Department of Electrical and Computer Engineering
Digital Speech Processing
Homework No. 7

Problem 1

Linear prediction analysis is used to obtain an eleventh-order all-pole model for a segment of voiced speech that was sampled at a rate of $F_S = 8000$ samples/second. The system function of the model is:

$$H(z) = \frac{G}{A(z)} = \frac{G}{1 - \sum_{k=1}^{11} \alpha_k z^{-k}} = \frac{G}{\prod_{i=1}^{11}(1 - z_i z^{-1})}$$

Table 1 shows five of the roots of the eleventh-order prediction error filter, $A(z)$.

i	$\lvert z_i \rvert$	$\angle z_i$ (rad)
1	0.2567	2.0677
2	0.9681	1.4402
3	0.9850	0.2750
4	0.8647	2.0036
5	0.9590	2.4162

Table 1: Root locations of eleventh-order prediction error filter in the z-plane.

(a) Determine where the other six zeros of $A(z)$ are located in the z-plane. If you cannot precisely determine the pole locations, explain where the pole might occur in the z-plane.
(b) Estimate the first three formant frequencies (in Hz) for this segment of speech.
(c) Which of the first three formant resonances has the smallest bandwidth? How is this determined?
(d) Plot and label the frequency response due to just the first three formants of the all-pole model for the (analog) frequency range $0 \leq f \leq F_S/2$.

Problem 2

An unvoiced speech signal segment can be modeled as a segment of a stationary random process of the form:

$$x[n] = w[n] + \beta w[n-1]$$

where $w[n]$ is a zero mean, unit variance, stationary white noise process and $|\beta| < 1$.
(a) What are the mean and variance of $x[n]$?
(b) What system can be used to **recover** $w[n]$ from $x[n]$?

(c) What is the normalized autocorrelation of $x[n]$ at a delay of 1 sample, i.e., what is $r_x[1] = \frac{R_x[1]}{R_x[0]}$?

Problem 3

A causal LTI system has system function:

$$H(z) = \frac{1 - 4z^{-1}}{1 - 0.25z^{-1} - 0.75z^{-2} - 0.875z^{-3}}$$

(a) Use the Levinson recursion to determine whether or not the system is stable.
(b) Is the system minimum phase?

Problem 4

A speech signal frame (windowed using a Hamming window) has energy:

$$E_n^{(0)} = \sum_m s_n^2[m] = 2000$$

Using the autocorrelation method of analysis on this speech frame, the first 3 PARCOR coefficients are computed and their values are:

$$\begin{aligned} k_1 &= -0.5 \\ k_2 &= 0.5 \\ k_3 &= 0.2 \end{aligned}$$

Find the energy of the linear prediction residual, $E_n^3 = \sum_m e_n^2[m]$ that would be obtained by inverse filtering $s_n[m]$ by the optimal third order predictor inverse filter, $A_3(z)$.

Problem 5

Write a MATLAB program to convert from a frame of speech to a set of linear prediction coefficients, using all 3 methods discussed in class, i.e., the Autocorrelation Method, the Covariance Method, and the Lattice Filter Method. Choose a section of a steady state vowel, and a section of unvoiced speech, and plot LPC spectra from the 3 methods along with the normal spectrum from the Hamming window weighted frame. Use $N = 300$, $p = 12$, with Hamming Window weighting for the autocorrelation method. Use the same parameters for the Covariance and Lattice Methods. Use the files ah.wav to get a vowel steady state sound beginning at sample 3000, and the file test_16k.wav to get a fricative beginning at sample 3000. (Dont forget that for the covariance and lattice methods, you also need to preserve p samples before the starting sample at $n = 3000$ for computing correlations, and error signals).