# Linear Prediction Analysis of Speech

# Linear Prediction Analysis of Speech

- LPC provides good model of speech
  - ✓ All pole model of LPC provides good approximation to vocal-tract spectral envelope for quasi-steady state voiced regions of speech.
  - ✓ Less effective for unvoiced and transition regions of speech

- LPC analysis leads source & vocal tract separation
  - ✓ LP synthesis model (Generative model)
  - ✓ LP analysis model (inverse filter model)

- LPC is analytically tractable model
  - ✓ Mathematically precise
  - ✓ Straightforward to implement either S/W or H/W
  - ✓ Computationally efficient over bank of filters

- LPC model works well in recognition applications

# The LPC Model

$$s(n) \approx a_1 s(n-1) + a_2 s(n-2) + \cdots + a_p s(n-p) \quad \text{—— (1)}$$

$$s(n) = \sum_{i=1}^{p} a_i s(n-i) + G u(n) \quad \text{—— (2)}$$

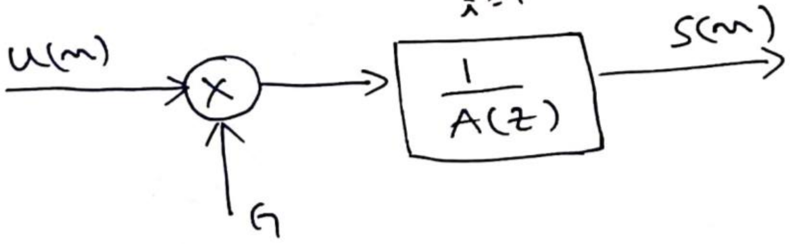
$u(n)$ —— Normalized Excitation

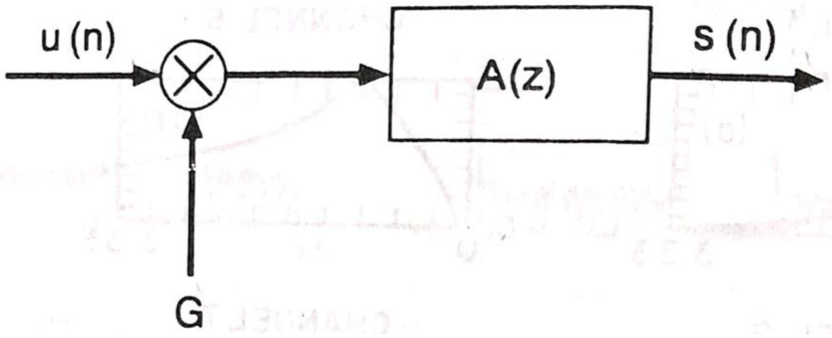
$G$ —— Gain of the Excitation

Applying z-transform on eq-(2)

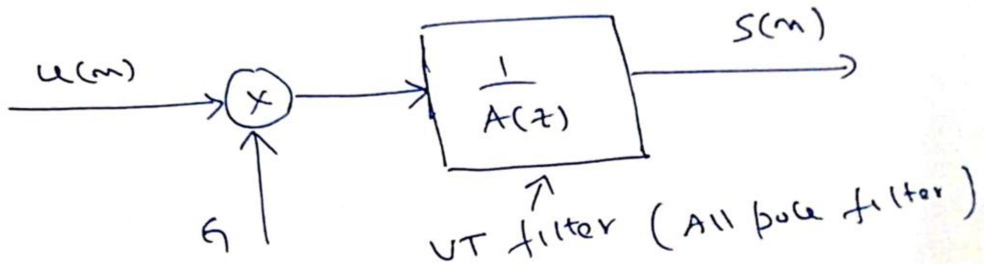
$$S(z) = \sum_{i=1}^{p} a_i z^{-i} S(z) + G U(z)$$

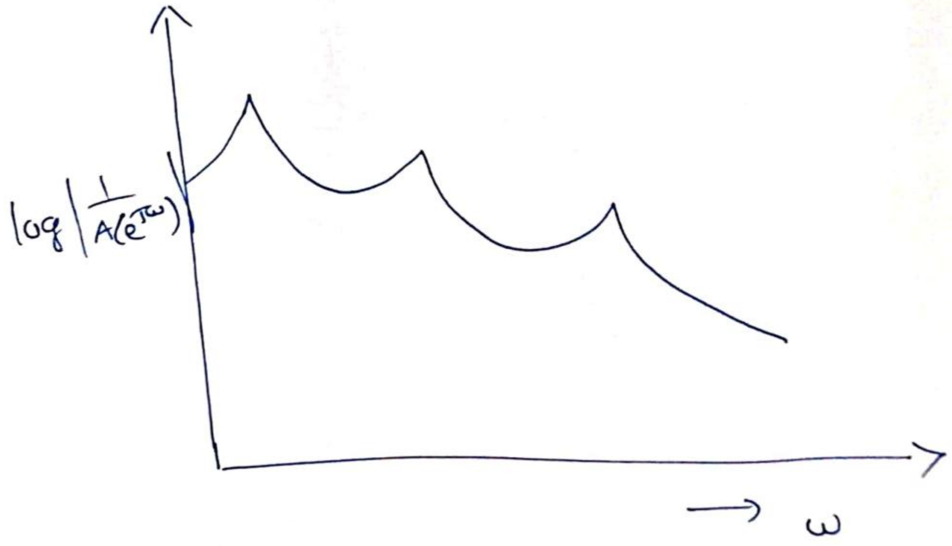
$$H(z) = \frac{S(z)}{G U(z)} = \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}} = \frac{1}{A(z)}$$

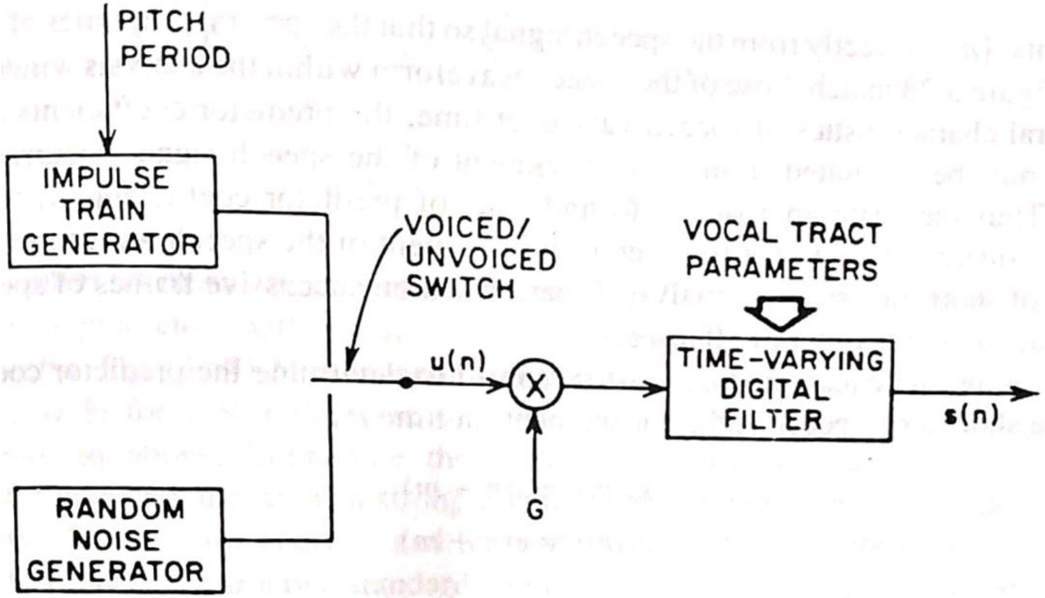


# Speech Synthesis Model Based on LPC Model



# Speech Analysis Model Based on LPC Model

<u>Speech Analysis Model</u>

$$\hat{s}(m) = \sum_{i=1}^{p} a_i \, s(m-i)$$

$\hat{s}(m)$ — predicted signal

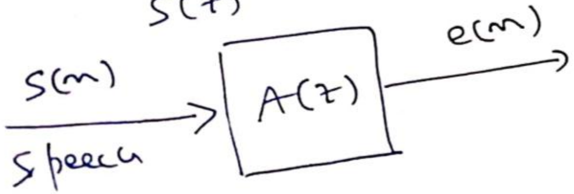
$s(m)$ — Actual Signal

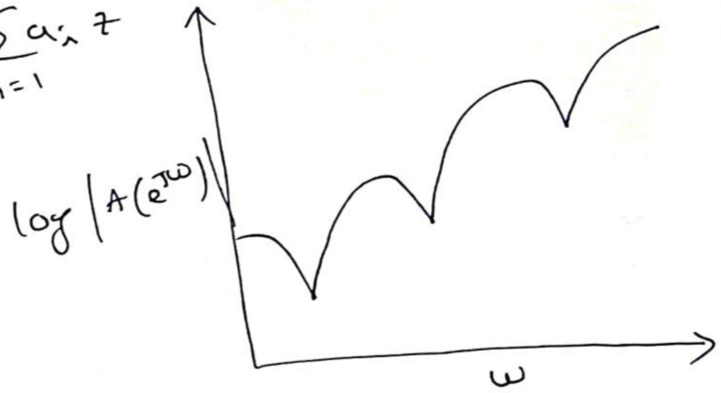
$e(m)$ — Error in prediction

$$e(m) = s(m) - \hat{s}(m) = s(m) - \sum_{i=1}^{p} a_i \, s(m-i)$$

$$E(z) = S(z)\left[1 - \sum_{i=1}^{p} a_i z^{-i}\right]$$

$$\frac{E(z)}{S(z)} = A(z) = 1 - \sum_{i=1}^{p} a_i z^{-i}$$




Inverse Filter formulation

# LP Analysis of Speech : Computation of LPCs

LP analysis of Speech (computation of LPCs)

$\hat{s}(n)$ — predicted speech signal $= \sum_{k=1}^{p} a_k s(n-k)$

$s(n)$ — Actual speech signal

$e(n)$ — Error in prediction of $n^{th}$ sample

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^{p} a_k s(n-k)$$

$E = \sum_{n=1}^{N} e^2(n)$ = Squared error for the frame of N samples

$$= \sum_{n=1}^{N} \left[ s(n) - \sum_{k=1}^{p} a_k s(n-k) \right]^2$$

For optimal values of $a_k$, $\frac{\partial E}{\partial a_k} = 0 \quad k = 1, 2, \ldots p$

$$\sum_{k=1}^{p} a_k \sum_{n} s(n-k)\, s(n-i) = -\sum_{n} s(n)\, s(n-i)$$

$$1 \le i \le p$$

# LP Analysis of Speech : Computation of LPCs

$$\sum_{k=1}^{p} a_k R(i-k) = R(i) \qquad i = 1, 2, \cdots p$$

$$\begin{bmatrix} R(0) & R(1) & R(2) & \cdots & R(p-1) \\ R(1) & R(0) & R(1) & \cdots & R(p-2) \\ \vdots & & & & \\ R(p-1) & R(p-2) & R(p-3) & \cdots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(p) \end{bmatrix}$$

Gauss Reduction Method
- $\frac{p^3}{3} + O(p^2)$ — operations
- $p^2$ — storage

Cholesky Decomposition
- $\frac{p^3}{6} + O(p^2)$
- $p^2/2$

Levinson & Robinson $p^2$ Recursion

Durbin Recursion
- $p^2 + O(p)$
- $2p$

# Error in LP Analysis of Speech

Total Error

$$E = \sum_m e^2(m) = \sum [s(m) - \hat{s}(m)]^2$$

$$E = \sum_m \left[s(m) - \sum a_k s(m-k)\right]^2$$

$$= \sum_m \left[s^2(m) + \left(\sum_{k=1}^{p} a_k s(m-k)\right)^2 - 2\sum_{k=1}^{p} a_k s(m-k)\, s(m)\right]$$

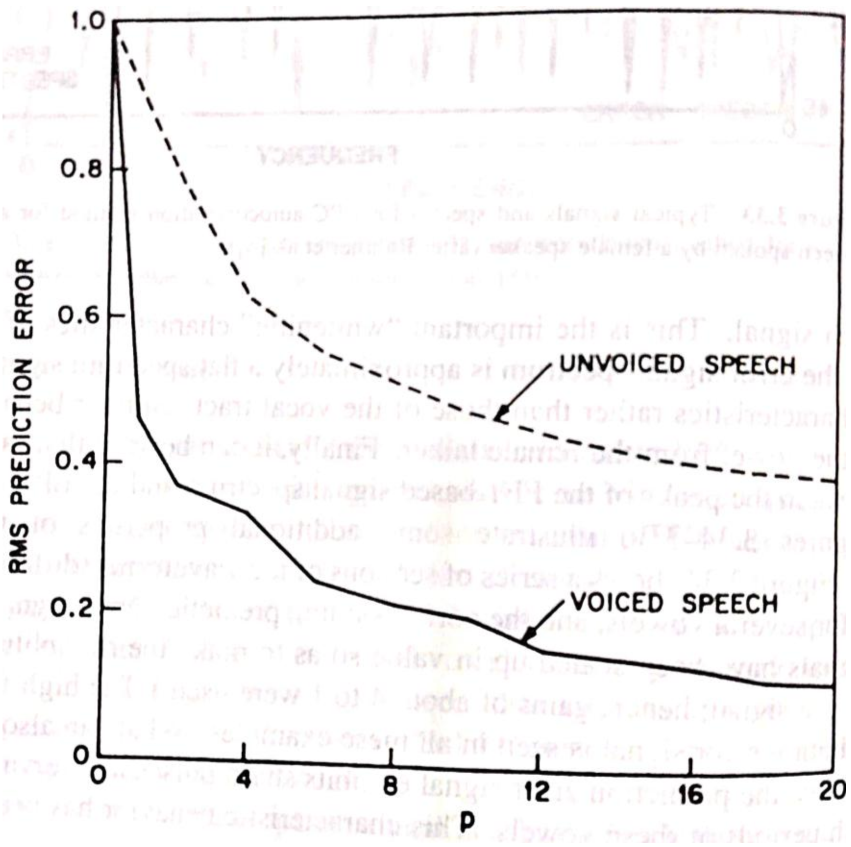
$$= \sum_m s^2(m) + \sum_m \sum_{k=1}^{p} \left(a_k s(m-k)\right)^2 - 2\sum_{k=1}^{p} a_k \sum_m s(m)\cdot s(m-k)$$

Substituting the result of p ordered equations

$$\sum_{k=1}^{p} a_k \sum_m s(m-k)\, s(m-i) = -\sum_m s(m)\, s(m-i) \qquad 1 \le i \le p$$

$$E = \sum_m s^2(m) + \sum_{k=1}^{p} a_k \sum s(m)\, s(m-k)$$

$$= R(0) - \sum_{k=1}^{p} a_k R(k)$$



# LP Analysis of Vowel ‘I’

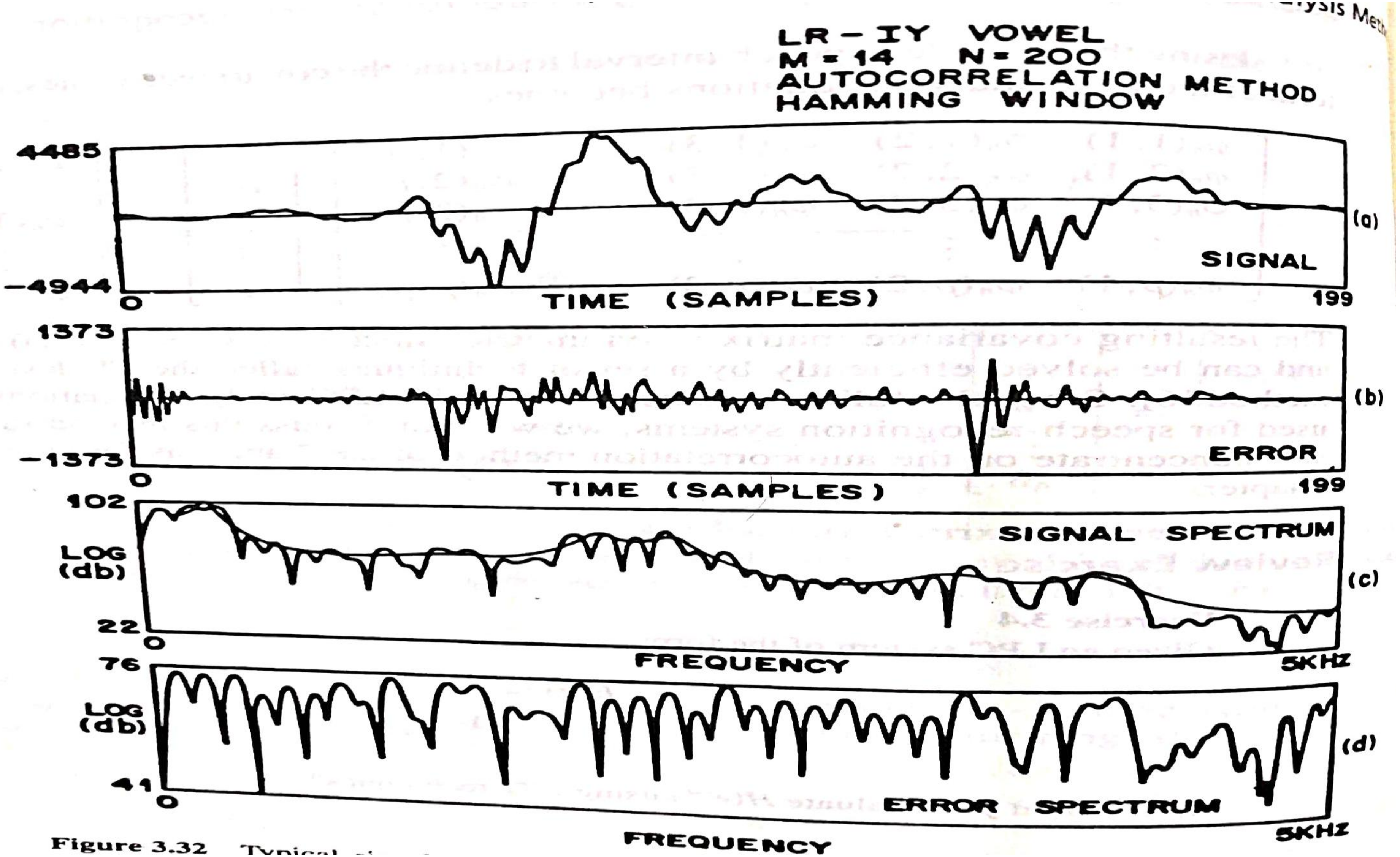
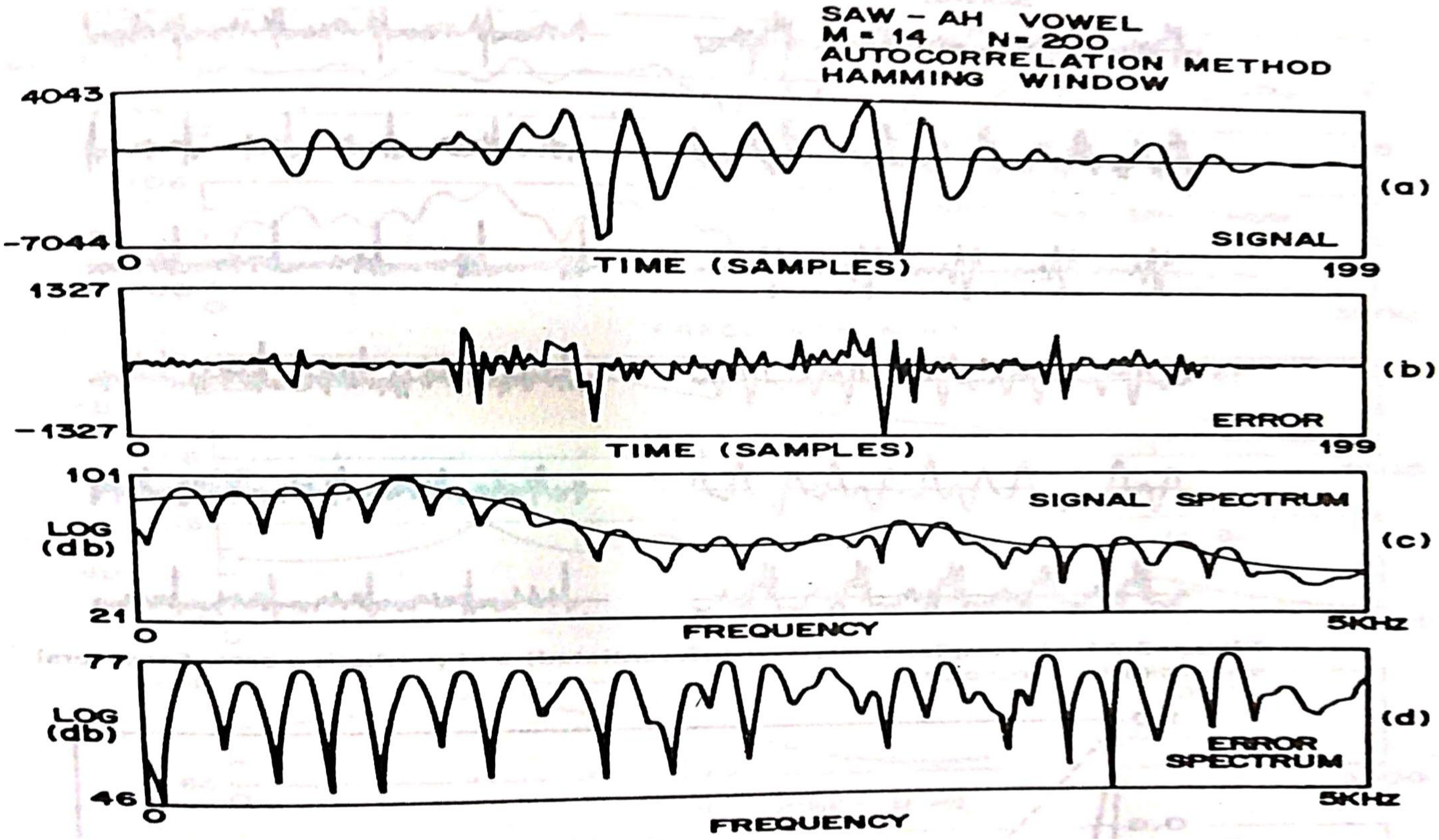
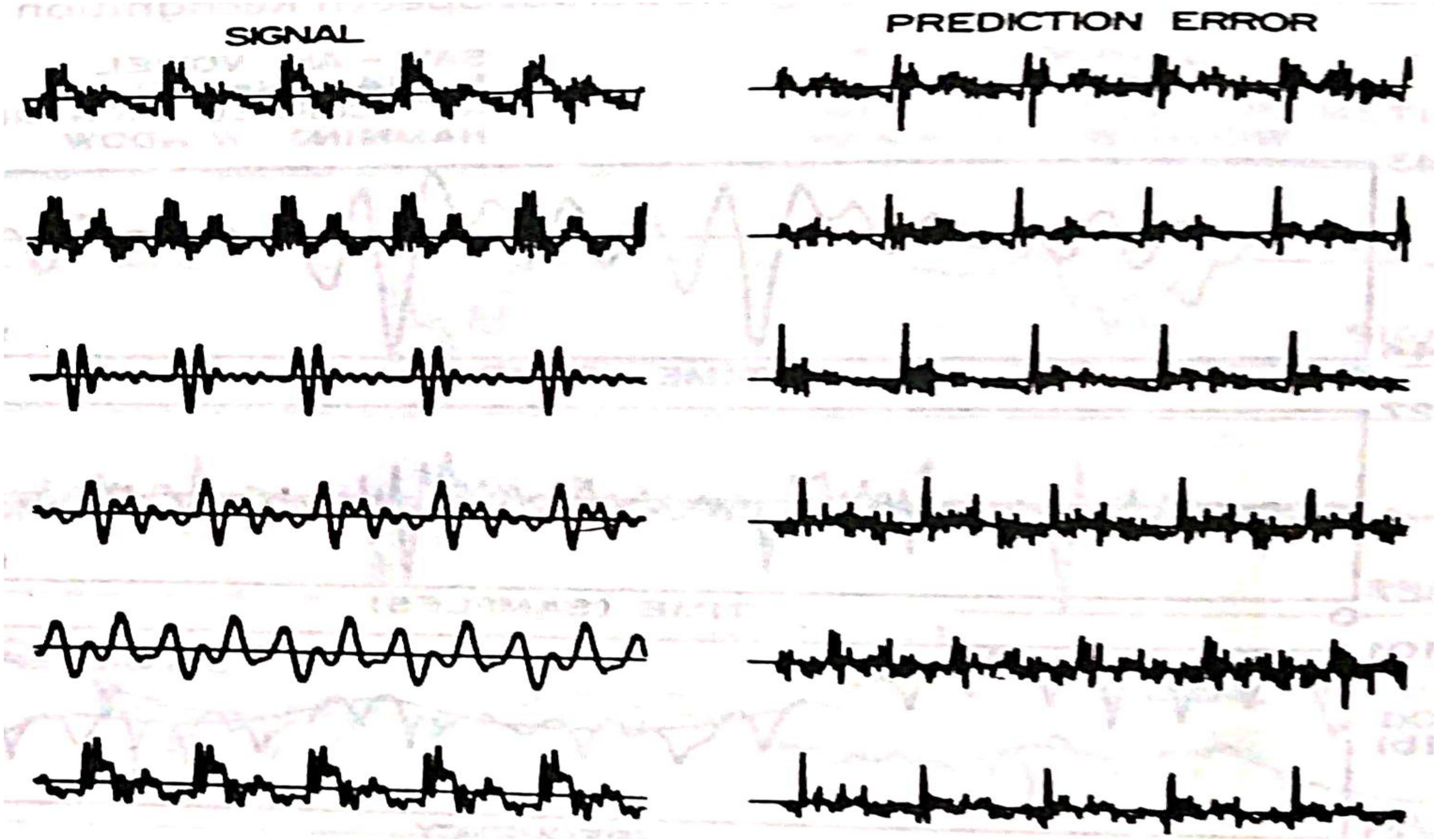



Figure 3.32 Typical...

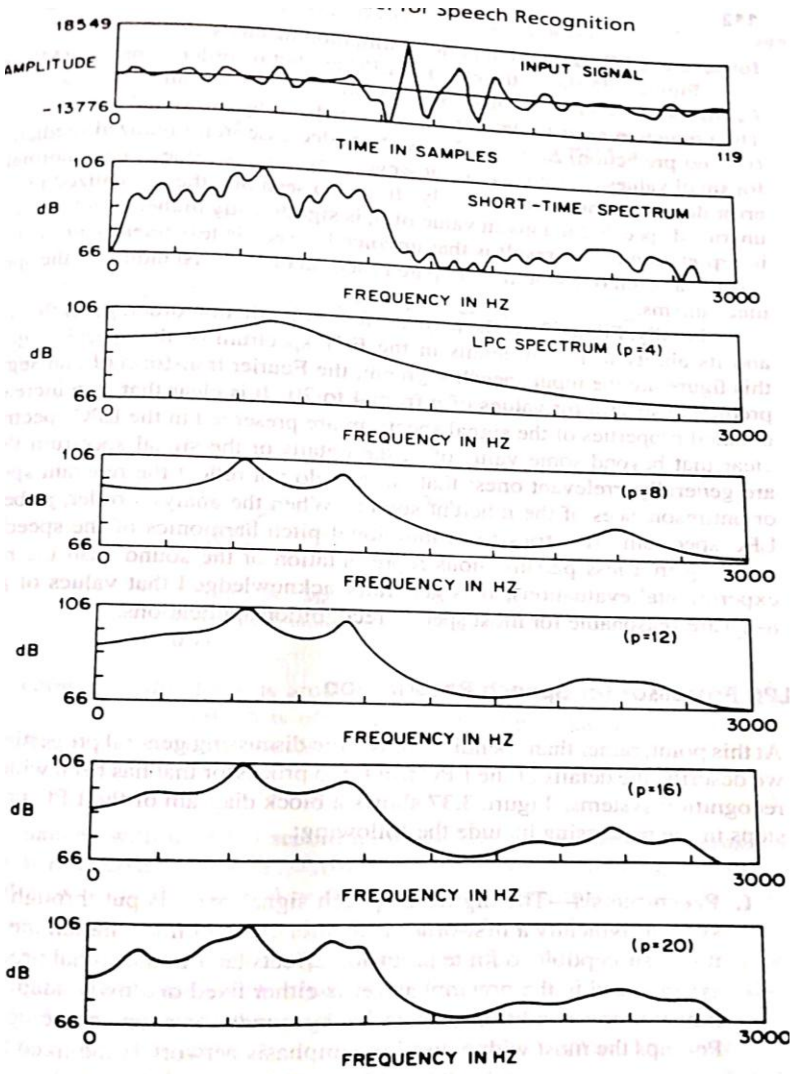
# LP Analysis of Vowel 'A'



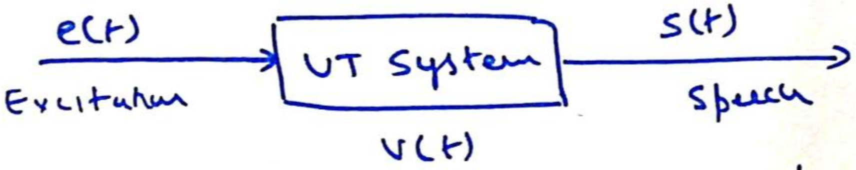
# LP Analysis of Vowels



# LP Analysis of a Vowel for different P values



# Cepstral Analysis of Speech




$$s(t) = e(t) * v(t); \qquad S(\omega) = |S(\omega)|\, e^{j\theta_s(\omega)}$$

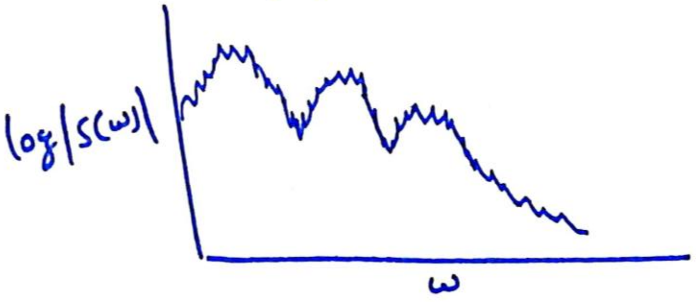
$$S(\omega) = E(\omega)\, V(\omega); \qquad E(\omega) = |E(\omega)|\, e^{j\theta_E(\omega)}$$

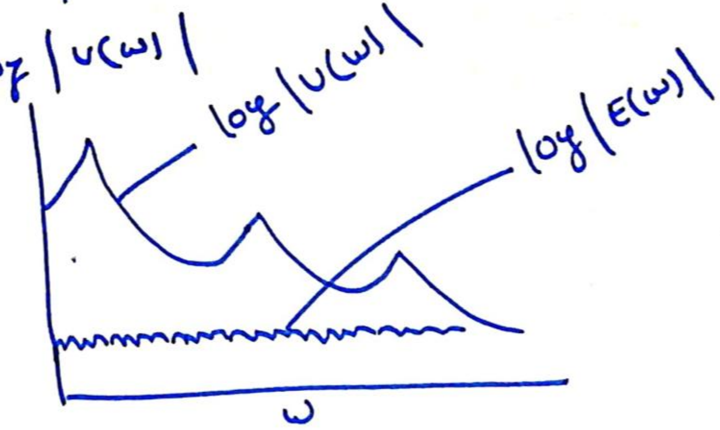
$$V(\omega) = |V(\omega)|\, e^{j\theta_V(\omega)}$$

$$\log S(\omega) = \log E(\omega) + \log V(\omega)$$

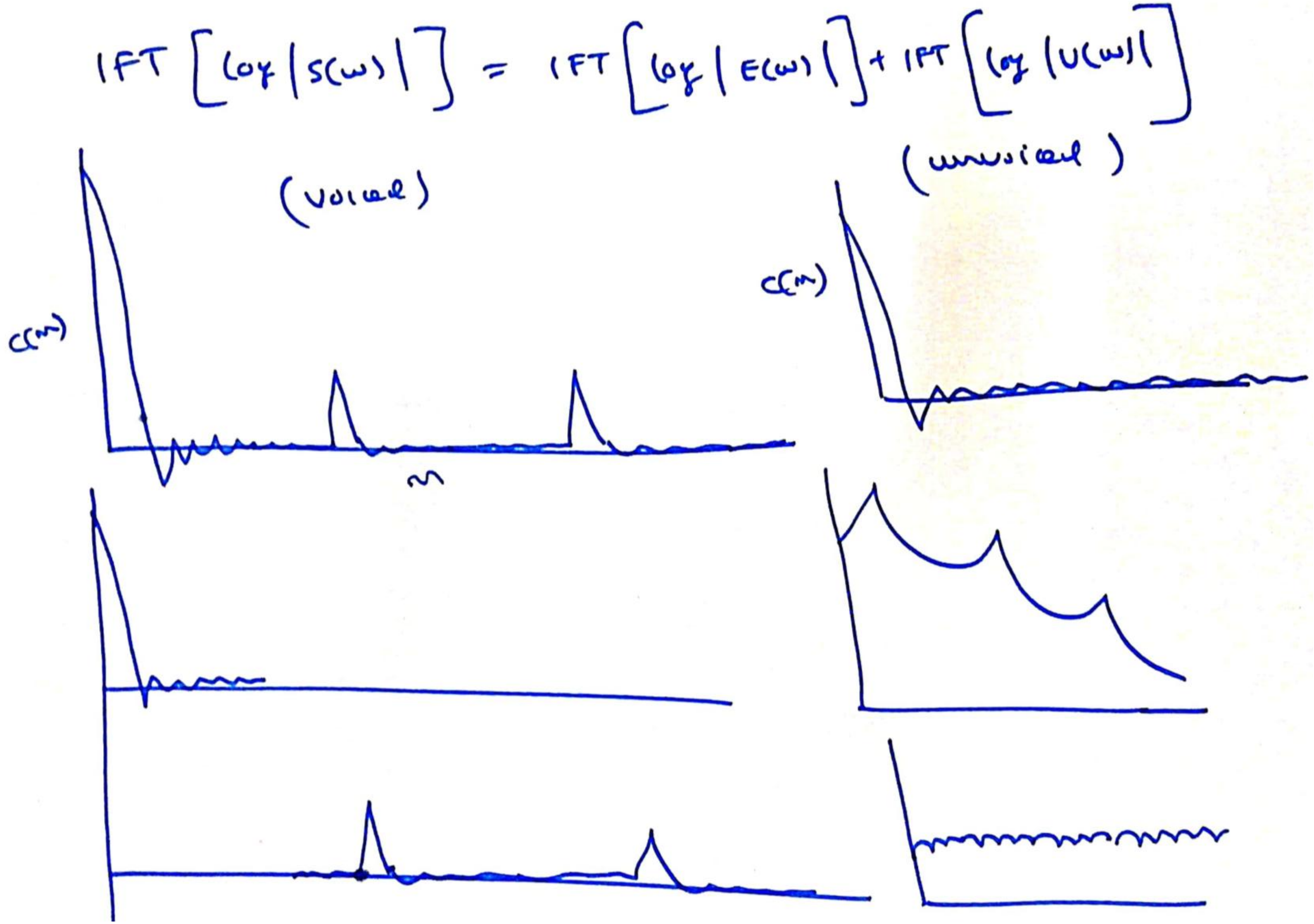
$$\log|S(\omega)| + j\theta_s(\omega) = \log|E(\omega)| + j\theta_e(\omega) + \log|V(\omega)| + j\theta_V(\omega)$$

$$\log|S(\omega)| = \log|E(\omega)| + \log|V(\omega)|$$



# Cepstral Analysis of Speech



# Extraction of Cepstral Features from LP Analysis

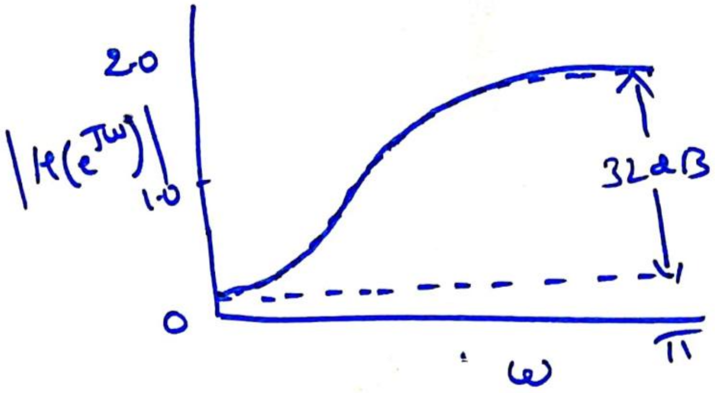
Features from LPC analysis of Speech

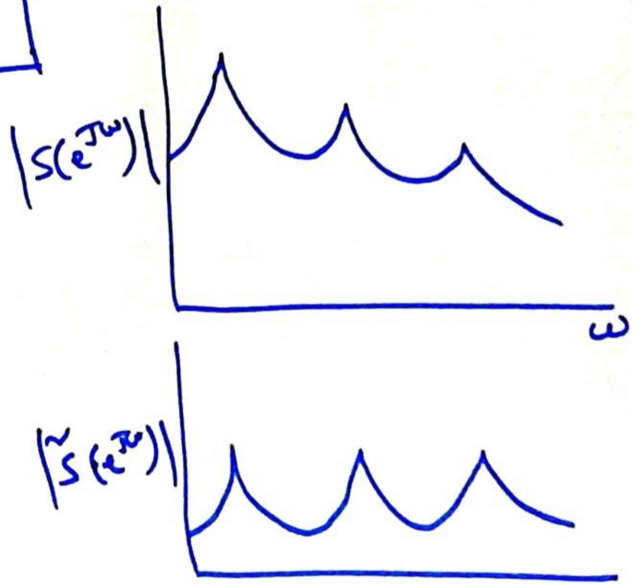
1. Preemphasis (First order FIR filter)

$$H(z) = 1 - a z^{-1} \qquad 0.9 \le a \le 1.0$$

$$\tilde{s}(n) = s(n) - a\,s(n-1) \ ; \quad a = 0.95$$

$$s(n) \longrightarrow \boxed{\text{Preemphasis filter } H(z)} \longrightarrow \tilde{s}(n)$$



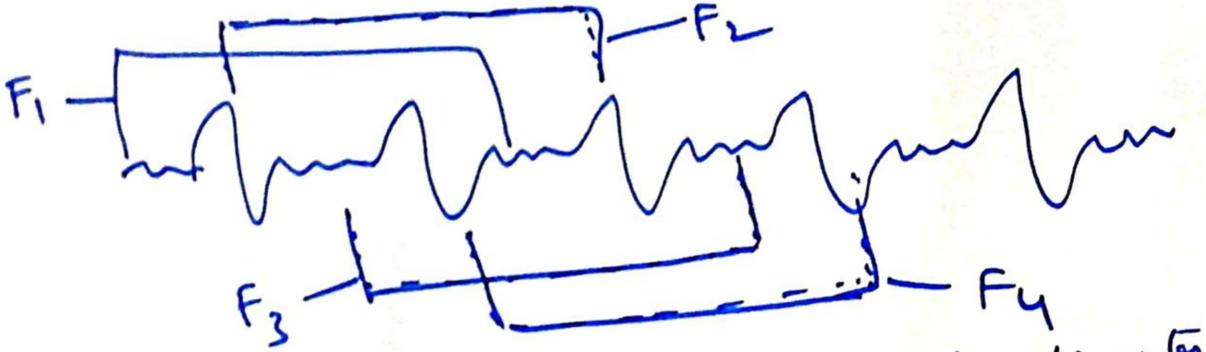



# Extraction of Cepstral Features from LP Analysis

2. Frame Blocking

Frame Size = N samples

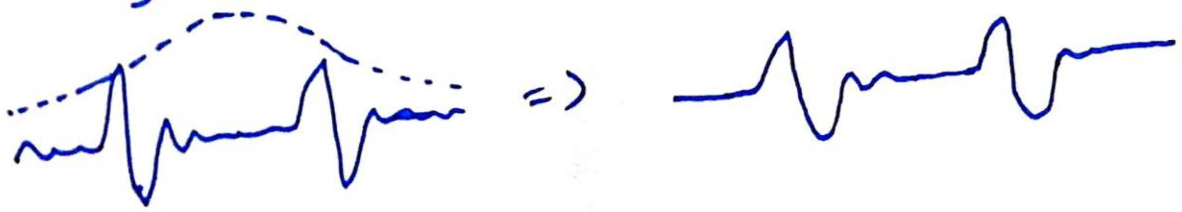
Adjacent frame separation = M samples $\approx \frac{1}{3} N$




Correlation between LPC spectral estimates

3. Windowing: Avoids signal discontinuities at begin & end

Hamming window $w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right)$



# Extraction of Cepstral Features from LP Analysis

4. Auto correlation Analysis

$$R(k) = \sum_{m=0}^{N-1-k} s(m)\, s(m+k) \qquad k = 0, 1, \cdots p$$

$$p = 8 \text{ to } 16$$

5. LPC Analysis (Durbin's Recursion)

LPC coefficients

Reflection coefficients

Log area ratio coefficient

Cepstral coefficients

6. LPC $\longrightarrow$ LPCCs

$\#$ LPCCs $\approx \frac{3}{2}(p)$

# Extraction of Cepstral Features from LP Analysis

7. Parameter Weighting (WLPCCs)

Low-order cepstral coeffs —— Spectral slope

High-order cepstral coeffs —— Noise

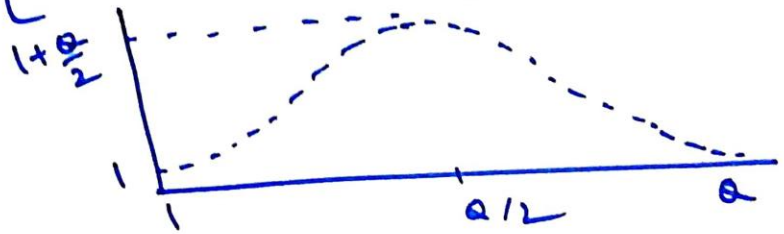
Weighting the cepstral coeffs with tapered window

Differential log-mag spectrum

- fixed spectral slope → constant
- prominent spectral peaks are preserved

Band pass filter in the cepstral domain

$$w_m = \left[1 + \frac{Q}{2}\sin\left(\frac{\pi m}{Q}\right)\right] \qquad 1 \le m \le Q$$



# Extraction of Cepstral Features from LP Analysis

8. Temporal Cepstral Derivative

Cepstral Representation — Local spectral prop..

$C_i = [c_1 \ c_2 \ c_3 \ \cdots \ c_Q]$ ; cepstral coeffs of $i^{th}$ frame

$$\Delta c_i(m) = \sum_{k=-N}^{N} k c_i(m+k) \Big/ \sum_{k=-N}^{N} k^2 \ ; \quad N = 3$$

$\{c_i\}$ — $Q$

$\{c_i, \Delta c_i\}$ — $2Q$

$\{c_i \ \Delta c_i \ \Delta\Delta c_i\}$ — $3Q$

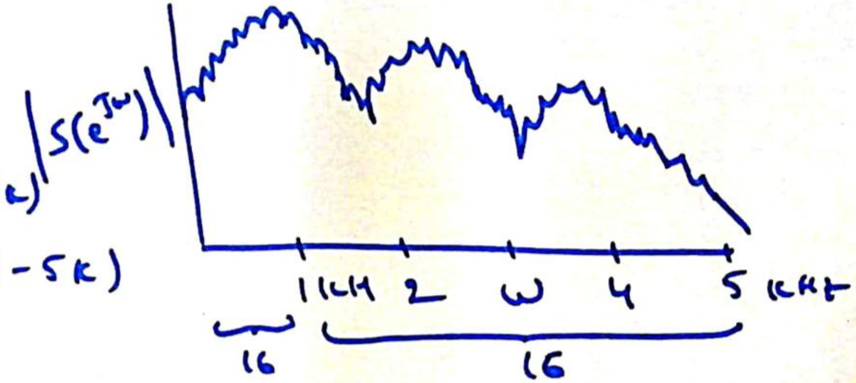
# Extraction of Mel-Frequency Cepstral Coefficients

## Spectral coefficients

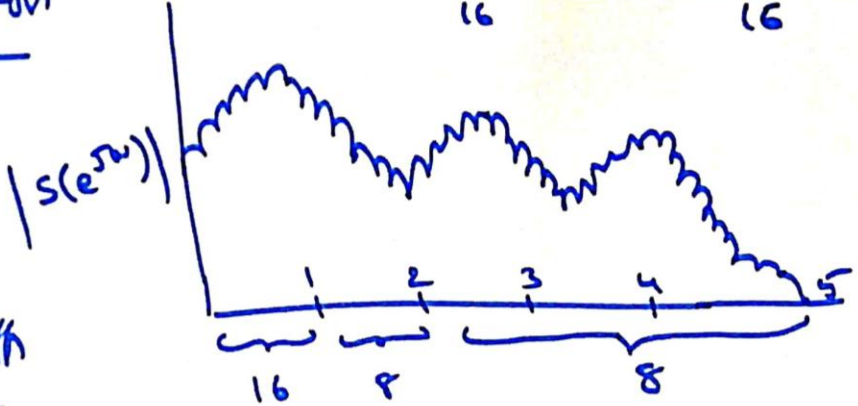
Speech frame $s(n)\big|_{n=0 \text{ to } N-1}$

$S(\omega) = \text{FFT}(s(n))$

Spectral coeffs $\left\langle \begin{array}{l} 0\text{–}15\ (0\text{–}1k) \\ 16\text{–}31\ (1\text{–}5k) \end{array} \right.$




## Mel-freq spectral coeffs




## Mel-freq cepstral coeffs

MFCCs $\quad$ IFT [Mel-spectral coeffs]

# Linear Prediction Analysis of Speech
# (Hands-on)