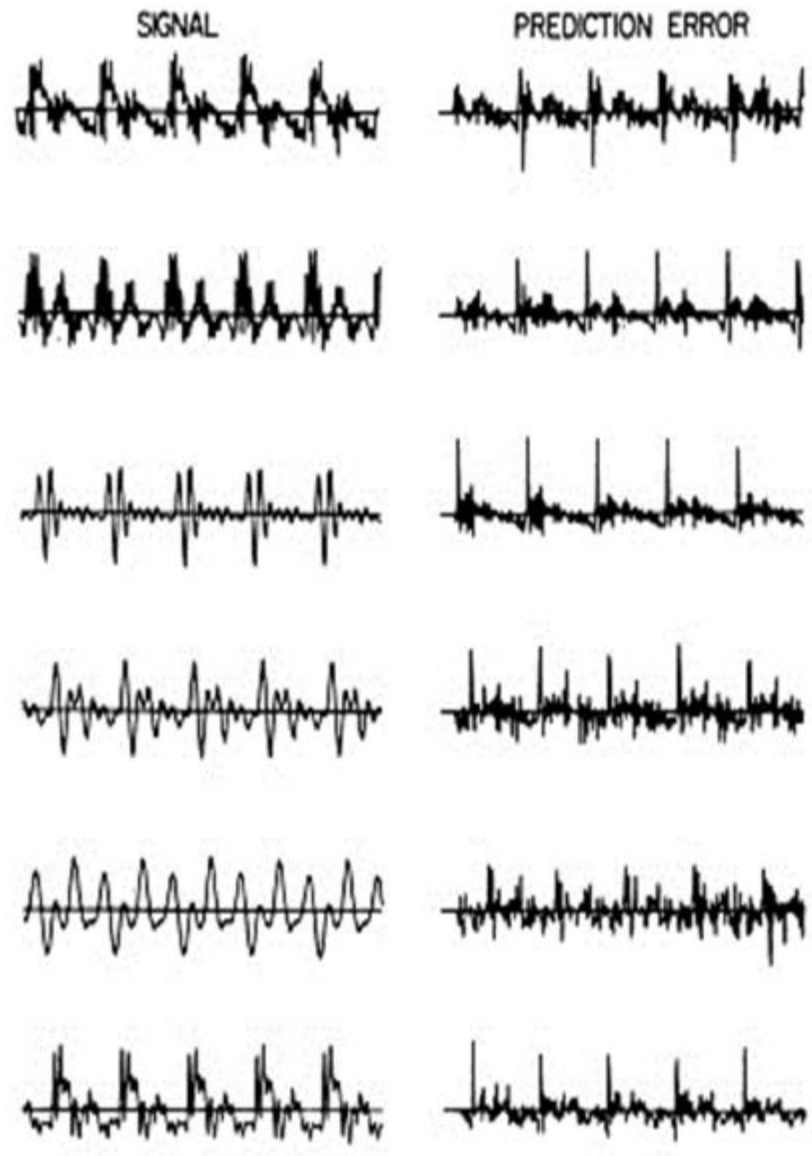
The Prediction Error Signal

Prediction Error Signal Behavior

- the prediction error signal is computed as

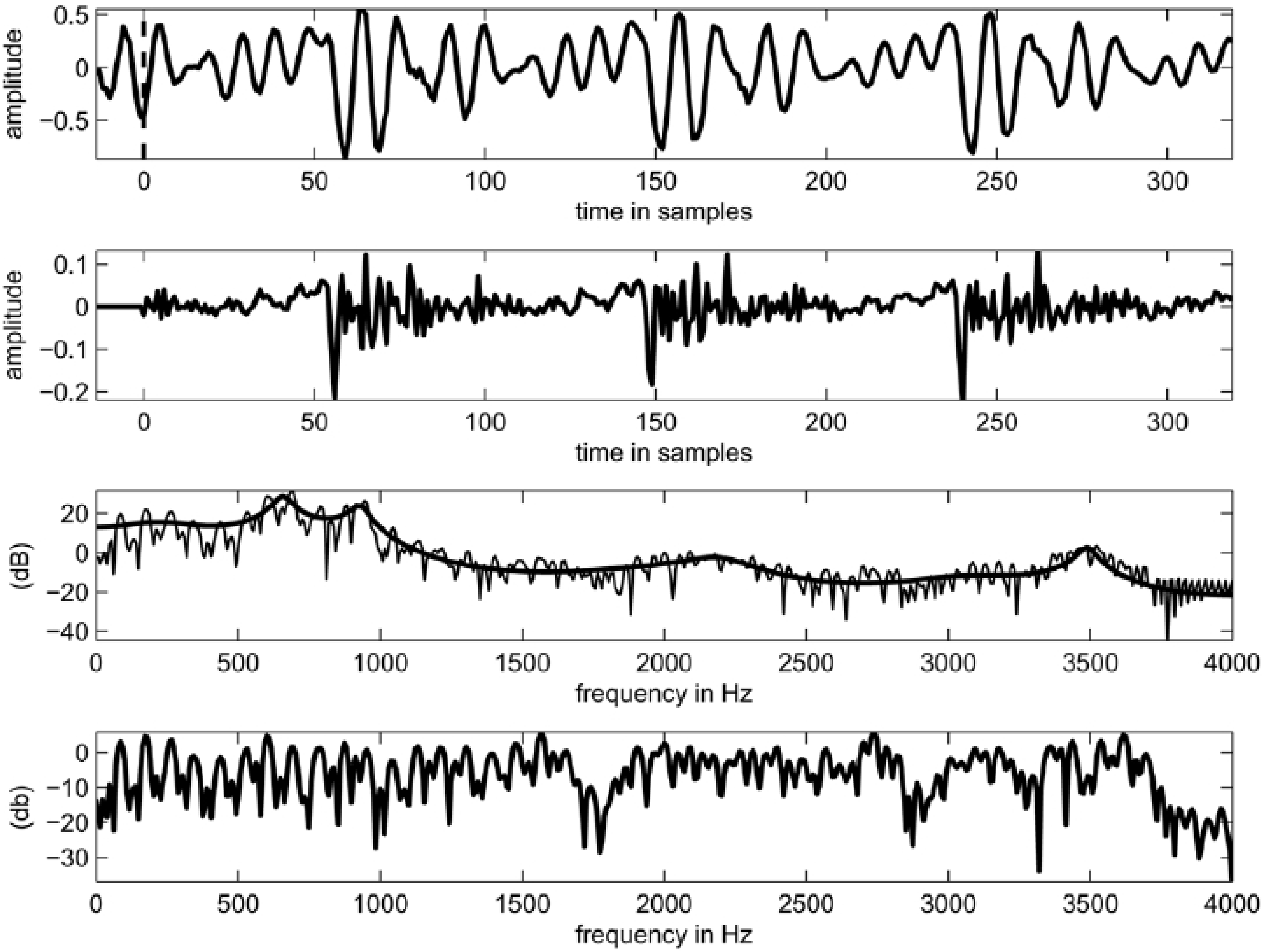
$$e(n) = s(n) - \sum_{k=1}^{p} \alpha_k s(n-k) = Gu(n)$$

- $e(n)$ should be large at the beginning of each pitch period (voiced speech) => good signal for pitch detection
- can perform autocorrelation on $e(n)$ and detect largest peak
- error spectrum is approximately flat-so effects of formants on pitch detection are minimized



LP Speech Analysis

file:s5, ss:11000, frame size (L):320, lpc order (p):14, cov method



Top panel:

speech signal

Second panel:

error signal

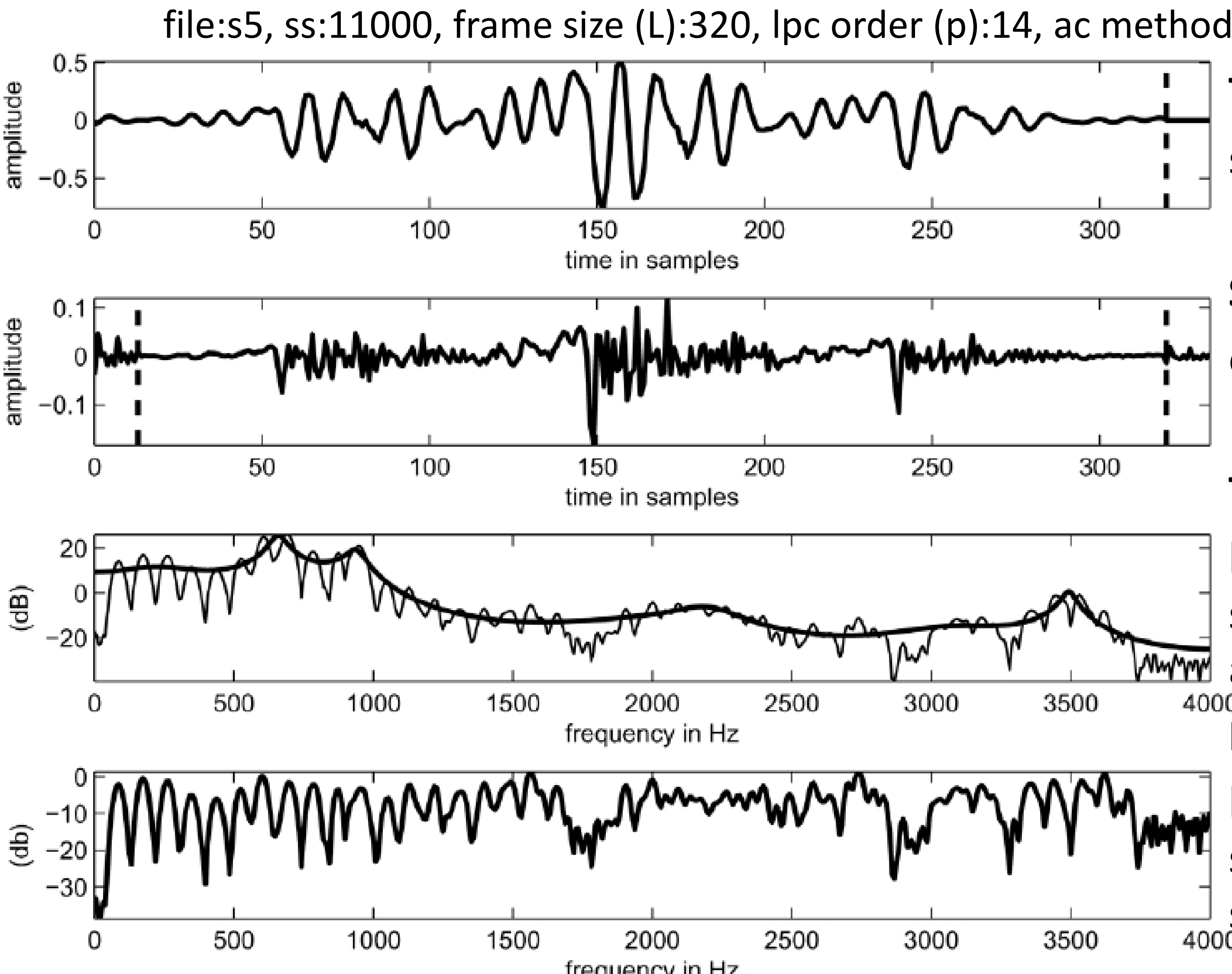
Third panel:

log magnitude spectra of signal and LP model

Fourth panel:

log magnitude spectrum of error signal

LP Speech Analysis



Top panel:

speech signal

Second panel:

error signal

Third panel:

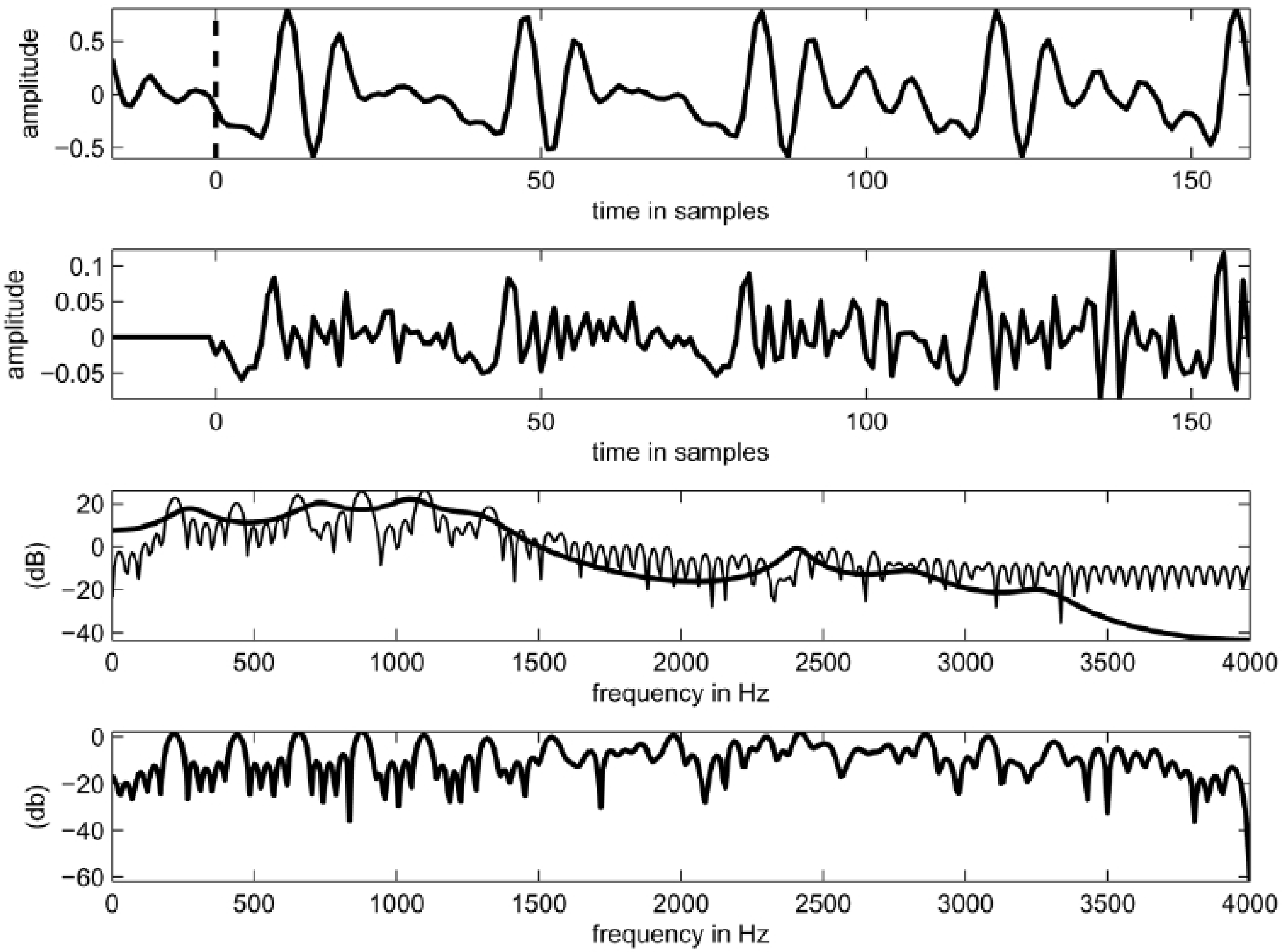
log magnitude spectra of signal and LP model

Fourth panel:

log magnitude spectrum of error signal

LP Speech Analysis

file:s3, ss:14000, frame size (L):160, lpc order (p):16, cov method



Top panel:

speech signal

Second panel:

error signal

Third panel:

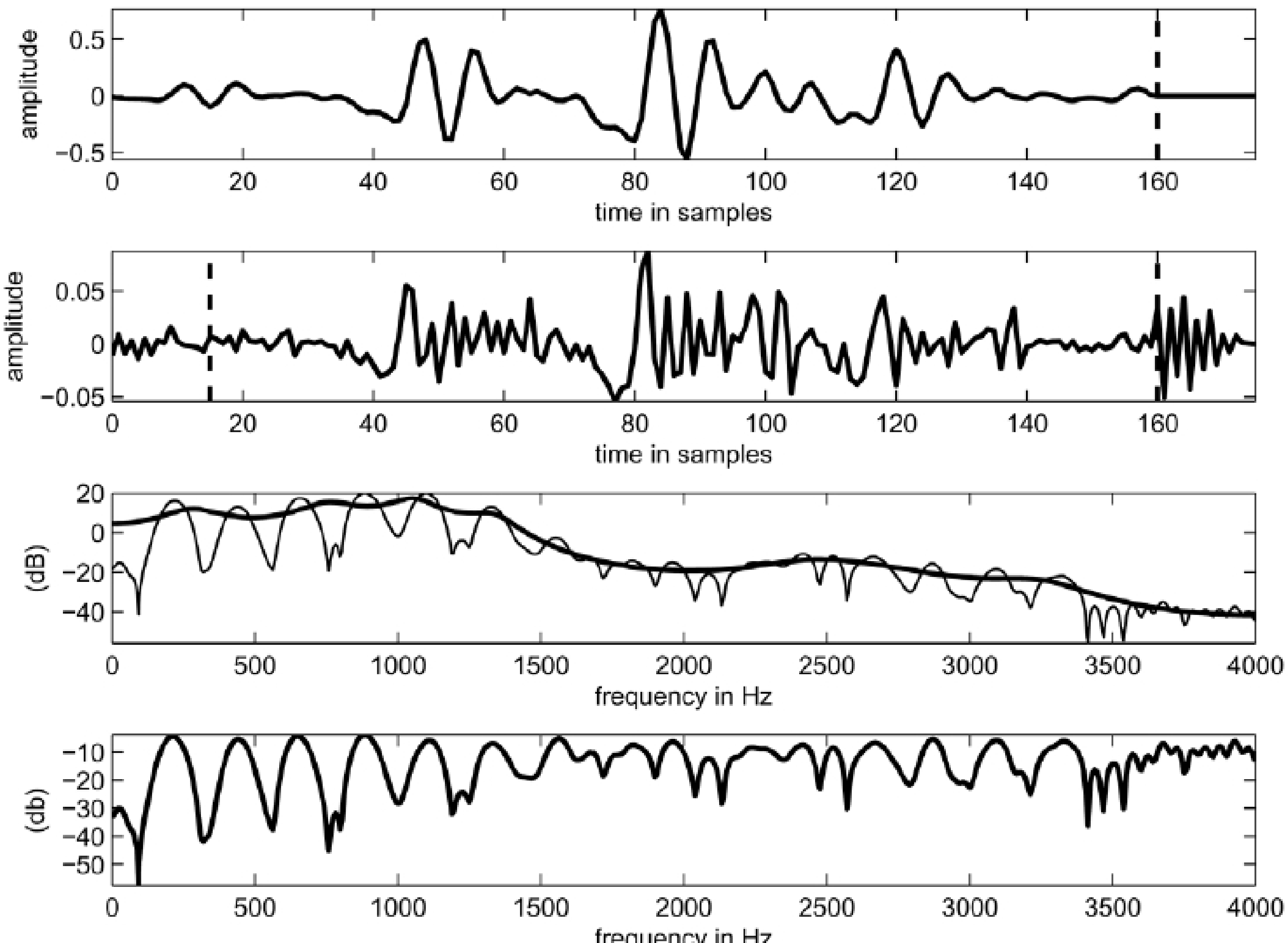
log magnitude spectra of signal and LP model

Fourth panel:

log magnitude spectrum of error signal

LP Speech Analysis

file:s3, ss:14000, frame size (L):160, lpc order (p):16, ac method



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

Properties of the LPC Polynomial

Minimum-Phase Property of $A(z)$

$A(z)$ has all its zeros inside the unit circle

Proof: Assume that z_o $(|z_o|^2 > 1)$ is a zero (root) of $A(z)$

$$A(z) = (1 - z_o z^{-1}) A'(z)$$

The minimum mean-squared error is

$$E_{\hat{n}} = \sum_{m=-\infty}^{\infty} e_{\hat{n}}[m]^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} | A(e^{j\omega}) |^2 | S_{\hat{n}}(e^{j\omega}) |^2 \, d\omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| 1 - z_o e^{-j\omega} \right|^2 \left| A'(e^{j\omega}) \right|^2 \left| S_{\hat{n}}(e^{j\omega}) \right|^2 d\omega > 0$$

$$\left| 1 - z_o e^{-j\omega} \right|^2 = \left| z_o \right|^2 \left| 1 - (1 / z_o^*) e^{-j\omega} \right|^2$$

Thus, $A(z)$ could not be the optimum filter because we could replace z_0 by $(1 / z_o^*)$ and decrease the error

PARCORs and Stability

- prove that $|k_i| \geq 1 \Rightarrow |z_j^{(i)}| \geq 1$ for some j

Proof: $A^{(i)}(z) = A^{(i-1)}(z) - k_i z^{-i} A^{(i-1)}(z^{-1}) = \prod_{j=1}^{i} (1 - z_j^{(i)} z^{-1})$

It is easily shown that $-k_i$ is the coefficient of z^{-i} in $A^{(i)}(z)$, i.e. $\alpha_i^{(i)} = k_i$.
Therefore

$$|k_i| = \prod_{j=1}^{p} z_j^{(i)}$$

If $|k_i| \geq 1$, then either all the roots must be **on** the unit circle or at least one of them must be **outside** the unit circle

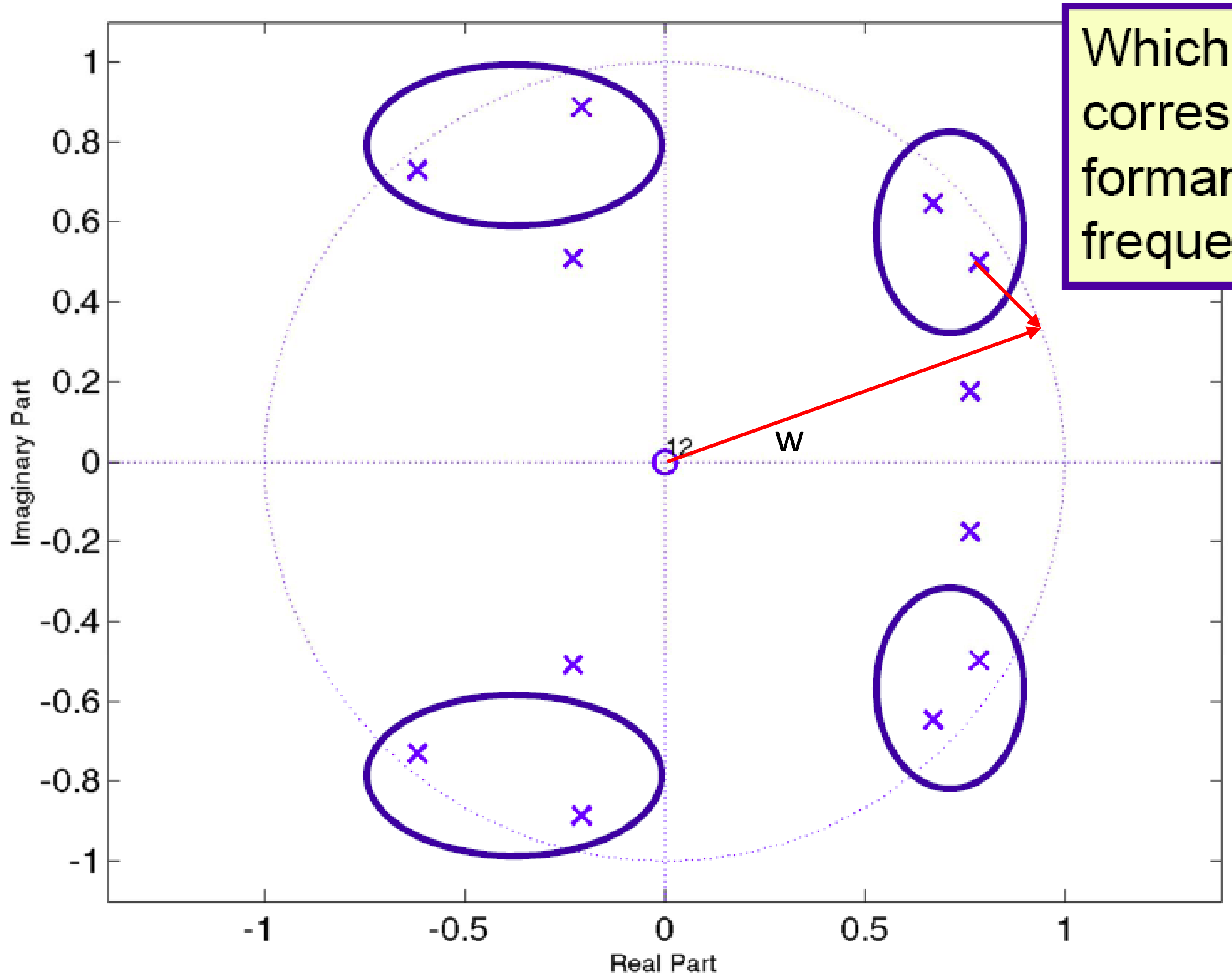
- $|k_i| < 1$ is a necessary and sufficient condition for A(z) to be a minimum phase system and 1/A(z) to be a stable system

Root Locations for Optimum LP Model

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

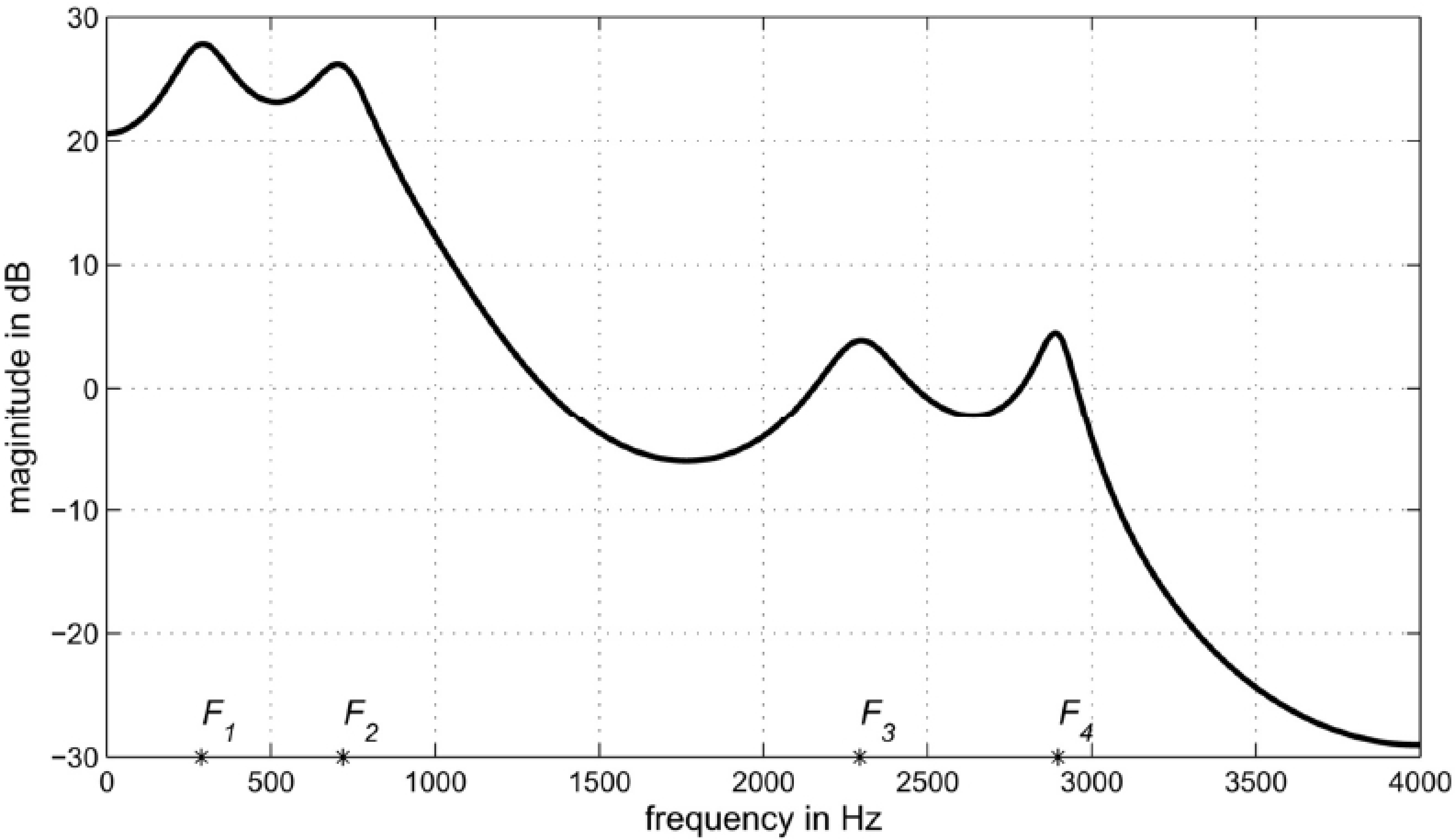
$$= \frac{G}{\prod_{i=1}^{p} (1 - z_i z^{-1})} = \frac{G z^p}{\prod_{i=1}^{p} (z - z_i)}$$

Pole-Zero Plot for Model



Which poles correspond to formant frequencies?

Pole Locations



Pole Locations (F_S=10,000 Hz)

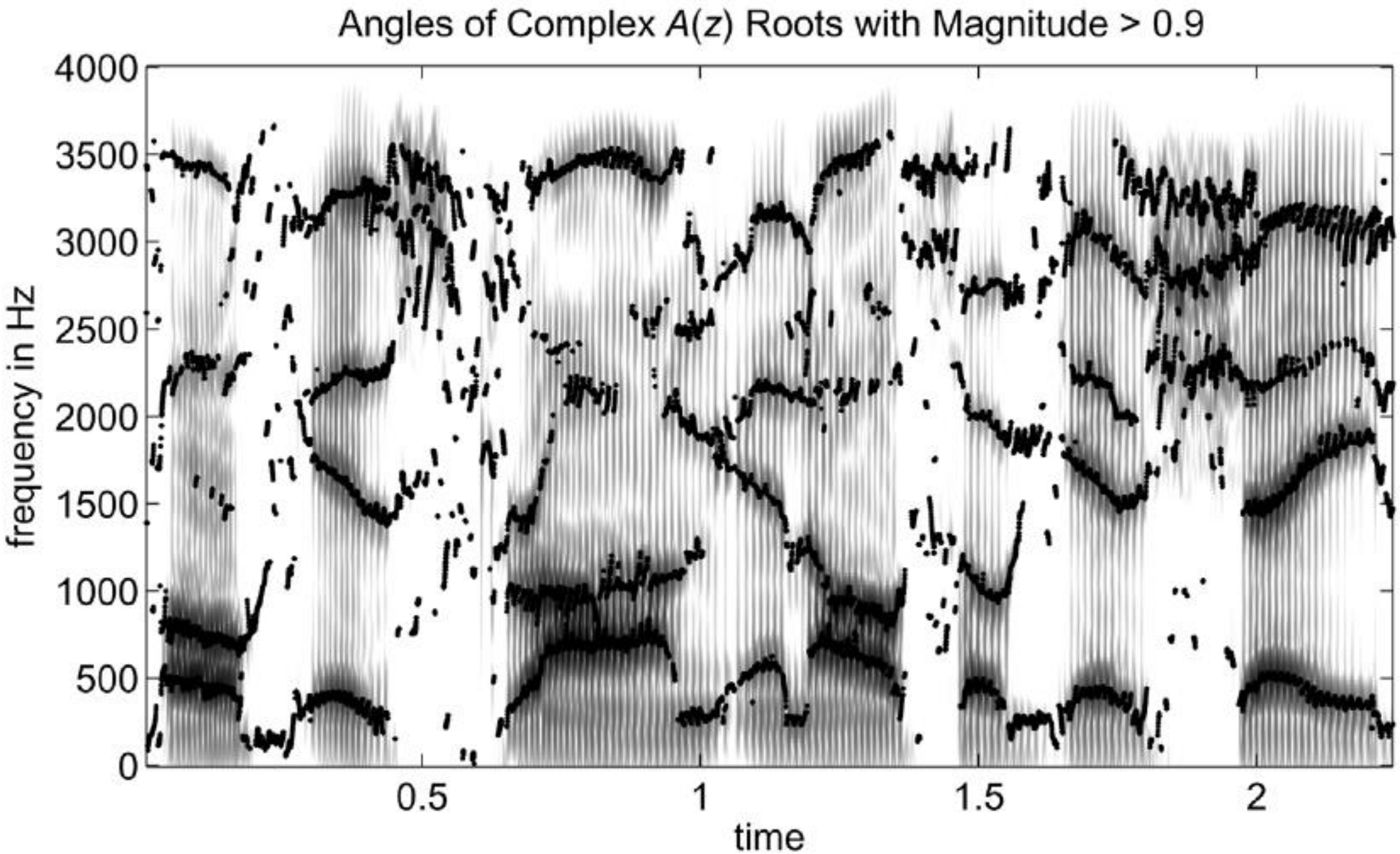
root magnitude	θ *root angle(degrees)*	*F root angle (Hz)*	*formant*
0.9308	10.36	288	F_1
0.9308	-10.36	-288	F_1
0.9317	25.88	719	F_2
0.9317	-25.88	-719	F_2
0.7837	35.13	976	
0.7837	-35.13	-976	
0.9109	82.58	2294	F_3
0.9109	-82.58	-2294	F_3
0.5579	91.44	2540	
0.5579	-91,44	-2540	
0.9571	104.29	2897	F_4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

$$F = (\theta / 180) \cdot (F_S / 2)$$

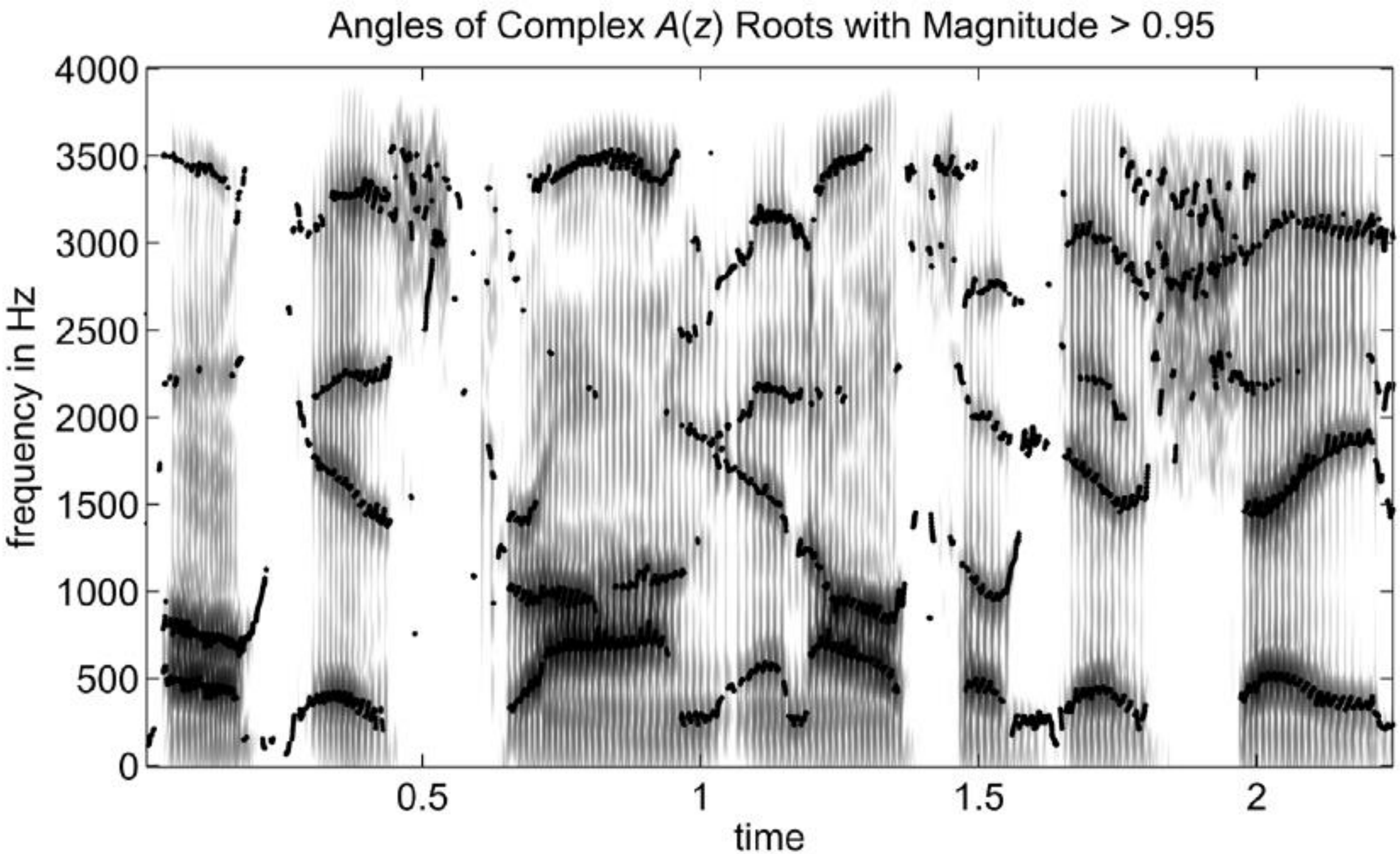
Estimating Formant Frequencies

- compute $A(z)$ and factor it
- find roots that are close to the unit circle.
- compute equivalent analog frequencies from the angles of the roots.
- plot formant frequencies as a function of time.

Spectrogram with LPC Roots



Spectrogram with LPC Roots



Alternative Representations of the LP Parameters

LP Parameter Sets

Parameter Set	*Representation*
LP Coefficients and Gain	$\{\alpha_k, 1 \leq k \leq p\}, G$
PARCOR Coefficients	$\{k_i, 1 \leq i \leq p\}$
Log Area Ratio Coefficients	$\{g_i, 1 \leq i \leq p\}$
Roots of Predictor Polynomial	$\{z_k, 1 \leq k \leq p\}$
Impulse Response of $H(z)$	$\{h[n], 0 \leq n \leq \infty\}$
LP Cepstrum	$\{\hat{h}[n], -\infty \leq n \leq \infty\}$
Autocorrelation of Impulse Response	$\{\tilde{R}(i), -\infty \leq i \leq \infty\}$
Autocorrelation of Predictor Polynomial	$\{R_a[i], -p \leq i \leq p\}$
Line Spectral Pair Parameters	$P(z), Q(z)$

PARCOR

- PARCORs to Prediction Coefficients
  - assume that k_i, i=1,2, …, p are given. Then we can skip the computation of k_i in the Levinson recursion.

$$
\begin{aligned}
&\text{for } i = 1,2,\ldots,p \\
&\qquad \alpha_i^{(i)} = k_i \\
&\qquad \text{if } i > 1, \text{ then for } j = 1,2,\ldots,i-1 \\
&\qquad\qquad \alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)} \\
&\qquad \text{end} \\
&\text{end} \\
&\alpha_j = \alpha_j^{(p)} \qquad j = 1,2,\ldots,p
\end{aligned}
$$

PARCOR

- Prediction Coefficients to PARCORs
  - assume that α_j, j=1,2, …, p are given. Then we can work backwards through the Levinson Recursion.

$$
\begin{aligned}
&\alpha_j^{(p)} = \alpha_j \quad \text{for } j = 1,2,\ldots,p \\
&k_p = \alpha_p^{(p)} \\
&\text{for } i = p, p-1, \ldots, 2 \\
&\qquad \text{for } j = 1,2,\ldots,i-1 \\
&\qquad\qquad \alpha_j^{(i-1)} = \frac{\alpha_j^{(i)} + k_i \alpha_{i-j}^{(i)}}{1-k_i^2} \\
&\qquad \text{end} \\
&\qquad k_{i-1} = \alpha_{i-1}^{(i-1)} \\
&\text{end}
\end{aligned}
$$

Log Area Ratio

- log area ratio coefficients from PARCOR coefficients

$$g_i = \log\left[\frac{A_{i+1}}{A_i}\right] = \log\left[\frac{1-k_i}{1+k_i}\right] \quad 1 \le i \le p$$

with inverse relation

$$k_i = \frac{1-e^{g_i}}{1+e^{g_i}} \qquad 1 \le i \le p$$

Roots of Predictor Polynomial

- roots of the predictor polynomial

$$A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k} = \prod_{k=1}^{p} \left(1 - z_k z^{-1}\right)$$

where each root can be expressed as a z-plane i.e.,

$$z_k = z_{kr} + j\, z_{ki}$$

- important for formant estimation

Impulse Response of H(z)

- IR of all pole system

$$h(n) = \sum_{k=1}^{p} \alpha_k h(n-k) + G\delta(n) \qquad 0 \le n$$

LP Cepstrum

- cepstrum of IR of overall LP system from predictor coefficients

$$\hat{h}(n) = \alpha_n + \sum_{k=1}^{n-1} \left(\frac{k}{n} \right) \hat{h}(k) \alpha_{n-k} \qquad 1 \le n$$

- predictor coefficients from cepstrum of IR

$$\alpha_n = \hat{h}(n) - \sum_{k=1}^{n-1} \left(\frac{k}{n} \right) \hat{h}(k) \alpha_{n-k} \qquad 1 \le n$$

where

$$H(z) = \sum_{n=0}^{\infty} h(n) z^{-n} = \frac{G}{1 - \sum_{k=1}^{p} \alpha_k z^{-k}}$$

Autocorrelation of IR

- autocorrelation of IR

$$\tilde{R}(i) = \sum_{n=0}^{\infty} h(n)h(n-i) = \tilde{R}(-i)$$

$$\tilde{R}(i) = \sum_{k=1}^{p} \alpha_k \tilde{R}(|i-k|) \qquad 1 \le i$$

$$\tilde{R}(0) = \sum_{k=1}^{p} \alpha_k \tilde{R}(k) + G^2$$

Autocorrelation of Predictor Polynomial

- autocorrelation of the predictor polynomial

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

with IR of the inverse filter

$$a(n) = \delta(n) - \sum_{k=1}^{p} \alpha_k \delta(n-k)$$

with autocorrelation

$$R_a(i) = \sum_{k=0}^{p-i} a(k)a(k+i) \qquad 0 \le i \le p$$

Line Spectral Pairs

- Quantization of LP Parameters
- consider the magnitude-squared of the model frequency response

$$\left|H(e^{j\omega})\right|^2 = \frac{1}{|A(e^{j\omega})|^2} = P(\omega, g)$$

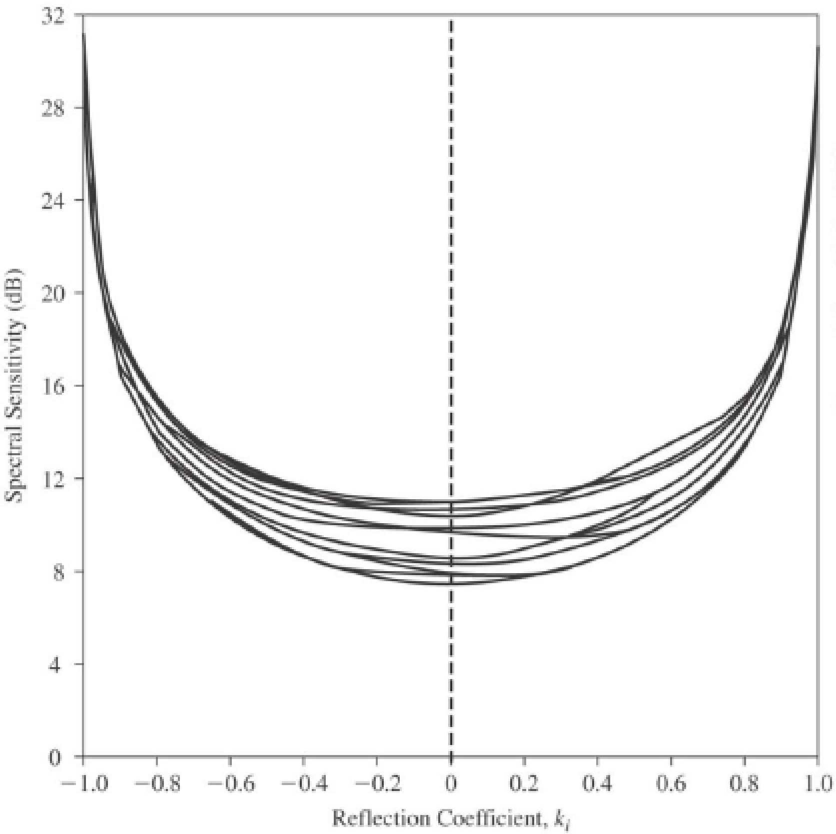
where g is a parameter that affects P.

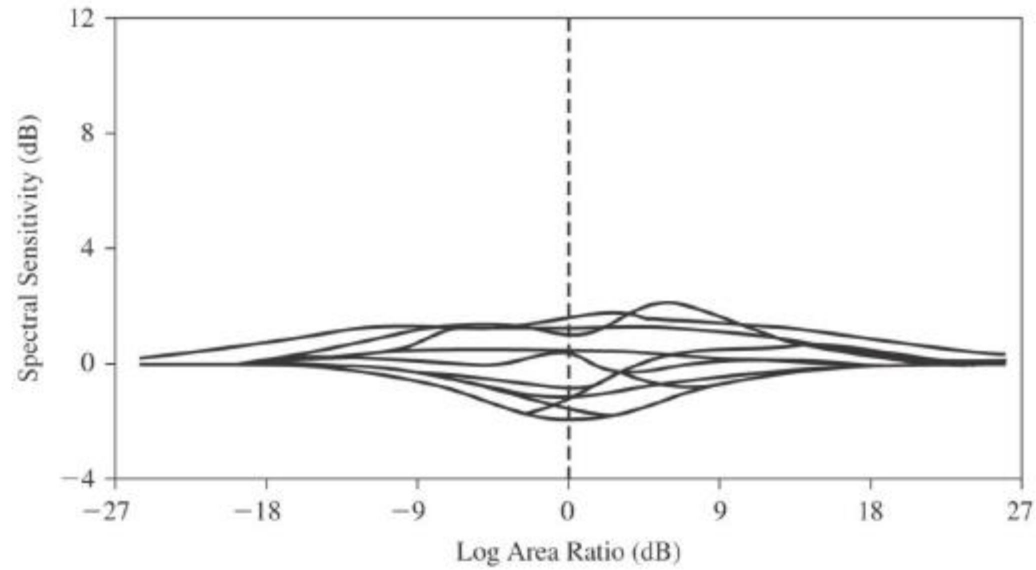
- spectral sensitivity can be defined as

$$\frac{\partial S}{\partial g_i} = \lim_{\Delta g_i \to 0} \left| \frac{1}{\Delta g_i} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \log \frac{P(\omega, g_i)}{P(\omega, g_i + \Delta g_i)} \right| d\omega \right] \right|$$

which measures sensitivity to errors in the g_i parameters

Line Spectral Pairs





spectral sensitivity for k_i parameters; low sensitivity around 0; high sensitivity around 1

spectral sensitivity for log area ratio parameters, g_i – low sensitivity for virtually entire range is seen

Line Spectral Pairs

$$A(z) = 1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \ldots + \alpha_p z^{-p}$$

$=$ all-zero prediction filter with all zeros, z_k, inside the unit circle

$$\tilde{A}(z) = z^{-(p+1)} A(z^{-1}) = \alpha_p z^{-1} + \ldots + \alpha_2 z^{-p+1} + \alpha_1 z^{-p} + z^{-(p+1)}$$

$=$ reciprocal polynomial with inverse zeros, $1/z_k$

- Consider the following

$$L(z) = \frac{\tilde{A}(z)}{A(z)} = \text{allpass system} \Rightarrow \left|L(e^{j\omega})\right| = 1, \text{ all } \omega$$

- Form the symmetric polynomial $P(z)$ as

$$P(z) = A(z) + \tilde{A}(z) = A(z) + z^{-(p+1)} A(z^{-1}) \Rightarrow P(z) \text{ has zeros for } L(z) = -1; \ (A(z) = -\tilde{A}(z))$$

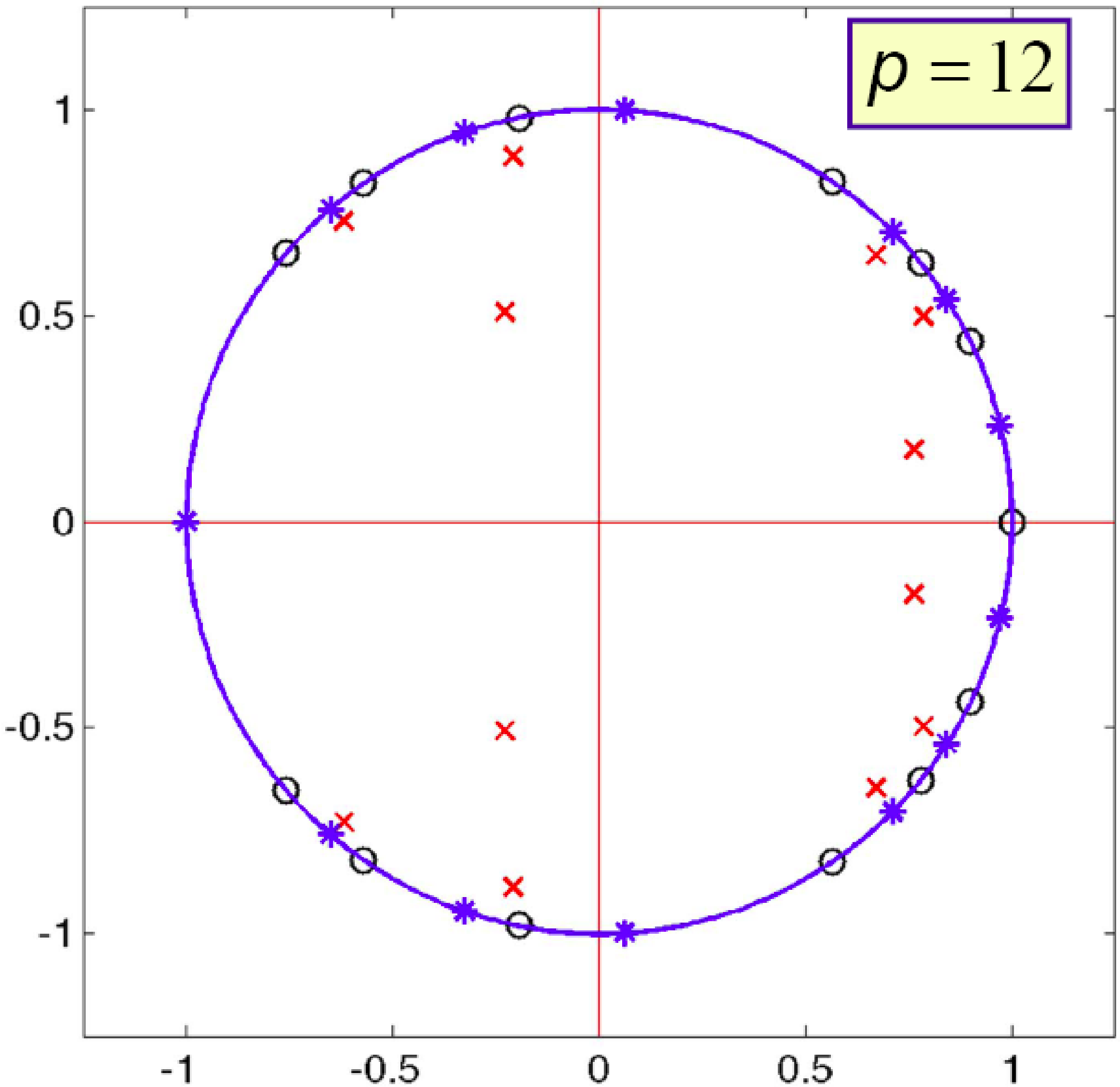
$$\Rightarrow \arg\left\{L(e^{j\omega_k})\right\} = (k + 1/2) \cdot 2\pi, \ k = 0, 1, \ldots, p-1$$

- Form the anti-symmetric polynomial $Q(z)$ as

$$Q(z) = A(z) - \tilde{A}(z) = A(z) - z^{-(p+1)} A(z^{-1}) \Rightarrow Q(z) \text{ has zeros for } L(z) = +1; \ (A(z) = \tilde{A}(z))$$

$$\Rightarrow \arg\left\{L(e^{j\omega_k})\right\} = k \cdot 2\pi, \ k = 0, 1, \ldots, p-1$$

LSP Example



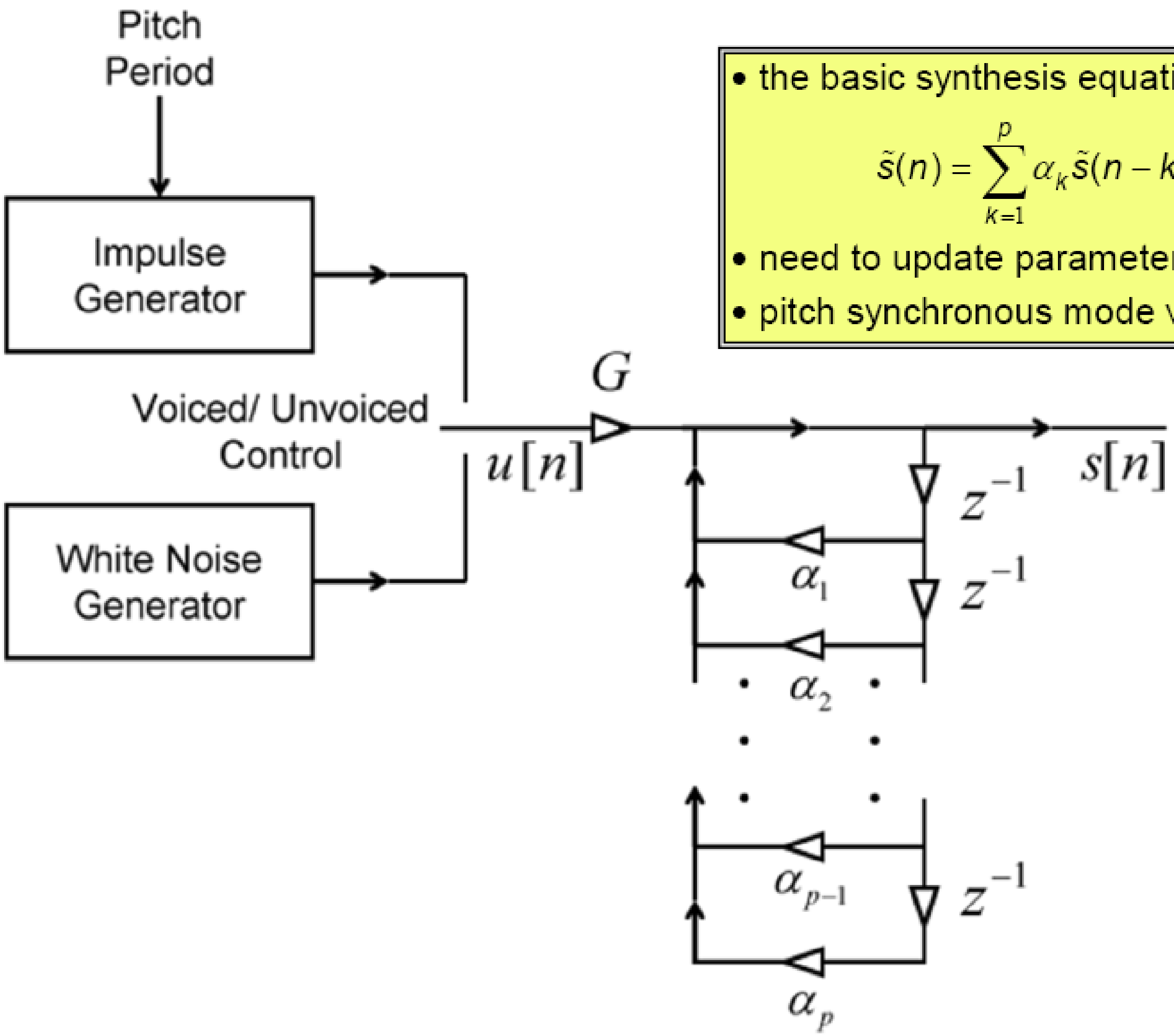
$*$ $P(z)$ roots
o $Q(z)$ roots
x $A(z)$ roots

Line Spectral Pairs

- properties of LSP parameters
  1. all the roots of $P(z)$ and $Q(z)$ are on the unit circle
  2. a necessary and sufficient condition for $|k_i| < 1$, $i = 1, 2, \ldots, p$ is that the roots of $P(z)$ and $Q(z)$ alternate on the unit circle
  3. the LSP frequencies get close together when roots of $A(z)$ are close to the unit circle

Applications

Speech Synthesis



- the basic synthesis equation is

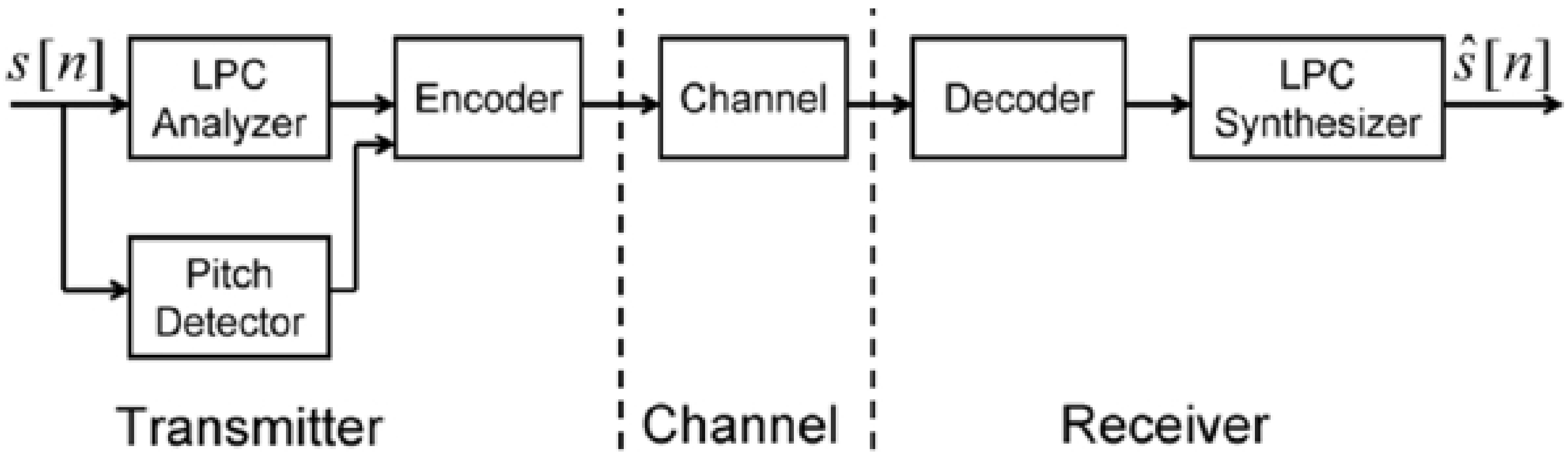
$$\tilde{s}(n) = \sum_{k=1}^{p} \alpha_k \tilde{s}(n-k) + Gu(n)$$

- need to update parameters every 10 msec or so
- pitch synchronous mode works best

Speech Coding

1. Extract α_k parameters properly
2. Quantize α_k parameters properly so that there is little quantization error
   – Small number of bits go into coding the α_k coefficients
3. Represent $e(n)$ via:
   – Pitch pulses and noise—LPC Coding
   – Multiple pulses per 10 msec interval—MPLPC Coding
   – Codebook vectors—CELP
      • Almost all of the coding bits go into coding of $e(n)$

LPC Vocoder



$s[n]$
LPC Analyzer
Pitch Detector
Encoder
Channel
Decoder
LPC Synthesizer
$\hat{s}[n]$
Transmitter
Channel
Receiver