homework solutions for:

Homework #4: Signal-to-Noise Ratio Estimation

submitted to:

Dr. Joseph Picone
ECE 8993 Fundamentals of Speech Recognition

May 3, 1998

submitted by:

Jonathan Hamaker

Institute for Signal and Information Processing
Department of Electrical and Computer Engineering
Mississippi State University
Box 9571, 216 Simrall, Hardy Rd.
Mississippi State, Mississippi 39762
Tel: 601-325-8335, Fax: 601-325-3149
Email: hamaker@isip.msstate.edu

Introduction

The signal-to-noise ratio (SNR), given by (1), is an important feature in determining the quality of audio data. This is particularly important in speech recognition technology since it is well known that recognition performance is strongly influenced by the SNR. Unfortunately, in most applications the SNR cannot be easily derived since the noise energy is not known. Further, the question arises as to what is “signal” and what is “noise”. For example, would a cough or breath noise be considered part of the “signal” in spontaneous speech? Does it convey information? With these problems in mind, we must define a statistically oriented method which makes a best estimate of the SNR given the a priori knowledge of the speech data. One such method which uses a short-term analysis of the speech signal to statistically characterize the signal and the noise is analyzed in this homework assignment.

$$SNR = 10\log\frac{\text{Signal Energy}}{\text{Noise Energy}} = 10\log\frac{E_S}{E_N} \tag{1}$$

The method we will use for the SNR estimation is based on the histogram analysis of a speech signal such as that shown in Figure 1. Ideally, we would expect to see two major modalities in the energy histogram as shown in Figure 2. These two modalities correspond to the nominal noise energy and the nominal energy of the signal in the presence of noise, respectively. Obtaining a cumulative histogram of this data, as shown in Figure 2, we can define thresholds which select the percentage of data points which we expect to correspond to noise energy and the percentage of data points which we expect to correspond to energy of the signal in the presence of noise. Typical thresholds used for the percent signal + noise / percent noise are 80%/20%, 85%/15% or 95%/ 15%. These values have been derived by experienced speech researchers based on analyses of many types of data. With this methodology we define (2) which specifies the estimated SNR based on short-term energy measures.

$$SNR = 10\log\frac{(E_S + E_N) - E_N}{E_N} \tag{2}$$

There is one detail that we have overlooked to this point: how do we get the short-term measurements. This requires that we decide on an optimum window and frame width to yield consistent and accurate SNR estimates for the given data set. The remainder of this paper is devoted to considering how each of the parameters described above effect the SNR estimate.

Problem Statement

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

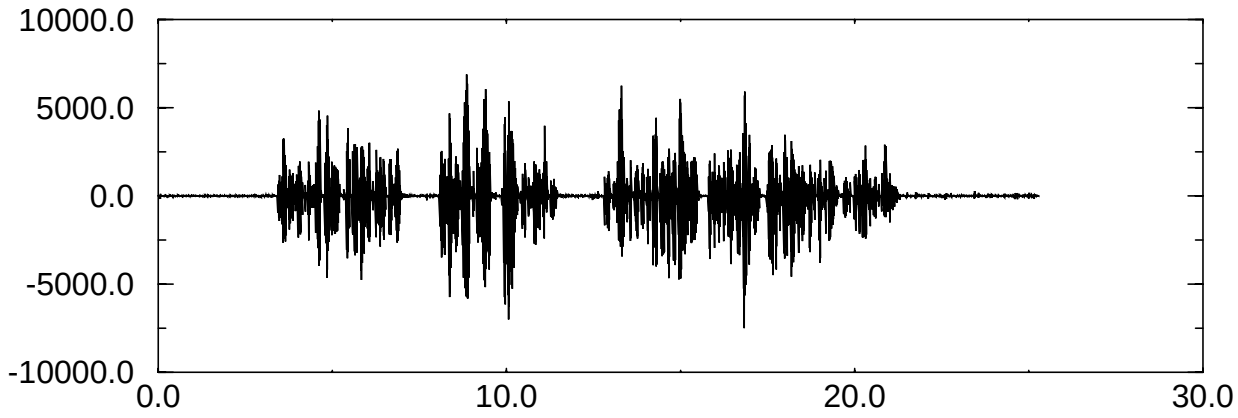

Figure 1. Speech signal with approximately 30dB SNR

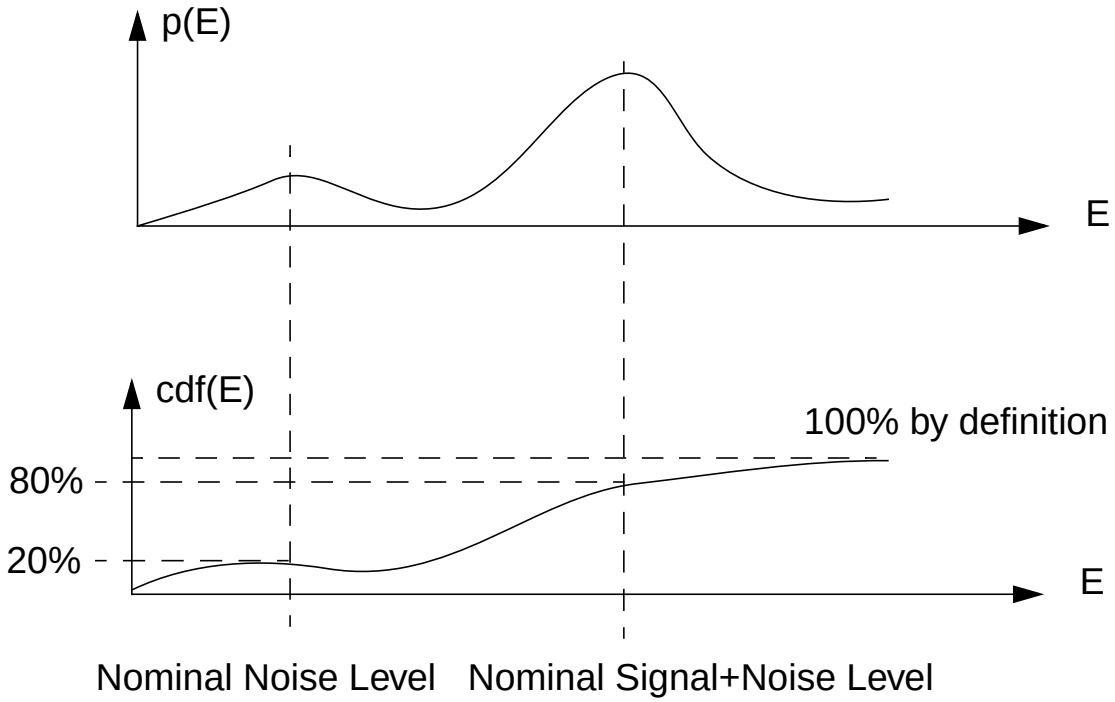

Figure 2. Energy probability density function (pdf) and energy cumulative density function (cdf). Note the two modalities shown in this ideal case.

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

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. Process a large chunk of the left side of a Switchboard conversation.

Methodology

For each frame of data, we apply a pre-emphasis filter to emphasize the high-frequency components of the signal. The pre-emphasis filter is given by $H(z) = 1 - \mu z^{-1}$ where μ is typically around 0.95. Secondly, we center the window about the frame to collect a windowed sample of the data. The default window used is rectangular, but can optionally be set to a Hamming window given by (3). The Hamming window is used to smooth abrupt discontinuities at the frame boundaries.

$$w(n) = 0.54 - 0.46\cos\frac{2\pi n}{M-1} \tag{3}$$

For each window we calculate the average energy using (4). We use the energy values for all windows in the speech data to generate the cumulative distribution of energies in the speech data using 10000 bins ranging from the minimum average energy found in the file to the maximum average energy found in the file. One might wonder why there would need to be such a large number of bins if the data is distributed as suggested by Figure 2. The reason for this is demonstrated in Figure 3. Notice that there is not the ideal noise and signal+noise modalities. Rather there is a smooth transition from low energy bands to high energy bands and the low energy bands are very dense, causing several orders of magnitude of energy values to fall in the noise region when a small number of bins are used.

$$E = \sum_{i=0}^{N-1} \frac{x_i^2}{N} \tag{4}$$

Results

To observe the variance of SNR relative to the design parameters, we hold the signal threshold constant at 80% and the noise threshold constant at 20% while varying the window length and the

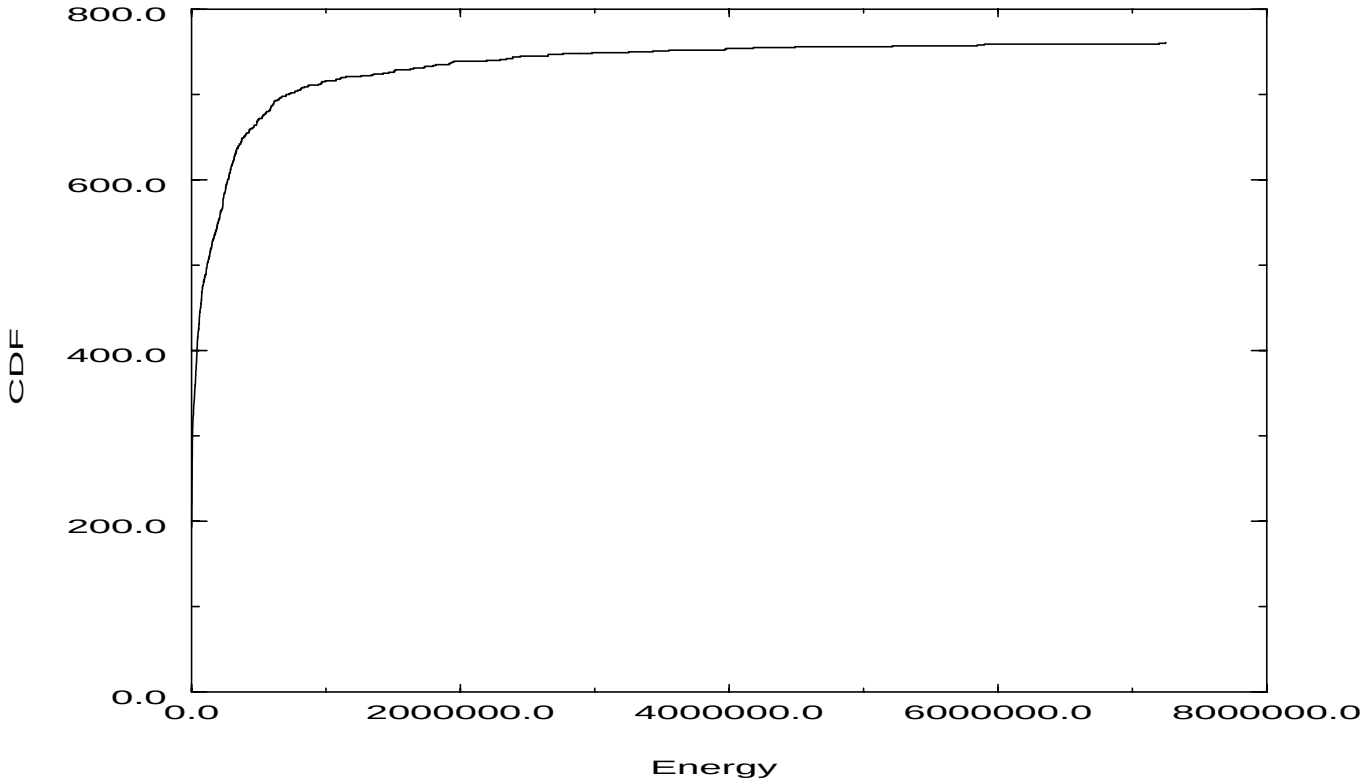

Figure 3. Energy cumulative density function (cdf) for one of the data samples studied in this experiment. Note that most energy is concentrated into a few lower valued bins and that there is not a clear indication of two modalities as was shown in the ideal case.

frame length. Results for each of the four files are shown in Tables 1-4. Table 5 gives the average SNR across all files and Figures 5 through 8 plot the SNR for each file across all window and frame lengths. Lastly, Table 6 gives the variance across all files. Consistent measures of SNR require that we minimize the variance of the SNR output across speech samples. Thus, from Table 6, we choose a window size of 20 ms and a frame size of 20 ms to be the optimal window and frame length.

Holding the window and frame size constant to these values, we now vary the signal+noise and the noise threshold. Results for this are shown in Tables 7-10. We note that as the signal and noise thresholds get further apart, the SNR increases. This is intuitive since this means that the energy assigned to the signal+noise is increasing relative to the energy assigned to the noise. Table 11 shows the variance of the data across files and Figures 9-12 give the average SNR for each file across each signal+noise and noise threshold. We see that a signal threshold of 80% coupled with a noise threshold of 10%, yields the lowest variance for a window duration of 20 ms and a frame duration of 20 ms. Thus for consistent performance, we would choose our SNR estimator to have these parameters.

As a final experiment, we use the SNR estimation routines to estimate the signal-to-noise ratio of

a the left side of the SWITCHBOARD conversation 2151. Results for this experiment are shown in Tables 12 and 13. We would expect the SNR to hover around 30 dB for telephone quality data. However, we see that the results are not quite this high. This can be attributed to the large spans of silence found in the file. Since silence is not considered to be part of the “signal”, it contributes to the noise energy — thus decreasing the average SNR across the file. The stem plots for varying window and frame duration as well as varying thresholds are shown in Figures 13 and 14 respectively.

Software

All software for this assignment was written using C++ and can be found at: *http://www.isip.msstate.edu/resources/courses/ece_8993_speech/1998/problem_04/hamaker/src/.* The command interface is shown in Figure 4.

```
name: snr_calculator
synopsis: snr_calculator [options]
descr: produces the overall signal-to-noise ratio of the speech signal
example: snr_calculator -num_chans 2 -sf 8000 -sig_thresh 0.80 -noise_thresh 0.1
-frame_dur 10 -window_dur 20 -input input_file.raw

arguments:
input_file.raw: 16-bit linear input file

options:
-num_chans:    number of channels in the audio file (default: 1)
-sf:           sample frequency of the data (default: 8000)
-swap_bytes:   flag indicating whether the byte ordering should be
                    swapped
-sig_thresh:   percent of energy considered as signal+noise
                    (default: 0.85)
-noise_thresh: percent of energy considered as noise (default: 0.15)
-frame_dur:    frame duration to use in computing the short duration
                    snr (default: 10 msec)
-window_dur:   window duration to use in computing the short duration
                    snr (default: 20 msec)
-use_hamming: flag indicating whether or not to use a hamming window
-use_pre_emph: flag indicating whether or not to pre-emphasize the data

man page: none
```

Figure 4. Command-line interface to the SNR calculation tool.

Frame Duration (ms)	Window Duration (ms)			
	10	20	30	60
5	15.81	16.22	16.26	16.67
10	16.34	16.46	16.23	16.64
20	—	16.22	16.68	16.60
40	—	—	—	17.01

Table 1: Estimated SNR as a function of frame and window durations for 710_b_8k.raw with a signal threshold of 80% and a noise threshold of 20%.

Frame Duration (ms)	Window Duration (ms)			
	10	20	30	60
5	30.68	30.26	30.39	30.73
10	30.67	30.15	30.32	30.70
20	—	29.97	30.06	30.70
40	—	—	—	30.80

Table 2: Estimated SNR as a function of frame and window durations for 710_s_8k.raw with a signal threshold of 80% and a noise threshold of 20%.

Frame Duration (ms)	Window Duration (ms)			
	10	20	30	60
5	9.27	9.90	9.94	9.91
10	9.42	9.76	10.09	9.88
20	—	9.84	9.89	9.83
40	—	—	—	9.81

Table 3: Estimated SNR as a function of frame and window durations for 711_g_8k.raw with a signal threshold of 80% and a noise threshold of 20%.

Frame Duration (ms)	Window Duration (ms)			
	10	20	30	60
5	10.15	10.34	10.38	10.46
10	10.20	10.22	10.44	10.45
20	—	10.24	10.16	10.44
40	—	—	—	10.50

Table 4: Estimated SNR as a function of frame and window durations for 712_f_8k.raw with a signal threshold of 80% and a noise threshold of 20%.

Frame Duration (ms)	Window Duration (ms)			
	10	20	30	60
5	16.48	16.68	16.74	16.94
10	16.66	16.65	16.77	16.92
20	—	16.57	16.70	16.89
40	—	—	—	17.03

Table 5: Average SNR across all files as a function of the window and frame size with a signal threshold of 80% and a noise threshold of 20%.

Frame Duration (ms)	Window Duration (ms)			
	10	20	30	60
5	98.05	90.26	91.09	93.88
10	96.84	90.37	89.53	93.80
20	—	**88.35**	89.21	94.08
40	—	—	—	94.79

Table 6: Variance of the SNR for varying values of window duration and frame duration with the signal+noise threshold fixed at 80% and the noise threshold fixed at 20%.

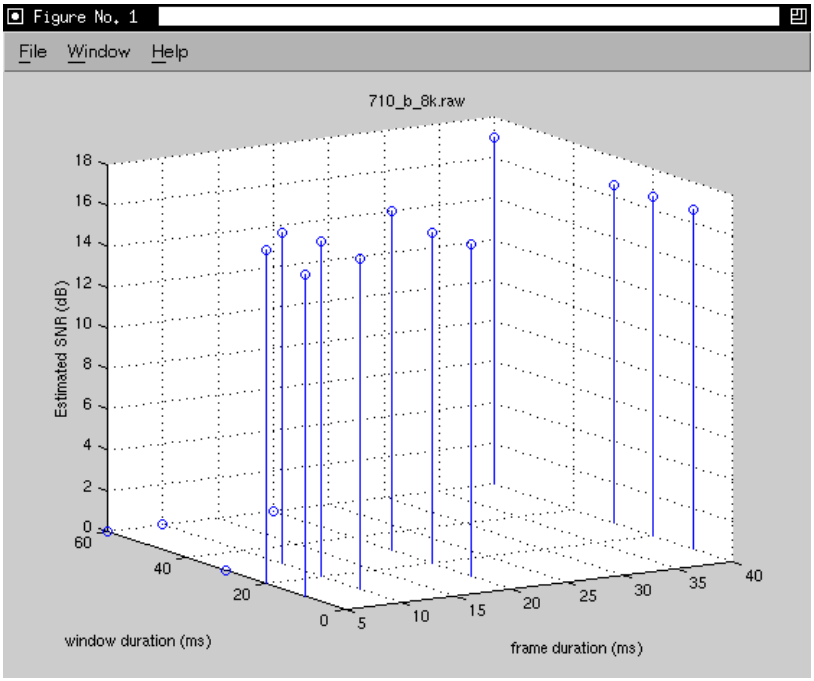

Figure 5. Stem plot of SNR vs window duration and frame duration for 710_b_8k.raw with the signal+noise and noise thresholds held constant at 80% and 20%, respectively.

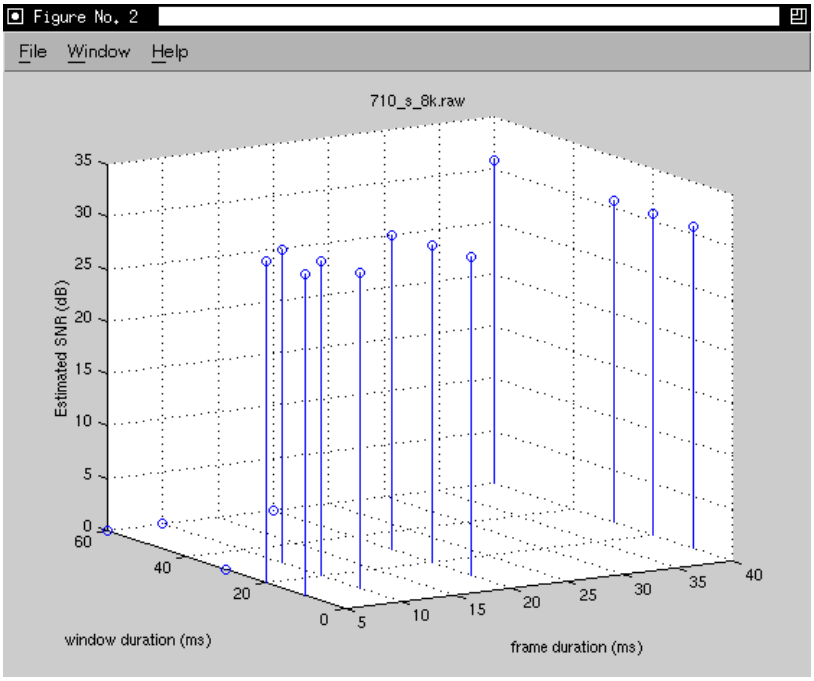

Figure 6.Stem plot of SNR vs window duration and frame duration for 710_s_8k.raw with the signal+noise and noise thresholds held constant at 80% and 20%, respectively.

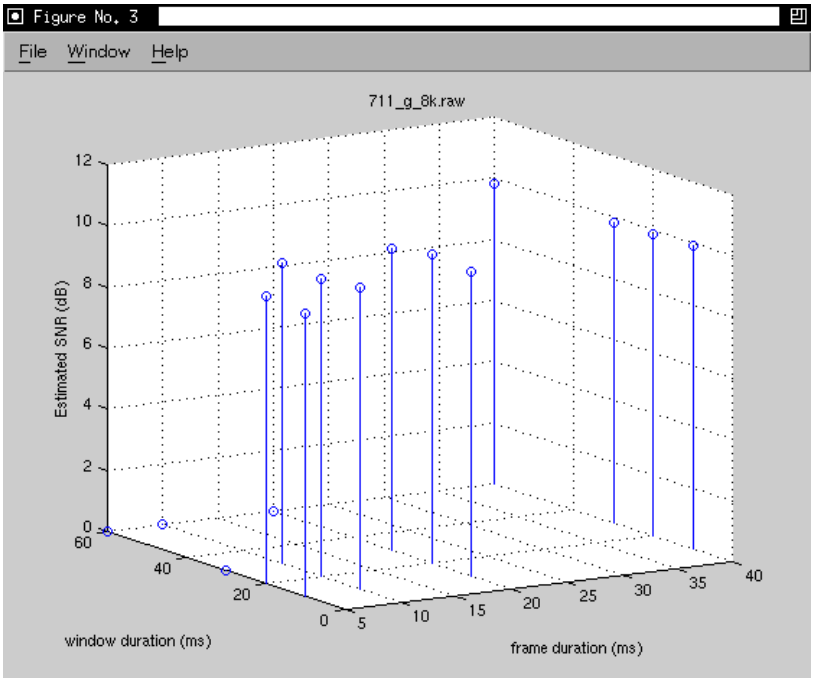

Figure 7.Stem plot of SNR vs window duration and frame duration for 711_g_8k.raw with the signal+noise and noise thresholds held constant at 80% and 20%, respectively.

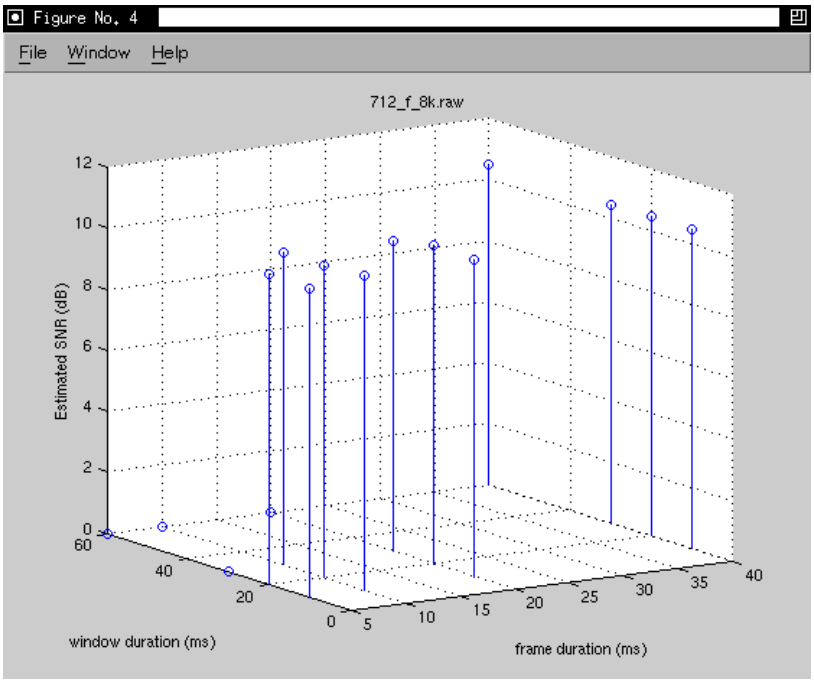

Figure 8. Stem plot of SNR vs window duration and frame duration for 712_f_8k.raw with the signal+noise and noise thresholds held constant at 80% and 20%, respectively.

Noise Threshold (%)	Signal+Noise Threshold (%)			
	80	85	90	95
10	16.84	18.37	20.34	22.96
15	16.84	18.37	20.34	22.96
20	16.22	17.76	19.73	22.36
25	15.60	17.14	19.11	21.74

Table 7: Estimated SNR as a function of signal and noise thresholds for 710_b_8k.raw with the frame duration set at 20 ms and the window duration fixed at 20 ms.

Noise Threshold (%)	Signal+Noise Threshold (%)			
	80	85	90	95
10	29.97	31.57	32.92	35.61
15	29.97	31.57	32.92	35.61
20	29.97	31.57	32.92	35.61
25	29.97	31.57	32.92	35.61

Table 8: Estimated SNR as a function of signal and noise thresholds for 710_s_8k.raw with the frame duration set at 20 ms and the window duration fixed at 20 ms.

Noise Threshold (%)	Signal+Noise Threshold (%)			
	80	85	90	95
10	10.47	11.49	13.12	14.71
15	10.25	11.27	12.91	14.50
20	9.84	10.91	12.55	14.15
25	9.51	10.54	12.20	13.81

Table 9: Estimated SNR as a function of signal and noise thresholds for 711_g_8k.raw with the frame duration set at 20 ms and the window duration fixed at 20 ms.

Noise Threshold (%)	Signal+Noise Threshold (%)			
	80	85	90	95
10	11.02	12.33	13.63	15.15
15	10.64	11.95	13.26	14.78
20	10.24	11.57	12.88	14.41
25	9.98	11.31	12.62	14.16

Table 10: Estimated SNR as a function of signal and noise thresholds for 712_f_8k.raw with the frame duration set at 20 ms and the window duration fixed at 20 ms.

Noise Threshold (%)	Signal+Noise Threshold (%)			
	80	85	90	95
10	**82.21**	86.01	84.98	95.39
15	84.75	88.61	87.55	98.18
20	88.35	91.94	90.76	101.35
25	91.13	94.92	93.51	104.01

Table 11: Variance of the SNR as a function of signal and noise thresholds across all data with the frame duration set at 20 ms and the window duration fixed at 20 ms.

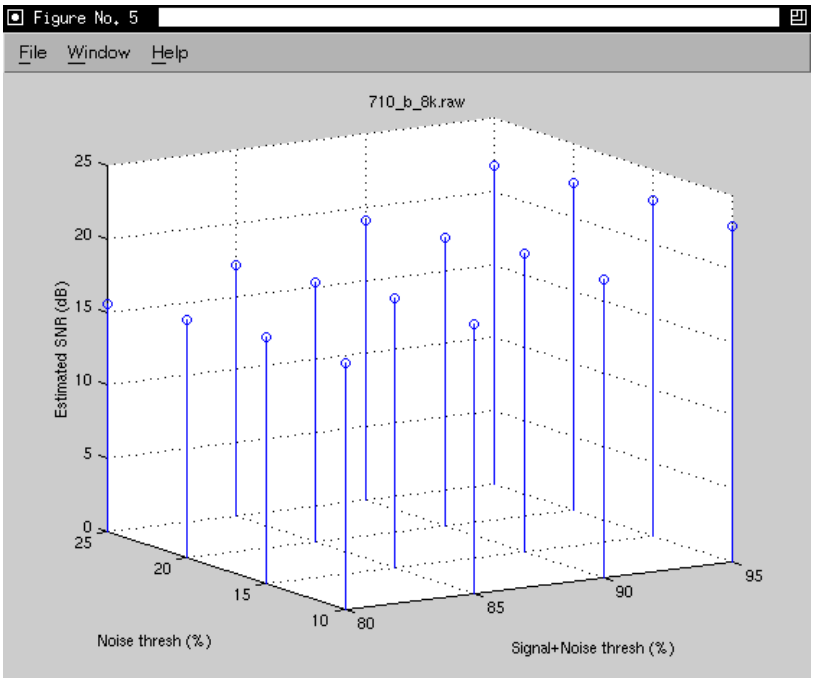

Figure 9. Stem plot of SNR vs signal+noise threshold and noise threshold for 710_b_8k.raw with the frame duration and window duration both held constant at 20 msecs.

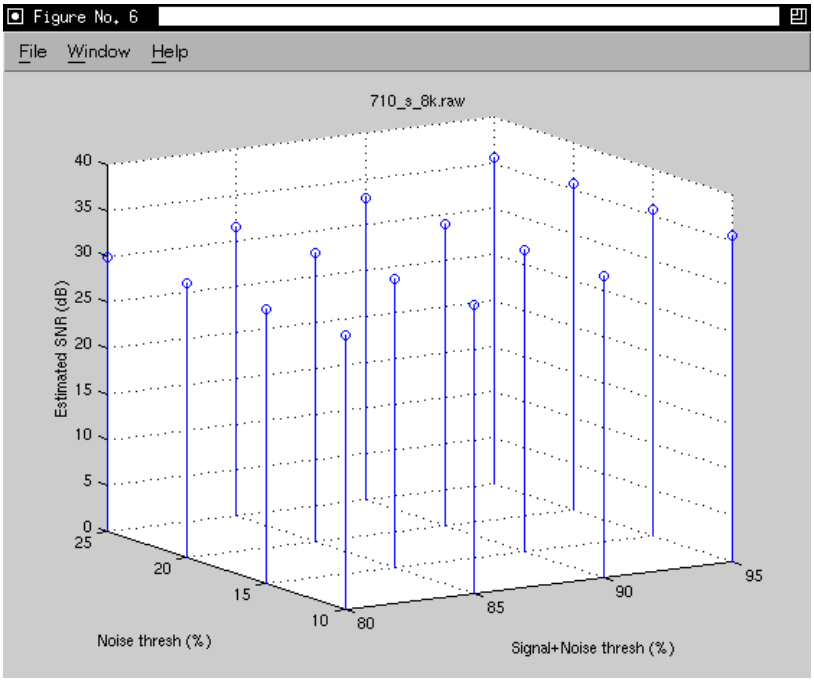

Figure 10.Stem plot of SNR vs signal+noise threshold and noise threshold for 710_s_8k.raw with the frame duration and window duration both held constant at 20 msecs.

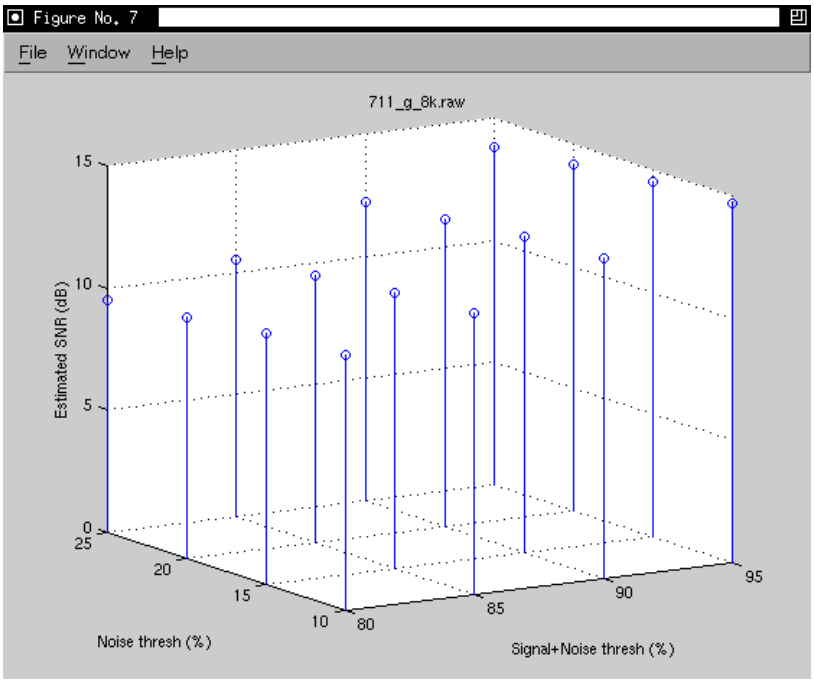

Figure 11. Stem plot of SNR vs signal+noise threshold and noise threshold for 711_g_8k.raw with the frame duration and window duration both held constant at 20 msecs.

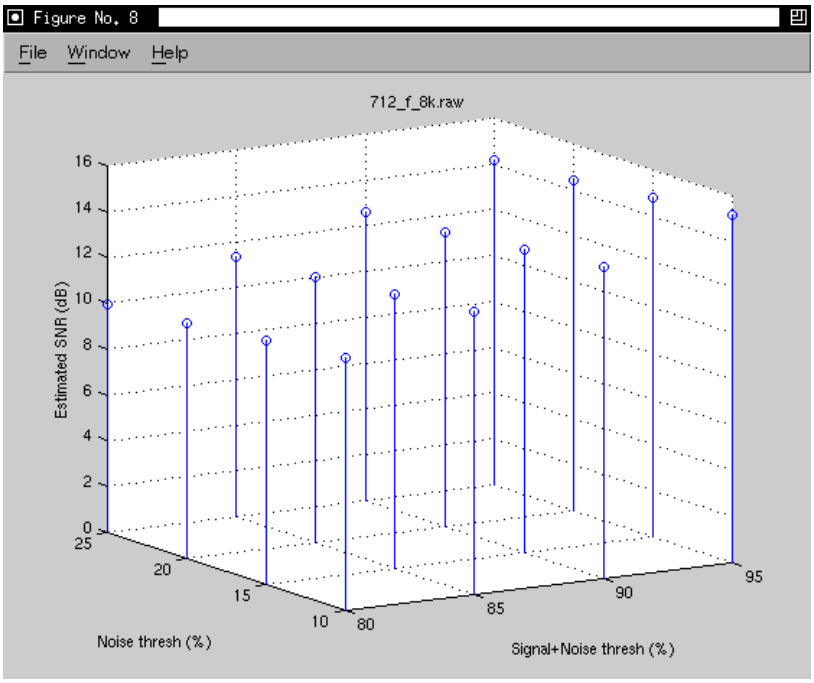

Figure 12.Stem plot of SNR vs signal+noise threshold and noise threshold for 712_f_8k.raw with the frame duration and window duration both held constant at 20 msecs.

Frame Duration (ms)	Window Duration (ms)			
	10	20	30	60
5	21.51	22.36	22.99	24.49
10	21.53	23.47	23.00	24.76
20	—	23.55	22.93	24.95
40	—	—	—	27.30

Table 12: SNR values for SWITCHBOARD conversation 2151 as a function of window duration and frame duration and the signal+noise threshold fixed at 80% and the noise threshold set at 20%.

Noise Threshold (%)	Signal+Noise Threshold (%)			
	80	85	90	95
10	23.55	25.19	27.33	29.96
15	23.55	25.19	27.33	29.96
20	23.55	25.19	27.33	29.96
25	23.55	25.19	27.33	29.96

Table 13: SNR values for SWITCHBOARD conversation 2151 as a function of signal+noise threshold and noise threshold with the window duration and frame duration both set to 20 msecs.

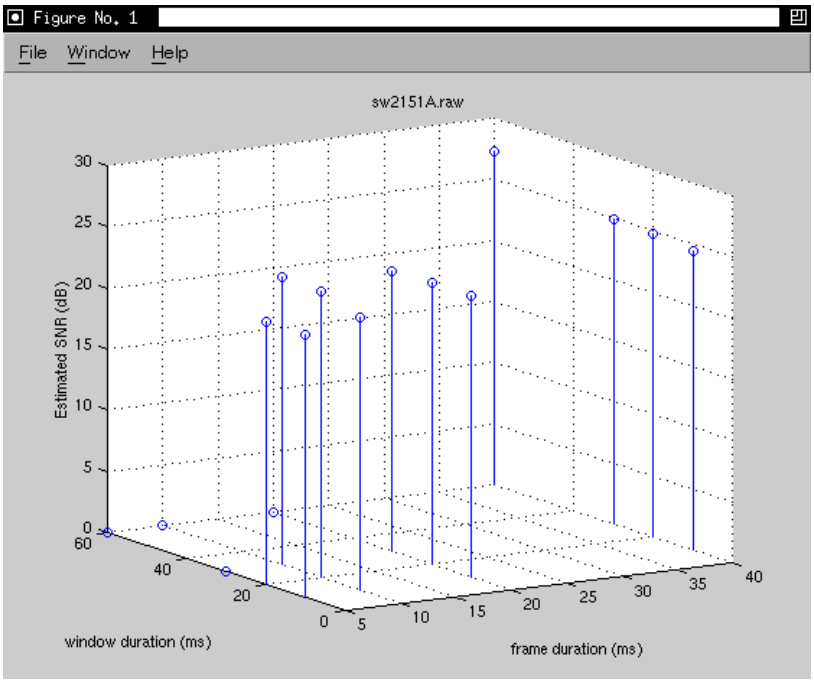

Figure 13. Stem plot of SNR vs frame and window durations for the left channel of sw2151.raw with a constant signal+noise threshold of 80% and a constant noise threshold of 20%.

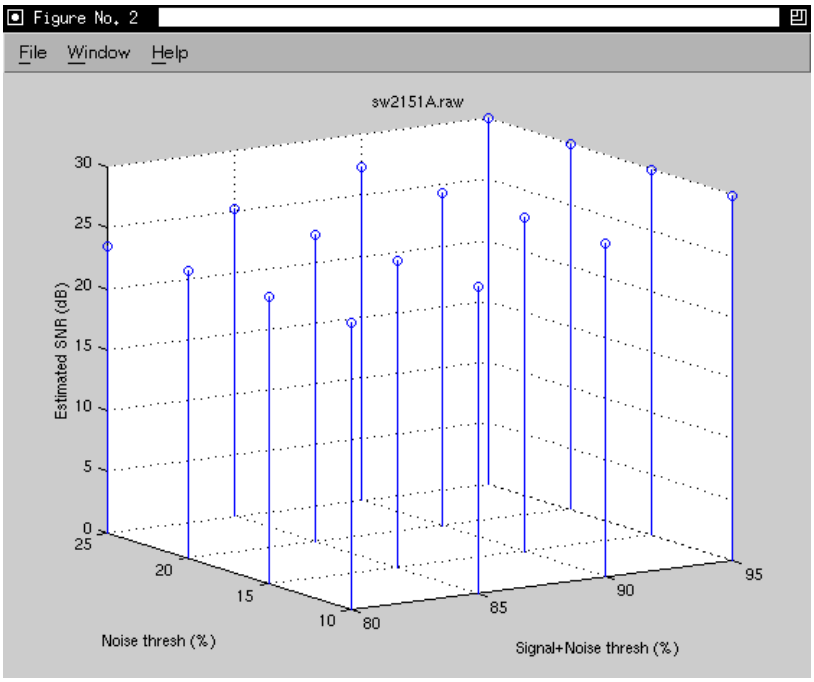

Figure 14.Stem plot of SNR vs signal+noise threshold and noise threshold for the left channel of sw2151.raw with the frame duration and window duration both held constant at 20 msecs.