Linear Predictive Analysis

- The basic assumption of **linear predictive** (**LP**) analysis:
  A speech sample can be approximated as a linear combination of past speech samples

- The LP model is based on the source-filter model in speech production, assuming that the filter is an all-pole LTI system.

- LP analysis generates a transformed representation for each short-time speech frame.

- LP analysis can produce reasonably accurate estimation of speech parameters at low computation cost.

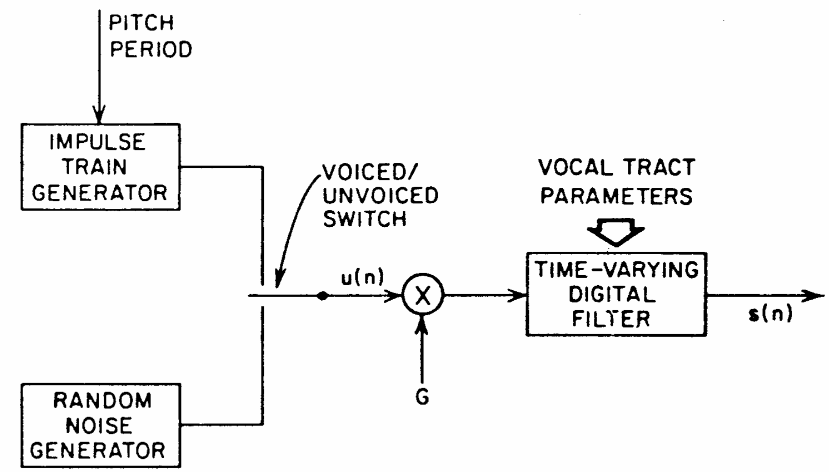
- LP analysis method is commonly used for
  1. estimating speech parameters, e.g. formants, pitch, spectrum, etc.
  2. low bit-rate transmission and storage of speech

LP model for speech signals

The vocal tract transfer function is

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

$A(z)$ is called “inverse filter”



In time-domain, the speech signal $s[n]$ can be expressed as

$$s[n] = \sum_{k=1}^{p} a_k s[n-k] + Gu[n]$$

The parameters of a complete LP model include:

Excitation

- voiced/unvoiced classification
- pitch period

Filter

- filter order p
- filter gain G
- filter coefficients a_k

LP analysis = derivation of the LP parameters from speech signals.

LP analysis is performed for each individual frame.

The objective is that the resultant model provides a good estimate of the short-time spectrum.

LP in frequency domain

- The LP model $H(z)$ is intended to model the function of vocal tract. Therefore, $H(e^{j\Omega})$ is characterizing the vocal tract frequency response.

- LP analysis is a method of estimating or approximating short-time spectrum.

Theoretically, it can be shown that

$$E_n = \sum_{m=n-N+1}^{n+p} e_n^2[m] = \frac{G^2}{2\pi} \int_{-\pi}^{\pi} \frac{|S_n(e^{j\Omega})|^2}{|H(e^{j\Omega})|^2} d\Omega$$

Given

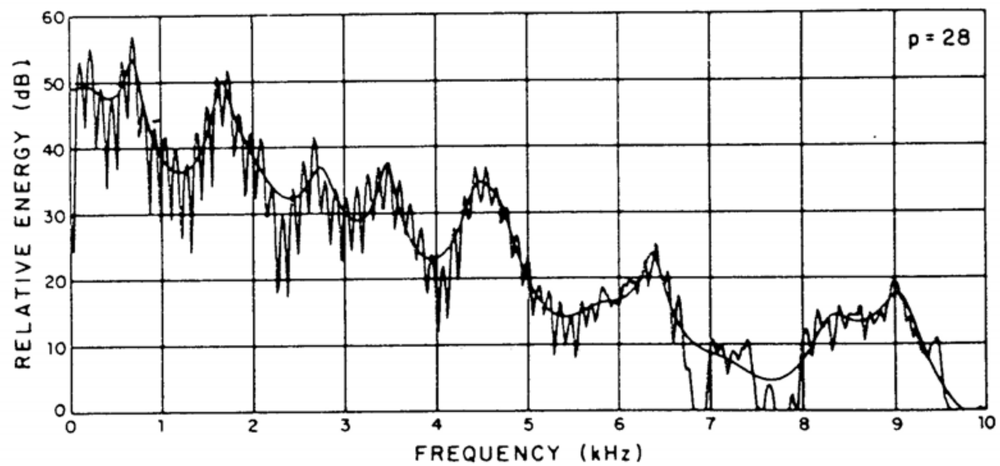
$$e[n] = Gu[n], U(z) = \frac{S(z)}{H(z)}$$

using Parseval's Theorem,

$$\sum_n |x[n]|^2 = \frac{1}{2\pi} \int_{2\pi} |X(e^{j\Omega})|^2 d\Omega$$

That is, minimizing E_n is equivalent to minimizing the integral of the ratio between signal spectrum and the LP frequency response.

$20\log_{10}|H(e^{j\Omega})|$ versus $20\log_{10}|S_n(e^{j\Omega})|$



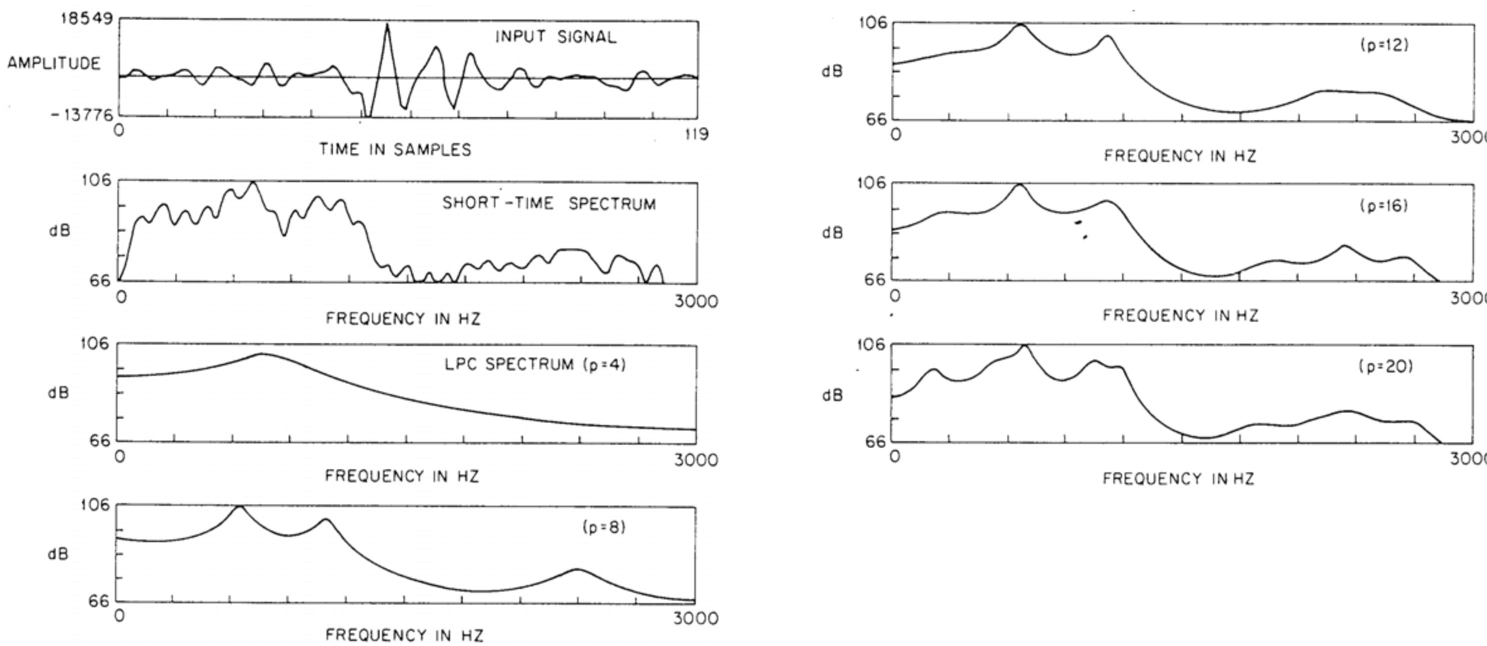
$20\log_{10}|S_n(e^{j\Omega})|$ computed by applying FFT with 20 msec window

$20\log_{10}|H(e^{j\Omega})|$ computed with $p = 28$

- LP spectrum doesn't show excitation details since it represents the vocal tract only.
- LP spectrum matches the signal spectrum more closely in the region of high signal energy (spectral peaks).

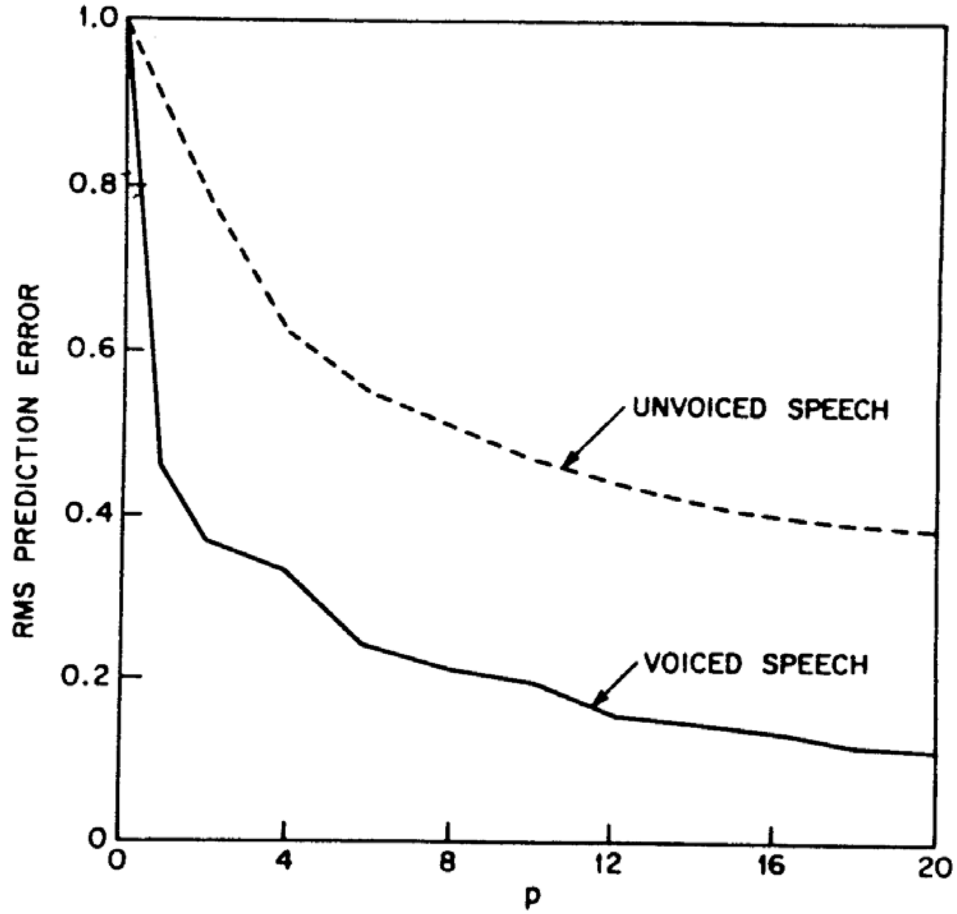
Selection of filter order

Vowel /a/, frame of 20 msec with 6 KHz sampling



- As p increases, the LP analysis preserves more spectrum details
- Usually, p is kept as small as possible provided that the general spectrum shape, i.e. formant peaks, is preserved.
- In practice, $p = 10$ to 20, depending on the nature of applications.

Variation of RMS prediction error with filter order *p*



The selection of window size

- N should be as small as possible to reduce the computation cost

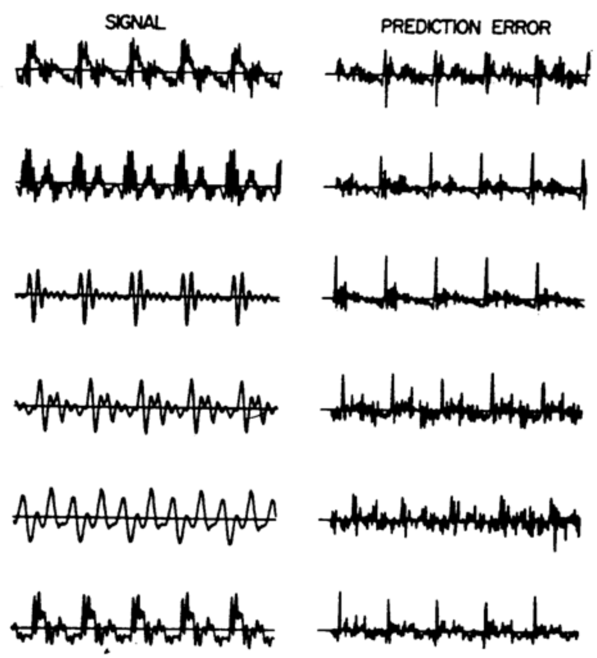
- For the auto-correlation method in which a tapering window is used, N must be in the order of several pitch periods
  e.g. $N = 100$ to 400 for $10kHz$ sampling
  large N is preferred if possible

- For the covariance method, N can be relatively small.

The prediction error

The prediction error signal is a by-product of LP analysis.

Since $e[n] = Gu[n]$, $e[n]$ is essentially an approximation of the input excitation. That is, for voiced speech, $e[n]$ would be large at the beginning of each pitch period (Woo! good for pitch detection !)



Original signal and linear prediction error for vowel segments /i/, /e/, /a/, /o/, /u/, /y/