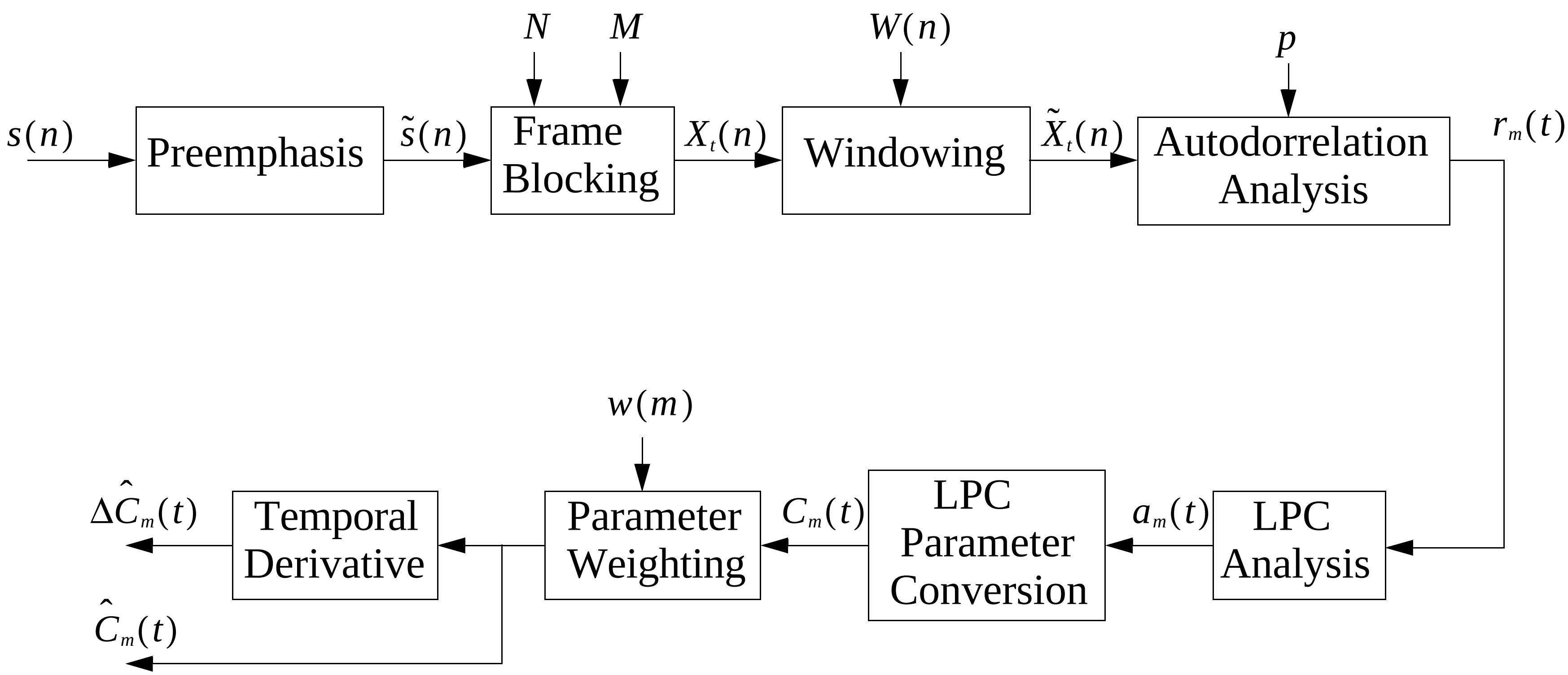


Figure 3.37 : Block diagram of LPC processor for speech recognition.

	Filter Banks	Cepstral Analysis	LPC
Reduced pitch effects	X	X	X
Exciation estimate		X	X
Direct access to spectra	X		
Less resolution at HF	X		
Orthogonal outputs		X	
Peak-hugging property			X
Reduced computation			X

Table 1 : Basic methods for spectral envelope estimation in speech.