1 Lecture Outline

Reading: Chapter 12

- Introduction to speech coding
- Scalar quantization (review of lecture 1)
- Quantizer noise: linear (additive) noise model
- Sampling and quantization demos
- Quantization of a sinusoid (tone)
- Quantization of speech (3 bit, 5 bit): time and frequency
- Introduction to nonuniform quantization
- Statistical models of speech

2 Introduction to Speech Coding

Definition: *Speech coding is defined to be any process that leads to a representation of the analog speech waveform by a sequence of bits.*

Goal: *Accurately represent the speech waveform with as few bits as possible.*

Using straight pulse-code modulation (PCM) for waveform encoding, several sample bit rates are given in Table 1. In these examples, the tradeoff is between "quality" (signal bandwidth and/or SNR) and bit rate.

Table 1: Data rates for uncompressed speech.

Sample rate, f_s	Resolution (bits/sample)	Bit rate [bits per second (bps)]
8,000	8	64,000
8,000	16	128,000
16,000	8	128,000
16,000	16	256,000

Low-bit rate coding is important since many channels over which speech is transmitted are bandwidth constrained. These include wireless and cellular channels as well as Internet-communications channels [Voice over Internet Protocol (VoIP)]. A basic digital transmission system for speech communications is illustrated in Fig. 1 (next page).

Speech coders are typically categorized into three groups:

1. Waveform coders: quantize the speech waveform (samples) directly and operate at high rates (16-64kbps). Ex. μ-law quantizer.
2. Voice coders (vocoders): quantize the parameters of a speech model (obtained through the speech waveform itself) and operate at low rates (1.2-4.8kbps). Ex. LPC, MELP, and CELP.
3. Hybrid coders: are partly waveform-based and partly model-based and operate at middle rates (2.4-16kbps). Ex. sinusoidal coders.

Figure 1: Figure 12.1: Speech communication over a digital transmission system

Note that quality of the coded waveform has many dimensions including naturalness, background artifacts, and intelligibility. Note that although useful in waveform coders, SNR (where the noise is the difference waveform between the original speech and the coded speech) is not always meaningful for vocoders because the phase of the original speech waveform may be modified, disallowing waveform differencing. Many objective and subjective tests have been developed to measure these different attributes of quality. Subjective testing is usually measured by the Mean Opinion Score (MOS) and objective measurements include the Perceptual Evaluation of Speech Quality (PESQ).

3 Scalar Quantization

See Fig. 2. As discussed in Lecture 2, the first step of the waveform encoding process is to map or quantize the sample value $x[n]$ to a finite set of amplitudes, $\hat{x}[n]$ each of which can be represented with B bits. With this representation, there are 2^B possible amplitudes. The second step of the process, maps or encodes $\hat{x}[n]$ to a finite set of symbols called the *codewords*, $c[n]$. The decoding process takes the sequence of received codewords, $c'[n]$ and transforms them back into a sequence of quantized samples.

Because each sample value is *individually* quantized, we call this *scalar quantization*. We will also discuss in this chapter, *vector quantization* whereby a group of samples are collectively quantized.

To date, we have discussed only *uniform quantizers*. See Fig. 3 for the mid-riser, mid-tread uniform quantizers.

Figure 2: Figure 12.3: Waveform coding

Figure 3: Mid-riser, mid-tread uniform quantizers

4 Scalar Quantization Analysis

In Lecture 2, we analyzed the uniform quantizer. We assumed a B-bit quantizer, maximum signal amplitude x_{max}, and an additive noise model

$$\hat{x}[n] = x[n] + e[n] \tag{1}$$

where $e[n]$ is the difference between the actual and quantized sample value or error sequence and

$$-\Delta/2 \leq e[n] \leq \Delta/2. \tag{2}$$

$e[n]$ is often called *granular distortion*. We note that another form of quantization noise, called *overload distortion*, is where the sample value is outside the assumed maximum range and thus is clipped to the assumed maximum. These clipped samples incur a quantization error in excess of $\pm\frac{\Delta}{2}$. Depending on the pdf of the signal, there are normally very few clipped samples and thus this source of error is usually ignored.

In the first-order analysis, we made the following assumptions

- $x[n]$, $e[n]$ are stationary random processes
- the probability density function (pdf) of $x[n]$ is uniform over $[-x_{\max}, x_{\max}]$ therefore $\mu_x = 0$ and $\sigma_x^2 = (2x_{\max})^2/12 = x_{\max}^2/3$
- the pdf of $e[n]$ is uniform over $[-\Delta/2, \Delta/2]$ therefore $\mu_e = 0$ and $\sigma_e^2 = \Delta^2/12 = x_{\max}^2/(3 \cdot 2^{2B})$
- $e[n]$ is white process, i.e. $E\{e[n]e[n+m]\} = \sigma_e^2\delta[m]$
- $e[n]$ is uncorrelated with $x[n]$, i.e. $E\{x[n]e[n+m]\} = 0$ for all m.

Note that the second assumption is not true (and several others are weak) if $x[n]$ is speech. The signal-to-quantization noise ratio (SNR) is the ratio of signal power (variance) to quantization noise variance:

$$\begin{aligned}
\text{SNR(dB)} &= 10\log_{10}\left(\frac{\sigma_x^2}{\sigma_e^2}\right) \\
&= 10\log_{10}\left(\frac{x_{\max}^2}{3} \cdot \frac{3 \cdot 2^{2B}}{x_{\max}^2}\right) \\
&= 10\log_{10}\left(2^{2B}\right) \\
&= 20B\log_{10}(2) \\
&\approx 6.02B.
\end{aligned} \tag{3}$$

Thus for each bit in our sample word length, we increase the SNR or dynamic range of the quantizer by 6dB "6dB per bit rule."

An alternate analysis does not assume the pdf of $x[n]$ is uniform and that $x_{max} \approx 4\sigma_x$. In this case $\sigma_x^2 = x_{max}^2/16$ and

$$\begin{aligned} \text{SNR(dB)} &= 10\log_{10}\left(\frac{\sigma_x^2}{\sigma_e^2}\right) \\ &= 10\log_{10}\left(\frac{x_{\max}^2}{16}\cdot\frac{3\cdot 2^{2B}}{x_{\max}^2}\right) \\ &= 10\log_{10}\left(2^{2B}\right) + 10\log_{10}\left(3/16\right) \\ &= 20B\log_{10}\left(2\right) - 7.27 \\ &\approx 6.02B - 7.27 \end{aligned} \tag{4}$$

which for large B, is roughly the same as the prior analysis result.

Fig. 4(a) contains a speech segment, $x[n]$ originally sampled with 16bits/sample. Fig. 4(b)-(c) illustrates the speech quantized with 5 bits/sample and the quantization error. Fig. 4(d)-(e) illustrates the speech quantized with 3 bits/sample and the quantization error.

5 Introduction to Nonuniform Quantization

Although uniform quantization is quite straightforward and appears to be a natural approach, it may not be optimal, i.e., the SNR may not be as small as could be obtained for a certain number of decision and reconstruction levels. To understand this limitation, suppose small sample values are more likely to occur than large sample values (either positive or negative values). Intuition, tells us to use more resolution (smaller quantization steps) for the small-valued samples (where the probability of occurrence is high) and less resolution (larger quantization steps) for the large-valued samples (where the probability of occurrence is low). See Fig. 5.

In theory, this would give us a lot of small quantization errors and a few large quantization errors. Thus it seems with the uniform quantizer, that we are not efficiently utilizing decision and reconstruction levels. Quantization in which reconstruction and decision levels do not have equal spacing is called *nonuniform quantization*.

6 Statistical Models of Speech

In order to design an optimal (nonuniform) quantizer for speech, it is necessary to know the pdf of speech. Normally, we estimate this pdf from a histogram from the speech waveform. In obtaining the histogram, we count up the number of occurrences of the value of each speech sample in different ranges, $x_k \leq x[n] \leq x_{k+1}$ where x_k defines a numerical boundary of the histogram bin. In MATLAB, we have

```
[x,fs,bits] = wavread('Speech.wav');
x = x - mean(x);  % force zero mean
x = x ./ sqrt(cov(x));  % force unit variance
bins = [-5.25:0.25:5.25];  % histogram bin centers
n = hist(x,bins);  % compute histogram of speech
A = sum(n)*0.25;n = n ./ A;  % normalize histogram/pdf
semilogy(bins,n);hold on;semilogy(bins,n,'*');hold off;axis([-5 5 10^-3 1]);  % pretty plots...
grid;title('Speech Histogram');ylabel('p_x(x)');xlabel('x');  % very pretty plots...
```

A histogram of Prof. De Leon's speech is given in Fig. 6

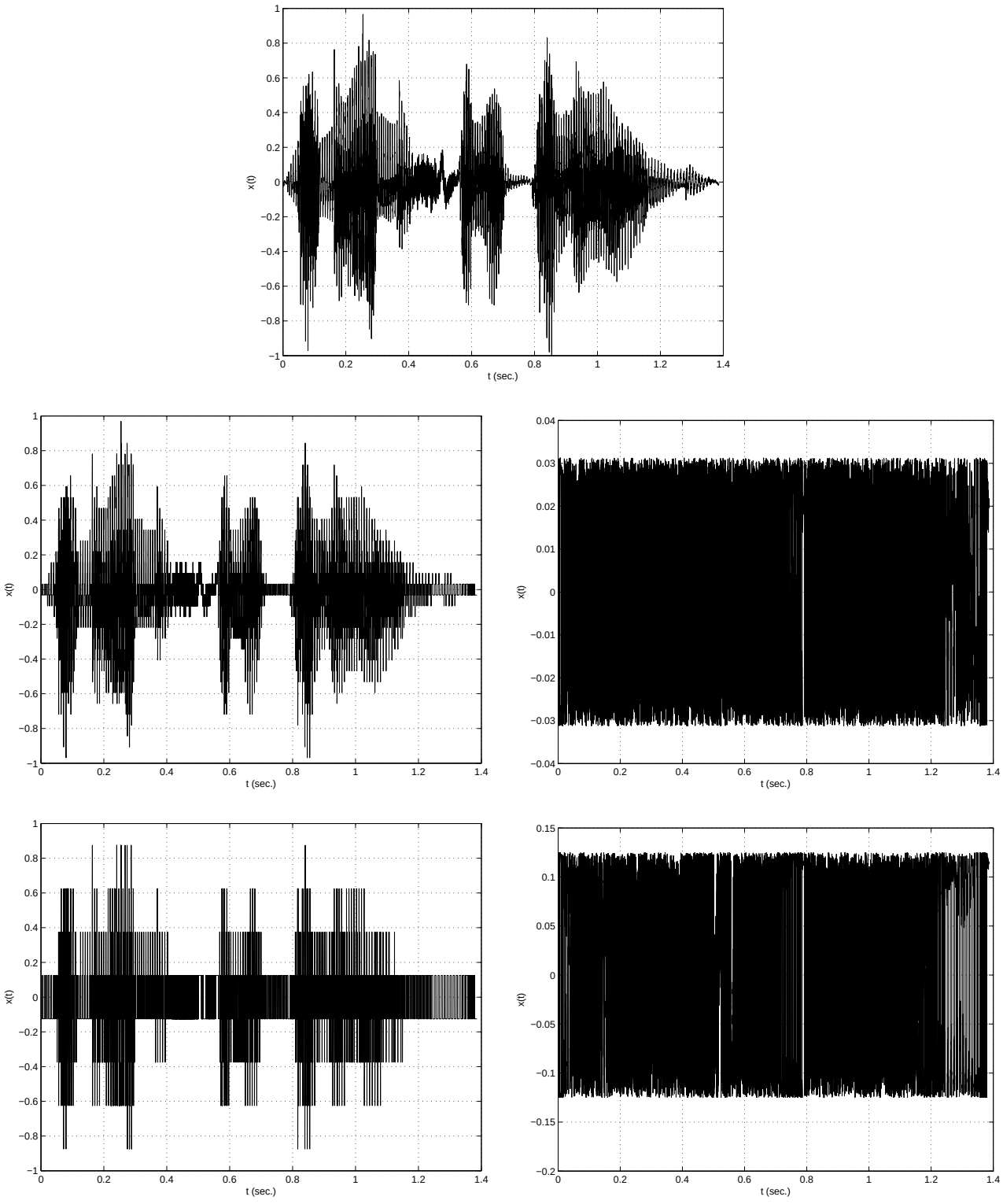

Figure 4: Speech signal with (a) 16 bit quantizer, (b) 5 bit quantizer, and (d) 3 bit quantizer. Quantization errors are shown in (c) and (e).

Figure 5: Figure 12.6b: Nonuniform quantization

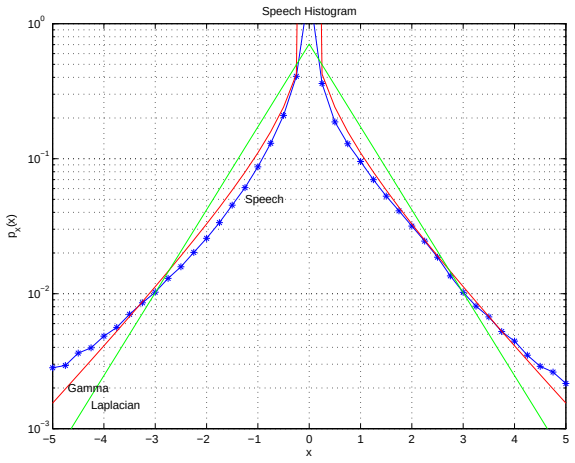

Figure 6: Histogram of Prof. De Leon's speech

Work done in the 1970's demonstrated that speech histograms can be approximated with a Laplacian pdf

$$p_x(x) = \frac{1}{\sqrt{2}\sigma_x} e^{-\frac{\sqrt{2}|x|}{\sigma_x}} \tag{5}$$

where σ_x is the standard deviation of the pdf. A better approximation can be obtained with a Gamma pdf

$$p_x(x) = \left(\frac{\sqrt{3}}{8\pi\sigma_x|x|}\right)^{1/2} e^{-\frac{\sqrt{3}|x|}{2\sigma_x}}. \tag{6}$$