CECS401
Fundamentals of Spoken Language Processing

Note-5

Tuesday 9/14/99

C. Linear predictive coding (LPC) of speech

Excitation model of voiced sounds

Example:

measured glottal volume velocity and sound pressure at the mouth

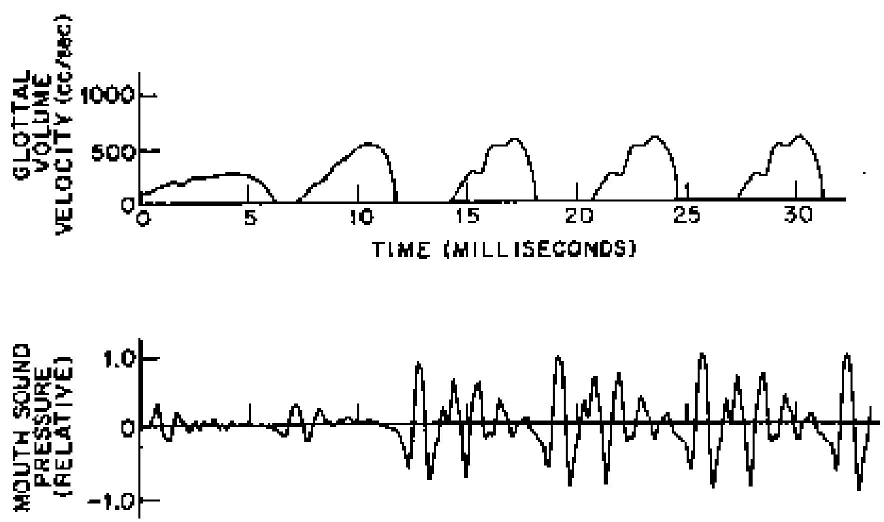


Figure 2.5 Glottal volume velocity and resulting sound pressure at the start of a voiced sound (after Ishizaka and Flanagan [4]).

Excitation model:

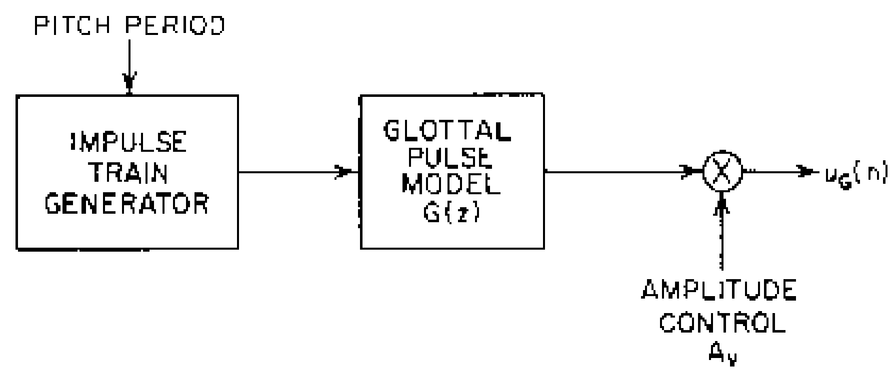
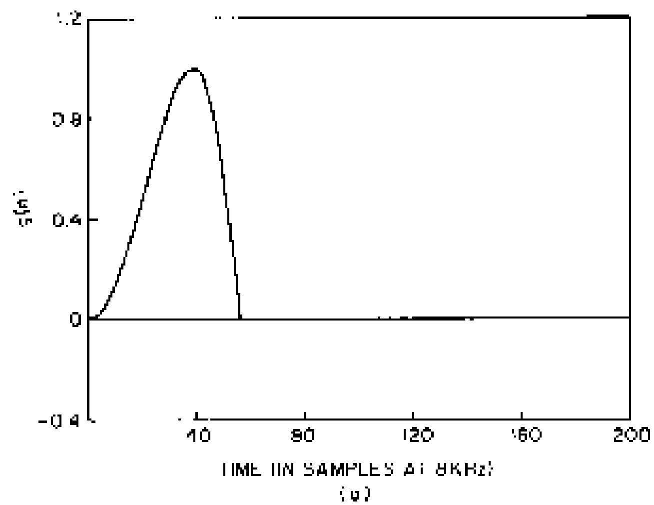


Fig. 3.48 Generation of the excitation signal for voiced speech.

An all-zero model of glottal pulse (A. E. Rosenberg):

$$g(n) = \begin{cases} \frac{1}{2}[1 - \cos(\pi n/N_1)] & 0 \le n \le N_1 \\ \cos(\pi(n - N_1)/2N_2) & N_1 \le n \le N_1 + N_2 \\ 0 & otherwise \end{cases}$$



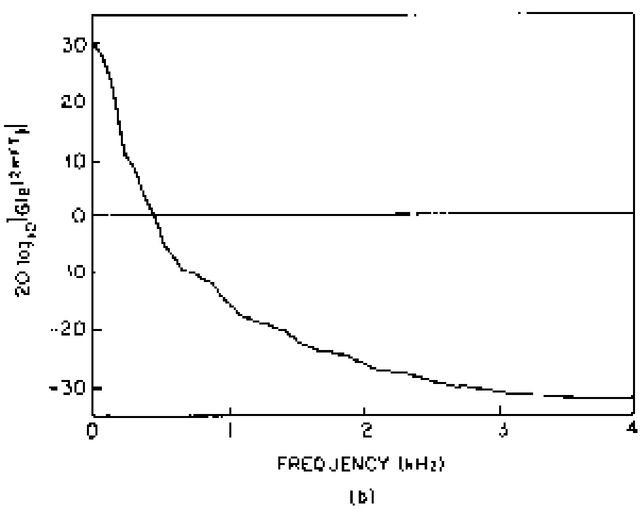


Fig. 3.49 (a) Rosenberg approximation to glottal pulse; (b) corresponding Fourier transform.

An all-pole model of glottal pulse:
(J. D. Markel and A. H. Gray, Jr.)

$$G(z) = \frac{1}{(1 - \mu z^{-1})^2}$$

In LPC of voiced speech, the input $u(n)$ is assumed as a periodic pulse train, and the system transfer function related to the complete digital model of speech becomes

$$H(z) = G(z)V(z)R(z)$$

Note that $G(z)$ has two poles and $R(z)$ has one zero ($R(z) = R_0(1 - z^{-1})$). In order to obtain vocal tract function $V(z)$, a first-order high-pass filter $1 - \gamma z^{-1}$ (*preemphasis*) is often applied to speech samples before LPC analysis.

LPC analysis equation

With excitation $u(n)$ unknown, the signal sample $s(n)$ is approximated as

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

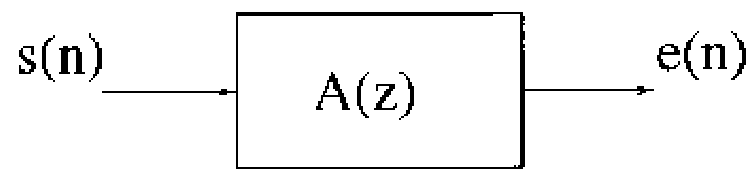
The prediction error is

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

In z-transform domain,

$$E(z) = S(z)\left(1 - \sum_{k=1}^{P} a_k z^{-k}\right) = S(z)A(z)$$

$A(z)$ is called *inverse* filter that produces the error sequence $e(n)$ from the input signal sequence $s(n)$.



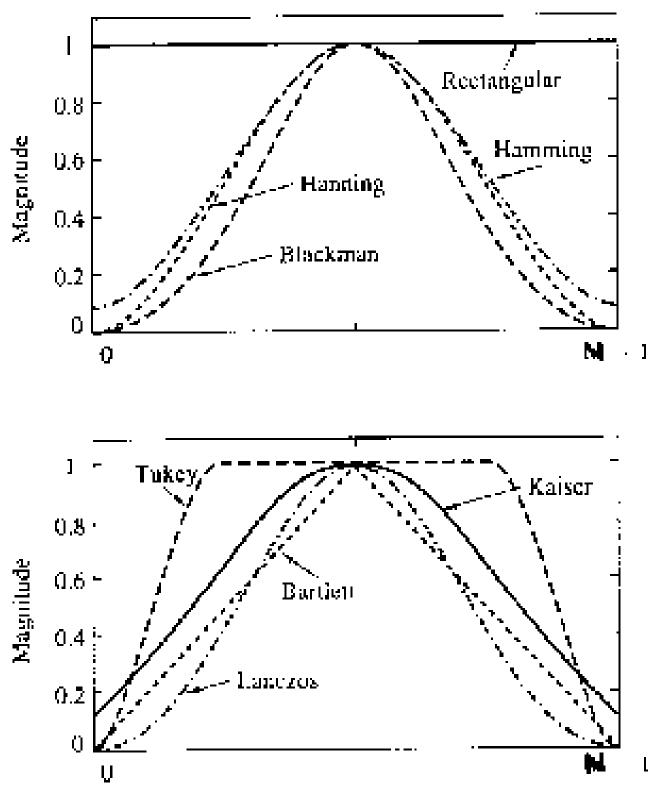
Estimation of LPC parameters (the autocorrelation method)

Definition of an analysis frame

$$s_n(m) = \begin{cases} s(n+m) \cdot w(m) & 0 \leq m \leq N-1 \\ 0 & otherwise \end{cases}$$

$w(m)$ is a window function that tapers signal samples at the frame ends of $m = 0$ and $m = N - 1$.

Shapes of several window functions:



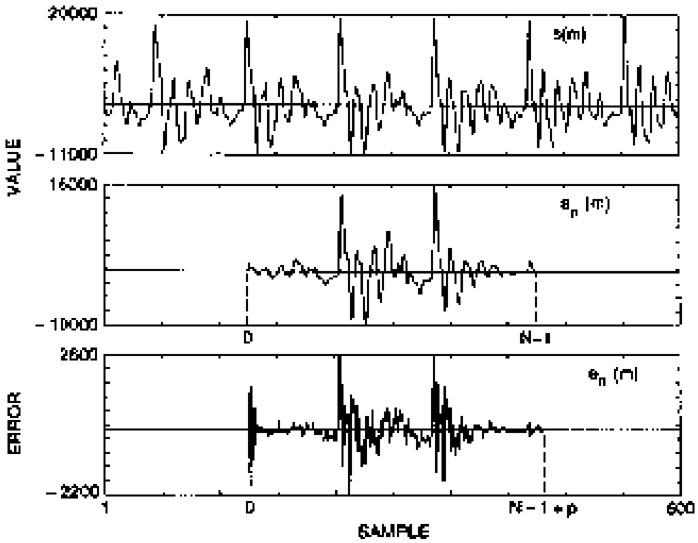


Figure 3.29 Illustration of speech sample, weighted speech section, and prediction error for voiced speech where the prediction error is large at the beginning of the section.

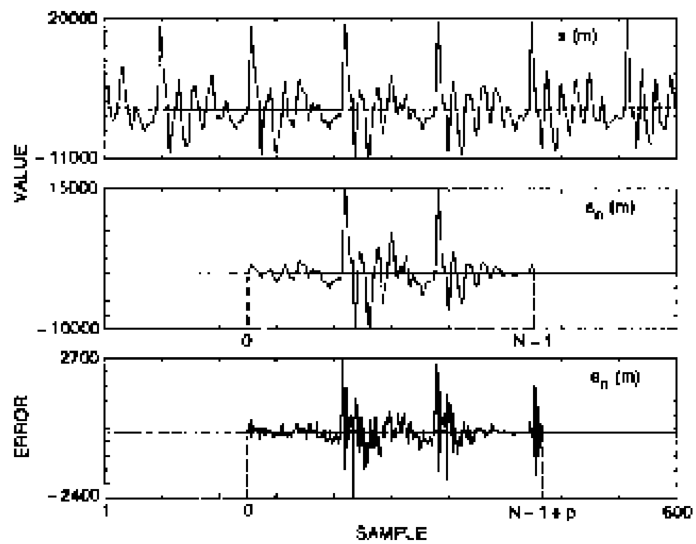


Figure 3.30 Illustration of speech sample, weighted speech section, and prediction error for voiced speech where the prediction error is large at the end of the section.

Mean-squared prediction error:

$$E_n = \sum_{m=0}^{N-1+p} e_n^2(m) = \sum_{m=0}^{N-1+p} \left(s_n(m) - \hat{s}_n(m)\right)^2 = \sum_{m=0}^{N-1+p} \left(s_n(m) - \sum_{k=1}^{p} a_k s_n(m-k)\right)^2$$

Minimum mean-squared error (MMSE) estimation of LPC parameters:

$$\frac{\partial E_n}{\partial a_k} = 0, \quad k = 1, 2, \cdots, p$$

solve the roots for $\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_p$.

Setting $\frac{\partial E_n}{\partial a_i} = 0,\ i = 1, 2, \cdots, p$ leads to

$$\sum_{m=0}^{N-1+p} s_n(m) s_n(m-i) = \sum_{k=1}^{p} \hat{a}_k \sum_{m=0}^{N-1+p} s_n(m-k) s_n(m-i)$$

Define

$$\phi_n(i,k) = \sum_{m=0}^{N-1+p} s_n(m-k) s_n(m-i)$$

then

$$\phi_n(i,0) = \sum_{k=1}^{p} \hat{a}_k \phi_n(i,k)$$

It can be shown by substitution of variables that

$$\phi_n(i,k) = \sum_{m=0}^{N-1-(i-k)} s_n(m)s_n(m+i-k) = r_n(i-k)$$

and also

$$\phi_n(i,k) = \phi_n(k,i) \rightarrow r_n(i-k) = r_n(k-i)$$

The LPC equation can then be expressed in terms of the autocorrelation coefficients $r_n(k)$'s as

$$r_n(i) = \sum_{k=1}^{p} \hat{a}_k r_n(|i-k|), \quad i = 1, 2, \cdots, p$$

The LPC equation in matrix form is

$$\begin{bmatrix} r_n(0) & r_n(1) & r_n(2) & \cdots & r_n(p-1) \\ r_n(1) & r_n(0) & r_n(1) & \cdots & r_n(p-2) \\ r_n(2) & r_n(1) & r_n(0) & \cdots & r_n(p-3) \\ \vdots & \vdots & \vdots & & \vdots \\ r_n(p-1) & r_n(p-2) & r_n(p-3) & \cdots & r_n(0) \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \vdots \\ \hat{a}_p \end{bmatrix} = \begin{bmatrix} r_n(1) \\ r_n(2) \\ r_n(3) \\ \vdots \\ r_n(p) \end{bmatrix}$$

or in short $R_n \cdot \hat{a} = r_n$ (normal equation).

Properties of R_n:

positive definite, symmetric, and Toeplitz (elements along each subdiagonals are identical).

Computation of $\hat{a}$:

- Direct method ($p^3/3 + O(p^2)$)
- Durbin's method ($p^2 + O(p)$)

Minimum mean-squared error (MMSE):

$$\hat{E}_n = \sum_{m=1}^{N-1+p} \left(s_n(m) - \sum_{k=1}^{p} \hat{a}_k s_n(m-k) \right)^2$$

$$= \sum_{m=1}^{N-1+p} (s_n(m))^2 - 2 \sum_{k=1}^{p} \sum_{m=1}^{N-1+p} \hat{a}_k s_n(m) s_n(m-k)$$

$$+ \sum_{k=1}^{p} \hat{a}_k \sum_{i=1}^{p} \hat{a}_i \sum_{m=1}^{N-1-p} s_n(m-k) s_n(m-i)$$

Applying

$$\phi_n(i,0) = \sum_{k=1}^{p} \hat{a}_k \phi_n(i,k)$$

one gets

$$\hat{E}_n = \phi_n(0,0) - \sum_{k=1}^{p} \hat{a}_k \phi_n(0,k) = r_n(0) - \sum_{k=1}^{p} \hat{a}_k r_n(k)$$