Lecture 13
Chapter 9. Linear Predictive Analysis of Speech Signals

DEEE725 Speech Signal Processing Lab
Gil-Jin Jang
Original slides from Lawrence Rabiner

Linear Prediction

- Objective: to approximate the output sequence as a linear combination of input samples, past output samples or both :

$$\hat{y}(n) = \sum_{j=0}^{q} b(j)\underbrace{x(n-j)}_{\text{Input samples}} - \sum_{i=1}^{p} a(i)\underbrace{y(n-i)}_{\text{Past output samples}}$$

  – The factors a(i) and b(j) are called predictor coefficients.

Linear Prediction

- Many systems of interest to us are describable by a linear, constant-coefficient difference equation :

$$\sum_{i=0}^{p} a(i)y(n-i) = \sum_{j=0}^{q} b(j)x(n-j)$$

$$Y(z) = H(z)X(z) \Leftrightarrow H(z) = \frac{Y(z)}{X(z)} = \frac{N(z)}{D(z)}$$

$$N(z) = \sum_{j=0}^{q} b(j)z^{-j} \text{ and } D(z) = \sum_{i=0}^{p} a(i)z^{-i}$$

Thus the predicator coefficient give us immediate access to the poles and zeros of $H(z)$.

Poles and Zeros of z-polynomials

$$H(z) = \frac{N(z)}{D(z)} = \frac{\sum_{j=0}^{q} b(j) z^{-j}}{\sum_{i=0}^{p} a(i) z^{-i}}$$

Poles: where the denominator $D(z) = 0$
Zeros: where the numerator $N(z) = 0$

Types of System Model

- There are two important variants :
  - **All-pole** model:
    - *autoregressive (AR)* model in statistics
    - The numerator $N(z)$ is a constant.
  - **All-zero** model :
    - *moving-average (MA)* model in statistics
    - The denominator $D(z)$ is equal to unity.
  - The mixed pole-zero model is called the ***autoregressive moving-average (ARMA)*** model.

Derivation of LP Equations

- Given a zero-mean signal $y(n)$, in the AR model

  - The error is :
$$\hat{y}(n) = -\sum_{i=1}^{p} a(i)\, y(n-i)$$
$$e(n) = y(n) - \hat{y}(n)$$
$$= \sum_{i=0}^{p} a(i)\, y(n-i)$$

  - To derive the predicator we use the ***principle of orthogonality***, the principle states that the desired coefficients are those which make the error orthogonal to the samples $y(n\text{-}1)$, $y(n\text{-}2)$,…, $y(n\text{-}p)$.

Derivation of LP equations

– Thus we require that

$$< y(n-j)e(n) >= 0 \quad \text{for } j = 1,\ 2,\ \ldots, p$$

- Or,

$$\left\langle y(n-j)\sum_{i=0}^{p} a(i)\,y(n-i)\right\rangle = 0$$

- Interchanging the operation of averaging and summing, and representing < > by summing over n (sample average), we have

$$\sum_{i=0}^{p} a(i)\sum_{n} y(n-i)\,y(n-j) = 0,\ j = 1,\ldots,p$$

- The required predicators are found by solving these equations.

Derivation of LP equations

- The orthogonality principle also states that resulting minimum error is given by

$$E = \langle e^2(n) \rangle = \langle y(n)e(n) \rangle$$

  - Or,

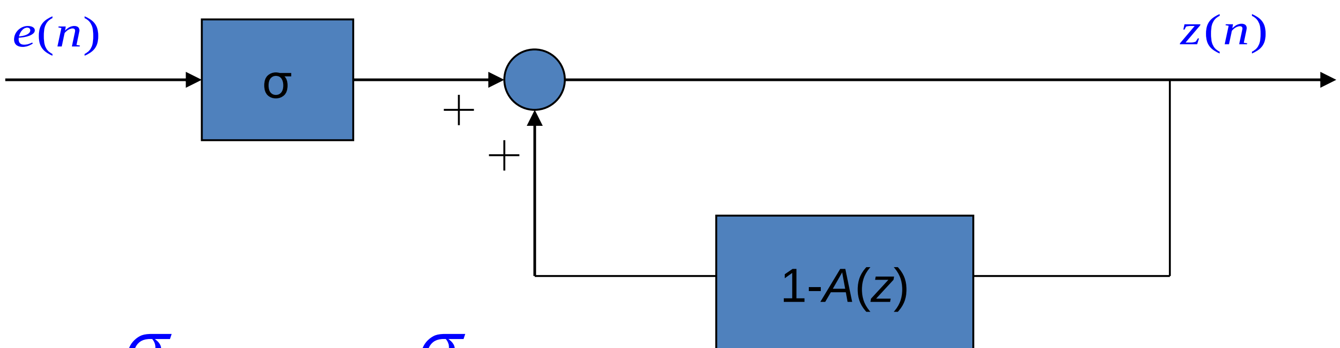
$$\sum_{i=0}^{p} a(i) \sum_{n} y(n-i)y(n) = E$$

- We can minimize the error over all time :
  -
$$\sum_{i=0}^{p} a(i) r_{i-j} = 0,\ j = 1,2,\ \ldots, p \qquad \sum_{i=0}^{p} a(i) r_i = E$$

  - where $r_i = \sum_{n=-\infty}^{\infty} y(n)y(n-i)$ ***autocorrelation***

Applications

- Autocorrelation matching :
  - We have a signal $y(n)$ with known autocorrelation $r_{yy}(n)$. We model this with the AR system shown below :



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

The LPC Model

$$s(n) \approx a_1 s(n-1) + a_2 s(n-2) + \ldots + a_p s(n-p),$$

Convert this to equality by including an excitation term:

$$s(n) = \sum_{i=1}^{p} a_i s(n-i) + Gu(n),$$

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

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

LPC Analysis Equations

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

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

The prediction error:

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

Error transfer function:

$$A(z) = \frac{E(z)}{S(z)} = 1 - \sum_{k=1}^{p} a_k z^{-k}.$$

Seek to minimize the mean squared error signal:

$$E_n = \sum_m e_n^2(m) = \sum_m \left[s_n(m) - \sum_{k=1}^{p} a_k s_n(m-k) \right]^2$$

The Autocorrelation Method

$$\frac{\partial E_n}{\partial a_i} = 0, \qquad i = 1,2,\ldots,p$$

$$\sum_m s_n(m-i)s_n(m) = \sum_{k=1}^{p} \hat{a}_k \sum_m S_n(m-i)S_n(m-k) \qquad (*)$$

Terms of short-term autocorrelation

$$r_n(i-k) \cong \sum_m S_n(m-i)S_n(m-k)$$

With this notation, we can write (*) as:

$$r_n(i) \cong \sum_{k=1}^{p} \hat{a}_k r_n(i-k) \qquad i = 1,2,\ldots,p$$

The Autocorrelation Method

Since the autocorrelation function is symmetric,

i.e. $r_n(-k) = r_n(k) \qquad so:$

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

and can be expressed in matrix form as :

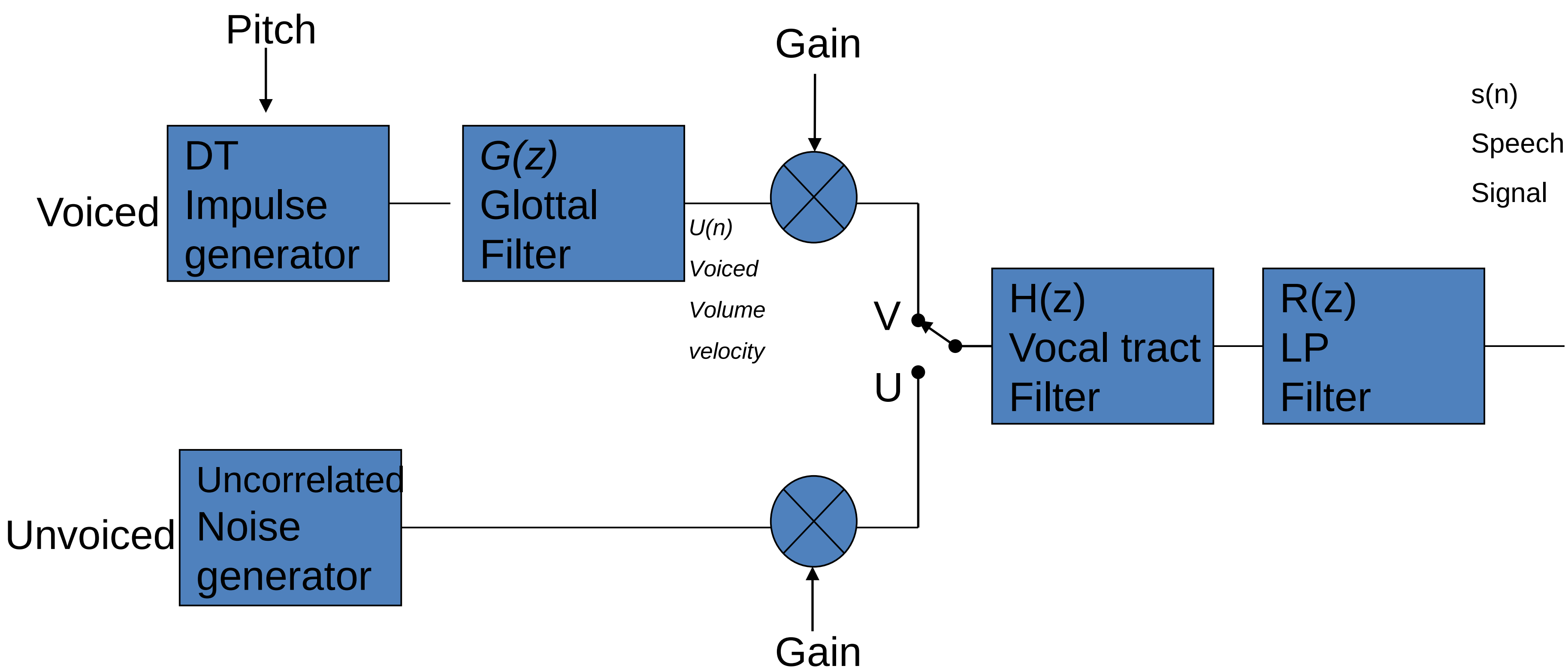
$$\begin{bmatrix} r_n(0) & r_n(1) & r_n(2) & \ldots r_n(p-1) \\ r_n(1) & r_n(0) & r_n(1) & \ldots r_n(p-2) \\ r_n(2) & r_n(1) & r_n(0) & \ldots r_n(p-3) \\ r_n(p-1) & r_n(p-2) & r_n(p-3) & \ldots \; r_n(0) \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_p \end{bmatrix} = \begin{bmatrix} r_n(1) \\ r_n(2) \\ r_n(3) \\ r_n(p) \end{bmatrix}.$$

The Autocorrelation Method

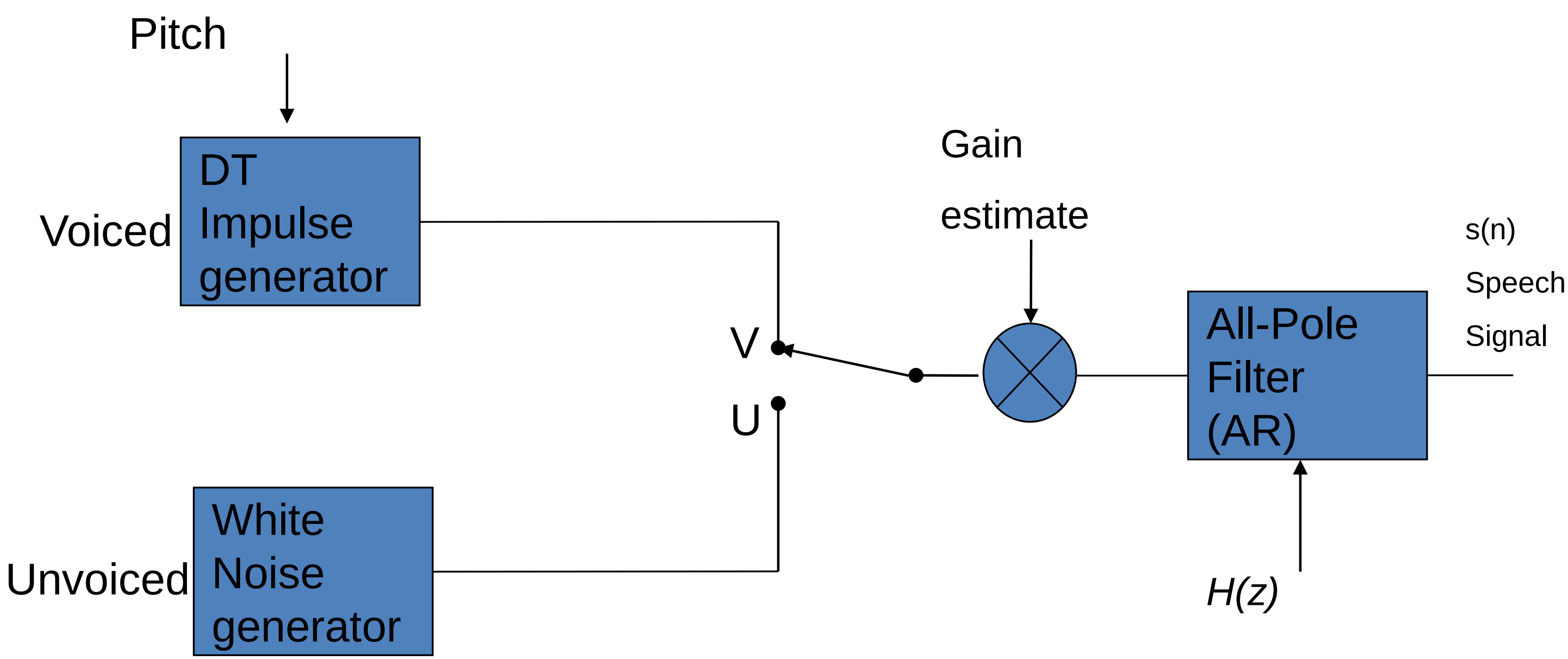
- The LP coefficients can be obtained by multiplying the inverse of the autocorrelation matrix
  - Matrix inversion is time consuming
    - Toeplitze matrix inversion by Levinson-durbin recursion (https://en.wikipedia.org/wiki/Levinson_recursion)

$$\begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_p \end{bmatrix} = \begin{bmatrix} r_n(0) & r_n(1) & r_n(2) & \ldots r_n(p-1) \\ r_n(1) & r_n(0) & r_n(1) & \ldots r_n(p-2) \\ r_n(2) & r_n(1) & r_n(0) & \ldots r_n(p-3) \\ r_n(p-1) & r_n(p-2) & r_n(p-3) & \ldots \; r_n(0) \end{bmatrix}^{-1} \begin{bmatrix} r_n(1) \\ r_n(2) \\ r_n(3) \\ r_n(p) \end{bmatrix}.$$

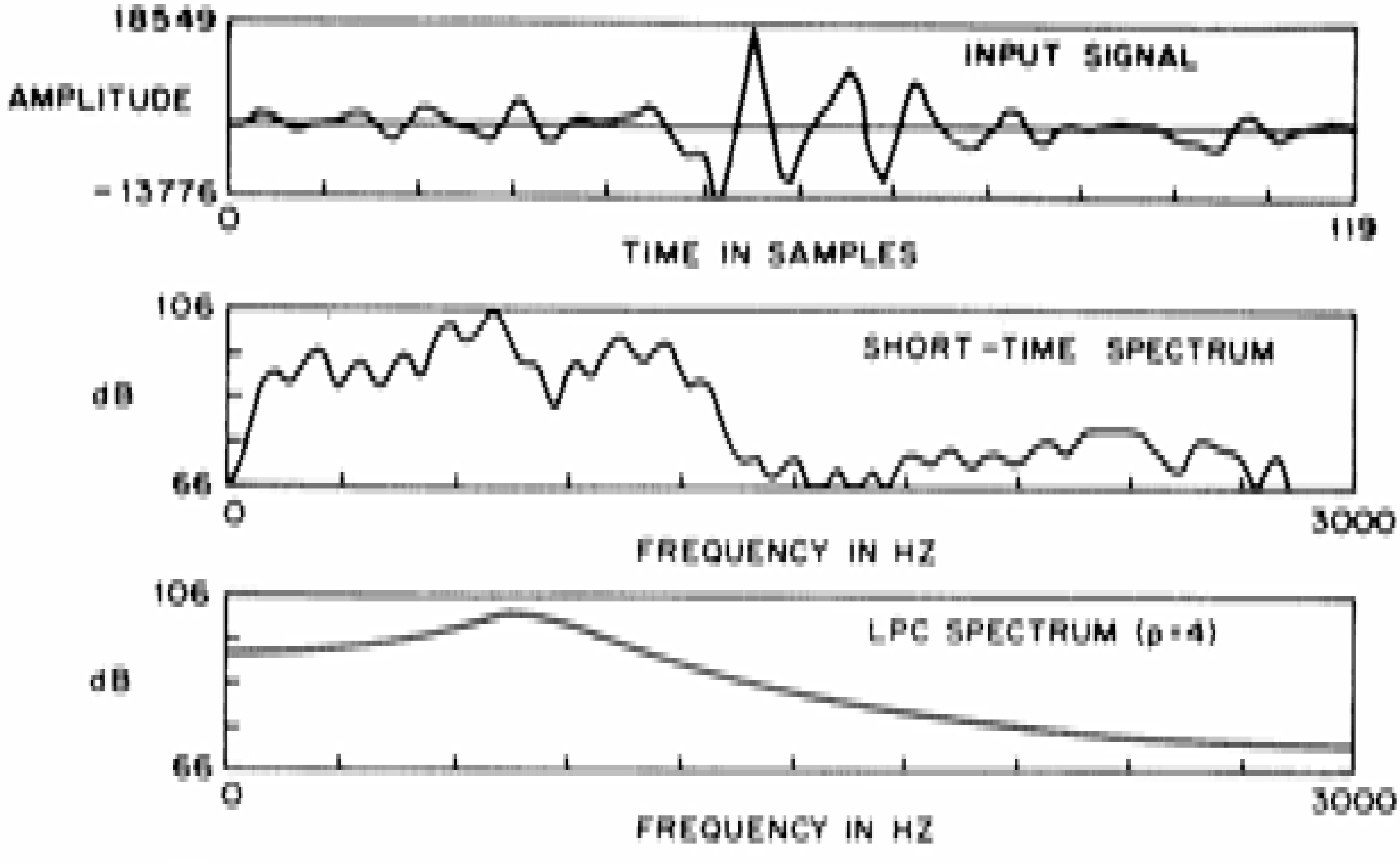
AR Modeling of Speech Signal: True Model



AR Modeling of Speech Signal: Using LP Analysis



Examples of LPC Analysis



Examples of LPC Analysis

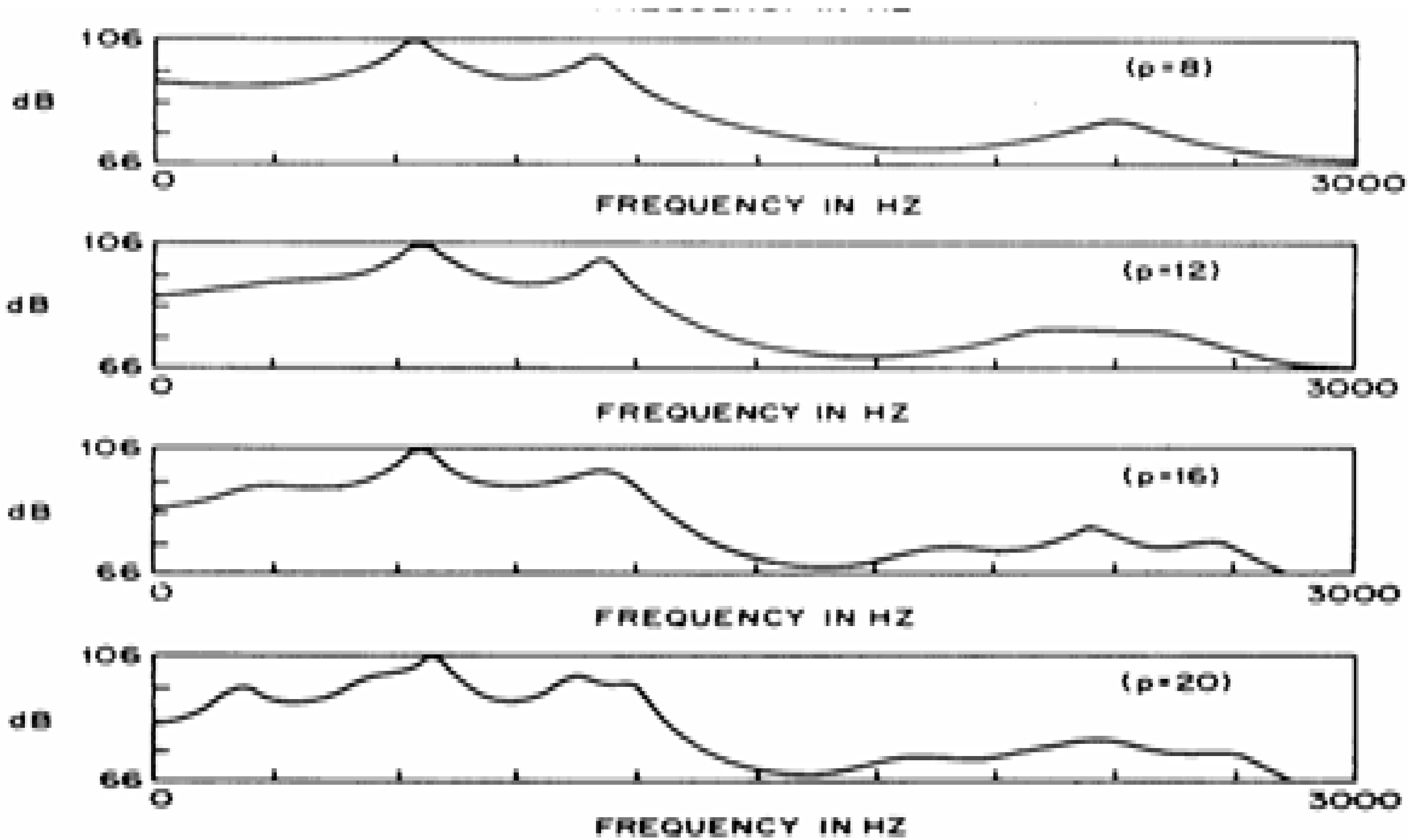


Figure 3.36 Spectra for a vowel sound for several values of predictor order, p.

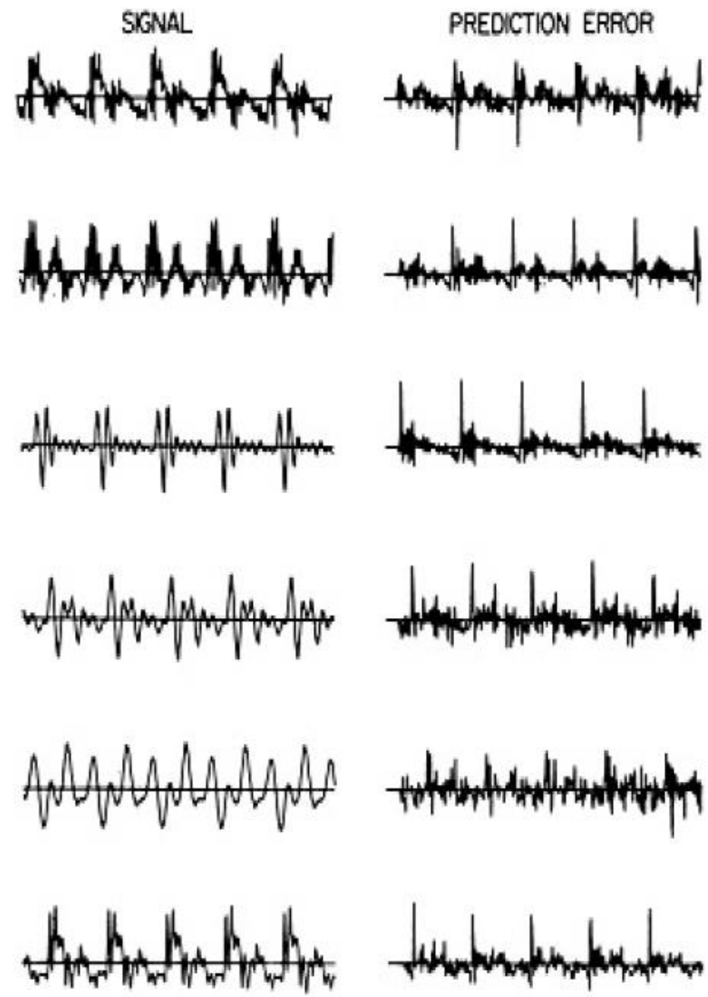
Reference: Chapter 6 of Introduction to Digital Speech Processing

Original slides from Lawrence Rabiner; Dan Ellis and Michael Mandel

MORE ON LINEAR PREDICTIVE CODING
LINE SPECTRAL PAIRS

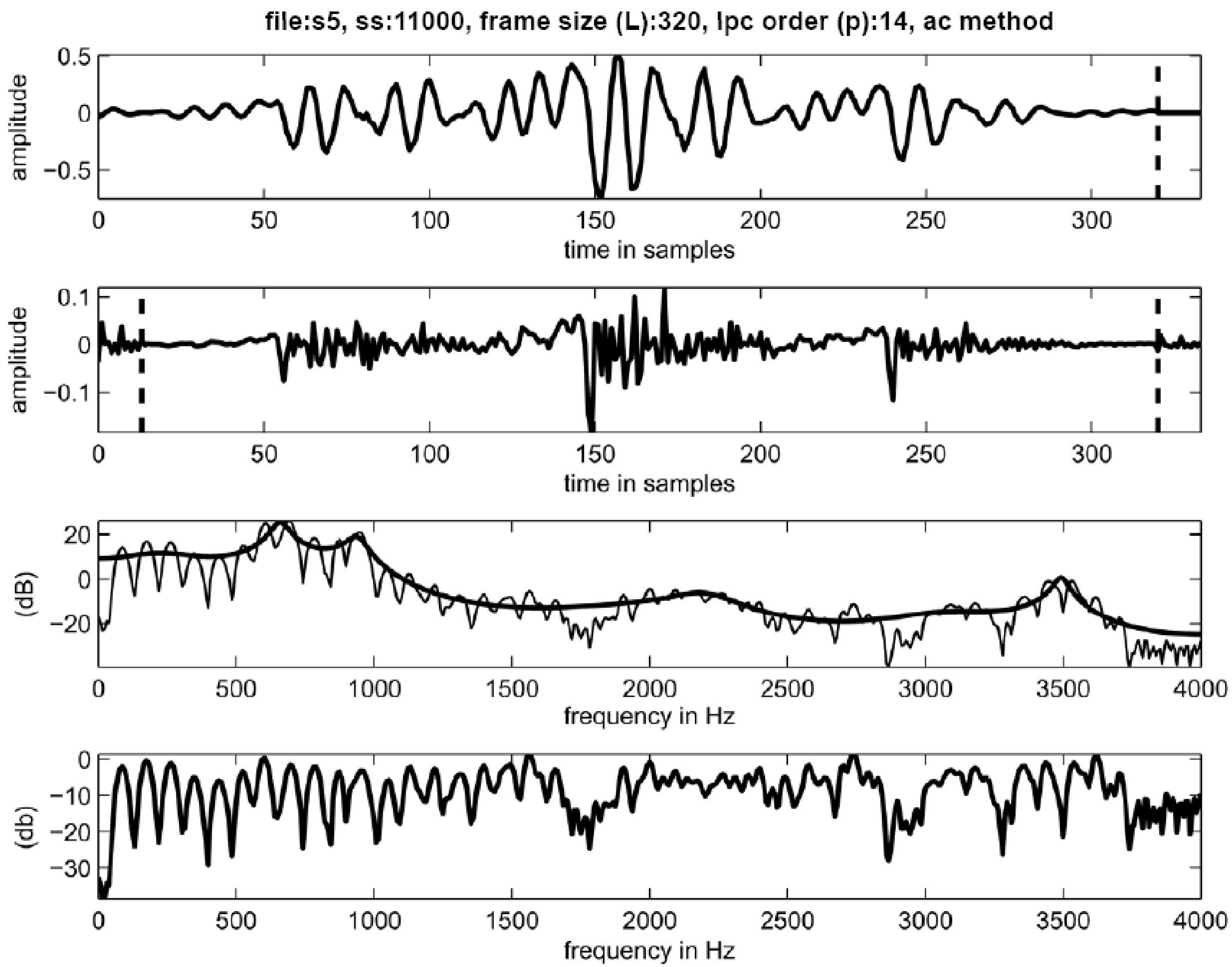
Prediction Error Signal Behavior

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



- The prediction error (residual) signal $e(n)$ should be large at the beginning of each pitch period in voiced speech, so it is good for pitch detection
- pitch can be detected at the largest peak of autocorrelation
- error spectrum is approximately flat, so effects of formants on pitch detection are minimized

The Prediction Error Signal 1



Top panel: speech signal

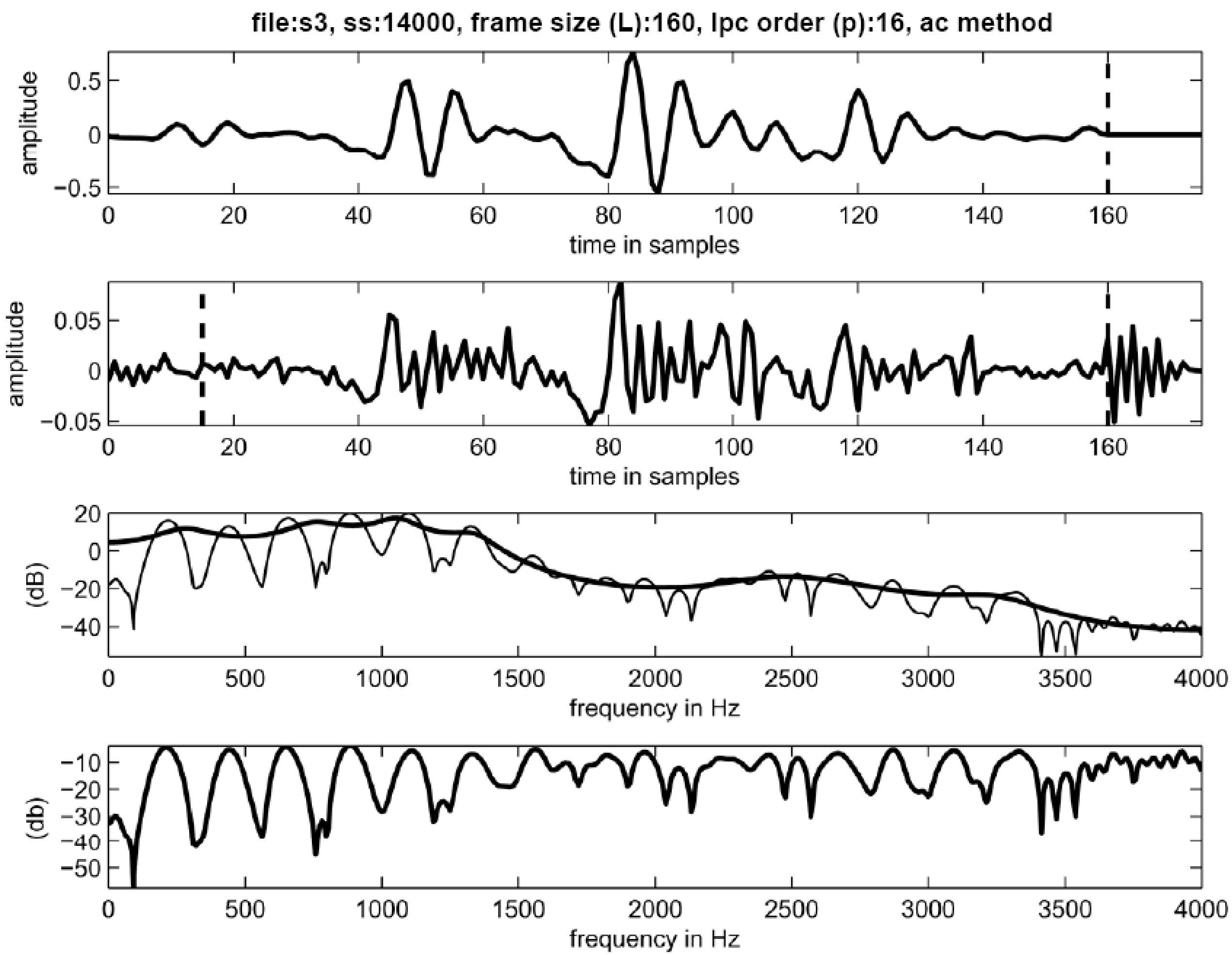
Second panel: error signal

Third panel: log magnitude spectra of signal and LP model

Fourth panel: log magnitude spectrum of error signal

31

The Prediction Error Signal 2



Top panel: speech signal

Second panel: error signal

Third panel: log magnitude spectra of signal and LP model

Fourth panel: log magnitude spectrum of error signal

33

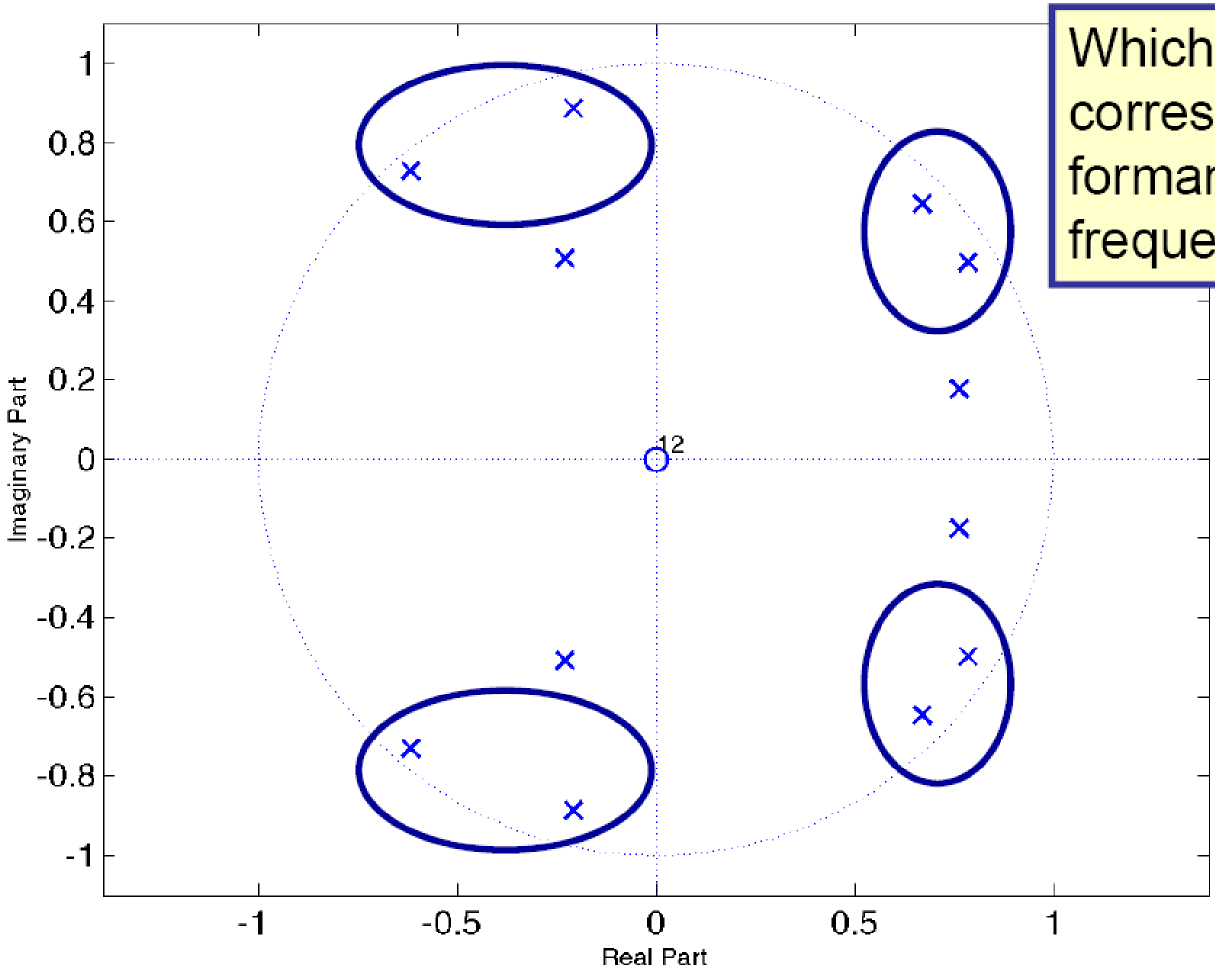
Formant Frequency Estimation

- How we can relate root locations, $\{z^k\}$, of the prediction error polynomial and formant frequencies?
  - Note that formant frequencies are roots on the unit circle
- Not all the roots can be assigned to formants, but some roots that are close to unit circle

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

$$= \frac{G}{\prod_{k=1}^{p}\left(1 - z_k z^{-1}\right)} = \frac{G z^{p}}{\prod_{k=1}^{p}\left(z - z_k\right)}$$

Pole Plots

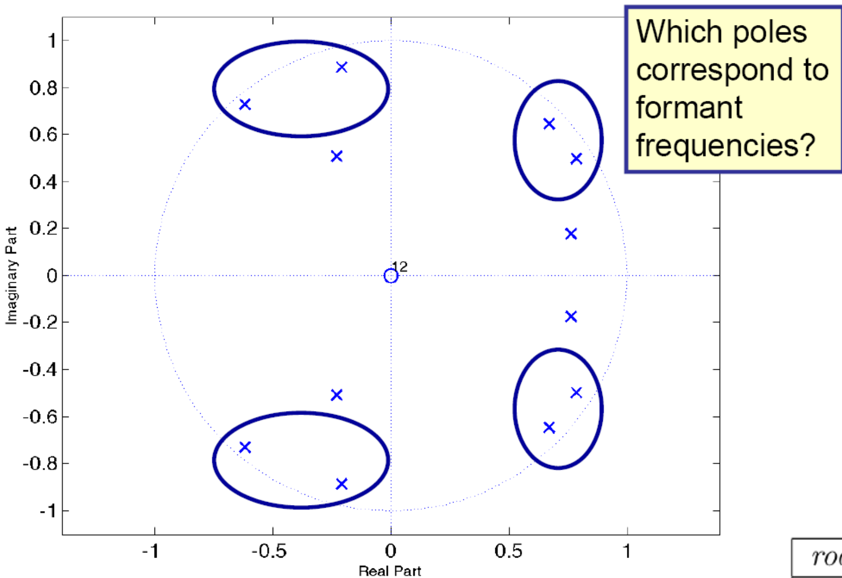


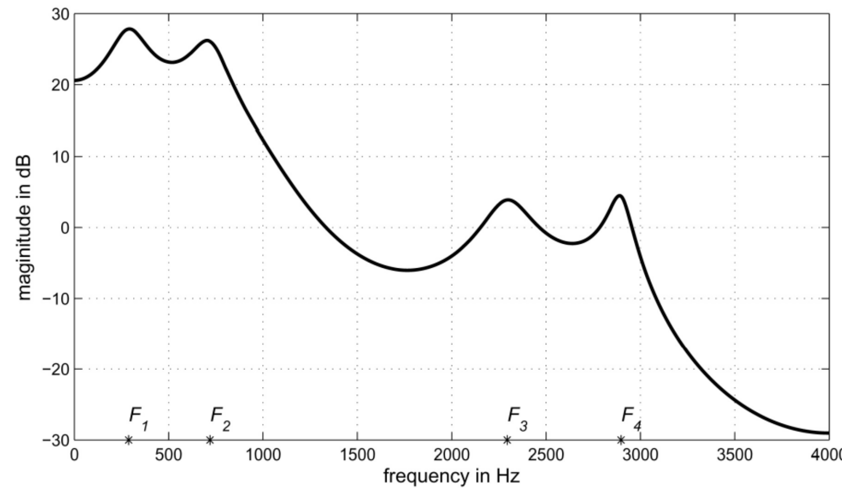
Which poles correspond to formant frequencies?

poles most likely to correspond to formants are the ones closest to the unit circle, since they produce resonance-like peaks when H(z) is evaluated on the unit circle

x: computed by roots(A) in MATLAB

Pole Locations



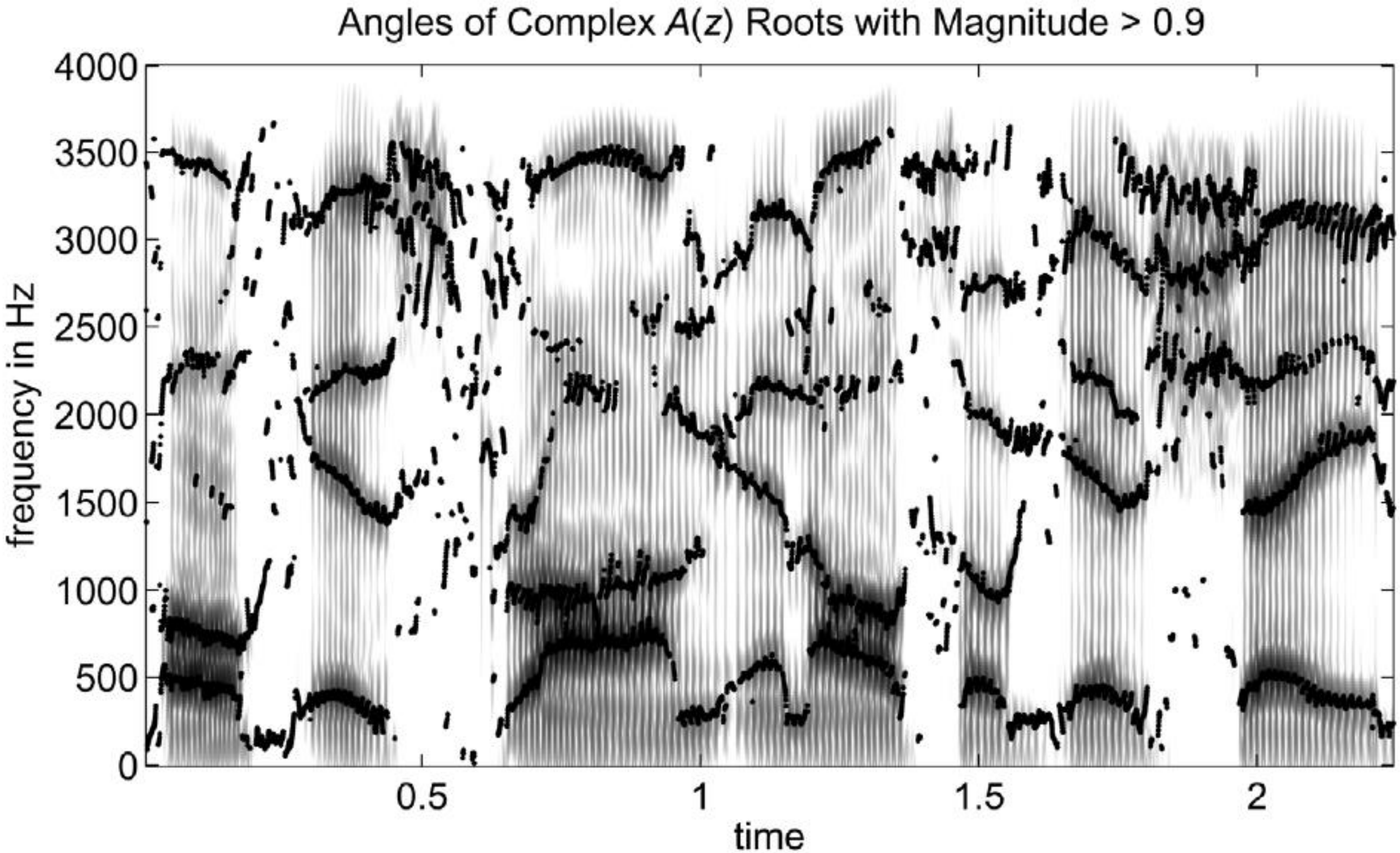


root magnitude	θ *root angle(degrees)*	F *root angle (Hz)*	*formant*
0.9308	10.36	288	F_1
0.9308	-10.36	-288	F_1
0.9317	25.88	719	F_2
0.9317	-25.88	-719	F_2
0.7837	35.13	976	
0.7837	-35.13	-976	
0.9109	82.58	2294	F_3
0.9109	-82.58	-2294	F_3
0.5579	91.44	2540	
0.5579	-91,44	-2540	
0.9571	104.29	2897	F_4
0.9571	-104.29	-2897	F_4

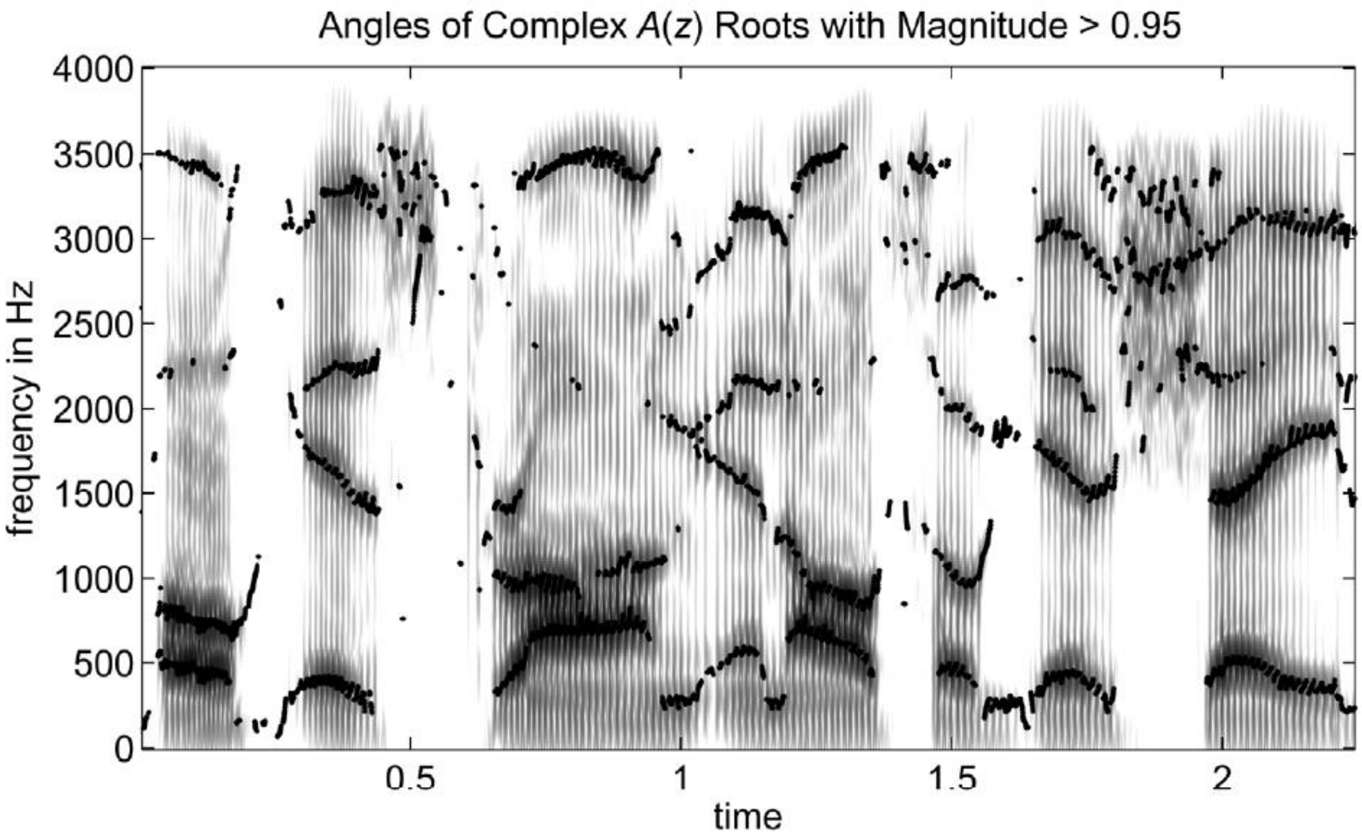
Estimating Formant Frequencies

- The assignment of $H(z)$ pole locations to formants should be based on two criteria:
  - **closeness of the complex pole to the unit circle**
    - compute A(z) and find roots that are close to the unit circle.
    - compute equivalent analog frequencies from the angles of the roots.
  - **temporal continuity of the pole locations**, since a valid formant frequency at a given time will be manifest over a range of times.
    - plot formant frequencies as a function of time.
    - find the stable formant streak.

Spectrogram with LPC Roots 1



Spectrogram with LPC Roots 2



Formant tracking:
-raise the magnitude threshold to an appropriate level
-remove isolated peak locations (defragmentation) by median filtering
-find a continuous streak and assign them as formant frequencies

LSP (line spectral pair)

LSF (line spectral frequency)

LINE SPECTRAL PAIR PARAMETERS = LINE SPECTRAL FREQUENCIES

Rational System Design

- $A(z)$ is an all-zero prediction filter with all zeros, z_k, inside the unit circle
- $\tilde{A}(z)$ is a reciprocal polynomial with inverse zeros, $1/z_k$.
- We can then create an all-pass rational system $F(z) = A(z)/\tilde{A}(z)$ with $|F(e^{j\omega})|=1$ for all ω.

$$A(z) = 1 - a_1 z^{-1} - a_2 z^{-2} - \ldots - a_p z^{-p}$$

$$\tilde{A}(z) = z^{-(p+1)} A(z^{-1}) = -a_p z^{-1} - \ldots - a_2 z^{-p+1} - a_1 z^{-p} + z^{-(p+1)}$$

$$F(z) = \frac{\tilde{A}(z)}{A(z)} = \frac{z^{-(p+1)} A(z^{-1})}{A(z)}$$

$$\left|F(e^{j\omega})\right| = 1, \quad \forall \omega$$

Line Spectral Pair

- A pair of a symmetric $P(z)$ and an anti-symmetric $Q(z)$ polynomials
- Characteristics of the roots
  - occurring on the unit circle in the z-plane, called ***line spectral frequencies*** (***LSFs***)
  - (phase) interlaced with each other
- Very useful in low bit rate coding

$$
\begin{aligned}
P(z) &= A(z) + \tilde{A}(z) \\
&= A(z) + z^{-(p+1)} A(z^{-1}) \\
Q(z) &= A(z) - \tilde{A}(z) \\
&= A(z) - z^{-(p+1)} A(z^{-1})
\end{aligned}
$$

$$
\begin{aligned}
P(z) = 0 &\Leftrightarrow A(z) = -\tilde{A}(z) \\
&\Leftrightarrow F(z) = -1, \arg\left\{F\left(e^{j\omega_k}\right)\right\} = \left(k + \tfrac{1}{2}\right)\cdot 2\pi \\
Q(z) = 0 &\Leftrightarrow A(z) = \tilde{A}(z) \\
&\Leftrightarrow F(z) = 1, \arg\left\{F\left(e^{j\omega_k}\right)\right\} = k \cdot 2\pi
\end{aligned}
$$

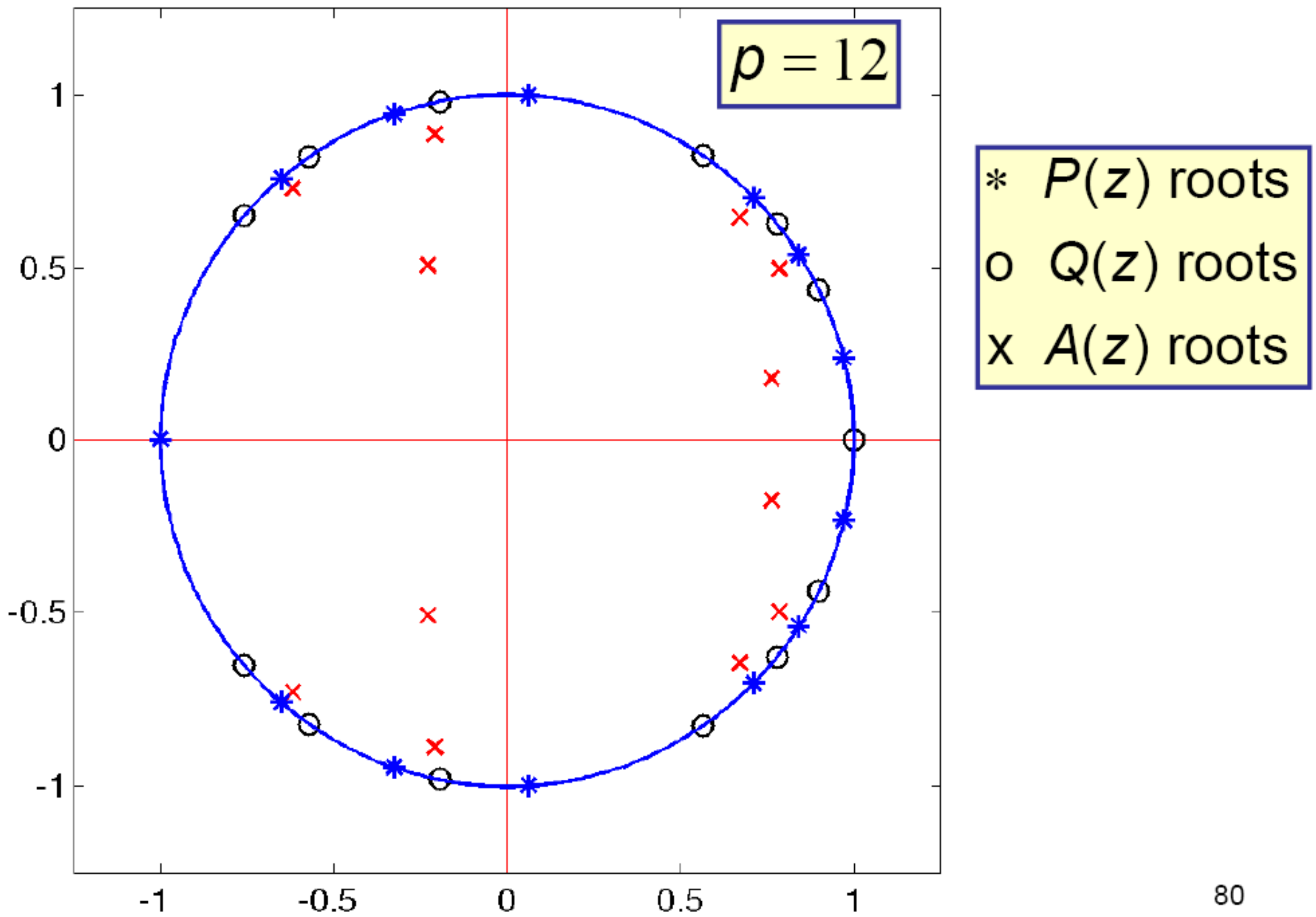
Properties of LSP parameters

- Stability of $H(z) = G/A(z)$ is guaranteed by quantizing LSF parameters
  - Errors in $P(z)$ and $Q(z)$ roots are proportional to the errors in $A(z)$ roots
- the LSP frequencies get close together when roots of $A(z)$ are close the unit circle, i.e., the roots of $P(z)$ are approximately equal to the formant frequencies

$$A(z) = \frac{P(z) + Q(z)}{2}$$

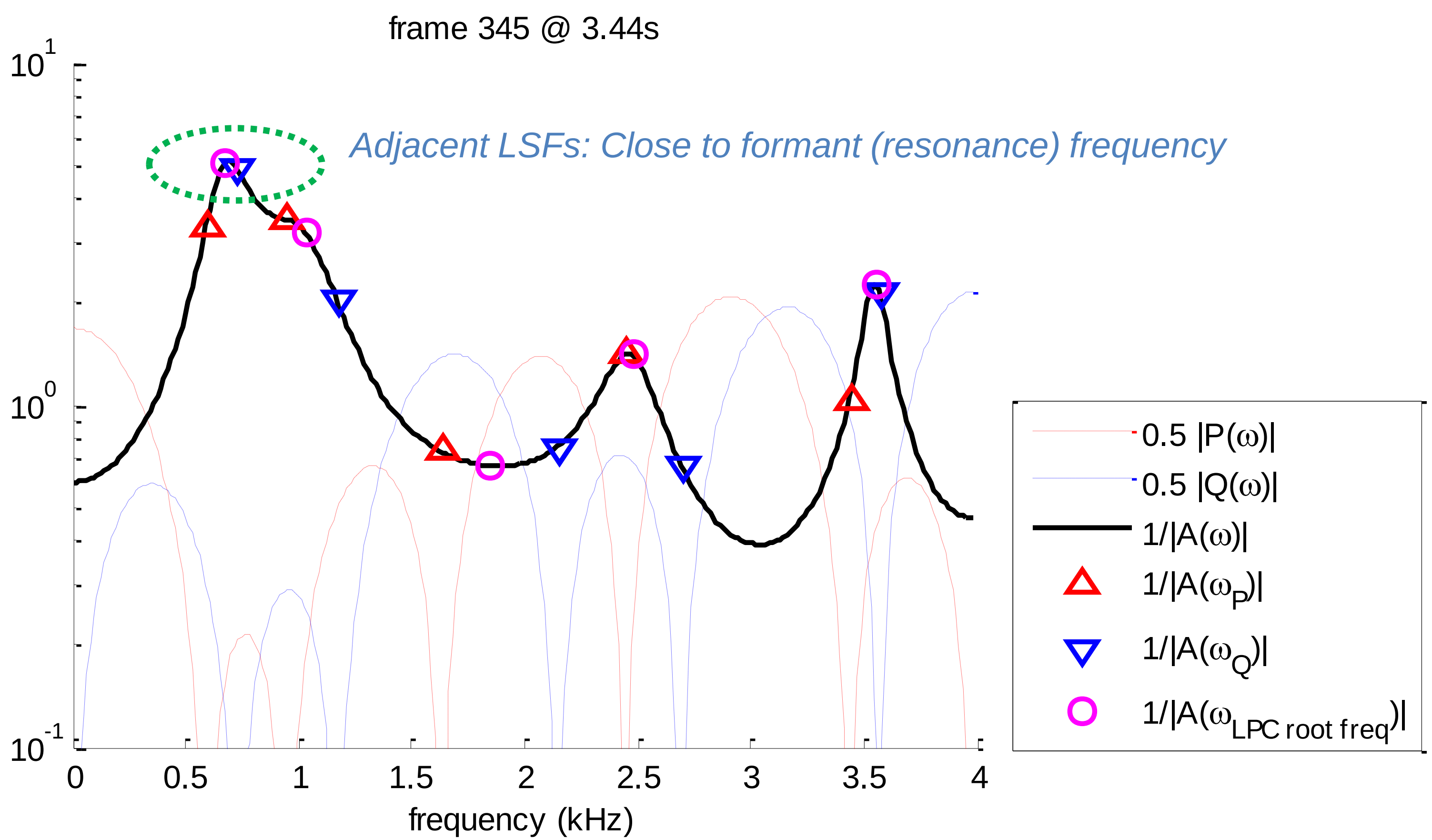
$$|A(z)| = \frac{|P(z)|^2 + |Q(z)|^2}{4}$$

LSP Example



80

Spectral Envelope at LSFs



Time Course of LSFs