Linear Prediction of Speech Wave

Estimation of LPC's from Speech

- **Given speech $s(n)$ is processed in blocks of 20 msec with a shift of 10 msec.**
- **Each block of 20 msec of speech $s'(n)$ is obtained by windowing $s(n)$.**

$$s'(n) = s(n).w(n) \qquad 0 \leq n \leq N-1 \tag{1}$$

- **For a given block of speech $s'(n)$, it's auto correlation sequence is computed using**

$$R(i) = \sum_{n=0}^{N-1-i} s'(n)s'(n+i) \qquad i \geq 0 \tag{2}$$

- **Computed values of $R(i)$ are used in the Levinson-Durbin's (L-D) recursive procedure for computing LPC's**
- **If p is the order of prediction, then we have p normal equations and we need $p+1$ auto correlation values i.e. $R(0), R(1), \cdots, R(p)$.**
- **At the end, the LPC's will be the output of L-D recursive procedure.**
- **This procedure is repeated for all blocks.**

Illustration of LPC's Computation

- **The order of prediction of LP coder is $p = 10$ i.e. we need 11 auto correlation values from $R(0)$ to $R(10)$.**

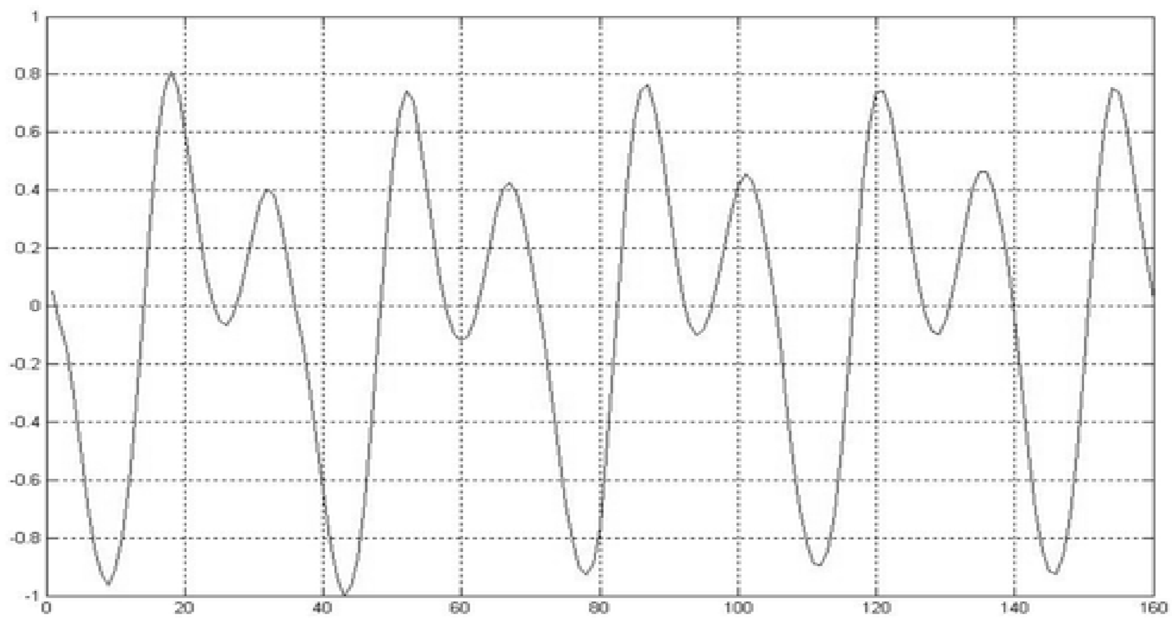


Figure 1: 20 msec windowed voiced speech segment.

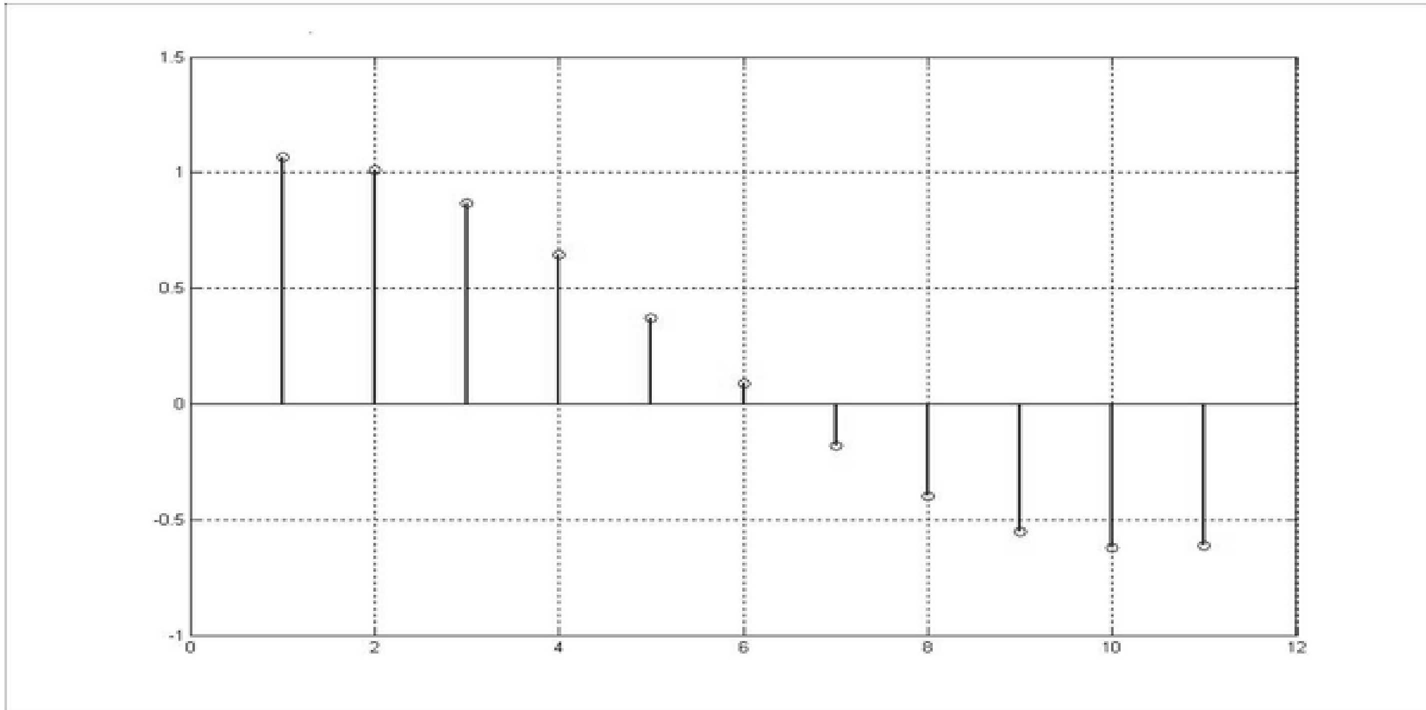


Figure 2: Auto correlation function of the widowed voiced segment for the LP order 10.

- **Length of the hamming window is 20 msec**
- **Order of linear prediction is 10**
- **Auto Correlation Values:**

  $R(k)$=**[1.0642, 1.0127, 0.8655, 0.6420, 0.3709, 0.0854,**

  -0.1813, -0.4003, -0.5507, -0.6218, -0.6136].

  E_0=$R(0)$=**1.0642.**

1. **For i=1**

   k_1=**-0.9516**

   a_1^1=**-0.9516**

   E_1=**-0.1004**

2. **For i=2**

k_2=**0.9783**

a_2^2=**0.9783** a_1^2=**-1.8826**

E_2=**0.00431**

3. **For i=3**

k_3=**-0.7577**

a_3^3=**-0.7577** a_1^3=**-2.6238** a_2^3=**2.4047**

E_3=**0.0018389**

4. **For i=4**

k_4=**-0.1856**

a_4^4=**-0.1856** a_1^4=**-2.4832** a_2^4=**1.9584**

a_3^4=**-0.2707**

E_4=**0.0018**

5. <u>For i=5</u>

k_5=**0.2866**

a_5^5=**0.2866** a_1^5=**-2.5364** a_2^5=**1.8808**

a_3^5=**0.2907** a_4^5=**-0.8974**

E_5=**0.0016297**

6. <u>For i=6</u>

k_6=**0.0742**

a_6^6=**0.0742** a_1^6=**-2.5151** a_2^6=**1.8142**

a_3^6=**0.3122** a_4^6=**-0.7579** a_5^6=**0.0985**

E_6=**0.0016207**

7. **For i=7**

k_7=**-0.0652**

a_7^7=**-0.0652** a_1^7= **-2.520** a_2^7=**1.8078**

a_3^7=**0.3616** a_4^7=**-0.7783** a_5^7=**-0.0197**

a_6^7=**0.2381**

E_7=**0.0016138**

8. **For i=8**

k_8=**0.032**

a_8^8=**0.032** a_1^8=**-2.522** a_2^8=**1.8154**

a_3^8=**0.3610** a_4^8=**-0.8032** a_5^8=**-0.0081**

a_6^8=**0.2959** a_7^8=**-0.1458**

E_8=**0.0016122**

9. <u>For i=9</u>

k_9=**0.1964**

a_9^9=**0.1964** a_1^9=**-2.5158** a_2^9=**1.7868**

a_3^9=**0.419** a_4^9=**-0.8048** a_5^9=**-0.1659**

a_6^9=**0.3669** a_7^9=**0.2107** a_8^9=**-0.4634**

E_9=**0.00155**

10. **For i=10**

k_{10}=**0.1127**

a_{10}^{10}=**0.1127** a_1^{10}=**-2.4936** a_2^{10}=**1.7346**

a_3^{10}=**0.4429** a_4^{10}=**-0.7634** a_5^{10}=**-0.1846**

a_6^{10}=**0.2762** a_7^{10}=**0.2580** a_8^{10}=**-0.2620**

a_9^{10}=**-0.0871**

E_{10}=**0.00153**

LP Coefficients

- **By L-D Recursive Approach:**

  a_k=[-2.4936, 1.7346, 0.4429, -0.7634, -0.1846, 0.2762, 0.2580, -0.2620, -0.0871, 0.1127].

- **By LPC Command:**

  a_k=[1, -2.4936, 1.7346, 0.4429, -0.7634, -0.1846, 0.2762, 0.2580, -0.2620, -0.0871, 0.1127].

  - a_0 is always 1 which is the multiplication factor for the current sample $s(n)$.

LPC's: Signal Processing Perspective

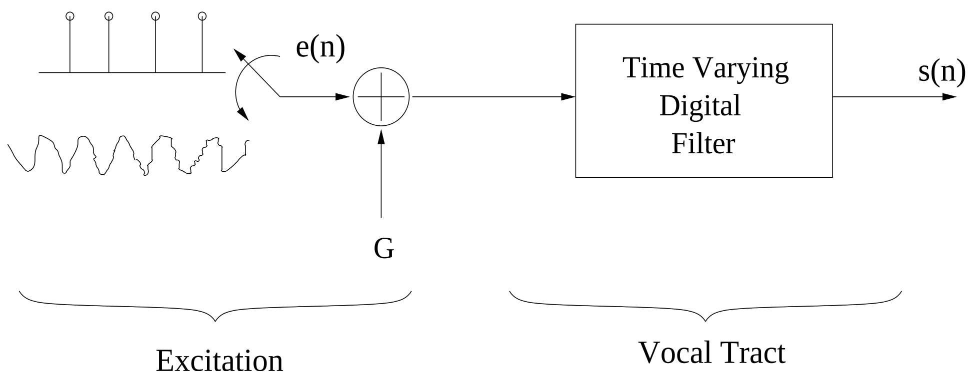


Figure 3: Digital speech production model.

$$s(n) = -\sum_{k=1}^{p} a_k . s(n-k) + G.e(n). \tag{3}$$

$$H(z) = \frac{G}{1 + \sum_{k=1}^{p} a_k z^{-k}} \tag{4}$$

- **As shown, $H(z)$ is assumed to be an all pole model.**

- **a_k's (LPC's) are the coefficients of denominator polynomial.**
- **Since there are no zeros, a_k's completely describe vocal tract response.**
- **For speech production the all pole filter represented by $H(z)$ should be a stable filter.**
  - **Roots of the denominator polynomial (poles) should lie inside unit circle.**

a_k=[1, -2.4936, 1.7346, 0.4429, -0.7634, -0.1846, 0.2762, 0.2580, -0.2620, -0.0871, 0.1127].

- Roots p of the Denominator polynomial are given by

p1=-0.6245+j0.1489	p2=-0.6245-j0.1489
p3=-0.3915+j0.5893	p4=-0.3915-j0.5893
p5=0.4577+j0.6779	p6=0.4577-j0.6779
p7=0.8933+j0.3701	p8=0.8933-j0.3701
p9=0.9188+j0.2045	p10=0.9188-j0.2045

- 10 poles and all are complex conjugate poles.

  →Practically need not be.

  →Some poles can be real also.

Check the positioning of roots p:

- **z=[];**
- **p=roots(a)**
- **zplane (z,p)**

z-Plane Representation

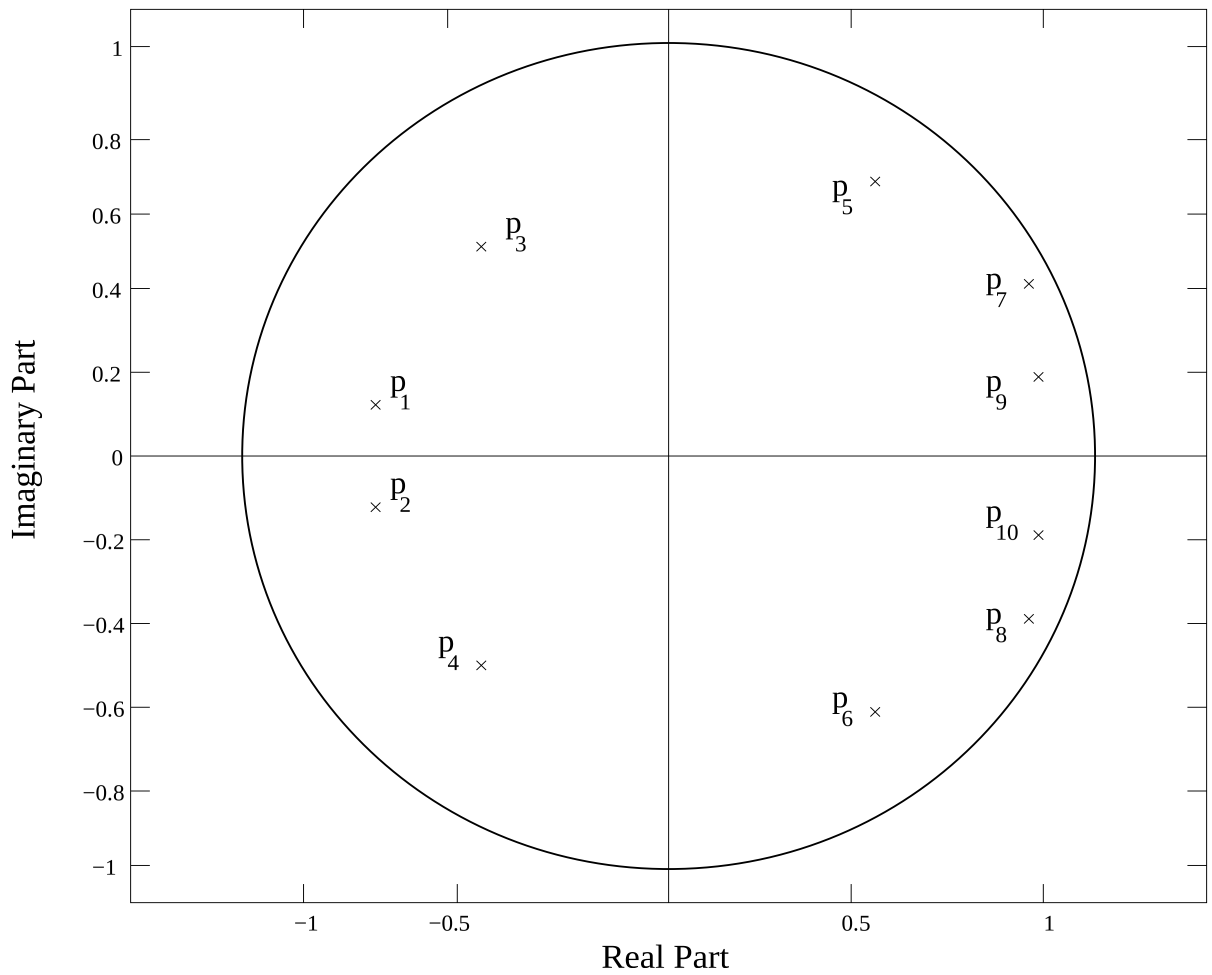


Figure 4: z-plane representation of all pole filter.

- **Since there are no zeros in $H(z)$, a_k's i.e. coefficients of denominator polynomial completely characterize the vocal tract system.**

- **Therefore for all practical tasks involving storage, telecommunication and pattern recognition, it is sufficient if we have a_k's to deal with vocal tract.**

Frequency Response (H(e^{jw}))

- **Pole-zero filter**

$$H(z) = \frac{S(z)}{E(z)} = \frac{1 + \sum_{k=1}^{q} b_k z^{-k}}{1 + \sum_{k=1}^{p} a_k z^{-k}} \tag{5}$$

- **All-pole filter**

$$H(z) = \frac{1}{1 + \sum_{k=1}^{p} a_k z^{-k}} \tag{6}$$

$\underline{z = e^{jw}}$

$$H(e^{jw}) = \frac{1}{1 + \sum_{k=1}^{p} a_k e^{-jwk}} \tag{7}$$

Frequency Response of LP filter

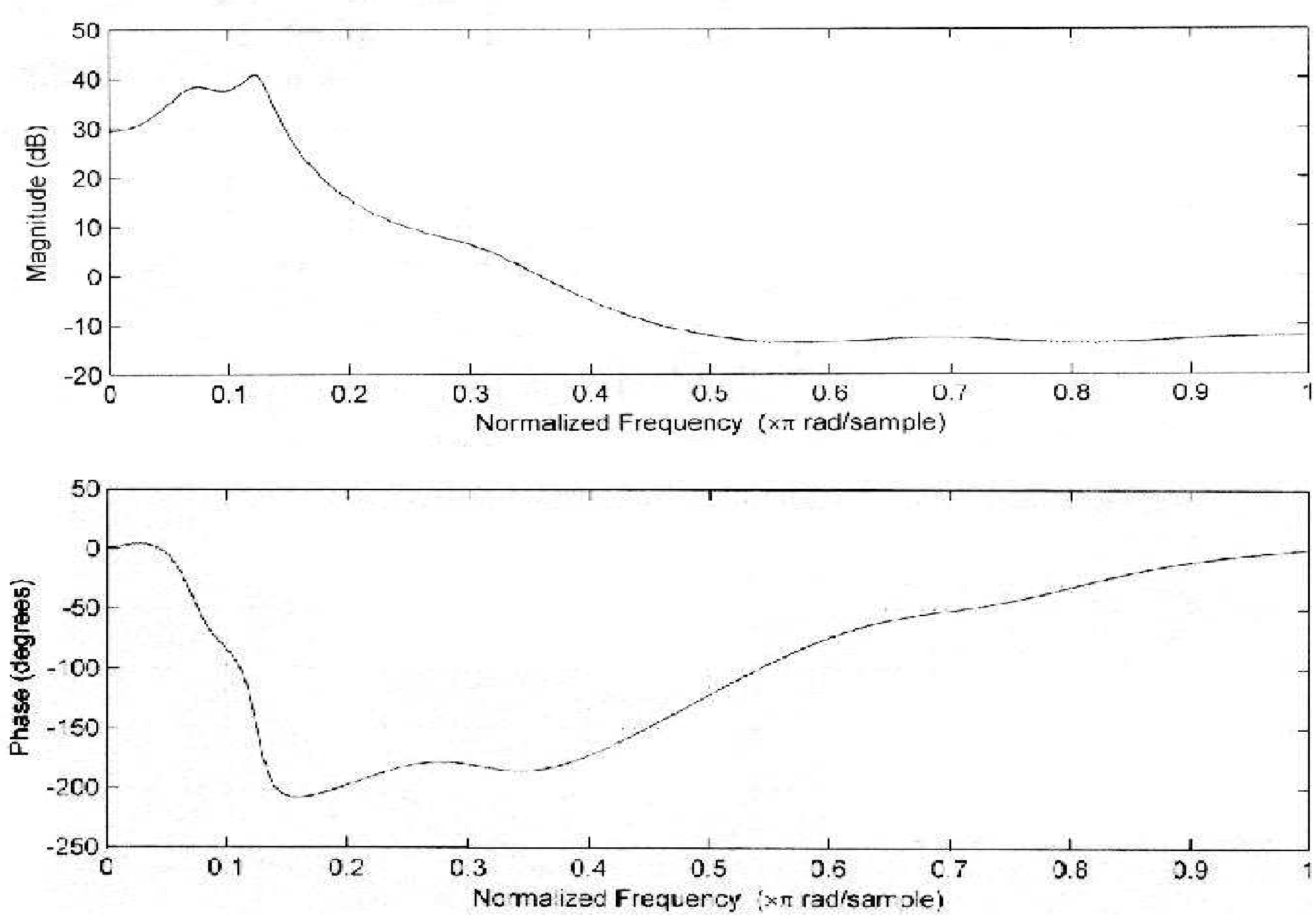
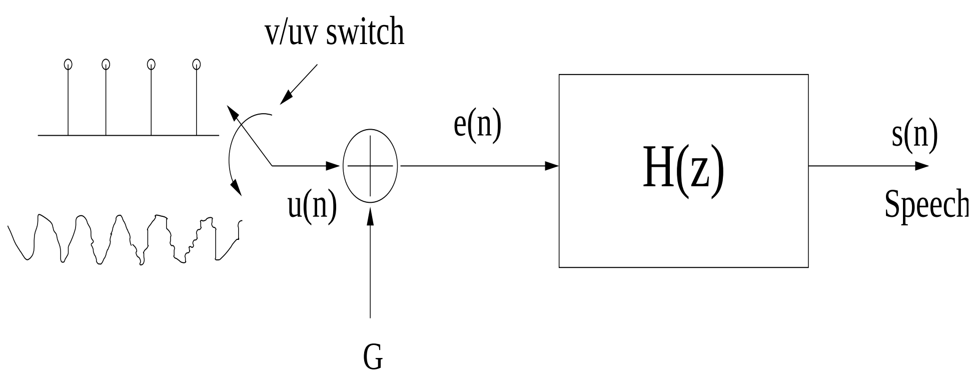


Figure 5: Frequency response of LP filter, (a) magnitude response, (b) phase response.

Excitation Information from LP Analysis

Digital Speech Production Model:



- **By all-pole modeling of speech, we have**

$$s(n) = -\sum_{k=1}^{p} a_k s(n-k) + Gu(n) \tag{8}$$

- **From the method of least squares for estimating LPC's, we have,**

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

$$s(n) = -\sum_{k=1}^{p} a_k s(n-k) + e(n) \tag{9}$$

- **By comparing equations (8)&(9), only input signal $u(n)$ that will result in the signal $s(n)$ as output is that where**

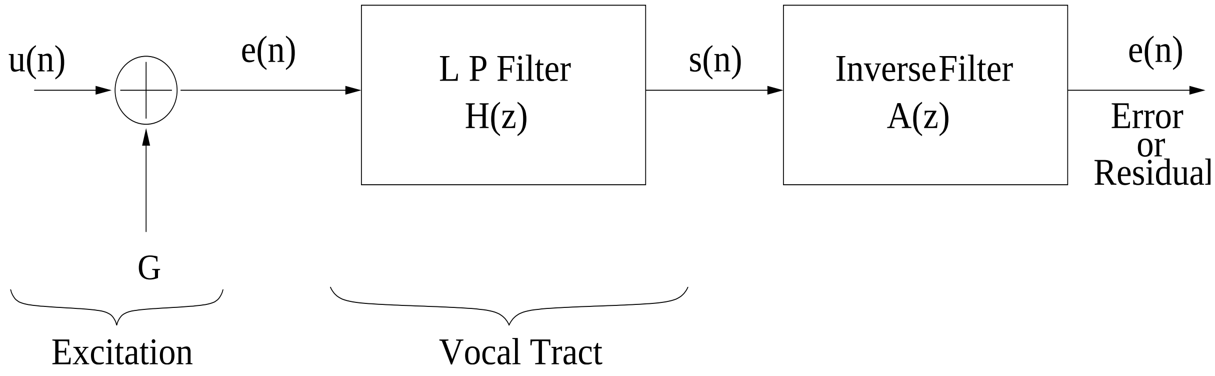
$$Gu(n) = e(n) \tag{10}$$

$\rightarrow$**The input signal is proportional to the error signal.**

$\rightarrow$**Total energy in input signal $Gu(n)$ must be equal to the total energy in error signal (E_p).**

Computation of Excitation Signal by Inverse Filter Formulation

- Excitation signal is proportional to the error between the actual signal and the predicted signal.
- Hence it is also termed as Linear Prediction (LP) error signal.
- It is more commonly termed as residual signal may be due to the fact that it is fraction or residual left without prediction.



$$H(z) = \frac{S(z)}{E(z)} = \frac{1}{1+\sum_{k=1}^{p} a_k z^{-k}} (or) \frac{G}{1+\sum_{k=1}^{p} a_k z^{-k}} = \frac{S(z)}{V(z)}.$$

Since, $A(z) = \frac{1}{H(z)}$

$$A(z) = 1 + \sum_{k=1}^{p} a_k z^{-k}$$

Frequency Response of Inverse Filter

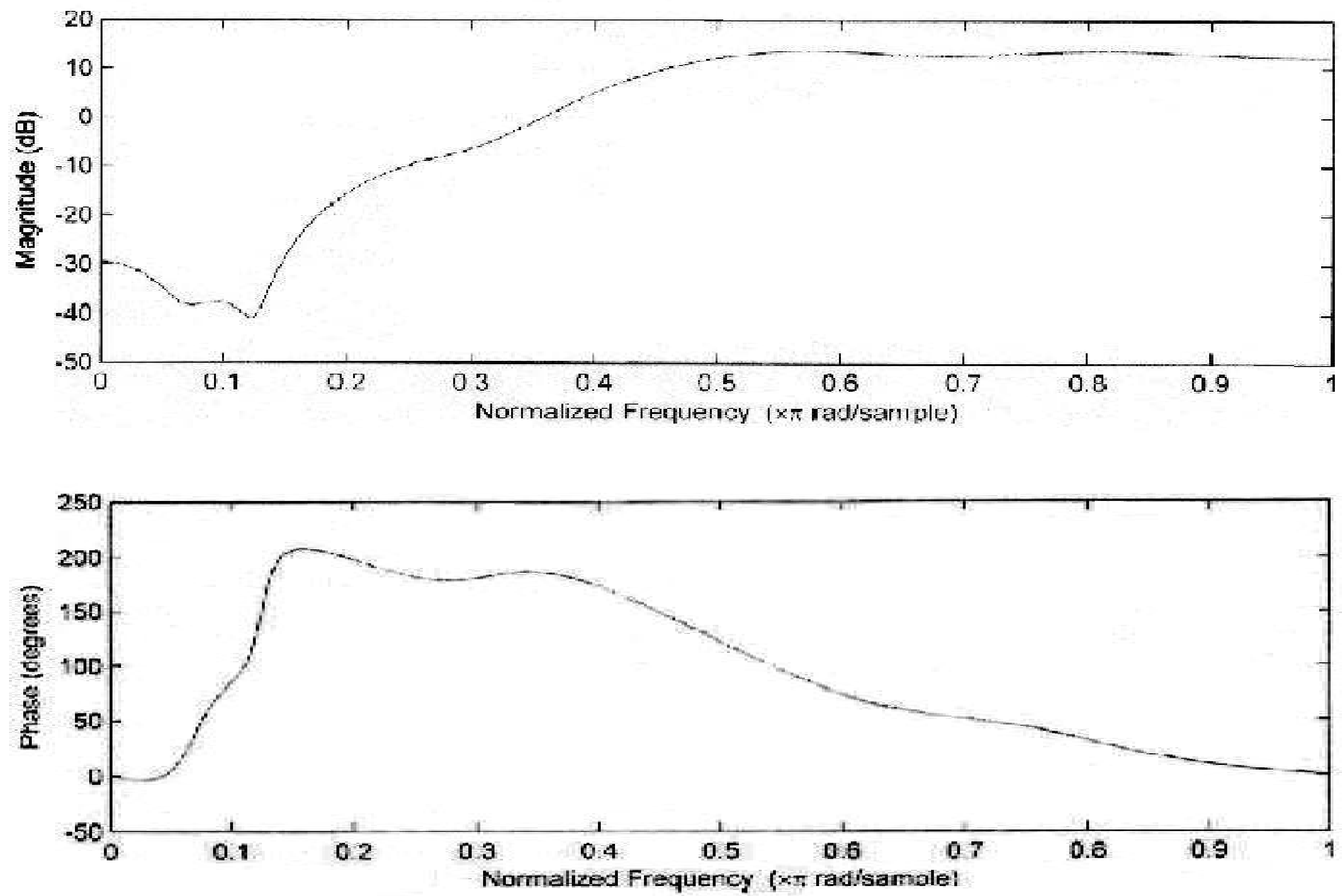


Figure 6: Frequency response of inverse filter, (a) magnitude response, (b) phase response.

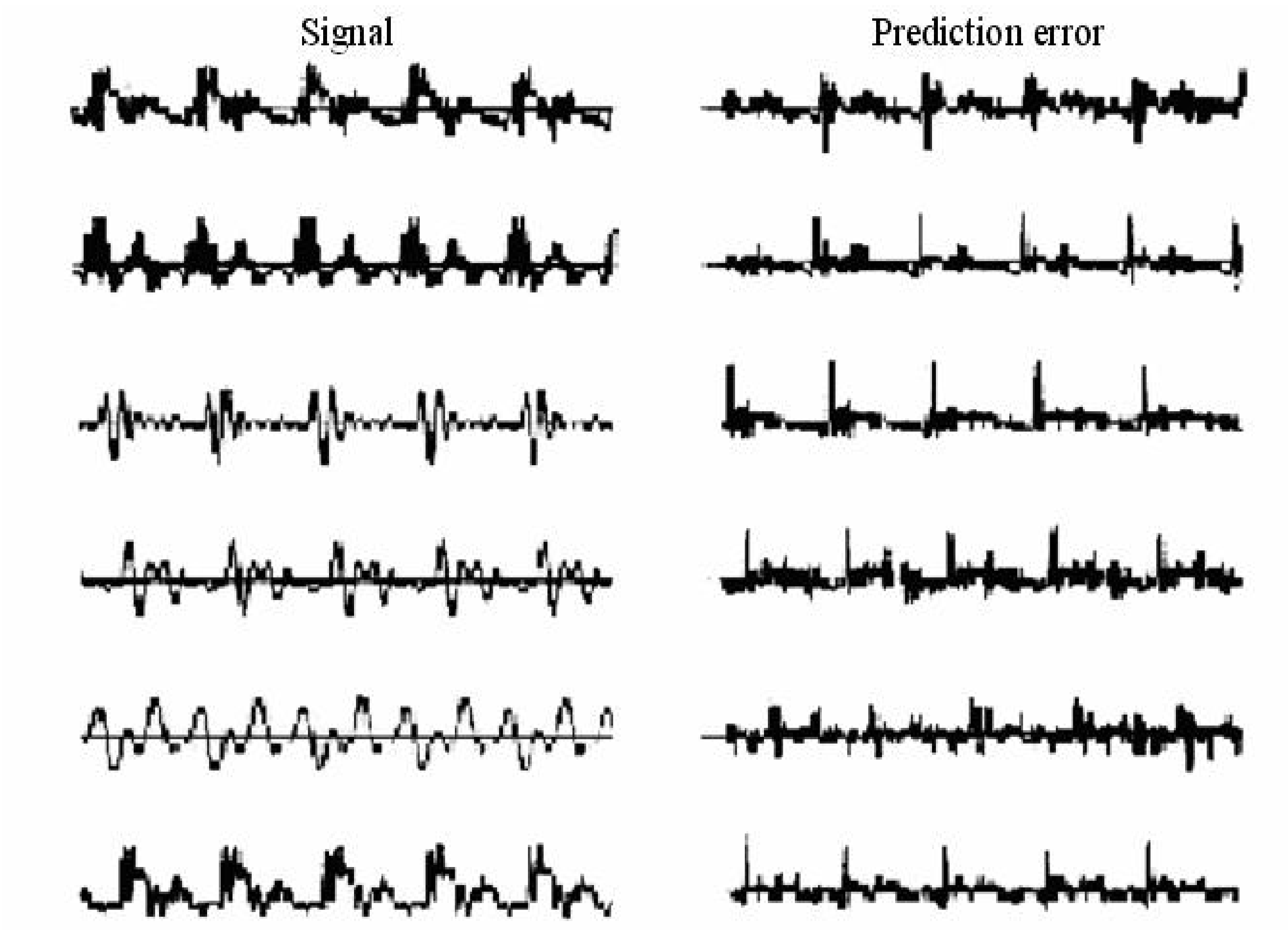


Figure 7: Examples of signal (differentiated) and prediction error for vowels (*i, e, a, o, u, y*). *(After Strube)*.

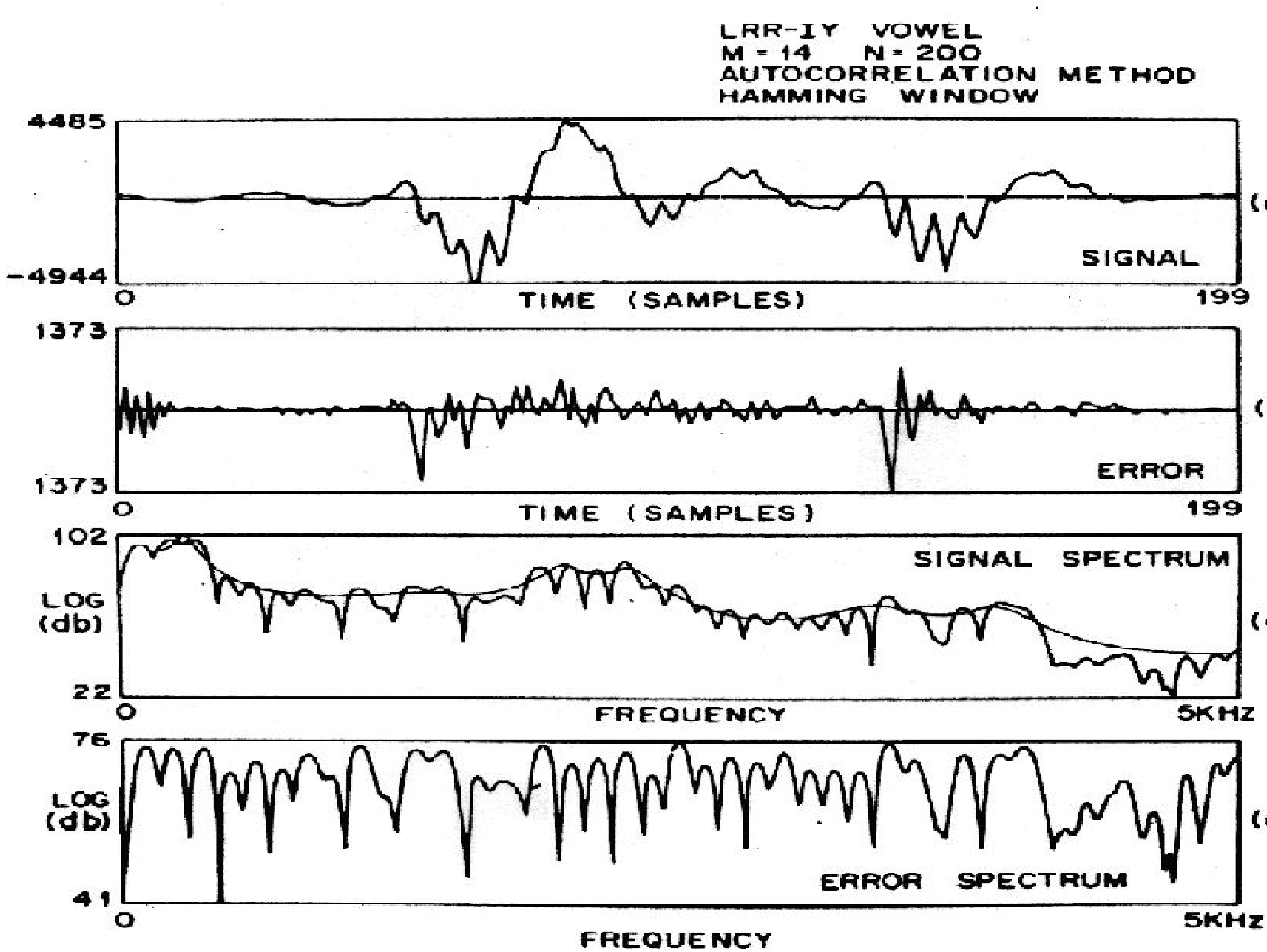


Figure 8: Typical signals and spectra for LPC autocorrelation method for a male speaker. (After Rabiner et al.).

1. Total Minimum Error (E_p):

$$E = \sum_n e^2(n) = \sum_n (s(n) + \sum_{k=1}^{p} a_k s(n-k))^2$$

$$E = \sum_n (s^2(n) + (\sum_{k=1}^{p} a_k s(n-k))^2 + 2s(n) \sum_{k=1}^{p} a_k s(n-k))$$

$$E = \sum_n s^2(n) + \sum_n (\sum_{k=1}^{p} a_k s(n-k))^2 + 2 \sum_{k=1}^{p} a_k \sum_n s(n) s(n-k)$$

Substituting from p-normal equations

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

$$\sum_n \left(\sum_{k=1}^{p} a_k s(n-k)\right)^2 = \sum_n \sum_{k=1}^{p} a_k s(n-k) \sum_{l=1}^{p} a_l s(n-i)$$

$$= \sum_{l=1}^{p} a_l \sum_{k=1}^{p} a_k \sum_n s(n-k)s(n-l)$$

$$= \sum_{l=1}^{p} a_l - \sum_n s(n)s(n-k) \qquad a \leq l \leq p$$

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

$$E_p = \sum_n s^2(n) + \sum_{k=1}^{p} a_k \sum_n s(n)s(n-k)$$

Using Autocorrelation definition

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

$$0 \le E_i \le E_{i-1}, E_0 = R(0) \qquad 1 \le i \le p.$$

- **Total minimum error decreases as the order increases.**

2. <u>Normalized Auto-Correlation Coefficients (r(i)):</u>

- $R(0), R(1), R(2), \cdots, R(p)$

  $R(i) \qquad 0 \le i \le p$

- **Normalizing with respect to** $R(i)$

  $$r(0), r(1), r(2), \cdots, r(p)$$

  when, $r(i) = \frac{R(i)}{R(0)}$

$$|r(i)| \leq 1 \qquad 0 \leq i \leq p$$

are termed as normalized autocorrelation values.

- **For scaling to a fixed point solution.**
- **LPC's remain unaltered when computed using** $r(i)$ **instead of** $R(i)$**.**

3. <u>Normalized Error (V_i):</u>

$$V_i = \frac{E_i}{R(0)} = \frac{R(0) + \sum_{k=1}^{p} a_k R(k)}{R(0)}$$

$$V_i = 1 + \sum_{k=1}^{p} a_k r(k)$$

Since, $0 \leq E_i \leq E_{i-1} \quad and \quad E_0 = R(0) \qquad 1 \leq i \leq p$

$$0 \leq \frac{E_i}{R(0)} \leq \frac{E_{i-1}}{R(0)} \quad and \quad V_0 = 1 \qquad 1 \leq i \leq p$$

$$0 \leq V_i \leq 1 \qquad i \geq 0$$

$$E_i = (1 - k_i^2)E_{i-1}$$

$$= (1 - k_i^2)(1 - k_{i-1}^2)E_{i-2}$$

$$= (1 - k_i^2)(1 - k_{i-1}^2)(1 - k_{i-2}^2)E_{i-3}$$

$$\vdots$$

$$E_i = \prod_{i=1}^{p}(1 - k_i^2)E_0$$

$$V_i = \frac{E_i}{E_0} = \prod_{i=1}^{p}(1 - k_i^2)$$

4. Reflection Coefficients (k_i):

- **Partial correlation coefficients.**
- **PARCOR coefficients.**
- K_i **gives the partial correlation between** $s(n)$ **and** $s(n+i)$.
- **Transmission line theory**

$$k_i = -\frac{[R(i)+\sum_{j=1}^{i-1} a_j^{(i-1)} R(i-j)]}{E_{i-1}}.$$

  - k_i **is considered as the reflection coefficient at the boundary between two sections with impedance** z_i **and** z_{i+1}.

$$k_i = \frac{z_{i+1}-z_i}{z_{i+1}+z_i}.$$

$$H(z) = \frac{z_{i+1}}{z_i} = \frac{1+K_i}{1-k_i}, \qquad 1 \le i \le p.$$

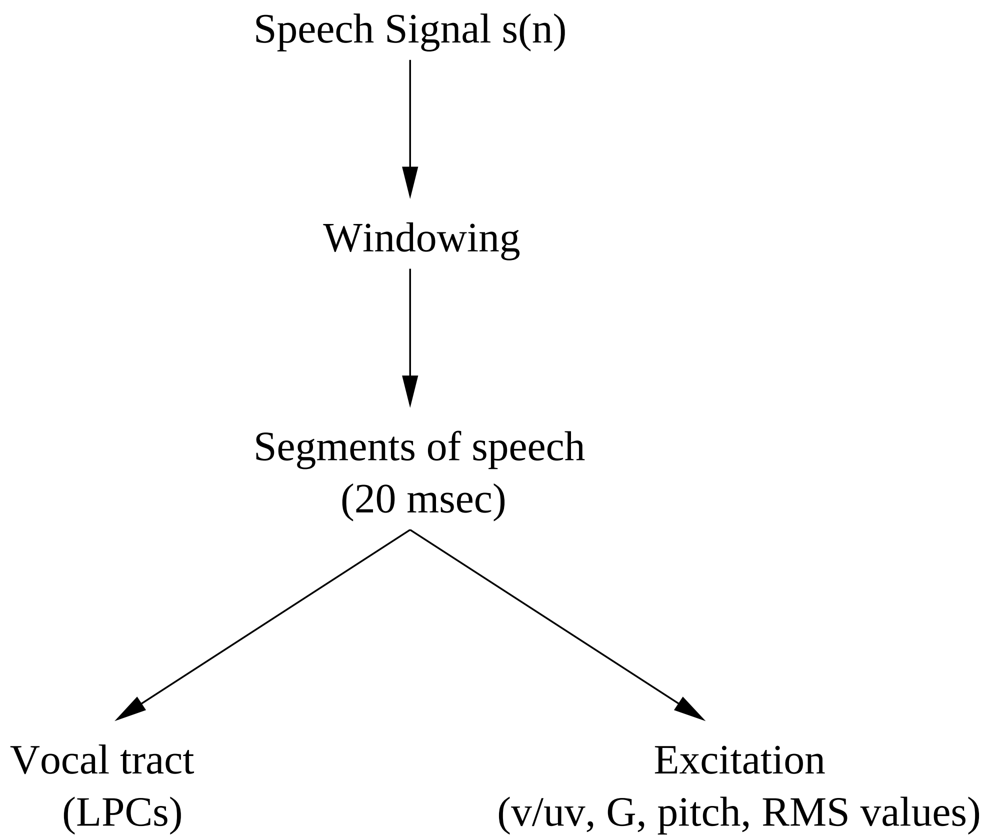
5. Filter Stability:

- **All roots inside unit circle**
- **Alternatively, successive errors are positive.**
  - $E_i > 0, \qquad 1 \le i \le p$ **is a necessary and sufficient conditions for stability of $H(z)$.**
  - **where $E_i = (1 - k_i^2)E_{i-1}$.**
  - **For stability $|k_i| < 1, \qquad 1 \le i \le p$.**

6. Prediction Order (p):

- **Where V_p becomes flat.**
- **Thumb rule, 10-14 for speech signal sampled at 8 kHz.**

Data Compression



- **For digital transmission, the LP parameters (LPC's, Excitation) are quantized.**

- **After quantization**

  1. **Filter $H(z)$ represented by the quantized parameters should be stable.**
  2. **Inherent order should be maintained.**
     $a_1, a_2, a_3, \cdots, a_p$ **if the order of a_1,a_2 changes they no-longer represent same $H(z)$.**
  3. **LPC's ensure ordering, but may be unstable after quantization.**
  4. **Reflection coefficients will not become unstable. Hence they are used.**

LPC to Cepstral Parameters Computation

$$H(z) = \frac{1}{1+\sum\limits_{k=1}^{p} a_k z^{-k}}$$

$$|H(e^{jw})| = \frac{1}{|1+\sum\limits_{k=1}^{p} a_k e^{-jwk}|}$$

$$c_n = \frac{1}{2\pi} \int\limits_{2\pi} log|H(e^{jw})|e^{jwn} dw$$

$$c_0 = log(E_0)$$

$$c_m = a_m + \sum_{k=1}^{m-1} (\frac{k}{m}) c_k a_{m-k} \qquad 1 \leq m \leq p$$

$$c_m = \sum_{k=1}^{m-1} (\frac{k}{m}) c_k a_{m-k} \qquad m \geq p$$

– **Usually order of LPCC $\simeq (3/2)$ order of LPC.**

If order of LPC $p = 8$, then the order of LPCC is $\simeq 12$.

a_k**=[-2.4936, 1.7346, 0.4429, -0.7634, -0.1846, 0.2762, 0.2580, -0.2620, -0.0871, 0.1127].**

k_i**=[-0.9516, 0.9783, -0.7577, -0.1856, 0.2866, 0.0742, -0.0652, 0.032, 0.1964, 0.1127].**

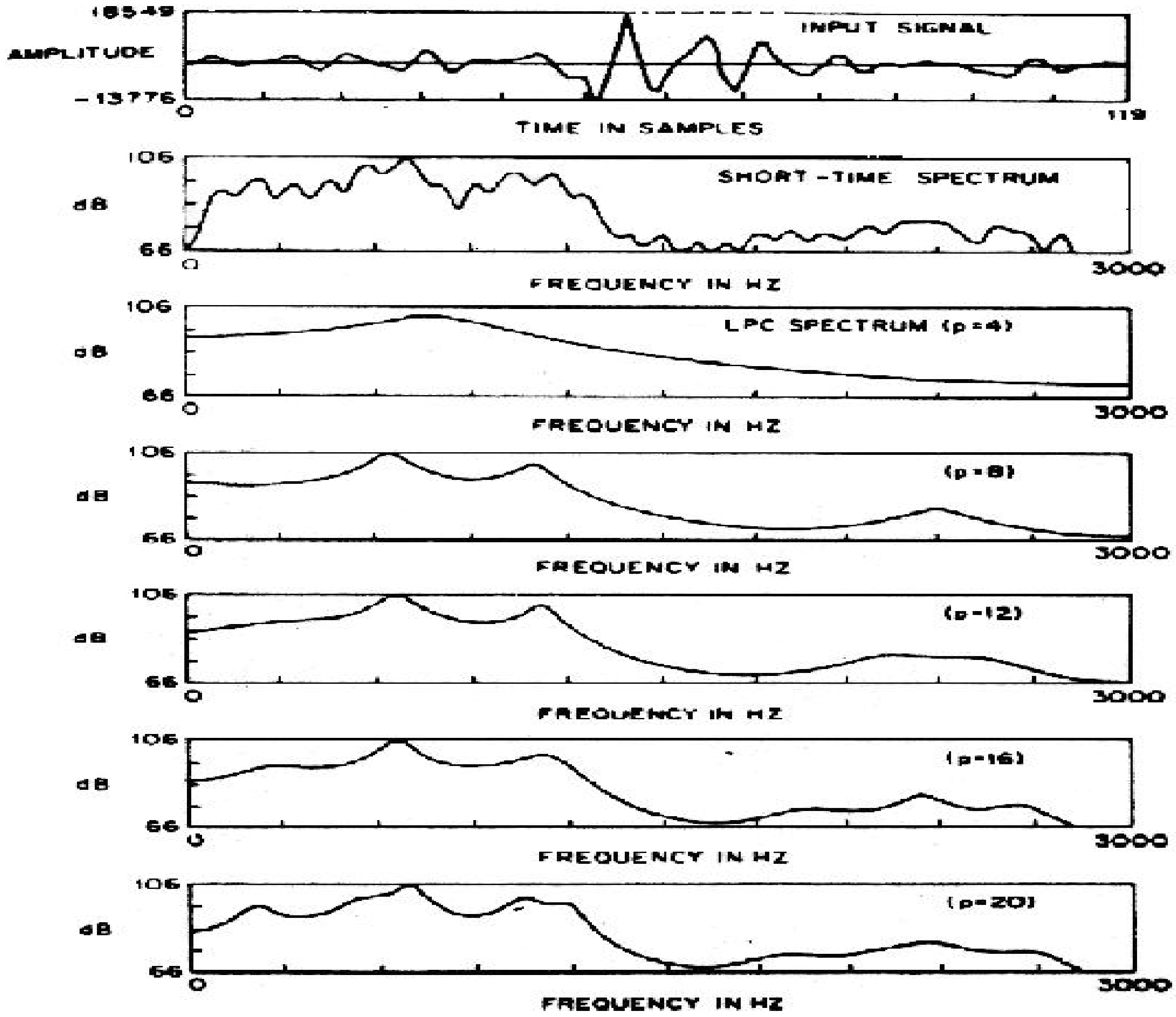


Figure 9: Spectra for a vowel sound for several values of predictor order, p.

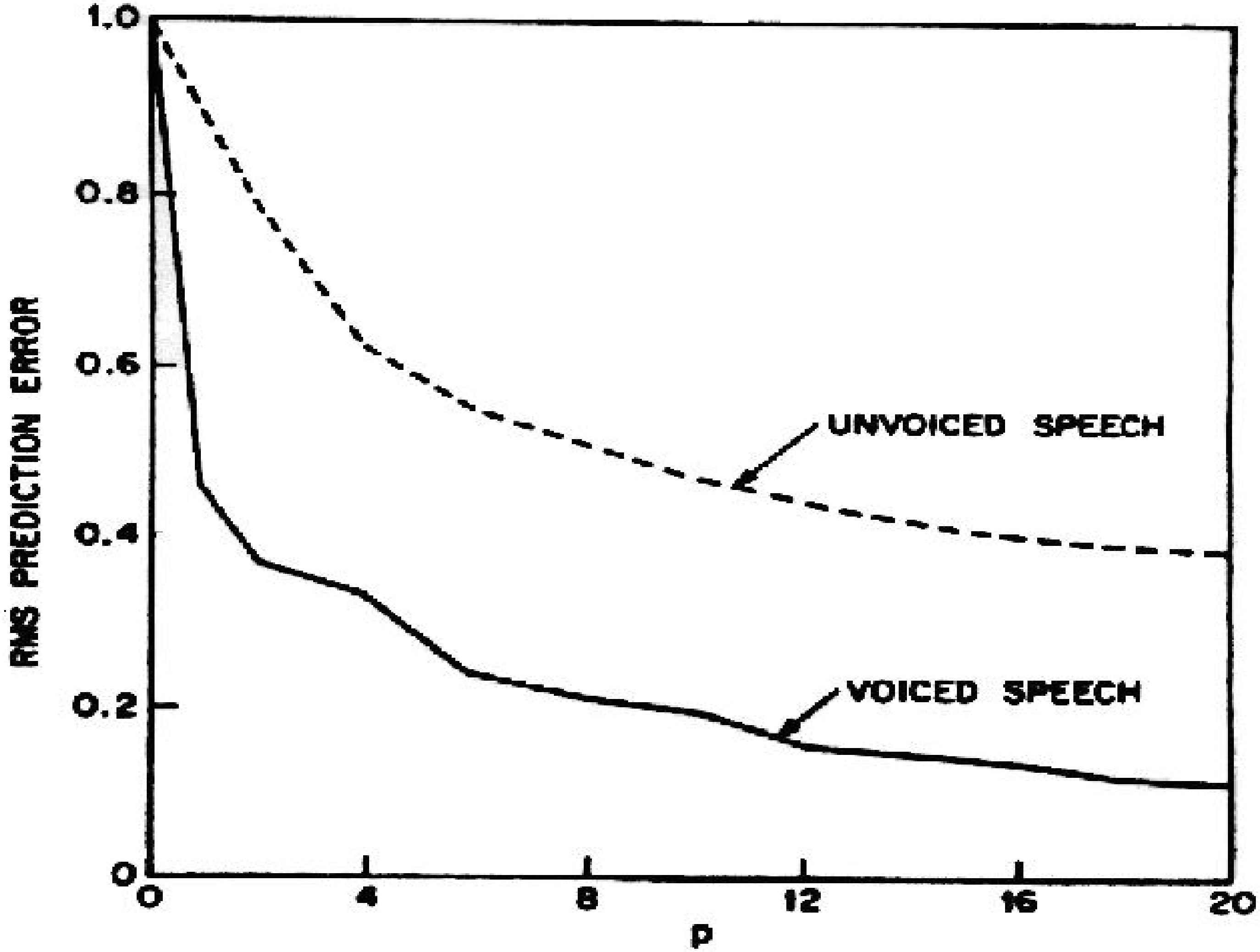


Figure 10: Variation of the RMS prediction error with the number of predictor coefficients, p (after Atal and Hanauer).

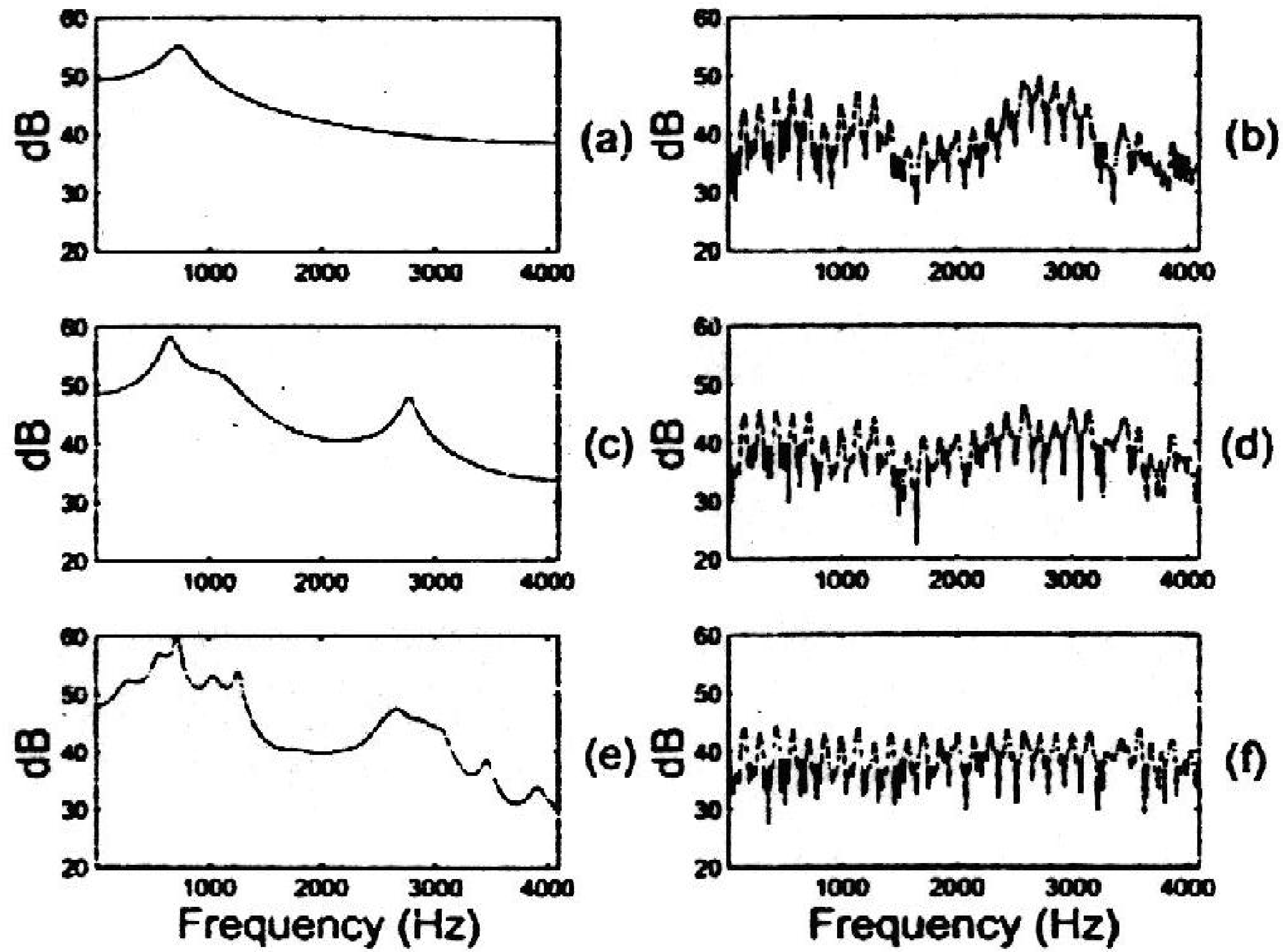


Figure 11: (a) LP spectrum and (b) residual spectrum for LP order 2. (c) LP spectrum and (d) residual spectrum for LP order 8. (e) LP spectrum and (f) residual spectrum for LP order 30.

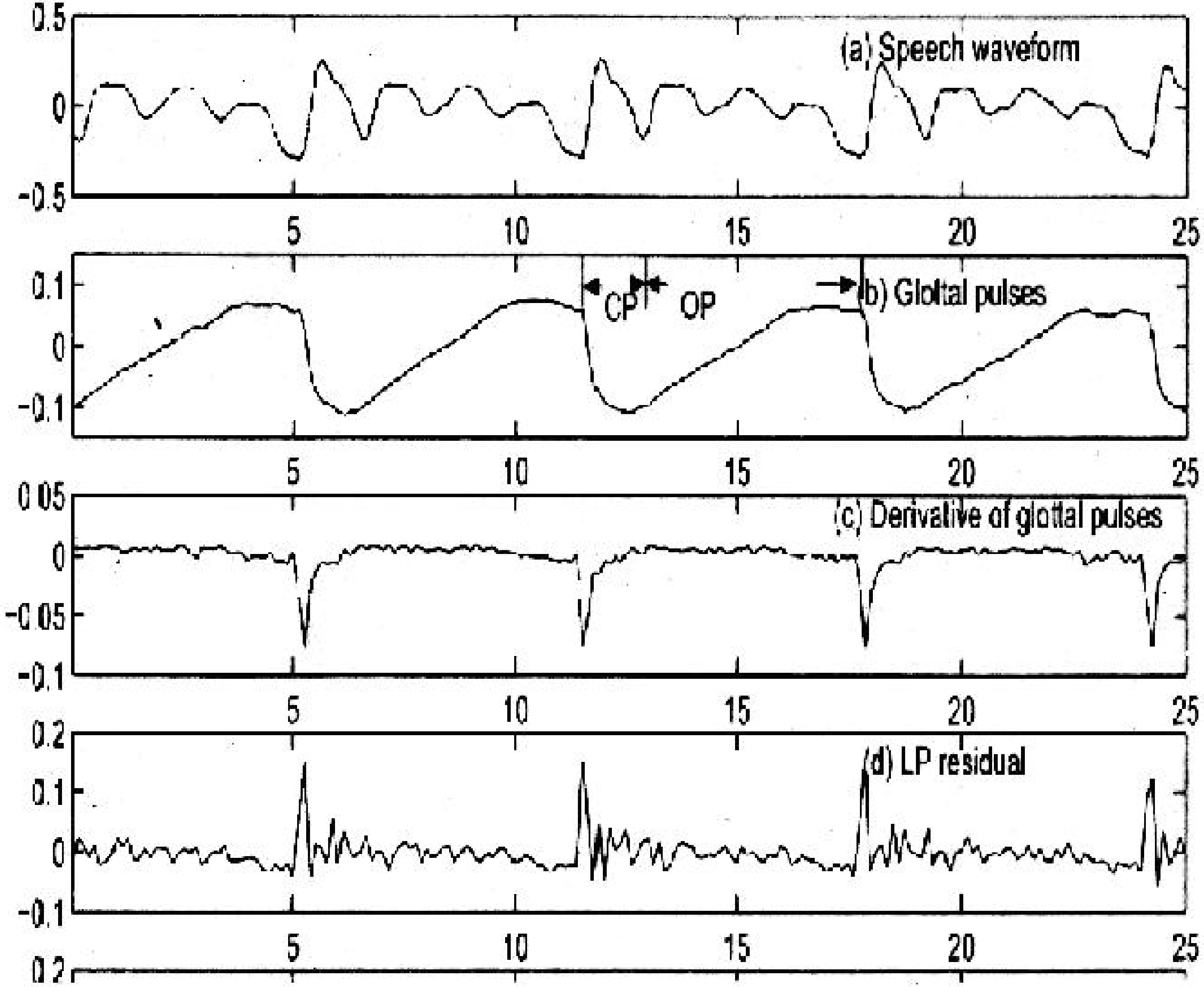


Figure 12: (a) Speech waveform, (b) glottal pulses, (c) derivative of glottal pulses, and (d) LP residual.

References

[1] B. S. Atal, and S. L. Hanauer, "Speech analysis and synthesis by LP of the speech wave," *JASA*, vol. 50, no. 2, pp. 637-655, 1971.

[2] J. Makhoul, "Linear prediction: A tutorial review," *Proc. of IEEE*, vol. 63, no. 4, pp. 561-580, Apr. 1975.

[3] L. R. Rabiner, and R. W. Schafer, *Digital Processing of Speech Signals*, Pearson Edn., 2006.

[4] L. R. Rabiner, and B. H. Juang, *Fundamentals of Speech Recognition*, Pearson Edn., 2006.