

Lab9 Digital Coding of Speech Signals-part1

张旭东 12011923

1.Introduction

Speech coding is to encode the analog speech signal and convert analog speech signal to digital signal, which can reduce the error rate of transmission and carry out digital transmission. The basic methods of speech coding can be divided into waveform coding and parameters coding. Waveform coding is to do sampling, quantization, coding on analog speech signal in time domain and convert them to digital voice signal. While parameters coding is to find out the characteristic parameters of the speech and encode the characteristic parameters based on the pronunciation mechanism of human language.

The purpose of this experiment is to test and understand the statistical model of speech signal and the process of uniform quantization.

2.Lab result and analysis

2.1 Problem1

The goal of this MATLAB exercise is to verify some aspects of the statistical model of speech.

In this exercise, three concatenated speech files, named `out_s1_s6.wav`, `out_male.wav` and `out_female.wav` are used.

- `out_s1_s6.wav` -the concatenation of six individual speech files(with beginning and ending silence regions removed), namely `s1.wav`, `s2.wav`, `s3.wav`, `s4.wav`, `s5.wav` and `s6.wav`.
- `out_male.wav` -the concatenation of four individual male speech files(with beginning and ending silence regions removed), namely `s2.wav`, `s4.wav`, `s5.wav`, `s6.wav`.

- `out_female.wav` -the concatenation of two individual female speech files(with beginning and ending silence regions removed), namely `s1.wav`, `s3.wav`.

The first step need to do is to obtain the three above speech files. According to the waveforms of six individual speech files, the voiced fragment is `s1(4537:22560)`, `s2(1335:18601)`, `s3(3693:24000)`, `s4(3623:20031)`, `s5(1276:19318)`, `s6(263:16504)`. Then the three above speech files are concatenated according to the description.

After that, the large speech file is treated as a source of statistical speech samples, the mean, variance, minimum and maximum of the speech signal are gotten and the histograms of the speech amplitudes using different bins are plotted.

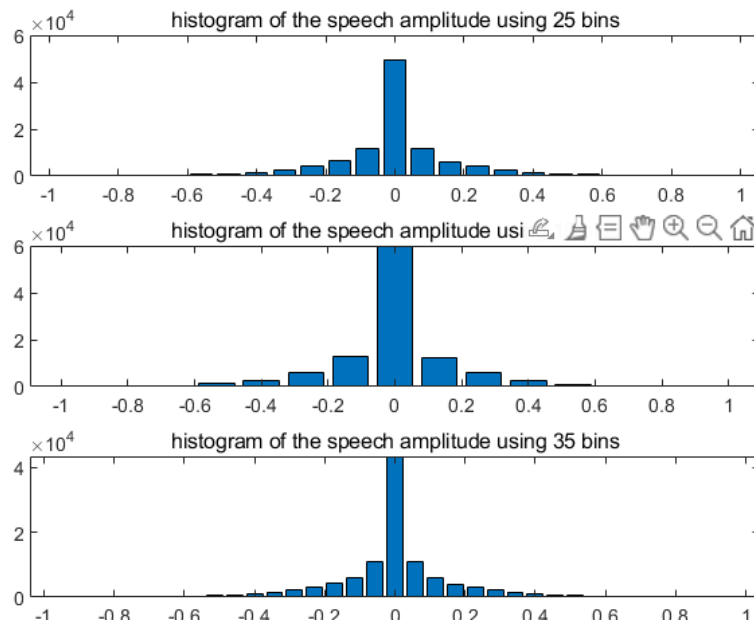


Fig.1 histograms of the speech amplitudes with different bins

The mean is 1.6186×10^{-4} and the variance is 0.0332. The minimum is -1.00 and the maximum is 1.00 .

From Fig 1, it is obvious that with the value of bins increasing, the number of partitions getting larger and larger and the result is more accurate.

Then, the m-file `pspect.m` is used to compute an estimate of the long-term average power spectrum of speech using the signal in the long concatenated file `out_s1_s6.wav`. A range of window durations is experimented to determine the effect of window size on the smoothness of the power spectrum. The result is as shown in Fig 2.

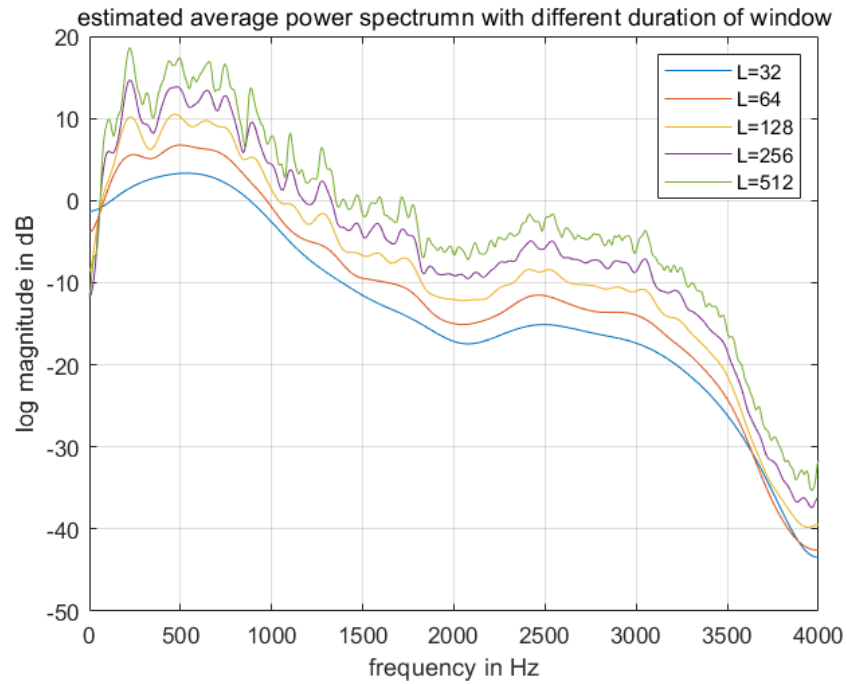


Fig.2 average power spectrum with different window durations

From the figure, what can be known is that with the duration of window increasing, the fluctuation is more violent and the more details are displayed. When the duration of window is longer, the frame of the short-time Fourier transform is longer and the accuracy in the frequency domain is higher. And the number of samples included in the frame is more, which cause the overall energy is larger.

Finally, the m-file `pspect.m` is used to compute an estimate of the long-term average power spectrum of speech using the male speech file and the female speech file for the window duration of 32 samples. The result is as shown in Fig 3.

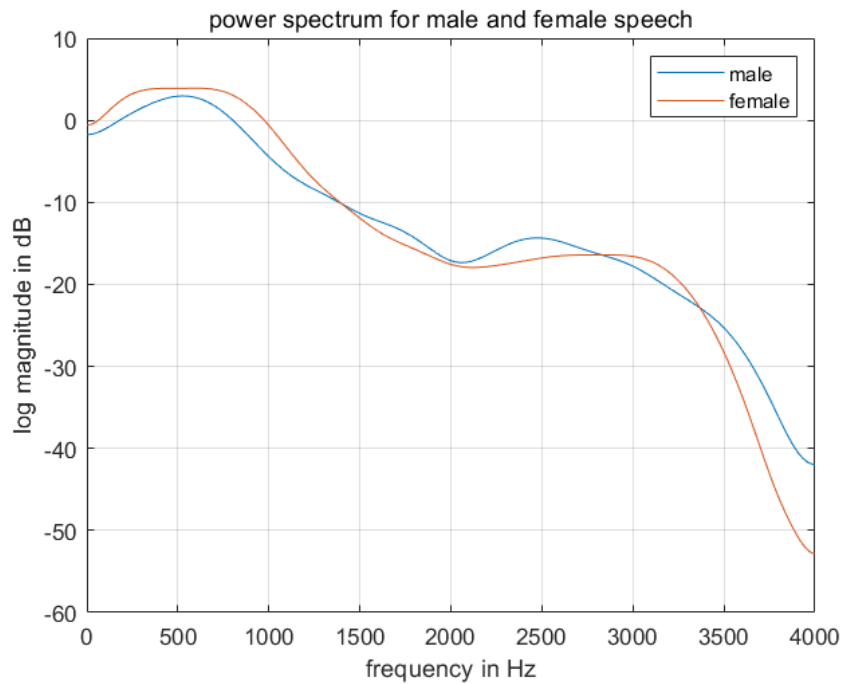


Fig.3 power spectrum for male and female speech file

From the above figure, it is easily known that when the frequency is smaller than 1400 Hz(estimated value), the energy contained in the female speech file is larger than that contained in the male speech file. What's more, the energy contained in the female speech file is less overall than that contained in the male speech file at the high frequency.

The whole code is as below:

```
[s1,fs1]=audioread('s1.wav');
[s2,fs2]=audioread('s2.wav');
[s3,fs3]=audioread('s3.wav');
[s4,fs4]=audioread('s4.wav');
[s5,fs5]=audioread('s5.wav');
[s6,fs6]=audioread('s6.wav');
out_s1_s6=[s1(4537:22560)' s2(1335:18601)' s3(3693:24000)' s4(3623:20031)'
s5(1276:19318)' s6(263:16504)'];
out_male=[s2(1335:18601)' s4(3623:20031)' s5(1276:19318)' s6(263:16504)'];
out_female=[s1(4537:22560)' s3(3693:24000)'];
audiowrite('out_s1_s6.wav',out_s1_s6,fs6);
audiowrite('out_male.wav',out_male,fs6);
audiowrite('out_female.wav',out_female,fs6);
```

```

mean_s6=mean(out_s1_s6);
disp(mean_s6);
var_s6=var(out_s1_s6);
disp(var_s6);
max_s6=max(out_s1_s6);
disp(max_s6);
min_s6=min(out_s1_s6);
disp(min_s6);
[hist1,center1]=hist(out_s1_s6,25);
[hist2,center2]=hist(out_s1_s6,15);
[hist3,center3]=hist(out_s1_s6,35);
figure;
subplot(3,1,1);bar(center1,hist1);
title('histogram of the speech amplitude using 25 bins');
subplot(3,1,2);bar(center2,hist2);
title('histogram of the speech amplitude using 15 bins');
subplot(3,1,3);bar(center3,hist3);
title('histogram of the speech amplitude using 35 bins');

figure;
pspect(out_s1_s6,fs,1024,32);
hold on;
pspect(out_s1_s6,fs,1024,64);
hold on;
pspect(out_s1_s6,fs,1024,128);
hold on;
pspect(out_s1_s6,fs,1024,256);
hold on;
pspect(out_s1_s6,fs,1024,512);
hold on;
title('estimated average power spectrumn with different duration of window');
legend('L=32','L=64','L=128','L=256','L=512');

[s_male,fs_male]=audioread('out_male.wav');
[s_female,fs_female]=audioread('out_female.wav');
figure;
pspect(s_male,fs_male,1024,32);

```

```
hold on;
pspect(s_female,fs_female,1024,32);
legend('male','female');
title('power spectrum for male and female speech');
```

2.2 Problem2

The goal of this MATLAB exercise is to experiment with the process of uniform quantization of a speech signal. This exercise uses the MATLAB function

$X = \text{fxquant}(s, \text{bits}, \text{rmode}, \text{lmode})$

- **s** is the input speech signal to be quantized
- **bits** is the total number of bits of the quantizer
- **rmode** is the quantization mode, where **rmode** is one of **round** for rounding to the nearest level, **trunc** for 2's complement overflow, **triangle** for a triangle limiter, and **none** for no limiter.
- **x** is the output speech signal

Firstly, a linearly increasing input vector going from -1 to $+1$ in increments of 0.001 and use the MATLAB function **fxquant** to plot the non-linear quantizer characteristic for the conditions **bits=4**, **lmode='sat'**, **rmode='round'** or **'trunc'**. The result is as shown in Fig4.

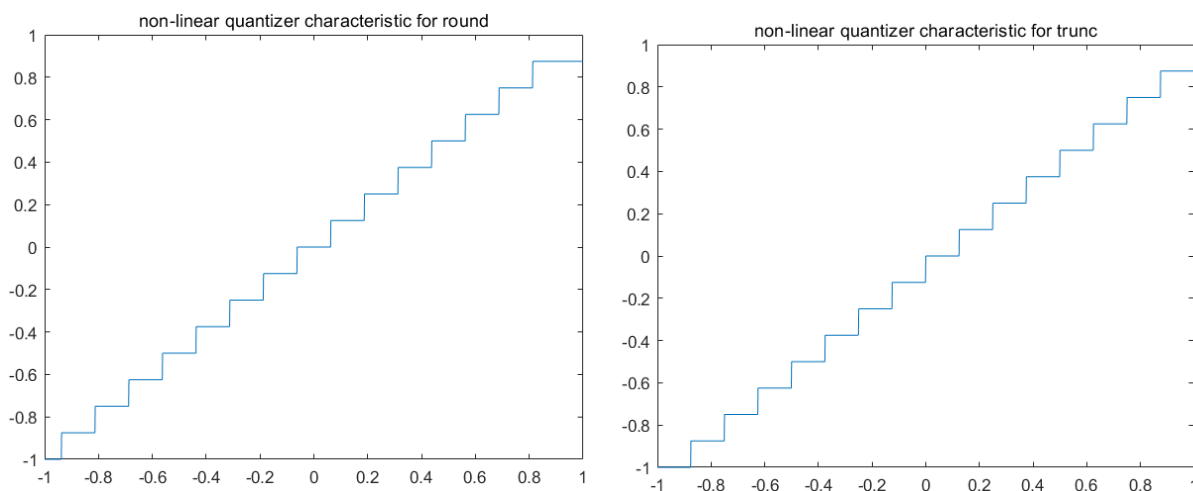
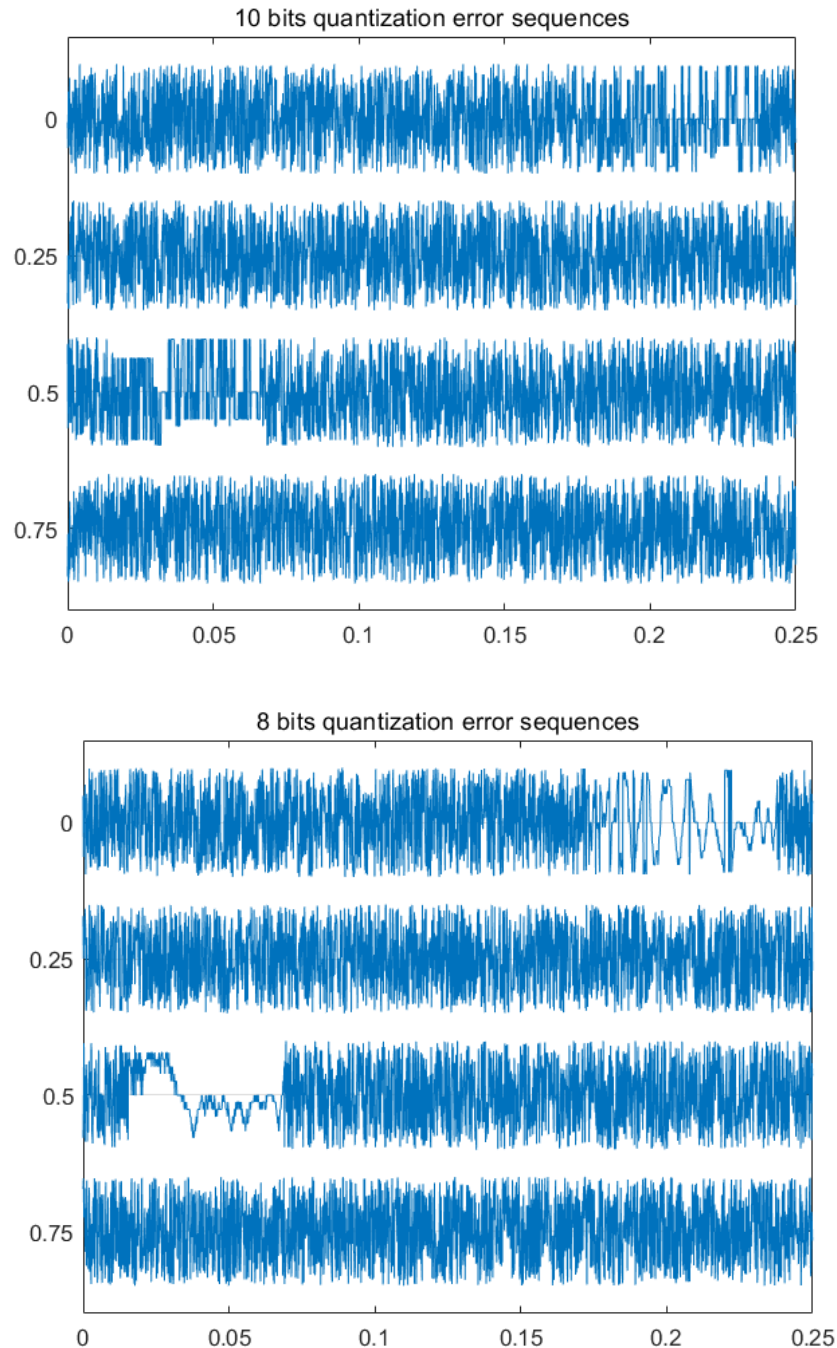


Fig.4 power spectrum for male and female speech file

For the conditions `bits=4`, `lmode='sat'`, `rmode='round'`, the range of values for $e[n]$ is -0.1250 to 0.0620 . While the range of values for $e[n]$ is -0.1250 to 0 when `rmode='trunc'`. From Fig 4, what can be found is that after quantization, the signal can be processed digitally. For `rmode='round'`, the 0 level is at the middle of the step. While the 0 level is at the edge of the step when the mode is `trunc`.

Then, `fxquant()` is used to quantize the speech samples of the file `s5.wav` between sample 1300 and sample 18800 with `rmode='round'` and `lmode='sat'`. Different numbers of bits for the quantizer is experimented, the quantization error sequences is computed and histograms of the quantization noise samples for each of these bit rates is plotted.



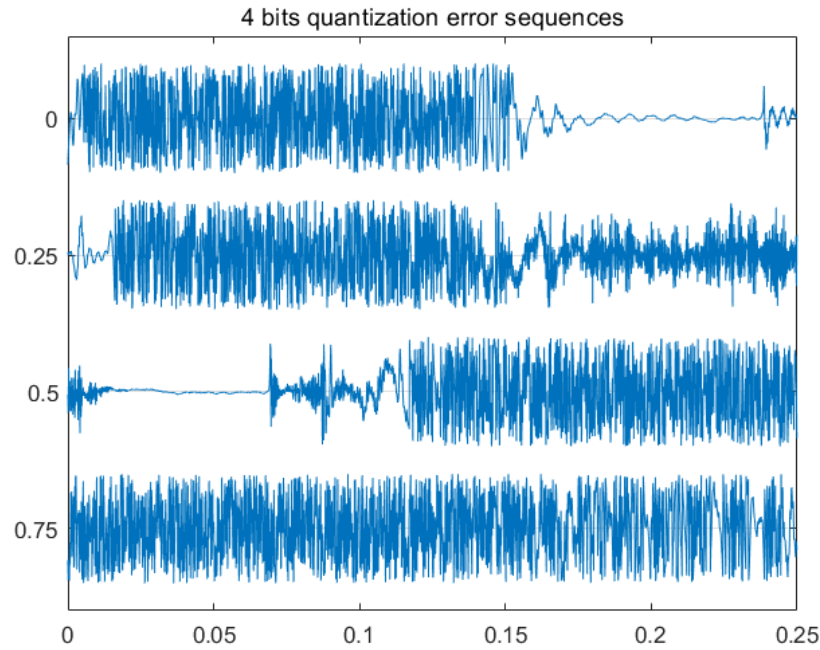


Fig.5 power spectrum for male and female speech file

From Fig 5, with the number of bits decreasing, the error seems more like the signal at some interval. The quantization error sequences of 10 bits seems like random noise and those of 8 bits and 4 bits seems more like original signal at some interval.

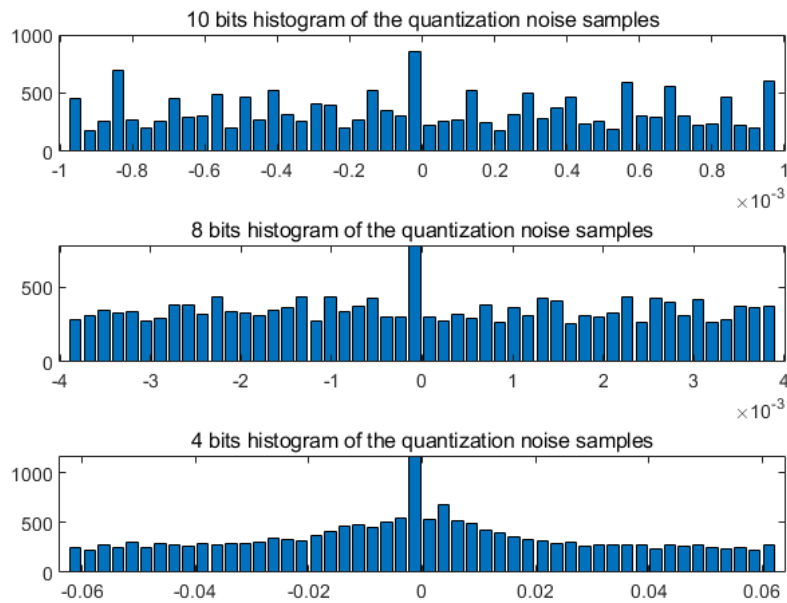


Fig.6 histograms of the quantization noise samples for each of these bit rates

From Fig 6, the histograms of the quantization noise samples for 4 bits shows a regular statistic distribution like original signal. While those of 10 bits and 8 bits fit white noise model more.

Finally, the power spectrum of the quantization noise sequences for each of the quantizers is computed and the result is shown in Fig 7.

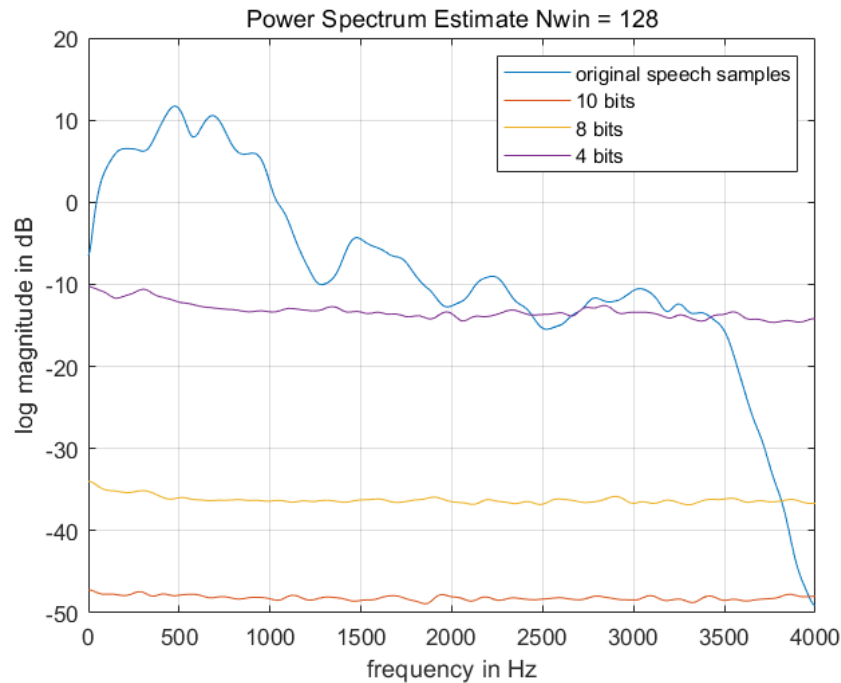


Fig.7 the power spectrum of the quantization noise sequences for each of the quantizers

It can be seen that the spectrum of original speech fluctuates the most with the frequency while that of the quantization noise sequences for 10 bits, 8 bits and 4 bits is relatively smooth in frequency domain, which supports the white noise assumption. With the number of bits increasing, the log magnitude of the power spectrum decreases. The approximate difference between the noise spectra for 10- and 8-bit quantization is -11.9336 .

The whole code is as below:

```
%(a)
xin=-1:0.001:1;
x1=fxquant(xin,4,'round','sat');
x2=fxquant(xin,4,'trunc','sat');
e1=x1-xin;
e2=x2-xin;

disp(min(e1));
disp(max(e1));
figure;
```

```

plot(xin,x1);
title('non-linear quantizer characteristic for round');

disp(min(e2));
disp(max(e2));
figure;
plot(xin,x2);
title('non-linear quantizer characteristic for trunc');

%(b)
[s,fs]=audioread('s5.wav');
s=s(1300:18800);
x1_s5_10=fxquant(s,10,'round','sat');
x1_s5_8=fxquant(s,8,'round','sat');
x1_s5_4=fxquant(s,4,'round','sat');
e1_10=x1_s5_10-s;
e1_8=x1_s5_8-s;
e1_4=x1_s5_4-s;

figure;
strips(e1_10(1:8000),2000/fs,fs);
title('10 bits quantization error sequences');
figure;
strips(e1_8(1:8000),2000/fs,fs);
title('8 bits quantization error sequences');
figure;
strips(e1_4(1:8000),2000/fs,fs);
title('4 bits quantization error sequences');

[hist10,center10]=hist(e1_10,50);
[hist8,center8]=hist(e1_8,50);
[hist4,center4]=hist(e1_4,50);

figure;
subplot(3,1,1);
bar(center10,hist10);
title('10 bits histogram of the quantization noise samples');

```

```

subplot(3,1,2);
bar(center8,hist8);
title('8 bits histogram of the quantization noise samples');
subplot(3,1,3);
bar(center4,hist4);
title('4 bits histogram of the quantization noise samples');

%(c)
figure;
pspect(s,fs,1024,128);
hold on;
pspect(e1_10,fs,1024,128);
hold on;
pspect(e1_8,fs,1024,128);
hold on;
pspect(e1_4,fs,1024,128);
hold on;
legend('original speech samples','10 bits','8 bits','4 bits');

[p_10,F_10]=pspect(e1_10,fs,1024,128);
[p_8,F_8]=pspect(e1_8,fs,1024,128);
difference=10*log10(p_10)-10*log10(p_8);
disp(mean(difference));

```

3.Conclusion

The main harvest of this experiment is to understand the statistical model of speech signal and the process of uniform quantization. Durations of window have an effect on the smoothness of the power spectrum. The longer the duration is, the more violent the fluctuation is. The number of bits for the quantizer affects statistic distribution of the quantization error sequences. The larger the number of bits for the quantizer, the more the histogram of the quantization error sequences fit white noise model.