pee

ECE 251A

Homework8: Speech Processing

Name: XIAOYE LIU

PID: A53217030

1. Objective

In this homework, we will analyze spectrum of different phonemes, and use autocorrelation method of linear prediction to model the formant structure of vowels and nasal sound.

2. Approach

In part A, for each of three phonemes, 256-point time series is plotted first. Then the corresponding 256-point FFT is computed and spectrum is plotted. In addition, 64-point of the time series block us taken into consideration. FFT is computed for 64-point case and spectrum is plotted.

In Part B, error power for inverse filter orders p = 2, 4, 6, 8, 10, 12, and 14 is determined and Ep vs. p is plotted. Then, for just order p = 14, zero fill the filter coefficient sequence out to 255, compute the N = 256-point FFT, and plot power spectrum of inverse filter coefficients,

3. Result

Part A Conventional Spectral Analysis (3 plots/ phoneme)

Figure 1-Figure 3 are the Conventional Spectral Analysis of phoneme /i/. Figure 1 plots 256-point (25.6 msc) time series of phoneme /i/. From the plot we can find pitch period of /i/ is around 90/fs*1000 = 9ms. Since voiced speech is periodic in nature, we expect some fundamental frequency and its harmonics in the spectrum of speech. Figure 2 shows normalized magnitude spectrum of /i/. As it can be observed in the spectrum, frequency components repeating at regular intervals indicating the presence of harmonic structure. In the frequency domain, the presence of this harmonic structure is the main distinguishing factor for voiced speech. Figure 3 shows normalized spectrum of 64-point (6.4 msc) /i/. Since 6.4 ms is smaller than one pitch period of /i/, which is 9ms, there is no harmonic structure shown in the spectrum. Figure 4 – Figure 6 are 3 plots of /u/ using conventional spectral analysis. Similar to /i/, there is harmonic structure in 256-point case, but not in 64-point case because pitch period is larger than 100 samples. Figure 7 – Figure 9 are 3 plots of /ng/ using conventional spectral analysis. Energy of nasal is lower than for vowels – in part because nasal membranes absorb the sound. Similarly, there is harmonic structure in 256-point case, but not in 64-point case

because pitch period is around 100 samples.

Part B Autocorrelation Method of Linear Prediction

Figure 10 shows spectrum of inverse filter coefficients for /i/. The first three formant peaks are 235Hz, 2188Hz and 3164Hz, which are close to the average values reported in the literature (f = 270, 2290, and 3010). In Figure 12, locations of three formant frequencies in 64-point case are similar to that in 256-point case. The error power of 64-point case in Figure 13 is larger that of 256-point case in Figure 11 when prediction order is larger than 2, due to not accurate as 256-point case. Figure 14 and Figure 16 respectively shows 256-point and 64-point spectrum of inverse filter coefficients of /u/. In Figure 14, the first two formant peaks are 312Hz and 860 Hz, which are very close to the average values in the literature (f = 300, 870). However, the third formant peak in Figure 14 is around 3203Hz, which is far from that is reported in the literature (f = 2240). The error power of 256-point case in Figure 15 is larger than that of 64-point case in Figure 17 when order is larger than 2. Figure 18 shows spectrum of inverse filter coefficient for /ng/. The velar nasal ‘ng’ is formed by an oral closure accompanied by an open nasal passage. The first three formant peaks are respectively 273Hz, 2461Hz and 2733 Hz. Since the locations of poles when using 64-point time series are not accurate, error power of it is larger than that f 256-point time series.

4. Summary

In this homework, I find different phonemes have different formant structures. More points take into consideration in linear prediction coefficients not necessarily means less error power.

5. Plot

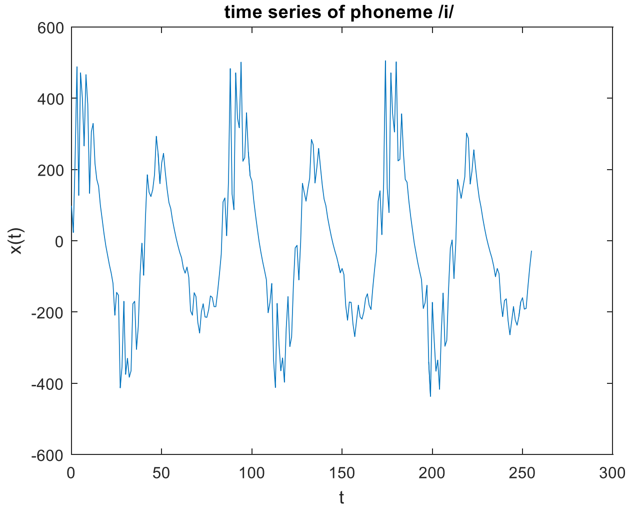

Figure 1

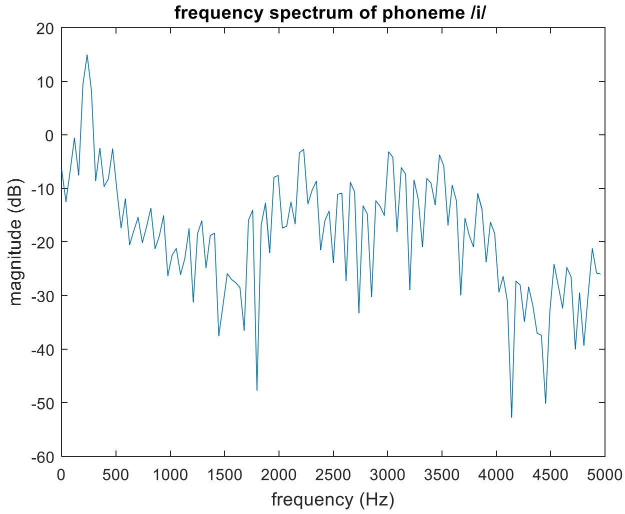

Figure 2

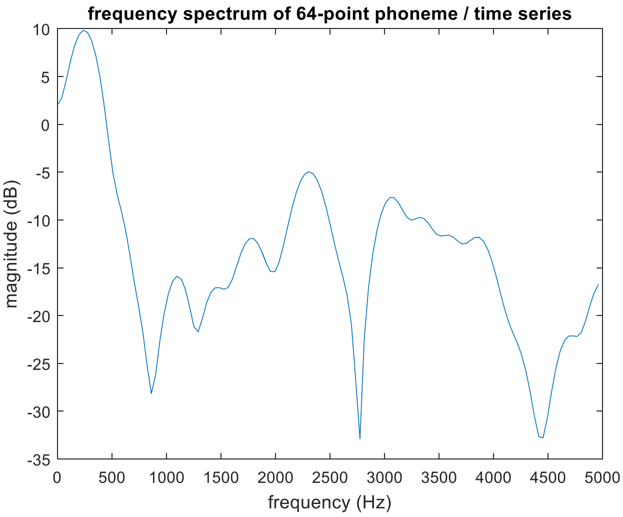

Figure 3

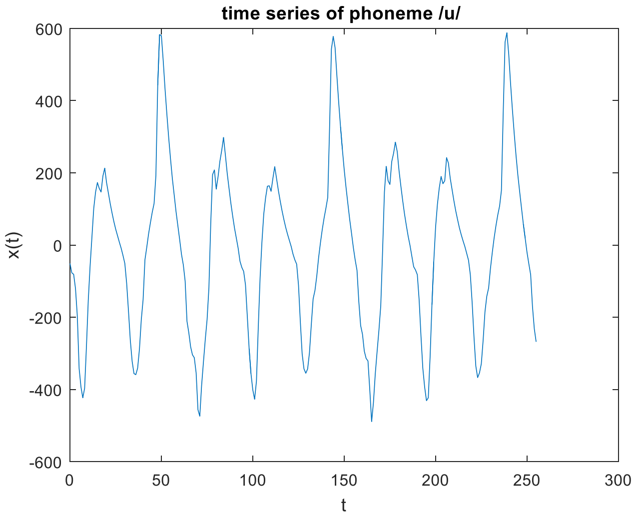

Figure 4

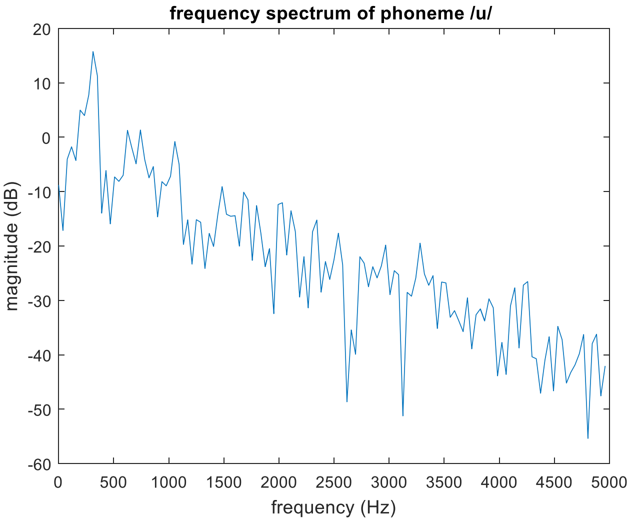

Figure 5

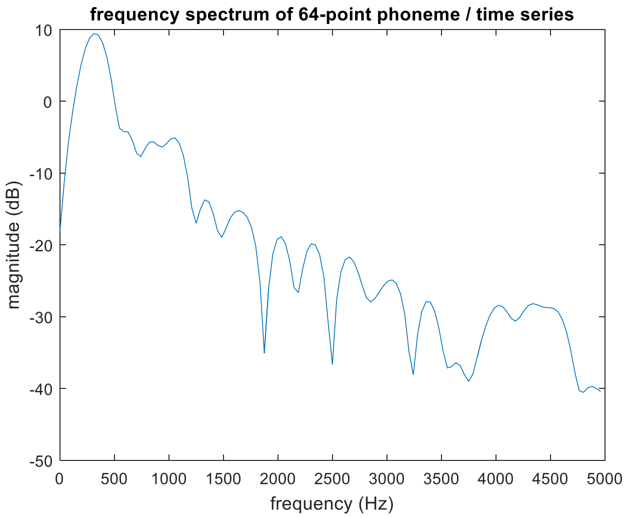

Figure 6

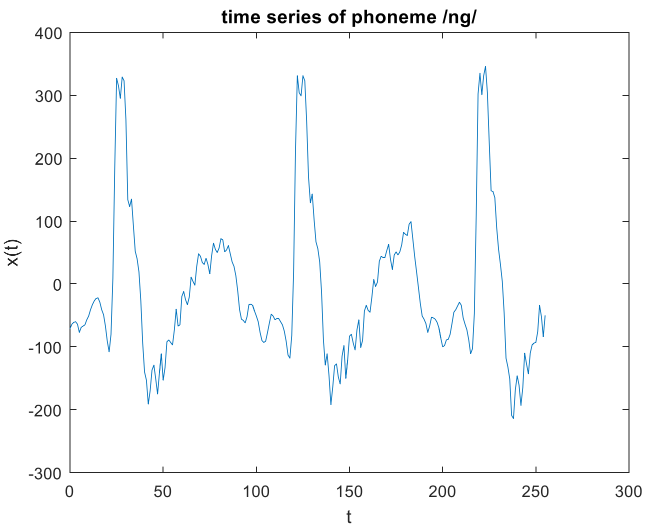

Figure 7

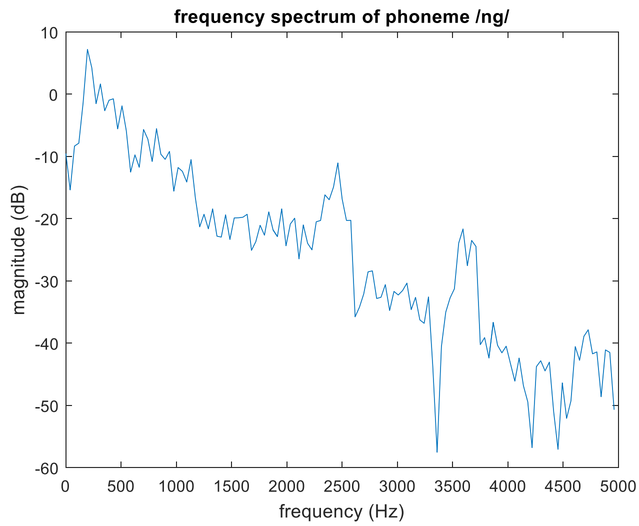

Figure 8

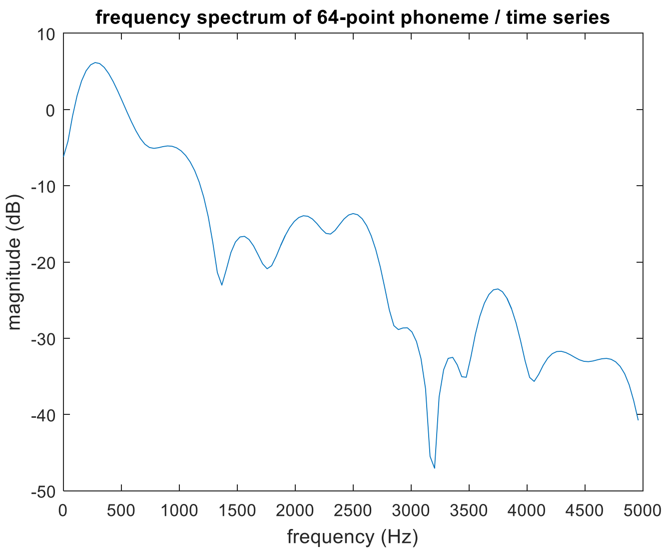

Figure 9

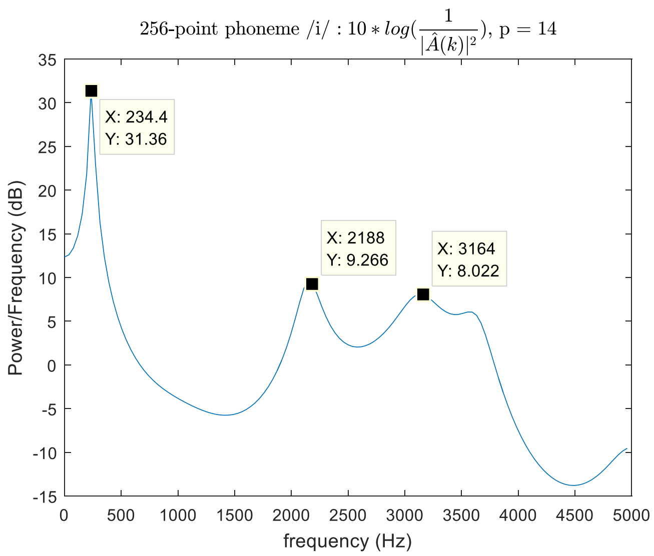

Figure 10

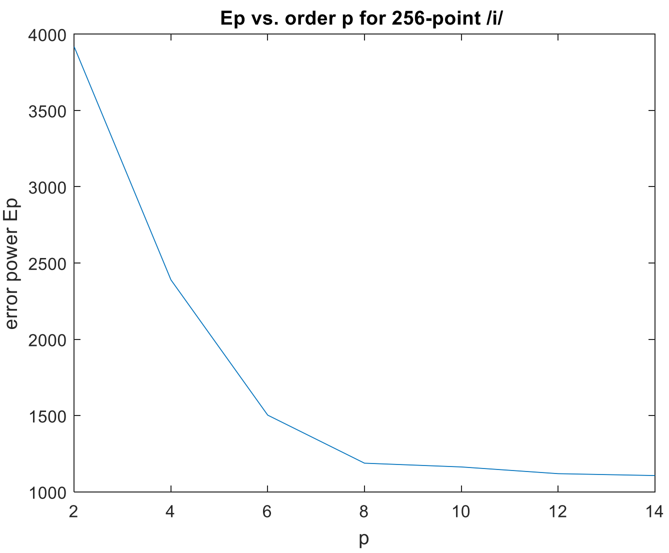

Figure 11

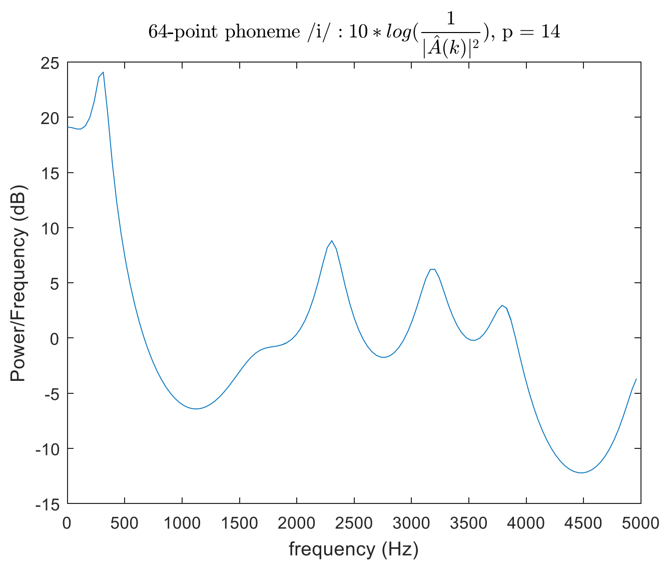

Figure 12

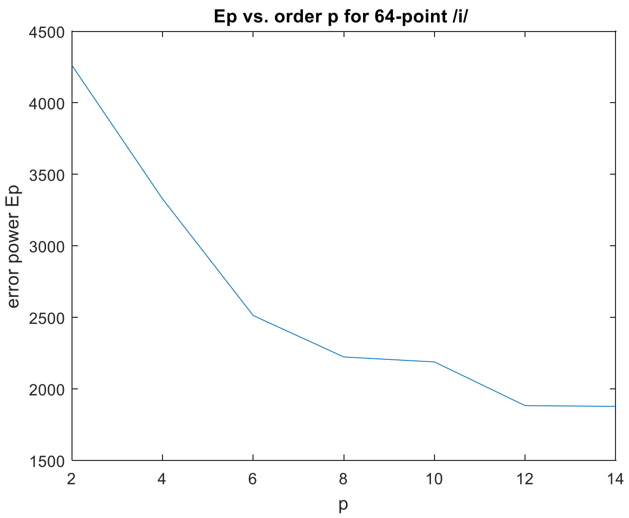

Figure 13

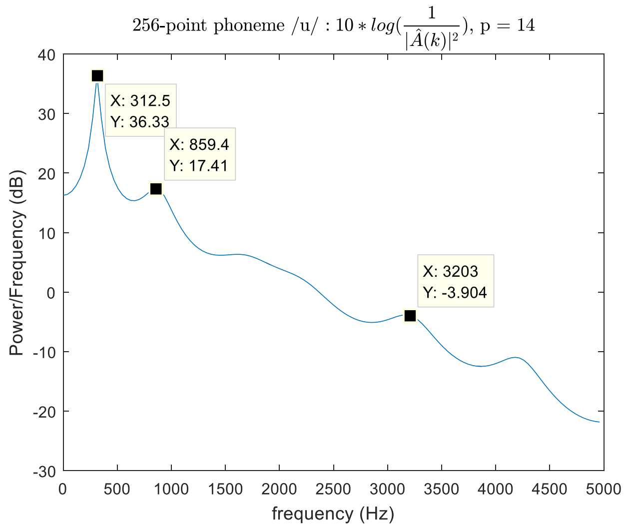

Figure 14

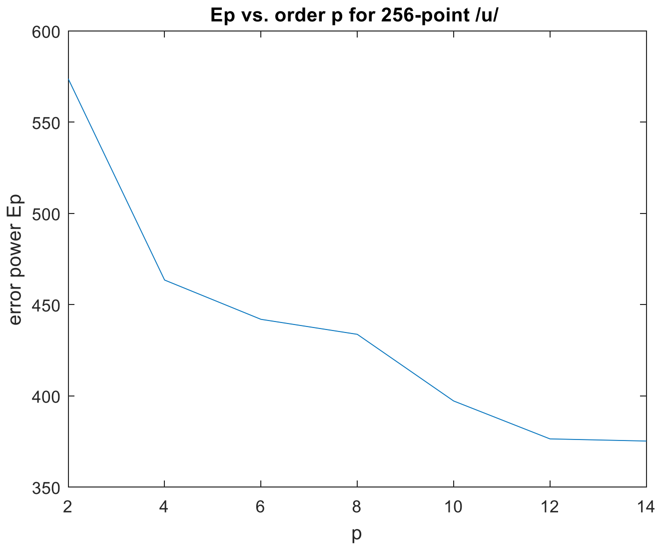

Figure 15

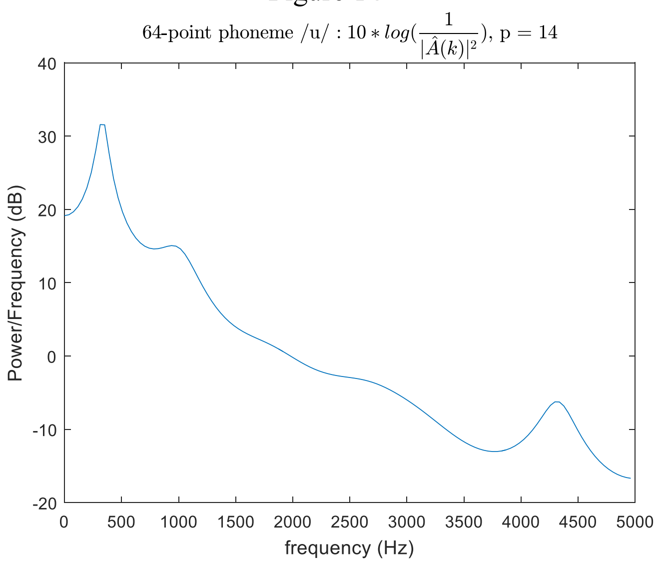

Figure 16

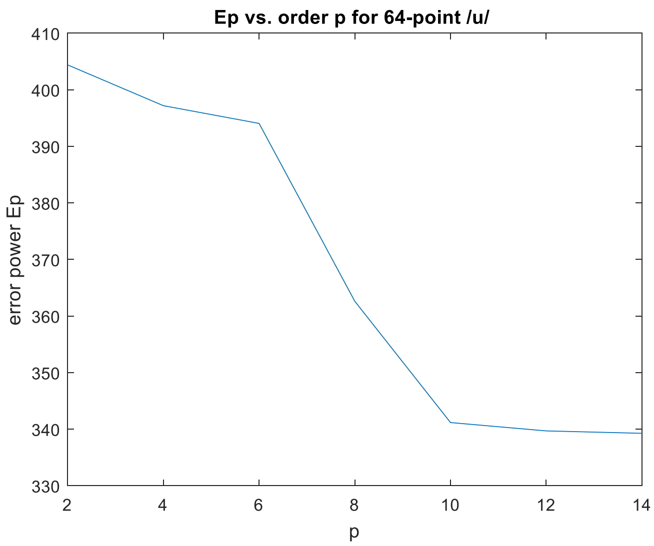

Figure 17

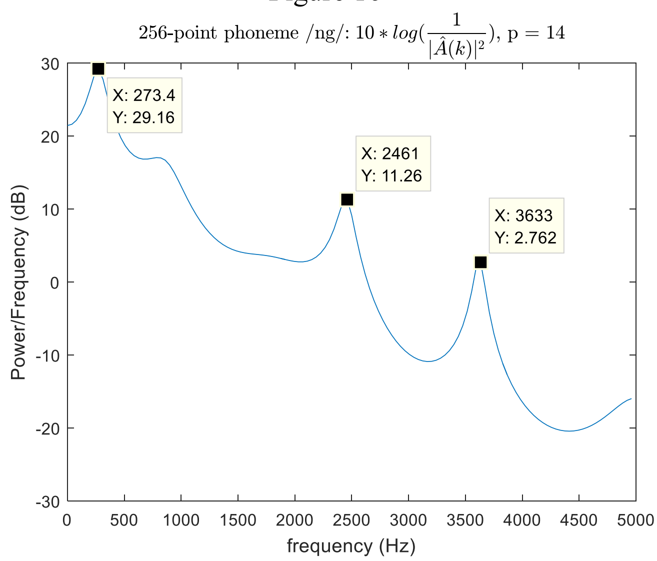

Figure 18

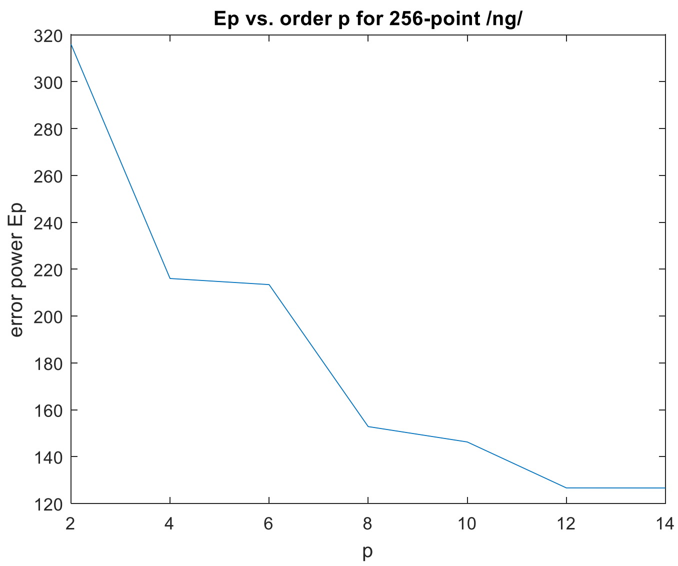

Figure 19

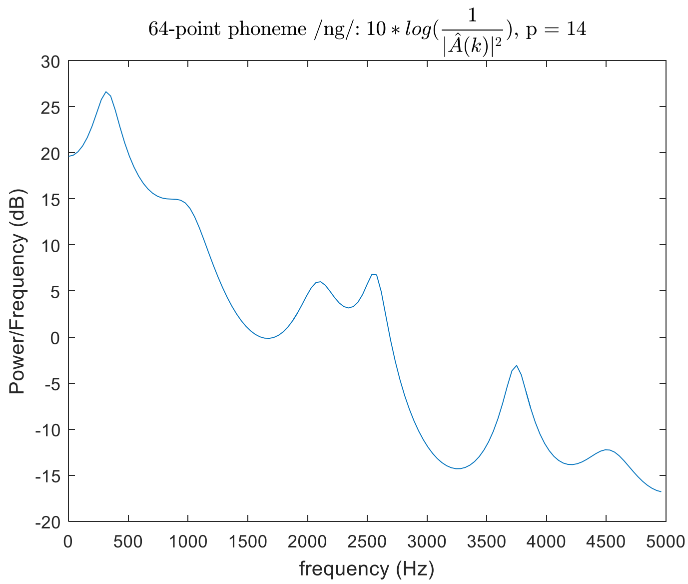

Figure 20

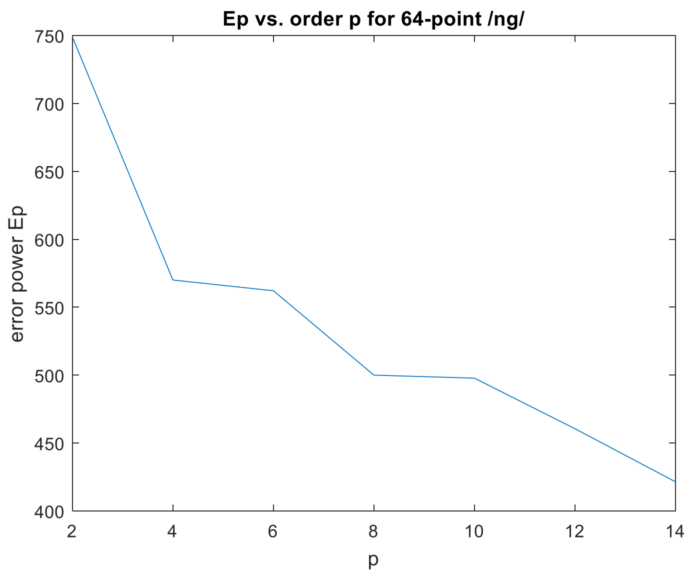

Figure 21

6. Appendix

List 1: Matlab Code for Part A

```
%% Part A
load -ascii bilu_single_col.txt
load -ascii bili_single_col.txt
load -ascii bilng_single_col.txt
load -ascii jcnwwa_single_col.txt
N = 256;
t = N-1;
close all;
input=[bili_single_col,bilu_single_col,bilng_single_col];
phoneme = ['/i/ '; '/u/ '; '/ng/'];
fs = 100000;
for i = 1:3
    aa=input(:,i);
    %/max(abs(input(i)));
    %aa=aa/1.1;
    U1 = sum(hanning(256).^2)*fs;
    U2 = sum(hanning(64).^2)*fs;
    figure((i-1)*3+1);
    plot(0:t,aa(1024:1279));
    title (['time series of phoneme ' phoneme(i,:)]);
    xlabel('t');ylabel('x(t)');
    freq = [0:5000/(N/2):5000-5000/(N/2)];
    figure((i-1)*3+2);
    Ak = fftshift(fft(aa(1024:1279).*hanning(256),256));
    plot(freq,10*log10(abs(Ak(129:end).^2)/U1));
    title(['frequency spectrum of phoneme ' phoneme(i,:)]);
    xlabel('frequency (Hz)');ylabel('magnitude (dB)');
    a2 = aa(1024:1087);
    figure((i-1)*3+3);
    Ak2 = fftshift(fft(a2.*hanning(64),256));
    plot(freq,10*log10(abs(Ak2(129:end).^2)/U2));
    title(['frequency spectrum of 64-point phoneme ' phoneme(i) ' 
time series']);
    xlabel('frequency (Hz)');ylabel('magnitude (dB)');
end
```

List 2: MATLAB Code for Part B

```
%% Part B
win1 = hanning(256);
win2 = hanning(64);
for i = 1:3
    aa=input(:,i);
    a1 = aa(1024:1279);
    a2 = a1(1:64);
    a1_win = a1.*win1;
    lpc_power(a1_win,14,10+(i-1)*4,phoneme(i,:));
    a2_win = a2.*win2;
    lpc_power(a2_win,14,12+(i-1)*4,phoneme(i,:));
end

function [] = lpc_power(input,p,no,phoneme)
N=256;
freq = [0:5000/(N/2):5000-5000/(N/2)];
[a,g] = lpc(input,p);
Ak =fftshift(fft(a,256));
Ak = Ak(129:end);
figure(no);
plot(freq,10*log10(1./(abs(Ak).^2)));
title([num2str(size(input,1)) '-point ' 'phoneme ' phoneme...
    '$$: 10*log(\frac{1}{|\hat{A}(k)|^2})$$' ', p = '
num2str(p)],'Interpreter','latex');
xlabel('frequency (Hz)')
ylabel('Power/Frequency (dB)')

Ep = zeros(1,7);
for i=[2,4,6,8,10,12,14]
    [a, Ep(i/2)] = lpc(input,i);
end
figure(no+1);
plot(2:2:14,Ep);
title(['Ep vs. order p for ' num2str(size(input,1)) '-point '
phoneme]);
xlabel('p');
ylabel('error power Ep');
end
```