EE679: Assignment 4

Ashwin Kachhara | 10d070048

Q1

We have to calculate the real cepstrum for a 30ms windowed segment for each of the phones. First of all, we pre-emphasize the original signal to emphasize the higher frequency energy in speech. We do so with a pre-emphasis coefficient of 0.975

Following pre-emphasis, we can calculate the real cepstrum. This is done by taking the inverse transform of log of the magnitude spectrum of the signal. The same has been implemented in the function realcepstrum.m, as shown below.

```
pre_em_S_a = zeros(length(S_a),1);
pre_em_S_a(1) = S_a(1);
pre_em_S_a(2:end) = S_a(2:end) - 0.975*S_a(1:end-1);
pre_em_Spec_a = abs(fft(pre_em_S_a(1:Lwin).*w, 1024));
N = 1024;
c_a = realcepstrum(pre_em_S_a,N);


function c = realcepstrum(s,N)
% USAGE: c = realcepstrum(s,N)
% Returns the N length real cepstrum of signal s
    Sw = abs(fft(s, N));
    C1 = log(Sw);
    c = real(ifft(C1));
end
```

Now we have the graphs of the cepstrum of each phone (\a\, \i\, \n\, \s\) sampled at 8kHz.

The cepstrum of each phone has a large peak at n=0 and several smaller peaks at equal period following that till N/2. Since the graph is symmetric, we are only interested in the n=1: N/2, where N =1024. This kind of periodic behavior implies that this might be related to the pitch of the signal.

In the cepstrum of \a\, \i\ and \n\ the multiple peaks are very prominently visible. This is because they are all voiced sounds and have a very strong source component.

In the cepstrum of \s\, we are unable to make out any distinctive peaks after the first one at zero. This is because, \s\ is an unvoiced sound, so the source component is not very strong. Also, \s\ is a fricative, which means that the vocal tract has anti-resonances.

Following are the graphs of the cepstrum:

Speech Cepstrum \a\

Speech Cepstrum \i\

Speech Cepstrum \n\

Speech Cepstrum \s\
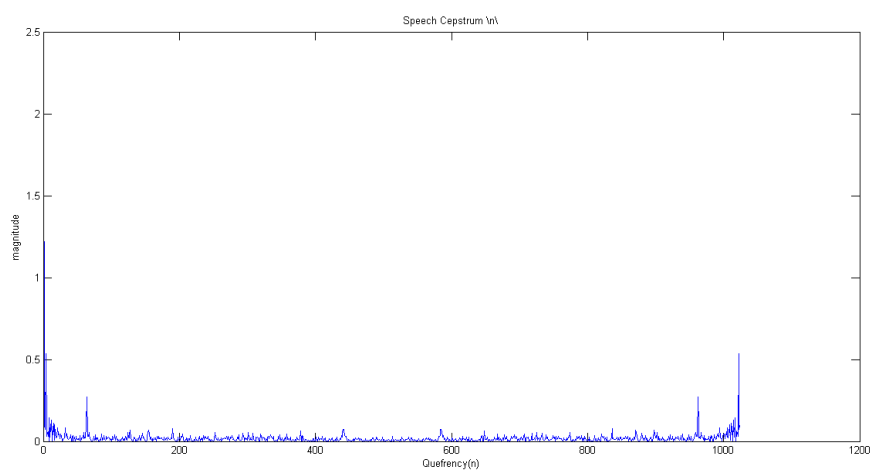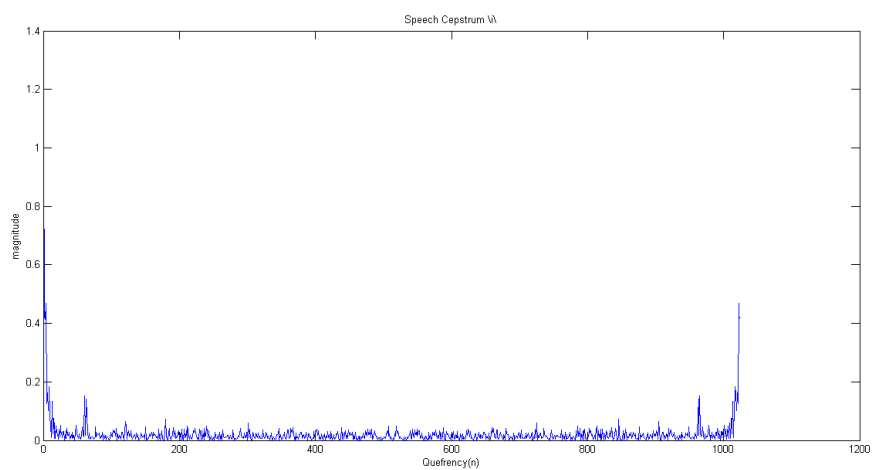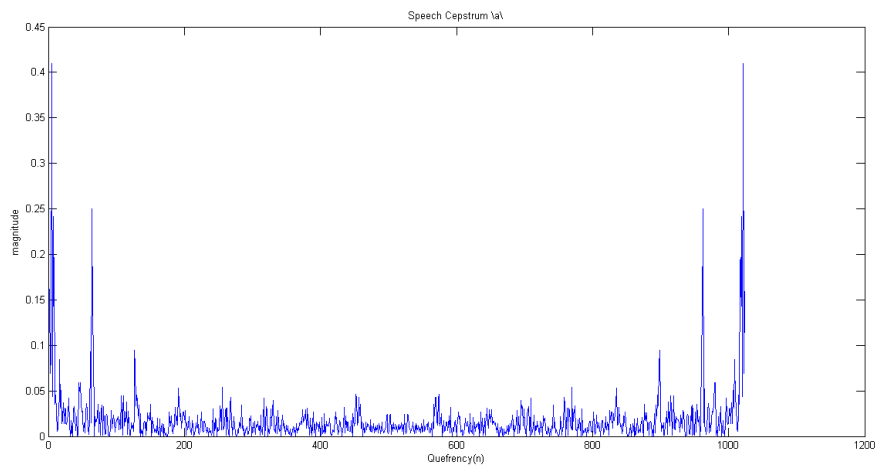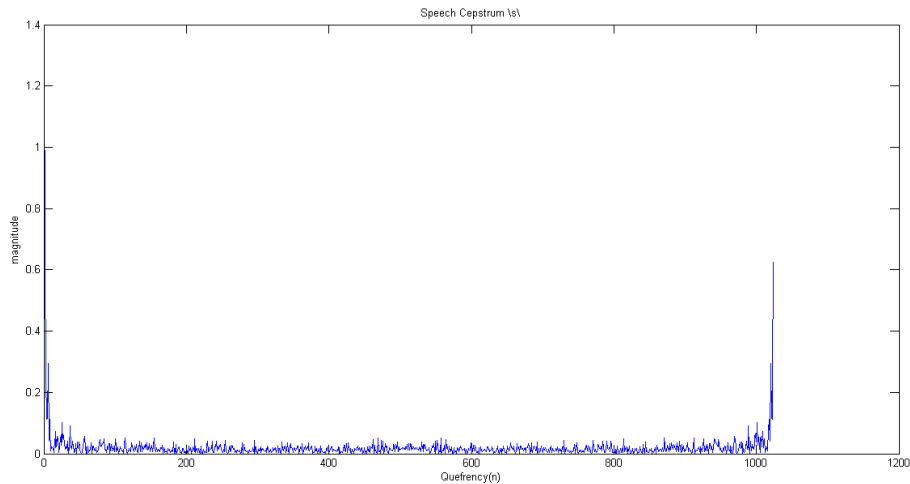
Q2

Now we will perform cepstral 'liftering' to obtain the vocal tract magnitude response for each phone.

```
filter = zeros(N,1);

filter(1:13) = 1;
c_a_filt13 = real(fft(c_a.*filter));
c_i_filt13 = real(fft(c_i.*filter));
c_n_filt13 = real(fft(c_n.*filter));
c_s_filt13 = real(fft(c_s.*filter));

filter(1:26) = 1;
c_a_filt26 = real(fft(c_a.*filter));
c_i_filt26 = real(fft(c_i.*filter));
c_n_filt26 = real(fft(c_n.*filter));
c_s_filt26 = real(fft(c_s.*filter));

filter(1:40) = 1;
c_a_filt40 = real(fft(c_a.*filter));
c_i_filt40 = real(fft(c_i.*filter));
c_n_filt40 = real(fft(c_n.*filter));
c_s_filt40 = real(fft(c_s.*filter));
```

After applying the filter to each cepstrum, we take the real part of its fft. This is the required vocal tract magnitude response, since it is the low quefrency part that we have taken. The pitch and harmonic info is omitted since it was in the higher quefrency.

We plot these graphs along with the original spectrum and the LP spectrum computed in the previous assignment.
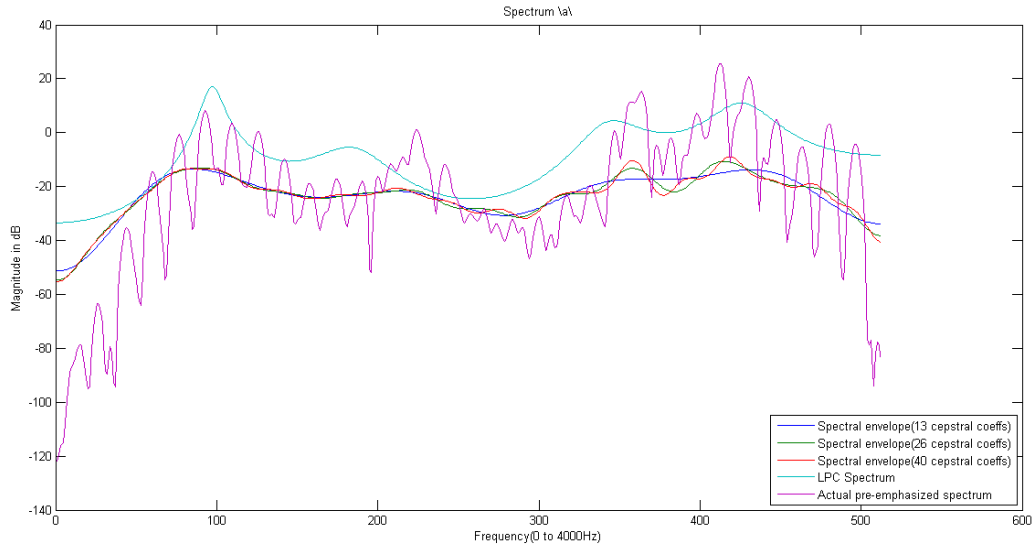
The spectral envelope obtained from the cepstrum models the main peaks in the original signal quite well. In fact, it seems to be more accurate than the spectrum calculated from the LP coefficients.

As we increase the filter length, some harmonic content seems to be appearing. Also, the peaks are getting better modeled with increasing filter length.
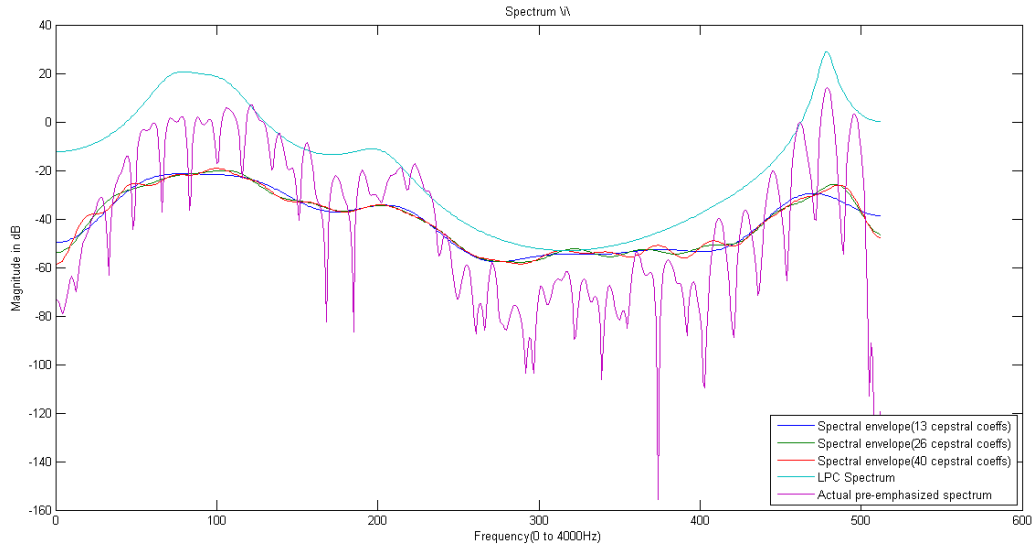
The spectral envelope obtained from the cepstrum models the main peaks in the original signal quite well. In fact, it seems to be more accurate than the spectrum calculated from the LP coefficients.

As we increase the filter length, some harmonic content seems to be appearing. Also, the peaks are getting better modeled with increasing filter length.

The filter length of 13 or 26 seems to be better that 40 in just modeling the vocal tract magnitude.
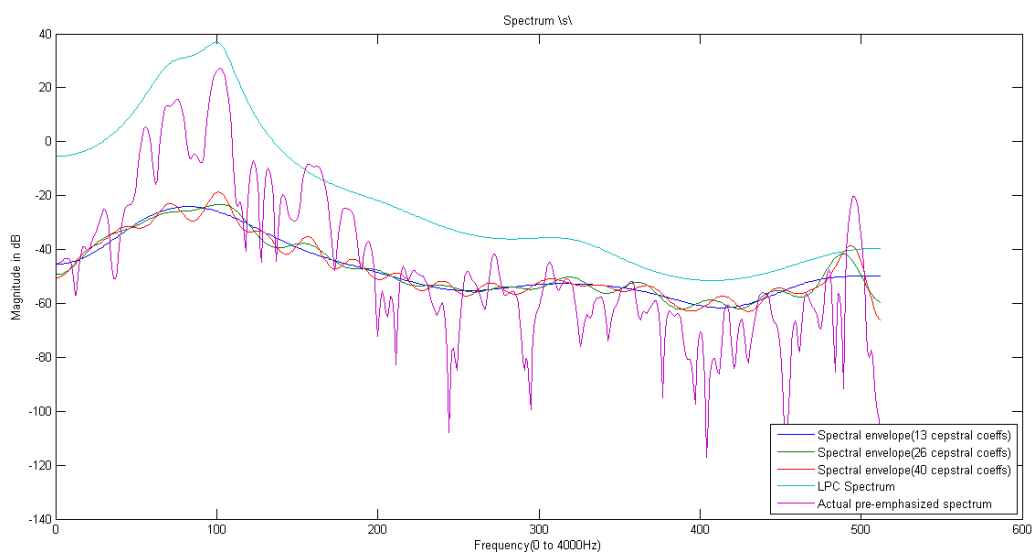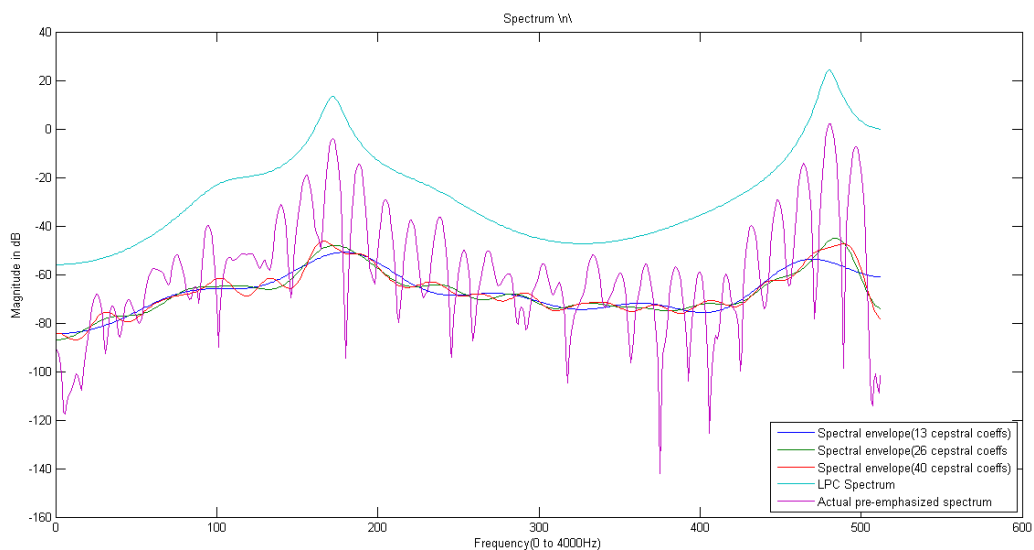
Spectrum \n\

Spectrum \s\

Overall, the cepstral coefficients seem to model the vocal tract pretty well. The LP spectrum is a peak hugging one, while the spectral envelope obtained from the cepstrum seems to model the up's and down's but does not follow the peaks closely. It is lower in level (dB) than the LP.

Q3

From the graphs in Q1, we see that second peaks in cepstrum of \a\, \i\, \n\ are at n= 64, 61 and 64 respectively. These correspond **to pitch of 8000/64 = 125Hz, 8000/61 = 131Hz, 8000/64 = 125Hz.** We cannot make out any pitch information from the cepstrum of \s\

Q4

In this part, we perform the cepstral analysis on the LP synthesized signals from the previous assignment.

First, we synthesize the phones \a\, \i\, \n\, \s\ from the LP coefficients as in the previous assignment. After that we calculate the cepstrum of the synthesize signal, as we did for the original signal in Q1. Further we will 'lifter' the cepstrum to try and obtain the vocal tract magnitude response and also the pitch period from the cepstrum.

So, the cepstrum graphs for the synthesized phones are as follows:

Speech Cepstrum \n\

Speech Cepstrum \s\
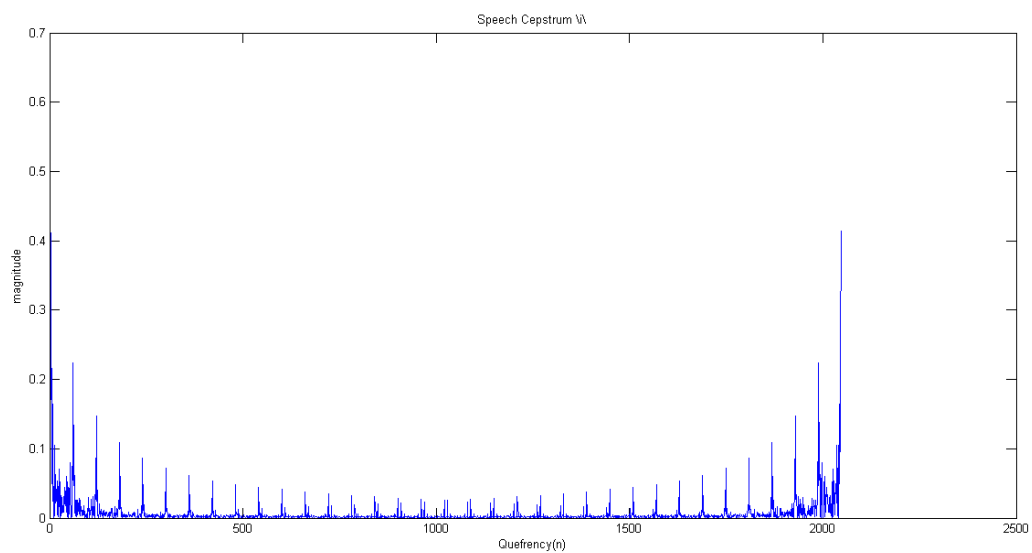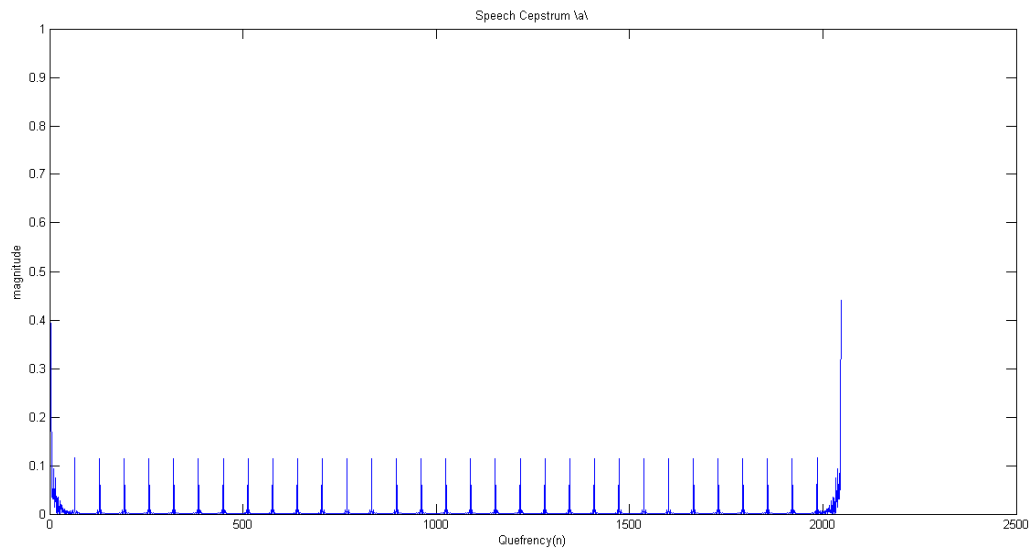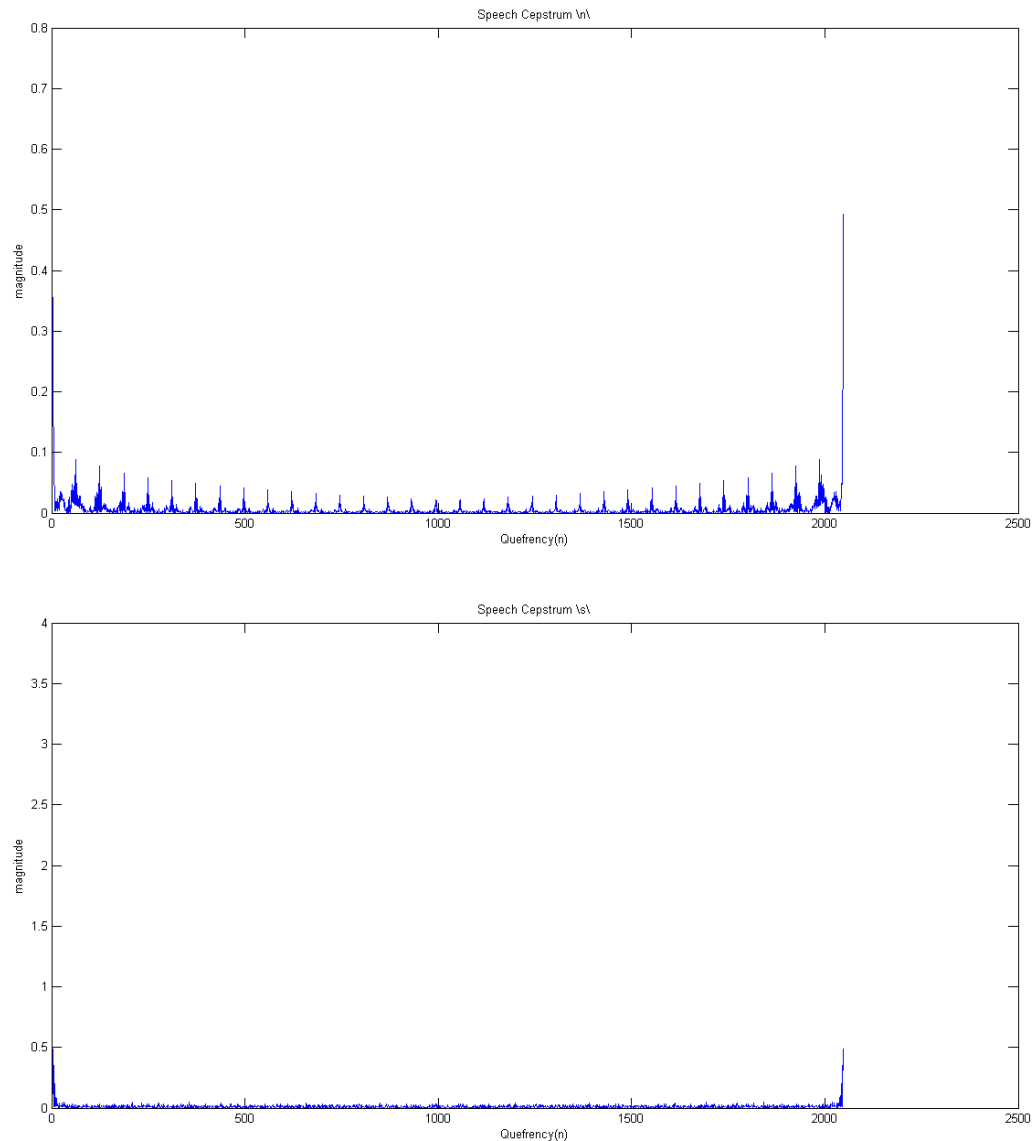
As you can see, the periodicity in this cepstrum does not die out as easily as in the original signal. In fact, it does not decay at all. This is because, in the original extracted components, there was a gradual decay term because of co-articulation. This is not there in the synthesized waveforms.

Also, while synthesizing waveforms, we have used the impulse as the vocal tract excitation. This is a very crude approximation which may be the reason for these prominent peaks. In actual speech, the excitation is more of a band-pass nature rather than an impulse.

Now, we plot the spectra from liftering the cepstrum and the LP coefficients.
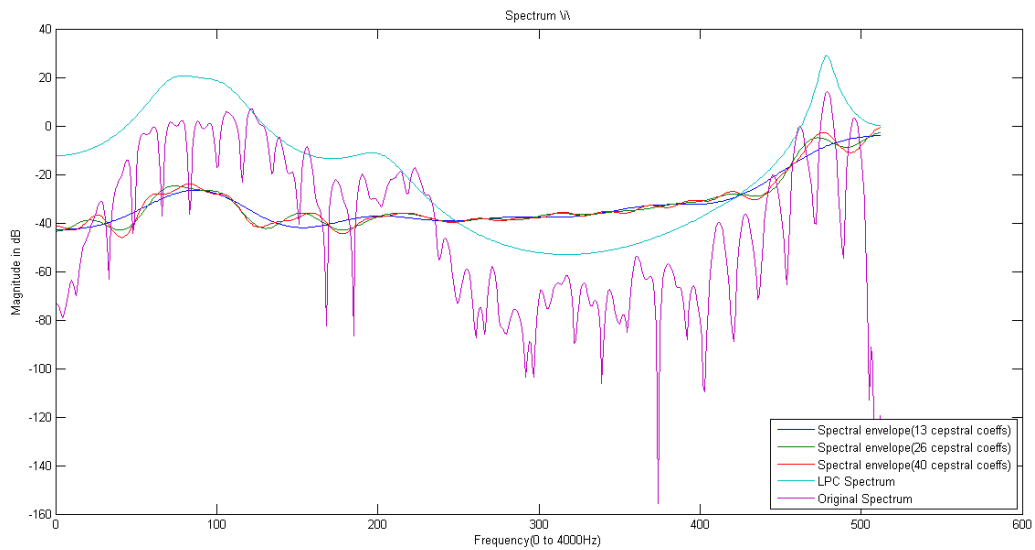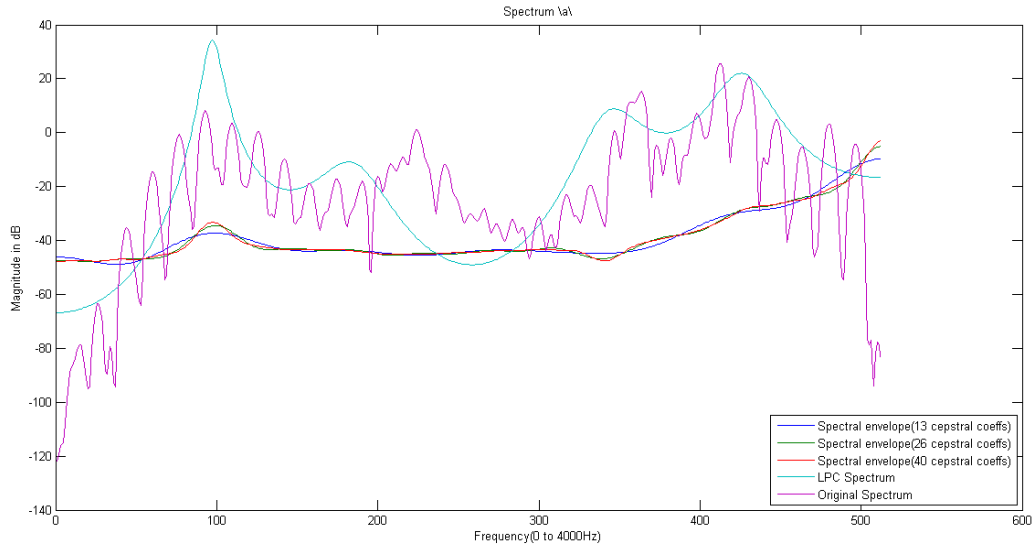
This time, we can see that the spectral envelope obtained from the cepstrum follows the wideband ups and downs of LP spectrum rather than the original spectrum. In many cases, it even overshoots the original spectrum. Since, the prominence of peaks in the LP Spectrum is not the best, this reflects on the cepstrum spectral envelope as well.

Otherwise the relationship with the filter length remains the same, more prominent peaks and harmonic nature in the higher filter length spectrum.
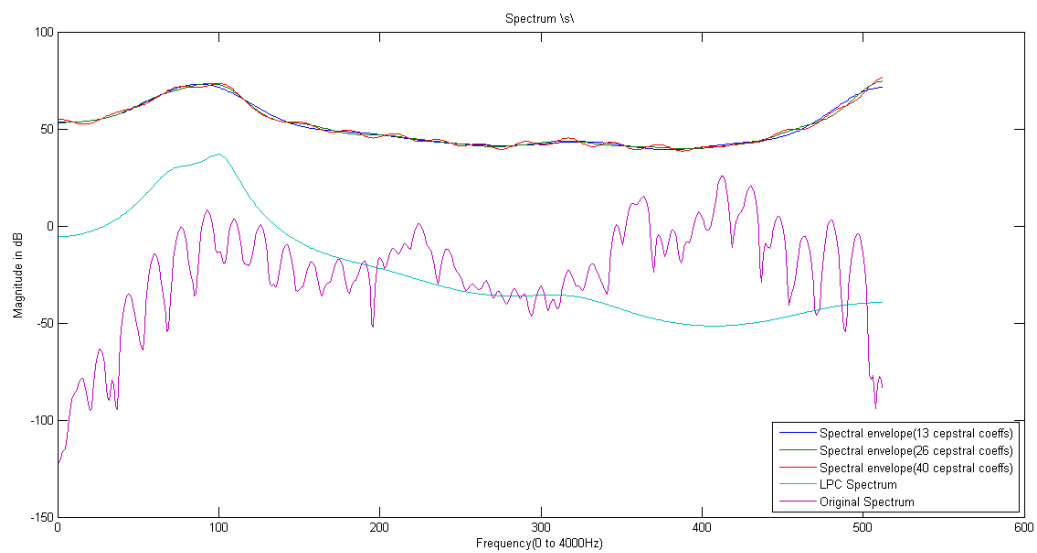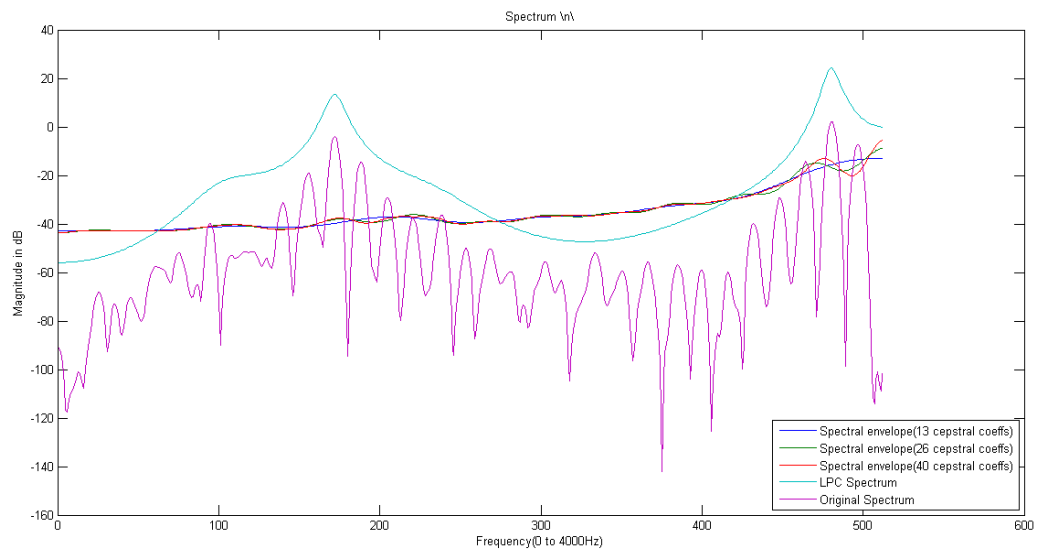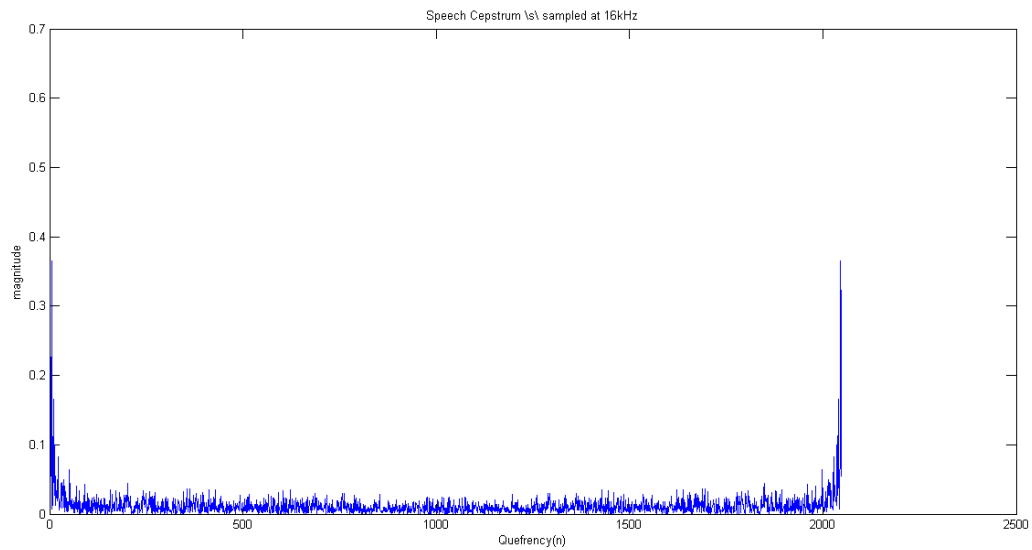
Spectrum \n\

Spectrum \s\

The spectrum for \s\ is way above both the original and the LP spectrum. It does follow the LP spectrum though.

Q5

In this section, we use the 16kHz sampled \s\ phone instead of the 8kHz sampled \s\ phone.
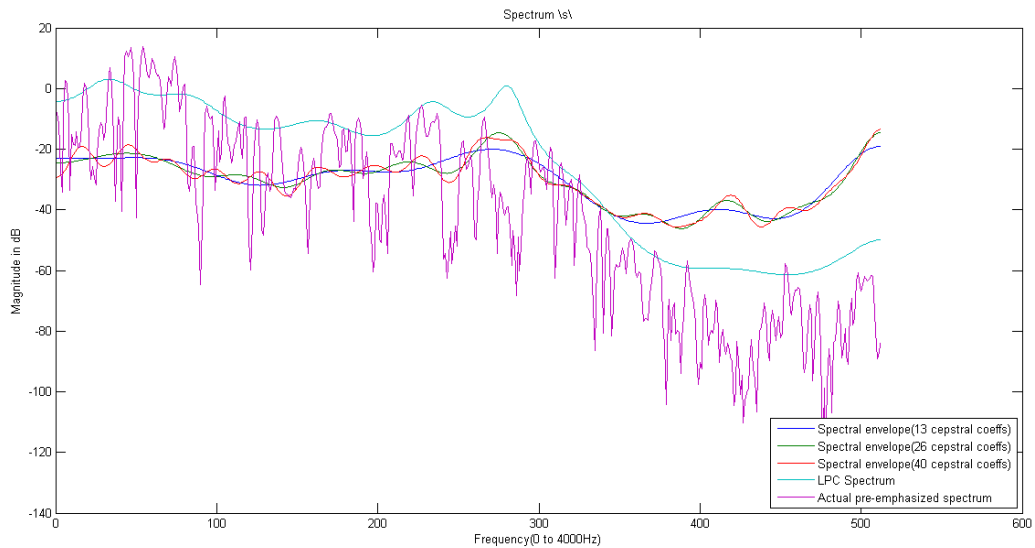
The steps are similar to that in Q1

The cepstrum plot is as follows:

Even in this plot, we are not able to make out any information about the pitch. There are no distinctive secondary peaks.

As for the spectrum,

The LP spectrum appears to be able to model better, but that is just because it is an order 20. The cepstral envelope was a better fit for the 8 kHz case, Here it overshot the original and LP spectrum by quite an amount, though the modeling of the peaks is accurate. Just the levels are off.

OPTIONAL PART:

Demonstration of Masking phenomenon.

We take a sinusoidal tone of frequency 1200 Hz and add other tones of frequencies from 100Hz to 2400Hz and listen to the sound in each of the case.

This gives us an idea of masking effect. The sounds (s100 to s2400) are included in the 'optional' directory.

```
Fs = 8000;        %# Samples per second
toneFreq = 1200;  %# Tone frequency, in Hertz
nSeconds = 2;     %# Duration of the sound
% Generating the main frequency component
y1 = 4*sin(linspace(0, nSeconds*toneFreq*2*pi, round(nSeconds*Fs)));

for i=1:24,
    % Generating the variable frequency ( this has lower amplitude)
    maskedFreq = 100*i;
    y2 = sin(linspace(0, nSeconds*maskedFreq*2*pi, round(nSeconds*Fs)));
    y = y1+y2;
    % Normalization after adding the 2 signals
    y = y/max(abs(y));
    % Writing to output file
    wavwrite(y,Fs,strcat('s', num2str(i*100)));
end
```

As expected, we are able to hear 2 distinct frequencies till s700.

From s800 to s1700 we are not able to distinguish 2 frequencies. We can only hear the 1200Hz one.

From s1700 to s2400, we are able to hear two distinct tones.