CS630: Speech Technology
LAB-5: Short-Time Spectrum Analysis

OBJECTIVE:

To study issues in short-time spectrum analysis of speech.

SEQUENCE OF STEPS:

(a) Record (about 2 sec) of vowel /a/ and fricative /s/.

(b) Study the effects of convolution and correlation.

Consider $x(n)$, 128 samples of /a/ and $h(n)$, 5 samples of unit values Perform

- Linear convolution of $x(n)$ and $h(n)$
- Circular convolution of $x(n)$ and $h(n)$ with (N=128)
- Linear correlation of $x(n)$ and h(n)
- Linear autocorrelation of $x(n)$
- Linear autocorrelation of $h(n)$

(c) Short-time spectrum - Effects of size and shape of window.

- Consider $x(n)$, 160 samples of /a/.
- Use 512 point DFT to get $X(k)$.
- Plot log spectrum log—X(k)—^{2}.
- Study the effect of size of window - 5 msec, 20 msec, 50 msec.
- Study the effect of shape of window on 20 msec data namely, Rectangular, Hamming and Hanning windows.

(d) Short-time spectrum of voiced and unvoiced speech.

- 20 msec of $x(n)$, Hamming window and 512 pt DFT
- Plot log spectrum for voiced and unvoiced segments

(e) Spectrograms.

- Record a short utterance
- Observe features of WB and NB spectrograms

(f) Write a brief note on the observations.

1 Procedure

(1) **Recording of required speech utterances**

- Record the sound units vowel /a/ and fricative /s/ using command
  brec -s 8000 -b 16 -t 2 -w filename1.wav
  where s is the sampling rate, b is the number of bits/sample,
  t is the duration of speech utterance and w is the fileformat.

- Take a segment of duration 200 msec of vowel /a/ and a segment of duration 200 msec of fricative /s/.
  Using Matlab, read the speech signal in $filename1.wav$ into an array.
  a=wavread('filename1.wav');
  plot(a);
  a200=a(10501:10660);
  s=wavread('filename2.wav');
  plot(s);
  s200=s(6501:6660);

(2) **Effect of Convolution and Correlation**
The convolution of two sequences x(n) and h(n) is defined as

$$y(n) = \sum_k x(k)h(n-k) \tag{1}$$

The convolution sum is used in filtering a signal where h(n) is a filter. The filter suppresses some frequency components of x(n).
Consider x(n) as 128 samples of /a/ given by
a128 = a(10501:10628); and h(n) as 5 samples of unit values given by
h5 = [1; 1; 1; 1; 1];

- **Linear convolution of x(n) and h(n)**
  The convolution of two vectors can be obtained by using the function conv as

linconv = conv(a128,h5);
plot(linconv);
The resulting vector is of length equal to[length(x(n))+length(h(n))-1] and is shown in Figure 1.

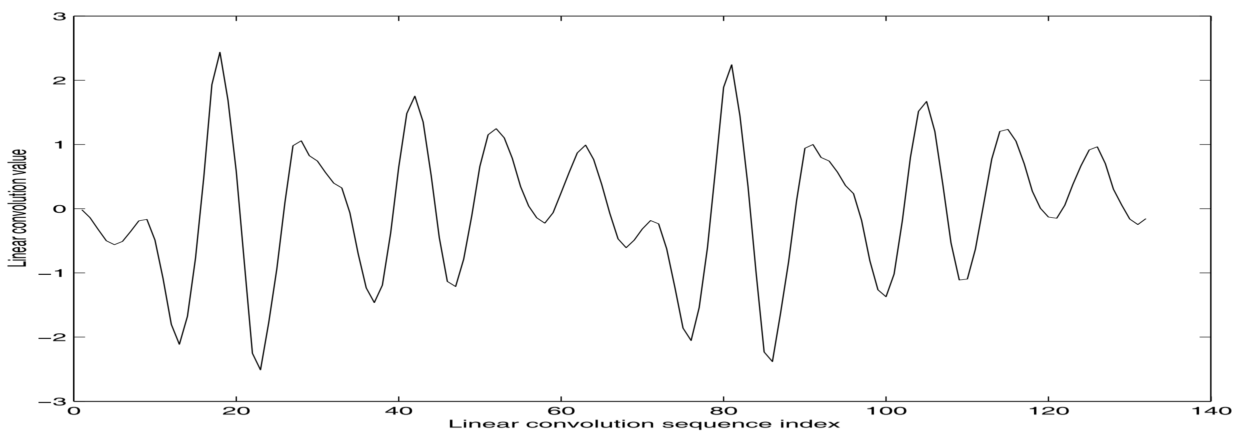

Figure 1: The linear convolution of x(n) and h(n).

- **Circular convolution of x(n) and h(n) by considering N=128**
  The circular convolution of two vectors are done by finding, the FFT of two vectors x(n) and h(n),
  followed by multiplying the two resulting vectors and finally finding the IFFT as given below.
  fftx=fft(a128,128);
  ffth=fft(h5,128);
  fftprod=fftx.*ffth;
  circonv=real(ifft(fftprod));
  plot(circonv);
  The result of circular convolution is shown in Figure 2.

  Observation: It is evident from figures 1 and 2 that the length of resulting vector in linear convoltion is 132 (length(x(n)) + length(h(n)) -1), where as the length of resulting vector in circular convolution is only 128 (max(length(x(n)),length(h(n)) if N is not specified). Circular convolution of x(n) and h(n) is the aliased version of the linear convolution if the sequence lengths are not properly extended.

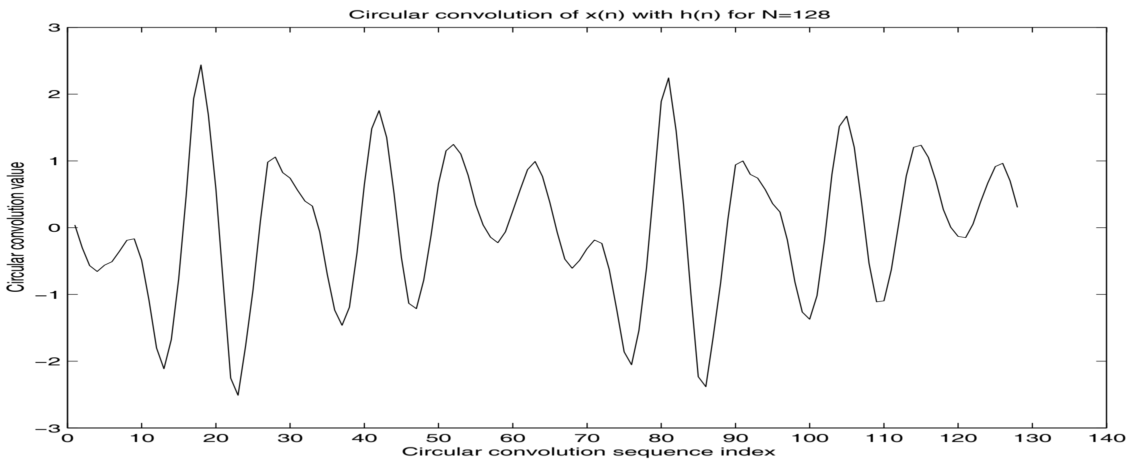

Figure 2: The circular convolution of x(n) and h(n).

- **Linear correlation of x(n) and h(n)** The correlation basically finds the relationship among two sequences. The correlation of two sequences x(n) and h(n) is given by
  $r(n) = \sum_k x(k)h(n+k)$
  If there exists no relationship among the sequences then r(0), r(1), r(2),......, are all will be zero. The correlation function is used for cross-correlation function estimation. The linear autocorrelatrion of x(n) and h(n), where x(n) is a vector of length 128 and h(n) is a vector of length 5 is obtained by using **'xcorr'** matlab builtin function. This function returns the length 255 cross-correlation sequence. Since h(n) is of length 5, it is zero padded to make its length equal to the length of x(n) which is 128. Thus in general, if largest length of either vectors is M then autocorrelation function returns a cross-correlation sequence of length $2*M-1$. The linear autocorrelation of x(n) and h(n) is computed as follows:
  lcrxnhn=xcorr(a128,h5);
  plot(lcrxnhn);
  The result of linear correlation is shown in Figure 3.

- **Linear autocorrelation of x(n)**
  The autocorrelation of x(n) is computed as
  lautocorrxn=xcorr(a128);

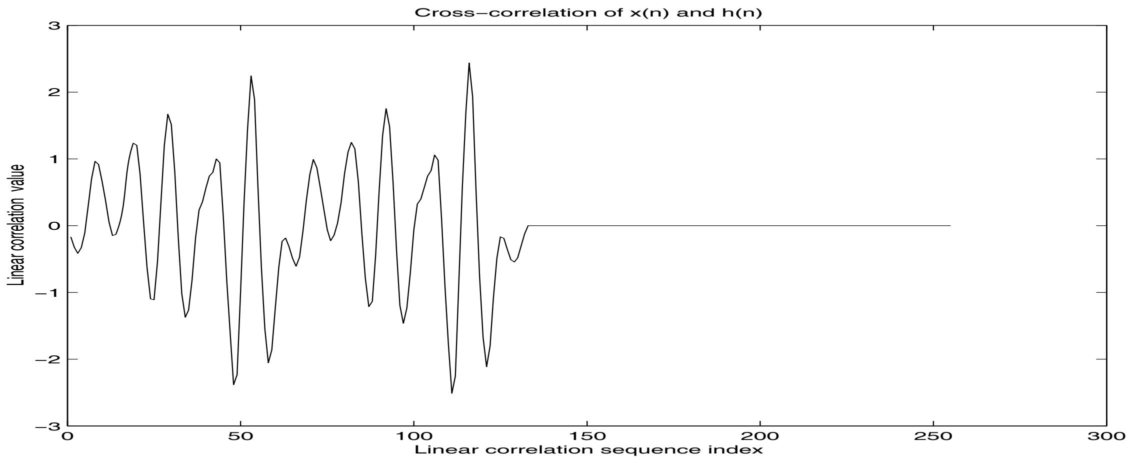

Figure 3: The cross-correlation of x(n) and h(n).

plot(lautocorrxn);
The result of linear autocorrelation of x(n) is shown in figure 4.
The autocorrelation of x(n) at n=0 gives energy of the signal.

- **Linear autocorrelation of h(n)**
  The autocorrelation of h(n) is computed using
  lautocorrhn=xcorr(h5);
  plot(lautocorrhn); The result of linear autocorrelation of x(n) is shown in figure 4.

(3) **Study the effect of size and shape of the window**

- **Effect of size of window a) 5 msec b) 20 msec c) 50 msec -**
  The effect of rectangular window of size 5 msec(assuming 512-point DFT) is illustrated by computing the short-time spectrum of the windowed signal as follows:
  a40 = a(10501:10540);
  X(k) = fft(a40,512);
  plot(10∗log10(abs(X(k).∗X(k))));
  The resulting log spectrum is shown in Figure 6. Since the spectrum of real valued signal is even symmetric, spectrum is plotted for only

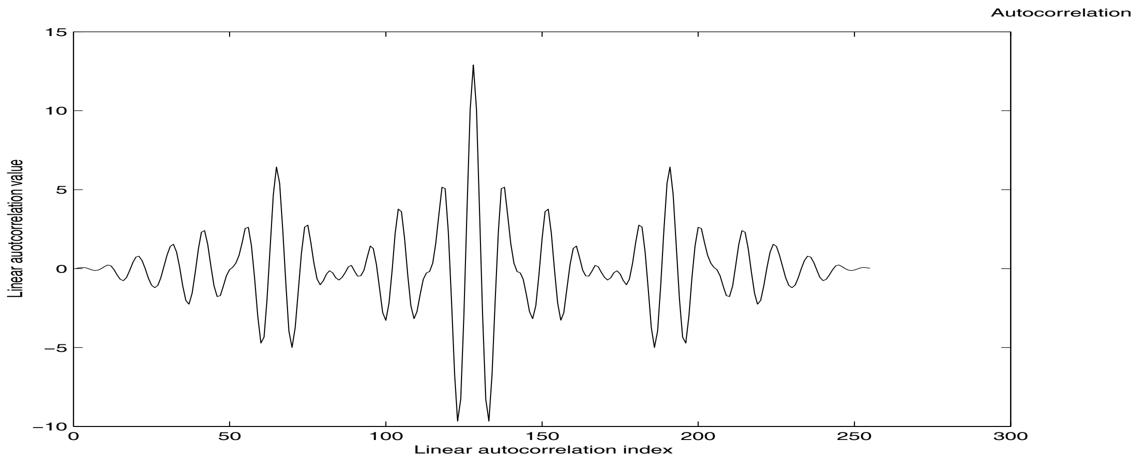

Figure 4: The autocorrelation of x(n).

half of the index values(k). This convention is followed in plotting the subsequent log spectrums. - The effect of rectangular window of size 20 msec(assuming 512-point DFT) is illustrated by computing the short-time spectrum of the windowed signal as follows:
a160 = a(10501:10660);
X(k) = fft(a160,512);
plot(10∗log10(abs(X(k).∗X(k))));
The resulting log spectrum is shown in Figure 7.

- The effect of rectangular window of size 50 msec(assuming 512-point DFT) is illustrated by computing the short-time spectrum of the windowed signal as follows:
a400 == a(10501:10900);
X(k) = fft(a400,512);
plot(10∗log10(abs(X(k).∗X(k))));
The resulting log spectrum is shown in Figure 8.

OBSERVATION:
The size of the window (rectangular in this case) determines the temporal resolution. Increase in size of the window decreases the temporal resolution as evident from the Figures 6, 7 and 8.

(4) **Effect of shape of the window**

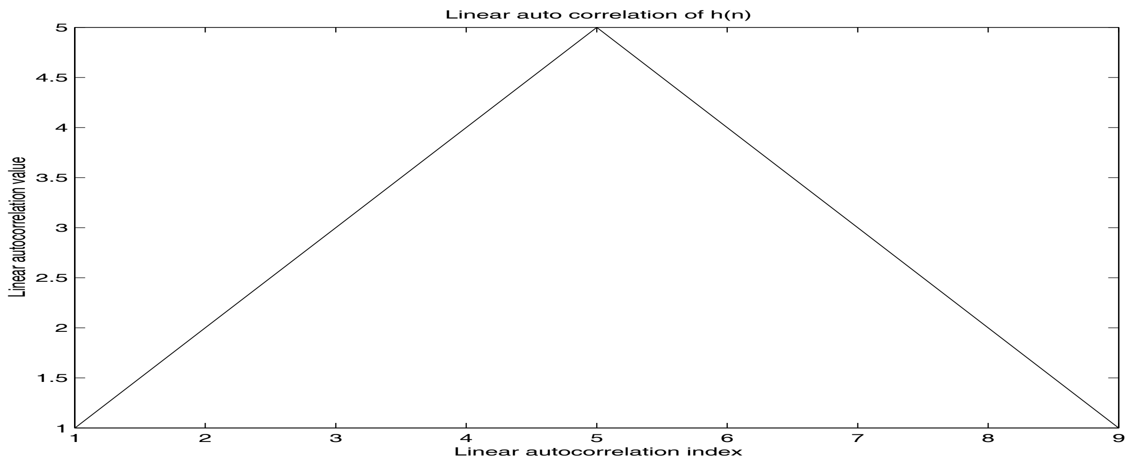

Figure 5: The autocorrelation of h(n).

Consider x(n) as 160 samples of /a/(20 msec).
a160 = a(10501:10660);
- rectangular window
The effect of rectangular window of 20 msec on the short-time spectrum is obtained as follows:
rectwindow = boxcar(160);
a160rectwindow= a160.*rectwindow;
X1= fft(a160rectwindow,512);
X=X1(1:256);
plot(10.∗log10(abs(X).*(X)));

The effect of rectangular window is shown in the Figure 9. **Hamming window**
The effect of hamming window on the short-time spectrum is obtained as follows:
hammingwindow = hamming(160);
a160hammingwindow = a160.*hammingwindow;
X2=fft(a160hammingwindow,512);
X=X2(1:256);
plot(10.*log10(abs(X.*X)));
The effect of hamming window on the log spectrum of a segment of speech signal is shown in Figure 10.

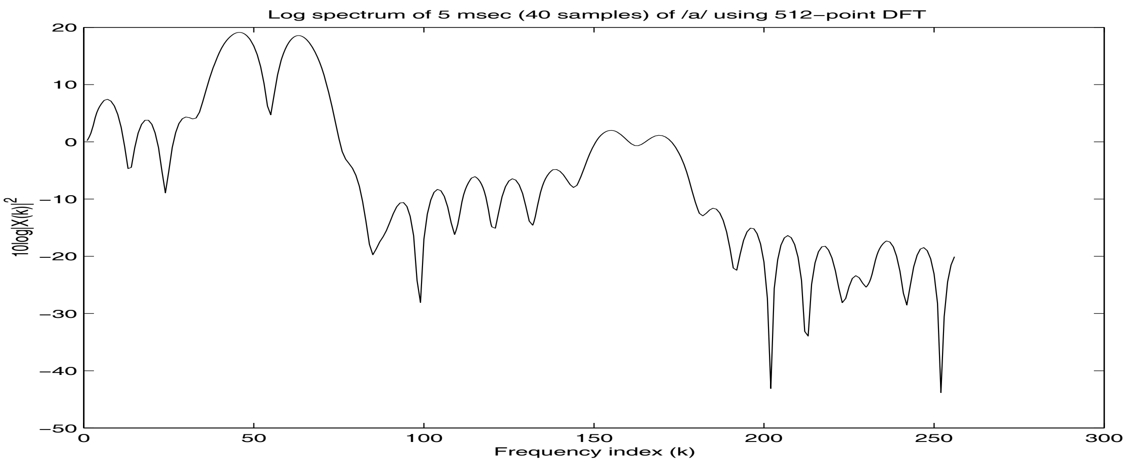

Figure 6: The Log spectrum of x(n), using rectangular window of size 5 msec.

Hanning window

The effect of hanning window on the short-time spectrum is obtained as follows:

hanningwindow = hanning(160);
a160hanningwindow = a160.*hanningwindow;
X2=fft(a160hanningwindow,512);
X=X2(1:256);
plot(10.*log10(abs(X.*X)));

The effect of hanning window on the log spectrum of a segment of speech siganl is shown in Figure 11.

OBSERVATION :

Sidelobe effect is dominant in the rectangular window.
Pitch harmonics can be clearly seen when Hamming or Hanning window is used.
Sidelobe attenuation is more with Hamming and Hanning windows and also the main lobe width increases.

(5) **Short-time spectra of voiced and unvoiced speech**

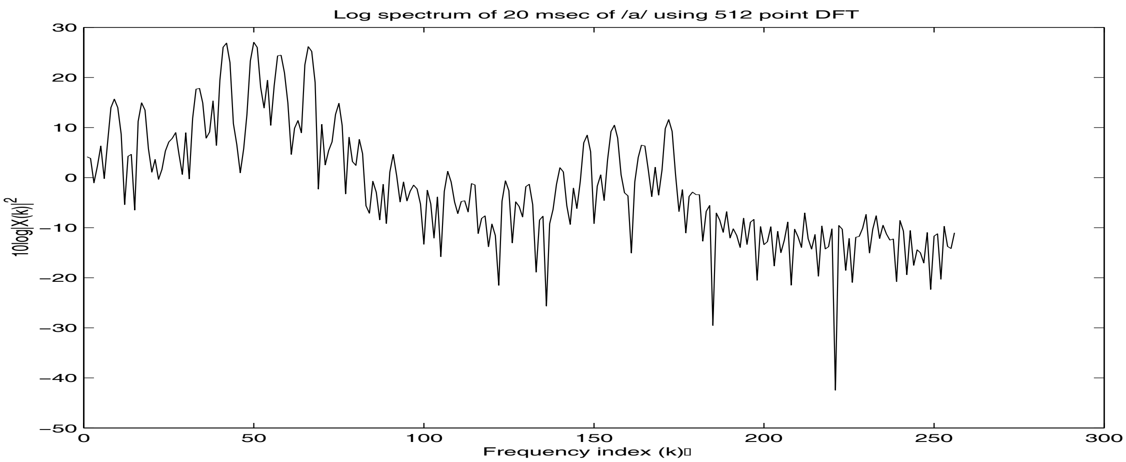

Figure 7: The Log spectrum of x(n), using rectangular window of size 20 msec.

The short-time spectra of 20 msec data of voiced speech segment of /a/, using Hamming window and 512-point DFT is computed as shown below.

```
hammingwindow = hamming(160);
a160hammingwindow = a160.* hammingwindow;
X=fft(a160hammingwindow,512);
X1 = X(1:256);
plot(10.*log10(abs(X1.*X1)));
```

The resulting log spectrum is shown in the Figure 12 (a).

The log spectrum of 20 msec data of unvoiced speech segment of /s/, using Hamming window and 512-point DFT is computed as follows.

```
s = wavread('s.wav');
s160 = s(6501:6660);
hammingwindow = hamming(160);
s160hamming = s160 .* hammingwindow;
X = fft(s160hamming, 512);
X1 = X(1:256);
plot(10.*log10(abs(X1.*X1))) ;
```

The resulting log spectrum is shown in the Figure 12(b).

OBSERVATION :

Voiced speech log spectrum shows clearly the harmonic (source feature)

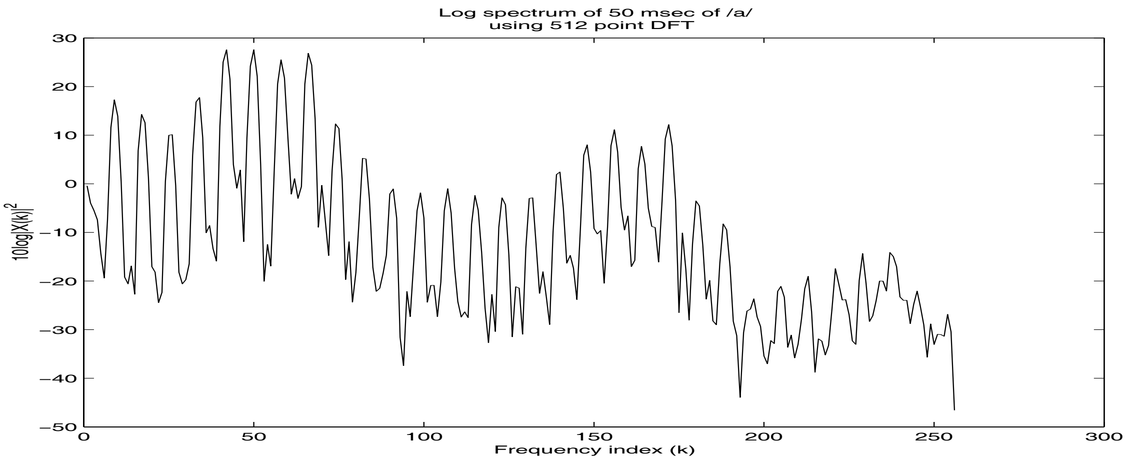

Figure 8: The Log spectrum of x(n), using rectangular window of size 50 msec.

and the formant structure (system feature) whereas there is no defined structure in the log spectrum of Unvoiced speech segment.

(6) **Spectrograms** Record a short utterance that contains both voiced and unvoiced components using the command :
brec -s 10000 -b 16 -t 2 -w filename.wav
where 's' is the sampling rate (Hz), 'b' is the number of bits/sample, 't' is the duration and w is the wave format.
The wideband and narrowband spectrograms for a given speech utterance is obtained by using the wavesurfer utility.
In wideband spectrogram a small time window (typically of duration 5 msec) is used. In Narrowband spectrogram
realtively larger time window (typically of duration 50 msec) is used. The wideband spectrogram is characterized by the vertical lines and the narrowband spectrogram by the horizontal lines.The resulting spectrograms are shown in Figure 13.

OBSERVATION Wideband spectrogram provides a better temporal resolution (vertical striations) and narrowband spectrogram provides a better spectral resolution (horizontal striations).

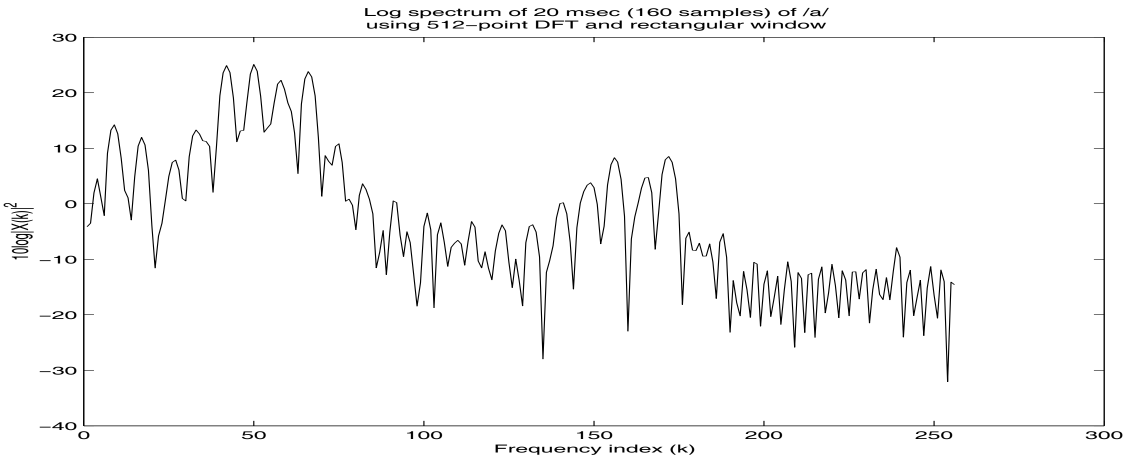

Figure 9: The Log spectrum of x(n), using rectangular window of size 20 msec.

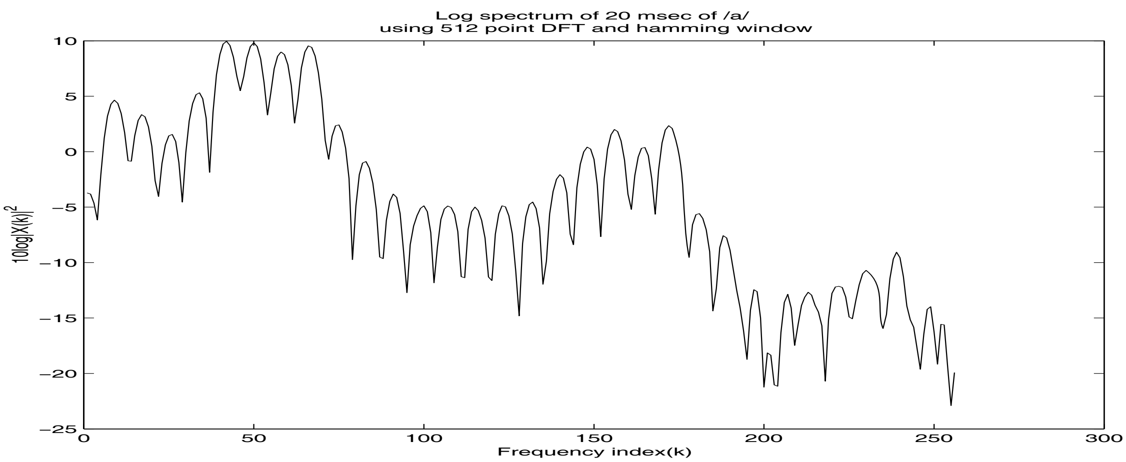

Figure 10: The Log spectrum of x(n), using hamming window of size 20 msec.

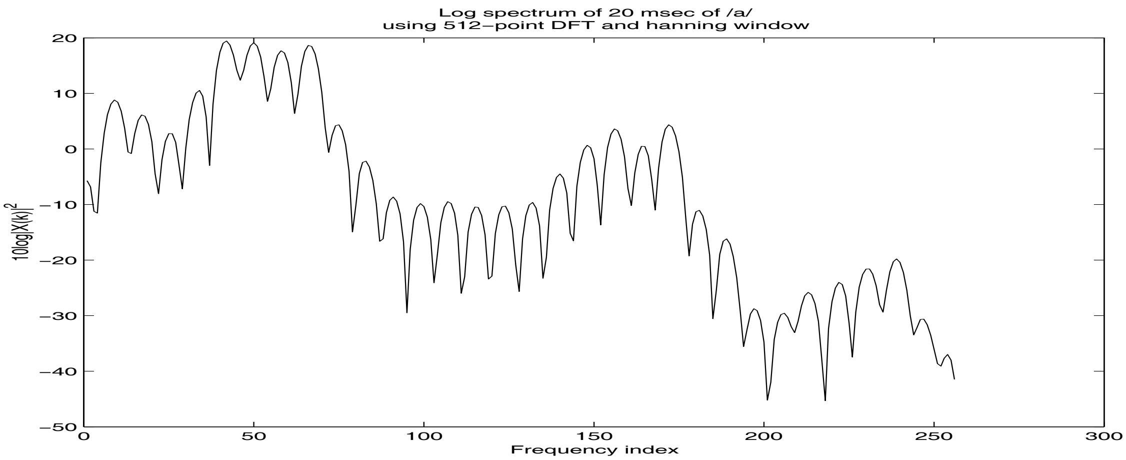

Figure 11: The Log spectrum of x(n), using hanning window of size 20 msec.

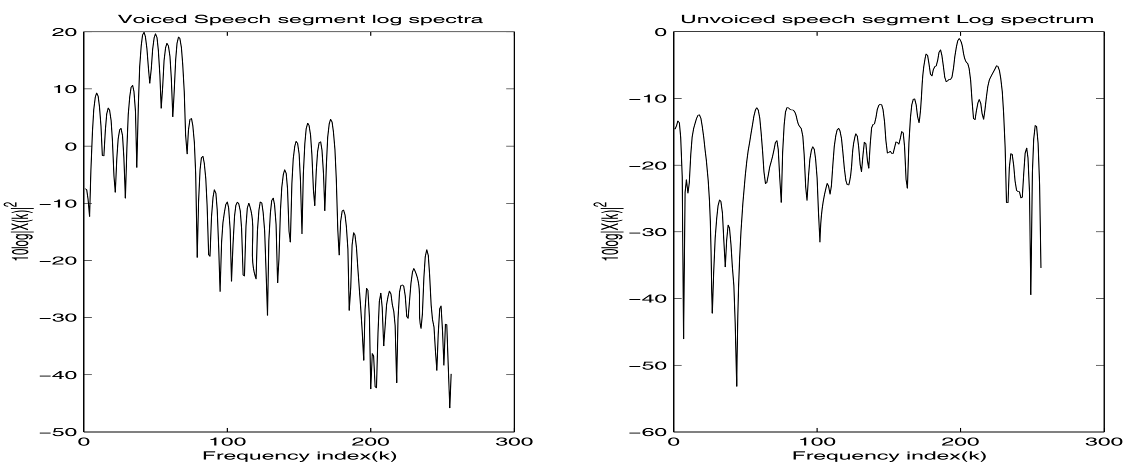

Figure 12: The Log spectrum of a) voiced speech segment b)Unvoiced speech segment.

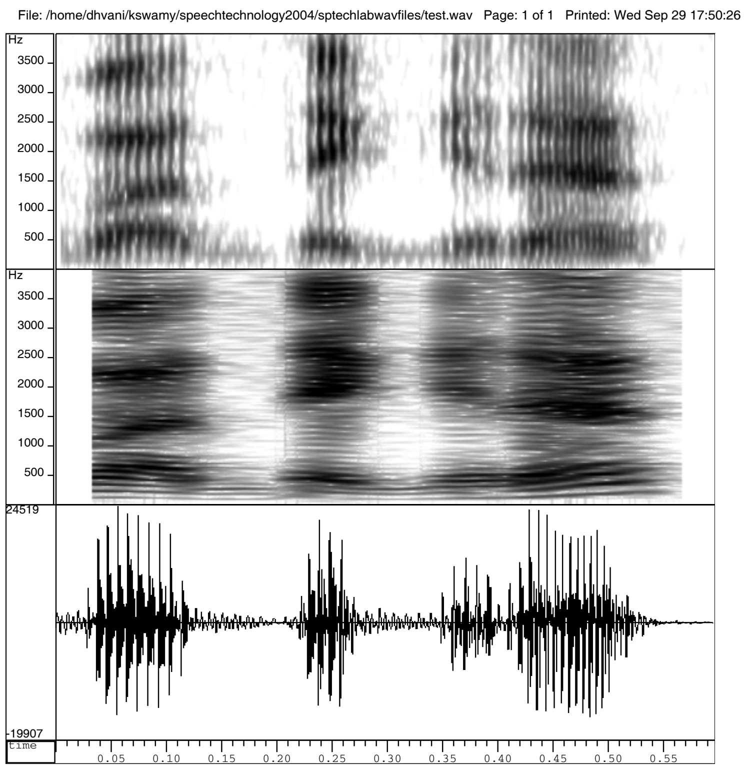

Figure 13: a. Wideband Spectrogram b. Narrowband spectrogram c. Segment of Speech signal