homework #4

Signal to Noise Ratio

EE 8993: Fundamentals of Speech Recognition

December 6, 1998

submitted to:

Dr. Joseph Picone

submitted by:

Suresh Balakrishnama

Institute for Signal and Information Processing
Department of Electrical and Computer Engineering
Mississippi State University
MS 39762, USA
Email: balakris@isip.msstate.edu

1. INTRODUCTION

Signal to noise ratio (SNR) is the most widely used measure for analog and waveform coding systems and also useful for assessing enhancement algorithms for broadband noise distortions. SNR measurements are only appropriate for coding or enhancement system that seek to reproduce the original input waveform.

2. Problem Description

Implement the algorithm described in class to compute the signal to noise ratio using a histogram of the energy distribution.

Validate this design by:

   1. Processing the four files below:

   ece_8993_speech/homework/1996/data/710_b_8k.raw
   ece_8993_speech/homework/1996/data/710_s_8k.raw
   ece_8993_speech/homework/1996/data/711_g_8k.raw
   ece_8993_speech/homework/1996/data/712_f_8k.raw

   and comparing your answers to the results from the class of 1996.

   First, plot the average SNR of the four files for the following
   conditions (do a scatter plot):

     - frame duration of 5, 10, 20, and 40 msec
     - window duration of 10, 20, 30, 60 msec

   Use a signal threshold of 80% and a noise threshold of 20%.

   Next, for the best set of parameters above, plot the average SNR as a

   function of the thresholds:

     - signal threshold 80%, 85%, 90%, 95%;
     - noise threshold 10%, 15%, 20%, 25%

   2. Processing a large chunk of Switchboard:

   /isip/d02/switchboard/data/sw2151_ec.raw

3. Implementation

The main objective here is to process each frame of data so that you cover all the samples when distributing each frame. Each frame of the speech signals for all channels is pre-emphasized using:

$$P(z) = 1 - \mu z^{-1} \tag{1}$$

where $\mu = 0.95$ in our signal to noise calculation. To pre-emphasize a signal means to apply a low pass filter that would increase the relative energy of the high-frequency spectrum. The energy of noise increases in proportional to the square of the channel frequency, by introducing a low pass filter, we would be able to get a more accurate signal to noise ratio. Furthermore, the use of pre-emphasis can eliminate the spectral contributions of the larynx and lips for analysis to seek parameters corresponding to the vocal tract only [1].

Then Hamming window is applied to the signal:

$$y(n) = 0.54 - 0.46 \cos \frac{2\pi n}{M - 1} \tag{2}$$

This is used to smooth the abrupt discontinuity at the window boundaries.

The energy is computed using:

$$E = \sum_{i=0}^{N} \frac{x_i^2}{N+1}. \tag{3}$$

The energy for each frame is stored until all the signals are processed. Then a probability density function (pdf) of the energy values are calculated. A total number of 10,000 bins are used to plot the energy histogram.

4. Results

frame	window			
	10	20	30	60
5	9.134826	8.996714	8.871572	8.542184
10	9.004938	9.030701	8.852352	**8.422152**
20	9.066294	8.928496	8.853244	8.511646
40	8.977531	9.085185	8.775523	8.622317

Table 1: Average signal to noise ratio (in dB) using different window and frame duration for the file 710_b_8k.raw

frame	window			
	10	20	30	60
5	9.134826	8.996714	8.871572	8.542184
10	9.004939	9.030701	8.852352	**8.422152**
20	9.066294	8.928496	8.853244	8.511646
40	8.977531	9.085185	8.775523	8.910591

Table 2: Average signal to noise ratio using different window and frame duration for the file 710_s_8k.raw

frame	window			
	10	20	30	60
5	9.355579	9.332850	9.317355	9.018757
10	9.380290	9.439940	9.363338	**9.009154**
20	9.314431	9.166645	9.301335	9.024555
40	9.632618	9.323182	9.099382	8.913860

Table 3: Average signal to noise ratio using different window and frame duration for the file
710_g_8k.raw

frame	window			
	10	20	30	60
5	9.617392	9.623659	9.514186	9.231970
10	9.677206	9.601563	9.580981	**9.207649**
20	9.745012	9.526969	9.626323	9.258477
40	9.632618	9.820978	9.419830	9.233452

Table 4: Average signal to noise ratio using different window and frame duration for the file
712_f_8k.raws

noise threshold(%)	signal threshold(%)			
	80	85	90	95
10	10.740457	10.740457	10.740457	10.740457
15	9.847071	9.847071	9.847071	9.847071
20	9.134826	9.134826	9.134826	9.134826
25	8.082599	8.082599	8.082599	**8.082599**

Table 5: Average signal to noise ratio (in dB) using different signal and noise threshold for the file 710_b_8k.raw

noise threshold(%)	signal threshold(%)			
	80	85	90	95
10	10.740457	10.740457	10.740457	10.740457
15	9.847071	9.847071	9.847071	9.847071
20	9.134826	9.134826	9.134826	9.134826
25	8.082599	8.082599	8.082599	**8.082599**

Table 6: Average signal to noise ratio (in dB) using different signal and noise threshold for the file 710_s_8k.raw

noise threshold(%)	signal threshold(%)			
	80	85	90	95
10	10.778791	10.778791	10.778791	10.778791
15	10.140963	10.140963	10.140963	10.140963
20	9.632618	9.632618	9.632618	9.632618
25	8.693860	8.693860	8.693860	**8.693860**

Table 7: Average signal to noise ratio(in dB) using different signal and noise threshold for the file 710_g_8k.raw

noise threshold(%)	signal threshold(%)			
	80	85	90	95
10	10.657279	10.657279	10.657279	10.657279
15	10.188071	10.188071	10.188071	10.188071

noise threshold(%)	signal threshold(%)			
	80	85	90	95
20	8.997402	8.997402	8.997402	8.997402
25	8.997402	8.997402	8.997402	**8.997402**

Table 8: Average signal to noise ratio (in dB) using different signal and noise threshold for the file 712_f_8k.raws
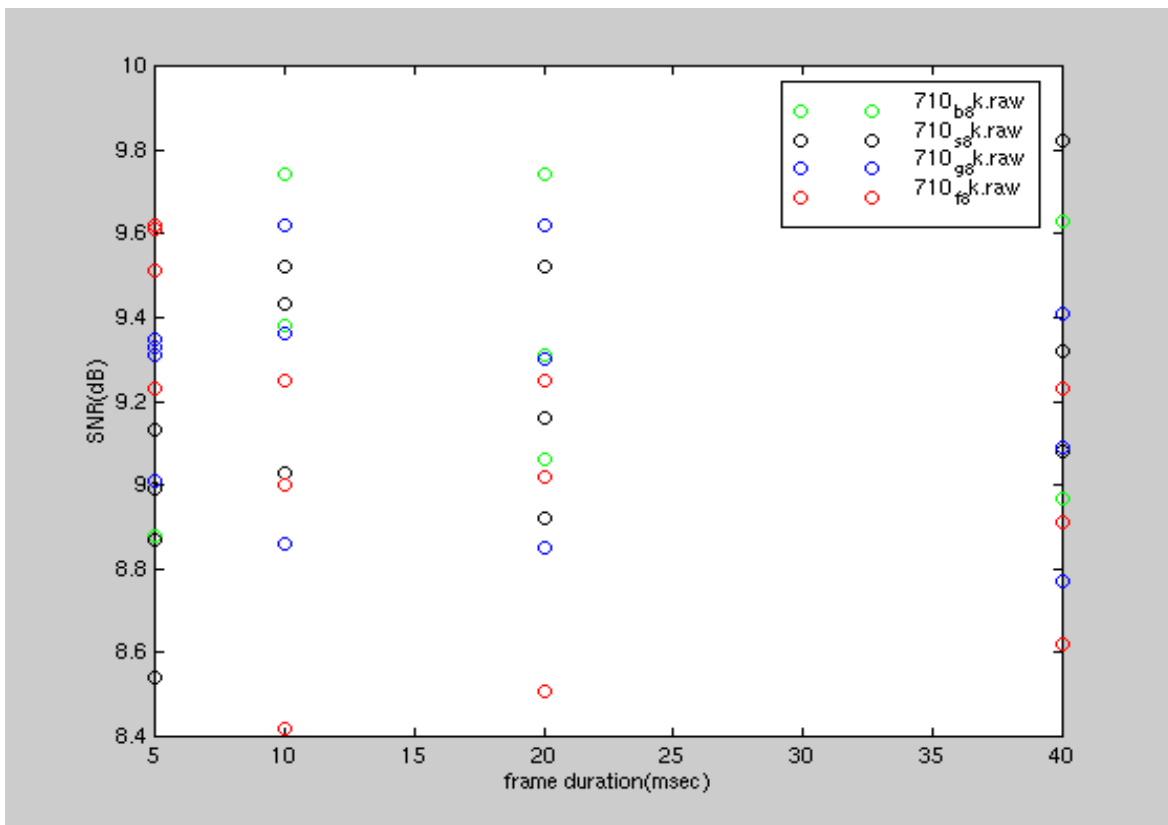
5. PLOTS

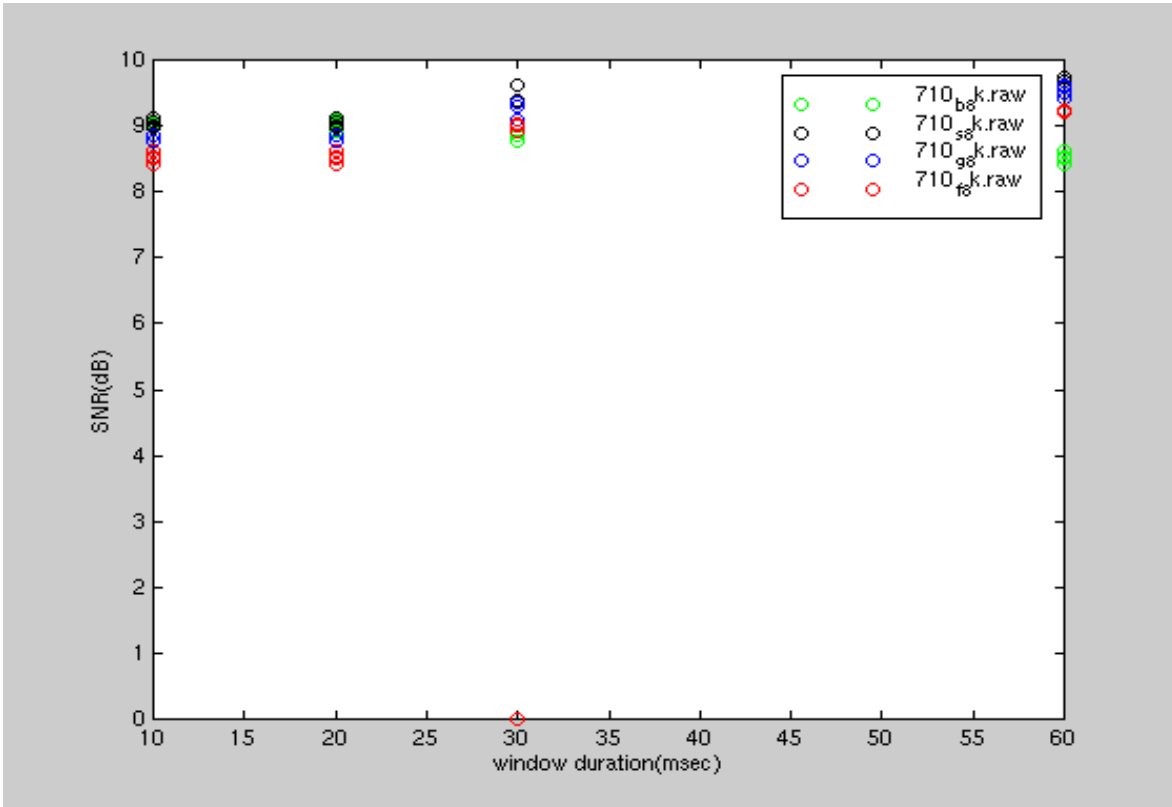Figure 1. Plot showing average SNR against frame duration(msec)

Figure 2. Plot showing average SNR against window duration(msec)
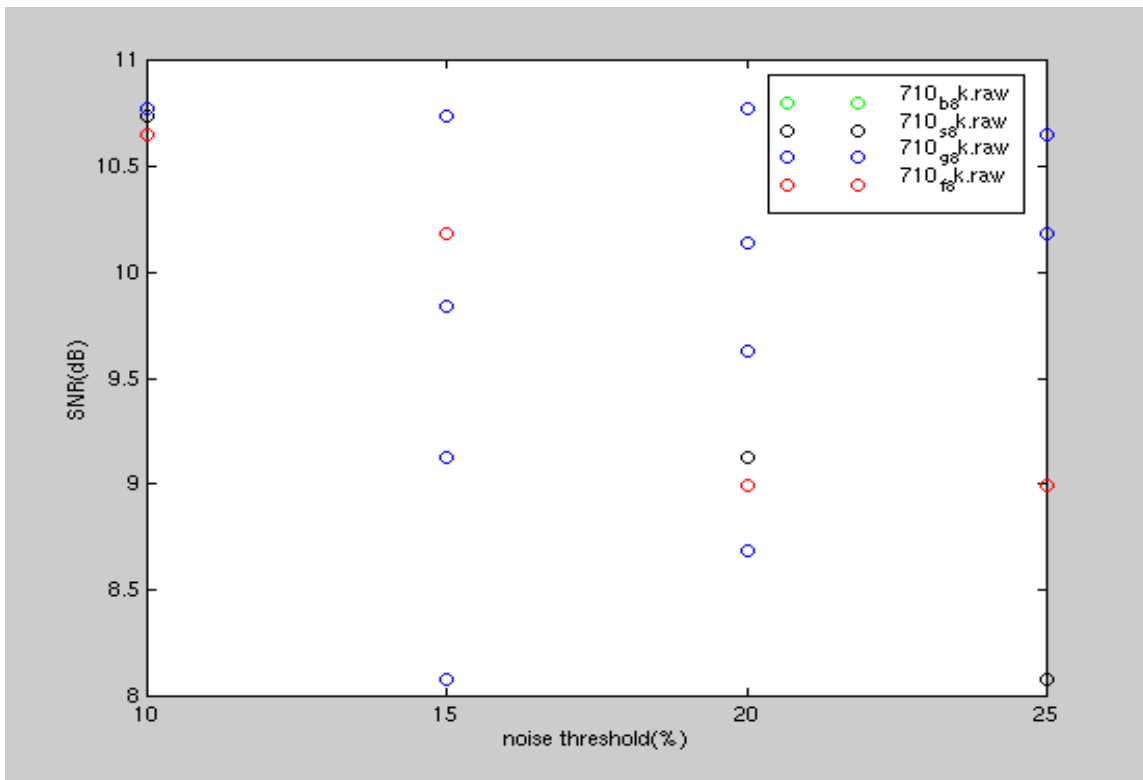

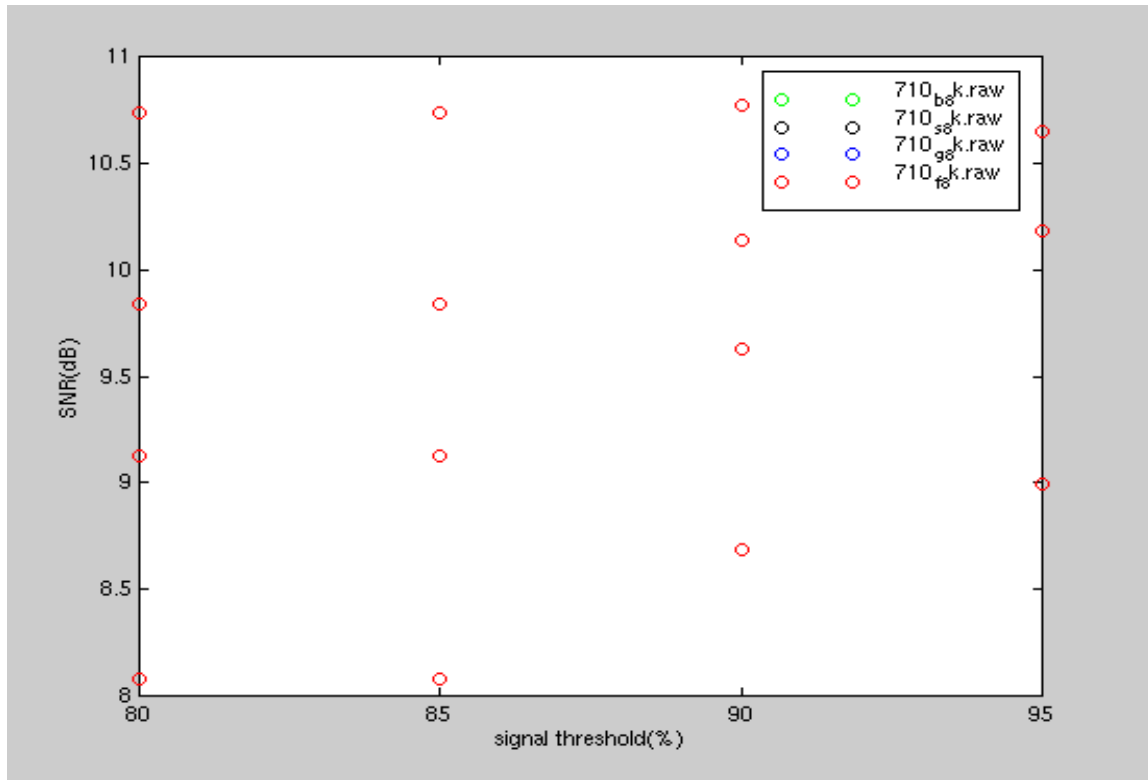Figure 3. Plot showing average SNR against noise threshold(%)

Figure 4.   Plot showing average SNR against signal threshold(%)

6. CONCLUSIONS

We have implemented the algorithm to compute the signal-to-noise ratio of a speech file. For any speech file a low signal-to-noise ratio is desirable.Table 1-8 show the SNR values highlighted corresponding to the optimum window and frame duration for the signal for each of the speech file given in the assignment. In next part, holding the frame and window duration to be constant, the noise and signal threshold are varied for each file. The SNR remains constant when signal threshold is increased whereas SNR decreases when noise threshold is increased. This means that the energy assigned to signal+noise is inversely related to energy of noise which is not an optimum SNR estimator and hence these window duration and frame duration selected do not infer perfect parameters for SNR estimator..

7. SOFTWARE

All Matlab code written for this project is available for public from our website at *www.isip.msstate.edu*

8. REFERENCES

[1]   F. Jelinek, Statistical Methods for Speech Recognition, *The MIT Press*, Cambridge, Massachusetts, USA.

[2]   J.Deller, J.G.Proakis and J.Hansen, "Discrete-time processing of speech signals", *Macmillan Publishing Company*, New York, USA.