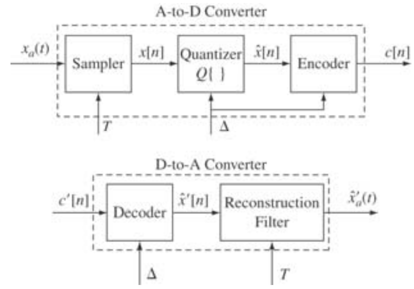
Digital Speech Processing—Lecture 15

Speech Coding Methods Based on Speech Waveform Representations and Speech Models—Uniform and Non-Uniform Coding Methods

1

Analog-to-Digital Conversion (Sampling and Quantization)



Class of "waveform coders" can be represented in this manner

2

Information Rate

• Waveform coder information rate, I_w, of the digital representation of the signal, $x_a(t)$, defined as:

$$I_w = B \cdot F_S = B / T$$

where B is the number of bits used to represent each sample and $F_S = \frac{1}{T}$, is the number of samples/second

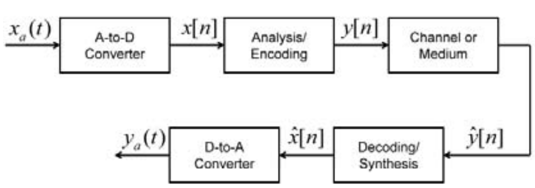
3

Speech Information Rates

- **Production level**:
  - 10-15 phonemes/second for continuous speech
  - 32-64 phonemes per language => 6 bits/phoneme
  - *Information Rate*=60-90 bps at the source
- **Waveform level**
  - speech bandwidth from 4 – 10 kHz => sampling rate from 8 – 20 kHz
  - need 12-16 bit quantization for high quality digital coding
  - *Information Rate*=96-320 Kbps => **more than 3 orders of magnitude** difference in Information Rates between the production and waveform levels

4

Speech Analysis/Synthesis Systems



• Second class of digital speech coding systems:
  o analysis/synthesis systems
  o model-based systems
  o hybrid coders
  o vocoder (voice coder) systems
• Detailed waveform properties generally not preserved
  o coder estimates parameters of a model for speech production
  o coder tries to preserve intelligibility and quality of reproduction from the digital representation

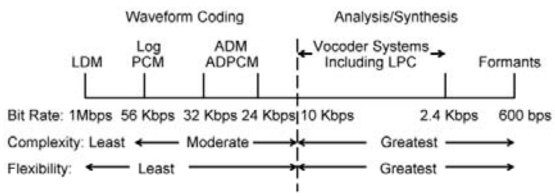
5

Speech Coder Comparisons

• Speech parameters (the chosen representation) are encoded for transmission or storage
  • analysis and encoding gives a data parameter vector
  • data parameter vector computed at a sampling rate much lower than the signal sampling rate
  • denote the "frame rate" of the analysis as F_{fr}
  • total information rate for model-based coders is:
  $$I_m = B_c \cdot F_{fr}$$
  • where B_c is the total number of bits required to represent the parameter vector

6

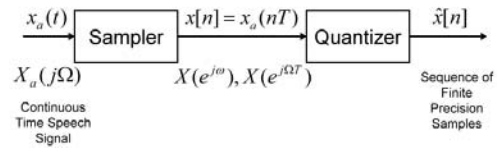
Speech Coder Comparisons



- waveform coders characterized by:
  - high bit rates (24 Kbps – 1 Mbps)
  - low complexity
  - low flexibility
- analysis/synthesis systems characterized by:
  - low bit rates (10 Kbps – 600 bps)
  - high complexity
  - great flexibility (e.g., time expansion/compression)

7

Introduction to Waveform Coding



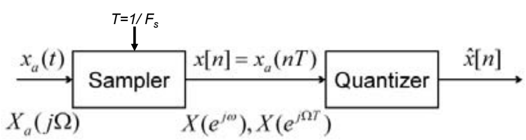
$$x(n) = x_a(nT)$$

$$X(e^{j\Omega T}) = \frac{1}{T}\sum_{k=-\infty}^{\infty} X_a(j\Omega + \frac{j2\pi k}{T})$$

$$\omega = \Omega T \Rightarrow X(e^{j\omega}) = \frac{1}{T}\sum_{k=-\infty}^{\infty} X_a(\frac{j\omega}{T} + \frac{j2\pi k}{T})$$

8

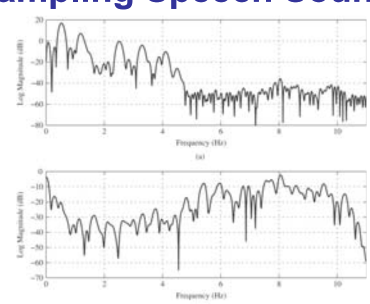
Introduction to Waveform Coding



- to ***perfectly recover*** $x_a(t)$ (or equivalently a lowpass filtered version of it) from the set of digital samples (as yet unquantized) we require that $F_s = 1/T$ > twice the highest frequency in the input signal
- this implies that $x_a(t)$ must first be lowpass filtered since speech is not inherently lowpass
  - for telephone bandwidth the frequency range of interest is 200-3200 Hz (filtering range) => F_s = 6400 Hz, 8000 Hz
  - for wideband speech the frequency range of interest is 100-7000 Hz (filtering range) => F_s = 16000 Hz

9

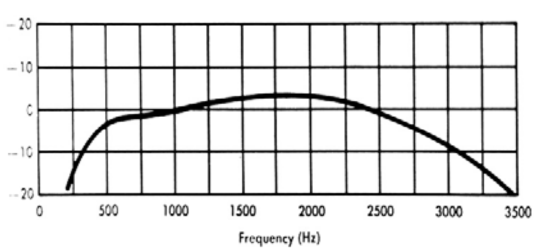
Sampling Speech Sounds



- notice high frequency components of vowels and fricatives (up to 10 kHz) => need F_s > 20 kHz without lowpass filtering
  - need only about 4 kHz to estimate formant frequencies
  - need only about 3.2 kHz for telephone speech coding

10

Telephone Channel Response



it is clear that 4 kHz bandwidth is sufficient for most applications using telephone speech because of inherent channel band limitations from the transmission path

11

Statistical Model of Speech

- assume $x_a(t)$ is a sample function of a continuous-time random process
- then $x(n)$ (derived from $x_a(t)$ by sampling) is also a sample sequence of a discrete-time random process
- $x_a(t)$ has first order probability density, $p(x)$, with autocorrelation and power spectrum of the form

$$\phi_a(\tau) = E\left[x_a(t)x_a(t+\tau)\right]$$

$$\Phi_a(\Omega) = \int_{-\infty}^{\infty} \phi_a(\tau)e^{-j\Omega\tau}d\tau$$

12

Statistical Model of Speech

- $x(n)$ has autocorrelation and power spectrum of the form:

$$\phi(m) = E\left[x(n)x(n+m)\right]$$
$$= E\left[x_a(nT)x_a(nT+mT)\right] = \phi_a(mT)$$
$$\Phi(e^{j\Omega T}) = \sum_{m=-\infty}^{\infty} \phi(m)e^{-j\Omega Tm}$$
$$= \frac{1}{T}\sum_{k=-\infty}^{\infty} \Phi_a(\Omega + \frac{2\pi k}{T})$$

- since $\phi(m)$ is a sampled version of $\phi_a(\tau)$, its transform is an infinite sum of the power spectrum, shifted by $\frac{2\pi k}{T}$ => aliased version of the power spectrum of the original analog signal

13

Speech Probability Density Function

- probability density function for $x(n)$ is the same as for $x_a(t)$ since $x(n)= x_a(nT)$ => the mean and variance are the same for both $x(n)$ and $x_a(t)$
- need to estimate probability density and power spectrum from speech waveforms
  - probability density estimated from long term histogram of amplitudes
  - good approximation is a gamma distribution, of the form:

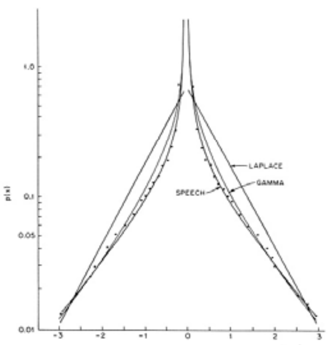
$$p(x) = \left[\frac{\sqrt{3}}{8\pi\sigma_x |x|}\right]^{1/2} e^{-\frac{\sqrt{3}|x|}{2\sigma_x}} \qquad p(0) = \infty$$

  - simpler approximation is Laplacian density, of the form:

$$p(x) = \frac{1}{\sqrt{2}\sigma_x} e^{-\frac{\sqrt{2}|x|}{\sigma_x}} \qquad p(0) = \frac{1}{\sqrt{2}\sigma_x}$$

14

Measured Speech Densities



- distribution normalized so mean is 0 and variance is 1 ($x=0$, $\sigma_x=1$)
- gamma density more closely approximates measured distribution for speech than Laplacian
- Laplacian is still a good model and is used in analytical studies
- small amplitudes much more likely than large amplitudes—by 100:1 ratio

15

Long-Time Autocorrelation and Power Spectrum

- analog signal

$$\phi_a(\tau) = E\left\{x_a(t)x_a(t+\tau)\right\} \Leftrightarrow \Phi_a(\Omega) = \int_{-\infty}^{\infty} \phi_a(\tau)e^{-j\Omega\tau}d\tau$$

- discrete-time signal

$$\phi(m) = E\left\{x(n)x(n+m)\right\} = E\left\{x_a(nT)x_a(nT+mT)\right\} = \phi_a(mT)$$
$$\Leftrightarrow \Phi(e^{j\Omega T}) = \sum_{m=-\infty}^{\infty} \phi(m)e^{-j\Omega Tm} = \frac{1}{T}\sum_{k=-\infty}^{\infty} \Phi_a(\Omega + \frac{2\pi k}{T})$$

- estimated correlation and power spectrum

$$\hat{\phi}(m) = \frac{1}{L}\sum_{n=0}^{L-1-m} x(n)x(n+m), \quad 0 \le |m| \le L-1,$$

L is large integer

$$\hat{\Phi}(e^{j\Omega T}) = \sum_{m=-M}^{M} w(m)\hat{\phi}(m)e^{-j\Omega Tm}$$

Estimates ≠ Expectations

- tapering with m due to finite windows
- need window on estimated correlation for smoothing because of discontinuity at $\pm M$

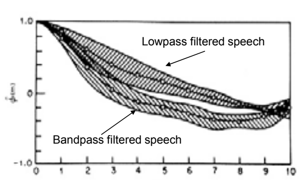
16

Speech AC and Power Spectrum

- can estimate long term autocorrelation and power spectrum using time-series analysis methods

$$-1 \le \hat{\rho}[m] = \frac{\hat{\phi}[m]}{\hat{\phi}[0]} = \frac{\frac{1}{L}\sum_{n=0}^{L-1-m} x(n)x(n+m)}{\frac{1}{L}\sum_{n=0}^{L-1-m} x(n)^2} \le 1,$$

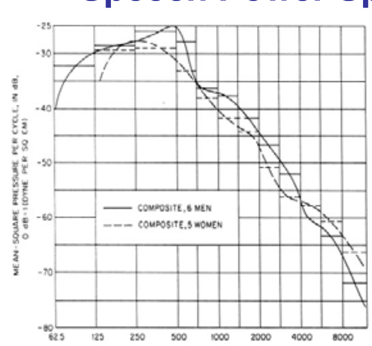
$0 \le m \le L-1$, L is window length



- 8 kHz sampled speech for several speakers
- high correlation between adjacent samples
- lowpass speech more highly correlated than bandpass speech

17

Speech Power Spectrum



- power spectrum estimate derived from one minute of speech
- peaks at 250-500 Hz (region of maximal spectral information)
- spectrum falls at about 8-10 db/octave
- computed from set of bandpass filters

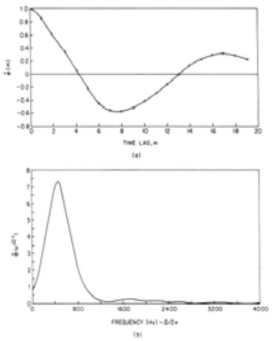
18

Alternative Power Spectrum Estimate

- estimate long term correlation, $\hat{\phi}(m)$, using sampled speech, then weight and transform, giving:

$$\hat{\Phi}(e^{j2\pi k/N}) = \sum_{m=-M}^{M} w(m)\hat{\phi}(m)e^{-j2\pi km/N}$$

- this lets us use $\hat{\phi}(m)$ to get a power spectrum estimate $\hat{\Phi}(e^{j2\pi k/N})$ via the weighting window, $w(m)$



Contrast linear versus logarithmic scale for power spectrum plots

Estimating Power Spectrum via Method of Averaged Periodograms

- Periodogram defined as:

$$P(e^{j\omega}) = \frac{1}{LU}\left|\sum_{n=0}^{L-1} x(n)w(n)e^{-j\omega n}\right|^2$$

- where $w[n]$ is an L-point window (e.g., Hamming window), and

$$U = \frac{1}{L}\sum_{n=0}^{L-1}(w(n))^2$$

- U is a normalizing constant that compensates for window tapering
- use the DFT to compute the periodogram as:

$$P(e^{j(2\pi k)/N}) = \frac{1}{LU}\left|\sum_{n=0}^{L-1} x(n)w(n)e^{-j(2\pi/N)kn}\right|^2, \quad 0 \le k \le N-1,$$

N is size of DFT

20

Averaging (Short) Periodograms

- variability of spectral estimates can be reduced by averaging short periodograms, computed over a long segment of speech
- using an L-point window, a short periodogram is the same as the STFT of the weighted speech interval, namely:

$$X_r[k] = X_{rR}(e^{j2\pi k/N}) = \sum_{m=rR}^{rR+L-1} x[m]w[rR-m]e^{-j2\pi km/N}, \quad 0 \le k \le N-1$$

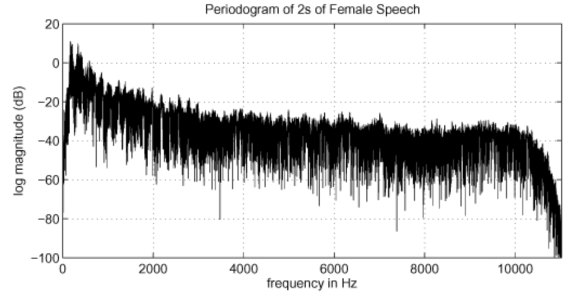
- where N is the DFT transform size (number of frequency estimates)
- Consider using N_S samples of speech (where N_S is large); the averaged periodogram is defined as:

$$\tilde{\Phi}(e^{j2\pi k/N}) = \frac{1}{KLU}\sum_{r=0}^{K-1}\left|X_{rR}(e^{j2\pi k/N})\right|^2, \quad 0 \le k \le N-1$$

- where K is the number of windowed segments in N_S samples, L is the window length, and U is the window normalization factor
- use $R = L/2$ (shift window by half the window duration between adjacent periodogram estimates)

21

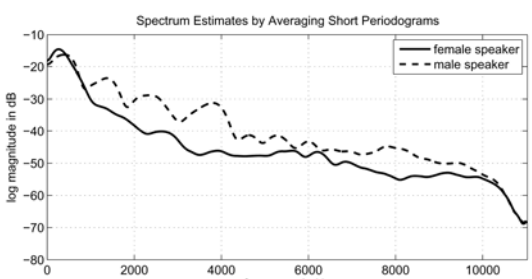
88,000-Point FFT – Female Speaker



Single spectral estimate using all signal samples

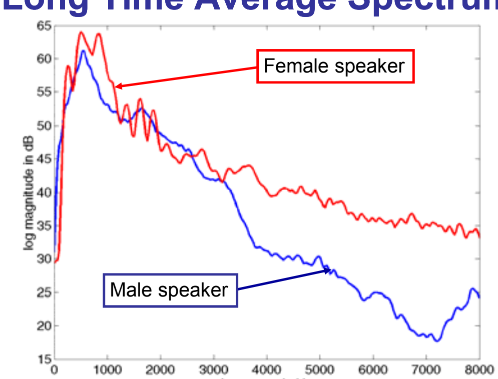
22

Long Time Average Spectrum



23

Long Time Average Spectrum



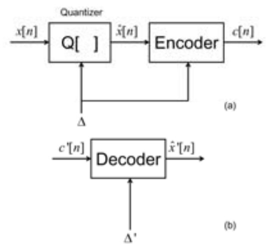
24

Instantaneous Quantization

- separating the processes of sampling and quantization
- assume *x(n)* obtained by sampling a bandlimited signal at a rate at or above the Nyquist rate
- assume *x(n)* is known to infinite precision in amplitude
- need to quantize *x(n)* in some suitable manner

25

Quantization and Coding



assume Δ'=Δ

Coding is a two-stage process
1. quantization process: $x(n) \to \hat{x}(n)$
2. encoding process: $\hat{x}(n) \to c(n)$

- where Δ is the (assumed fixed) quantization step size

Decoding is a single-stage process
1. decoding process: $c'(n) \to \hat{x}'(n)$

- if $c'(n) = c(n)$, (no errors in transmission) then $\hat{x}'(n) = \hat{x}(n)$
  $\hat{x}'(n) \neq x(n) \Rightarrow$ coding and quantization loses information

26

B-bit Quantization

- use *B*-bit binary numbers to represent the quantized samples => 2^B quantization levels
- *Information Rate of Coder*: $I = B\,F_S$ = total bit rate in bits/second
  - B=16, F_S= 8 kHz => I=128 Kbps
  - B=8, F_S= 8 kHz => I=64 Kbps
  - B=4, F_S= 8 kHz => I=32 Kbps
- goal of waveform coding is to get the highest quality at a fixed value of I (Kbps), or equivalently to get the lowest value of I for a fixed quality
- since F_S is fixed, need most efficient quantization methods to minimize I

27

Quantization Basics

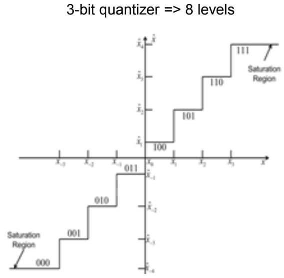
- assume $|x(n)| \le X_{max}$ (possibly ∞)
  - for Laplacian density (where $X_{max} = \infty$), can show that 0.35% of the samples fall outside the range $-4\sigma_x \le x(n) \le 4\sigma_x$ => large quantization errors for 0.35% of the samples
  - can safely assume that X_{max} is proportional to σ_x

28

Quantization Process

- quantization => dividing amplitude range into a finite set of ranges, and assigning the same bin to all samples in a given range

3-bit quantizer => 8 levels



$$0 = x_0 < x(n) \le x_1 \Rightarrow \hat{x}_1 \;(100)$$
$$x_1 < x(n) \le x_2 \Rightarrow \hat{x}_2 \;(101)$$
$$x_2 < x(n) \le x_3 \Rightarrow \hat{x}_3 \;(110)$$
$$x_3 < x(n) < \infty \Rightarrow \hat{x}_4 \;(111)$$
$$x_{-1} < x(n) \le x_0 = 0 \Rightarrow \hat{x}_{-1} \;(011)$$
$$x_{-2} < x(n) \le x_{-1} \Rightarrow \hat{x}_{-2} \;(010)$$
$$x_{-3} < x(n) \le x_{-2} \Rightarrow \hat{x}_{-3} \;(001)$$
$$-\infty < x(n) \le x_{-3} \Rightarrow \hat{x}_{-4} \;(000)$$

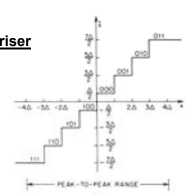
range | level | codeword

- codewords are arbitrary!! => there are good choices that can be made (and bad choices)

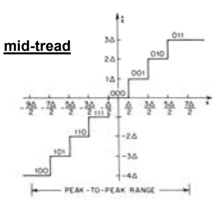
Uniform Quantization

- choice of quantization ranges and levels so that signal can easily be processed digitally

mid-riser



mid-tread



$$x_i - x_{i-1} = \Delta$$
$$\hat{x}_i - \hat{x}_{i-1} = \Delta$$
Δ = quantization step size

30

Mid-Riser and Mid-Tread Quantizers

- mid-riser
  - origin (*x=0*) in middle of rising part of the staircase
  - same number of positive and negative levels
  - symmetrical around origin
- mid-tread
  - origin (*x=0*) in middle of quantization level
  - one more negative level than positive
  - one quantization level of 0 (where a lot of activity occurs)
- code words have direct numerical significance (sign-magnitude representation for mid-riser, two's complement for mid-tread)

- for mid-riser quantizer:

$$\hat{x}(n) = \frac{\Delta}{2} sign\left[c(n)\right] + \Delta c(n)$$

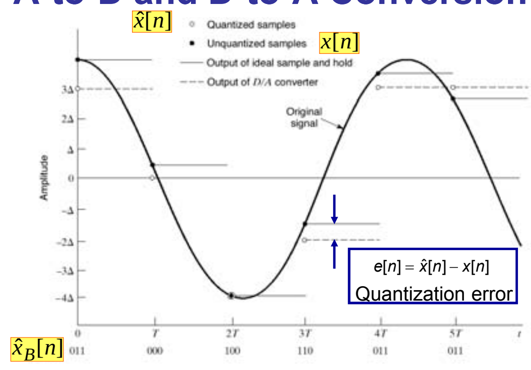
where $sign\left[c(n)\right] = +1$ if first bit of $c(n) = 0$

$= -1$ if first bit of $c(n) = 1$

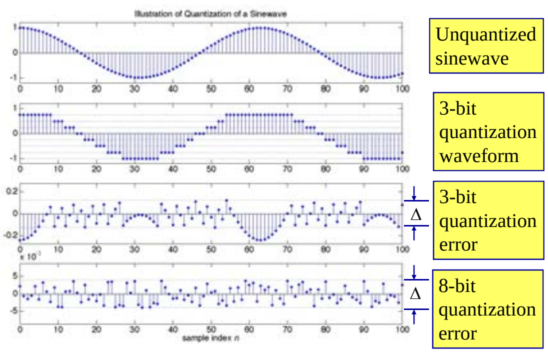
- for mid-tread quantizer code words are 3-bit two's complement representation, giving

$$\hat{x}(n) = \Delta c(n)$$

A-to-D and D-to-A Conversion

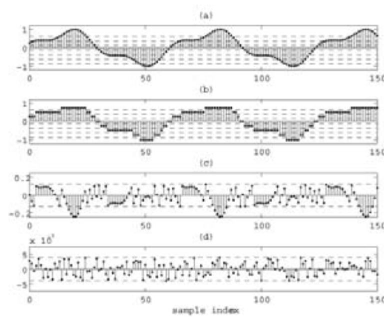


Quantization of a Sine Wave



Quantization of Complex Signal

$$x[n] = \sin(0.1n) + 0.3\cos(0.3n)$$



(a) Unquantized signal

(b) 3-bit quantized $x[n]$

(c) 3-bit quantization error, e[n]

(d) 8-bit quantization error, e[n]

Uniform Quantizers

- Uniform Quantizers characterized by:
  - number of levels—2^B (*B* bits)
  - quantization step size-*Δ*
- if $|x(n)| \le X_{max}$ and *x(n)* is a symmetric density, then

$$\Delta 2^B = 2 X_{max}$$
$$\Delta = 2 X_{max} / 2^B$$

- if we let

$$\hat{x}(n) = x(n) + e(n)$$

- with *x(n)* the unquantized speech sample, and *e(n)* the quantization error (noise), then

$$-\frac{\Delta}{2} \le e(n) \le \frac{\Delta}{2}$$

(except for last quantization level which can exceed X_{max} and thus the error can exceed *Δ/2*)

Quantization Noise Model

1. quantization noise is a zero-mean, stationary white noise process

$$E\left[e(n)e(n+m)\right] = \sigma_e^2, \quad m = 0$$
$$= 0 \quad otherwise$$

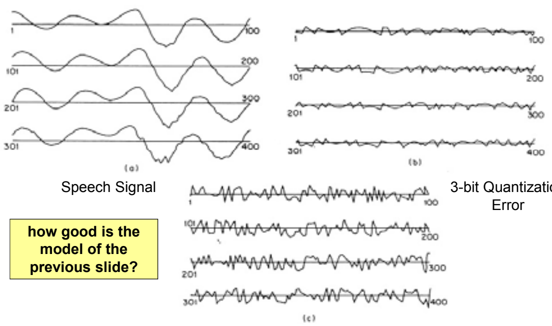
2. quantization noise is uncorrelated with the input signal

$$E\left[x(n)e(n+m)\right] = 0 \quad \forall m$$

3. distribution of quantization errors is uniform over each quantization interval

$$p_e(e) = 1/\Delta \quad -\Delta/2 \le e \le \Delta/2 \Rightarrow \bar{e} = 0, \ \sigma_e^2 = \frac{\Delta^2}{12}$$
$$= 0 \quad otherwise$$

Quantization Examples



Speech Signal

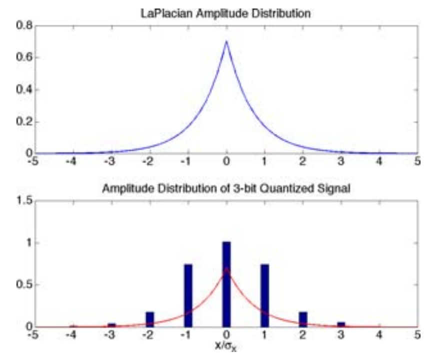
3-bit Quantization Error

8-bit Quantization Error (scaled)

how good is the model of the previous slide?

37

Typical Amplitude Distributions

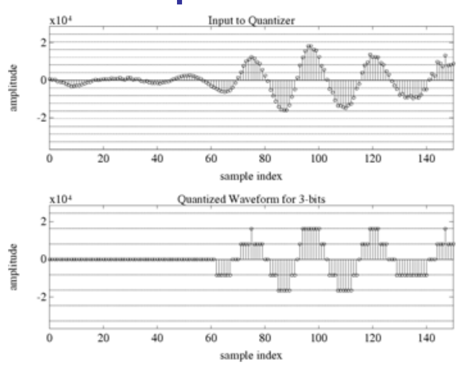


Laplacian distribution is often used as a model for speech signals

A 3-bit quantized signal has only 8 different values

38

3-Bit Speech Quantization

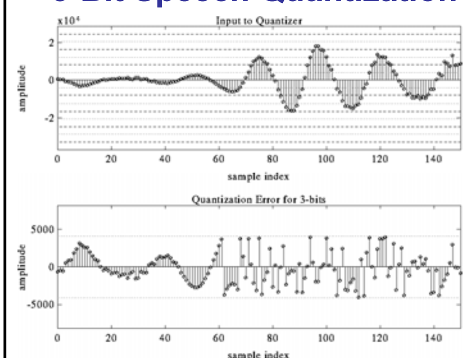


Input to quantizer

3-bit quantization waveform

39

3-Bit Speech Quantization Error

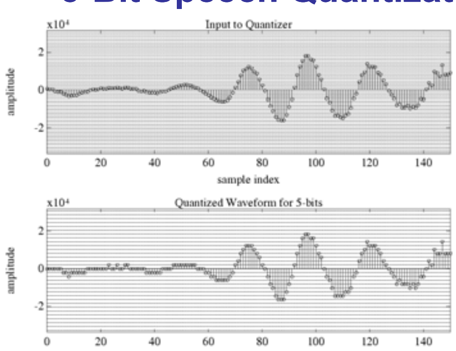


Input to quantizer

3-bit quantization error

40

5-Bit Speech Quantization

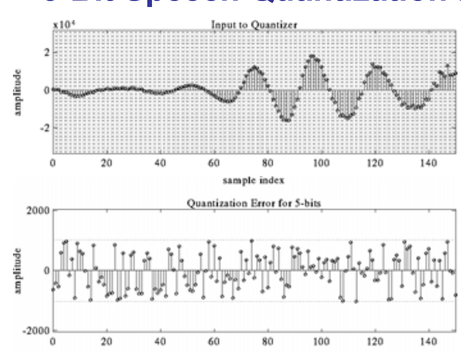


Input to quantizer

5-bit quantization waveform

41

5-Bit Speech Quantization Error

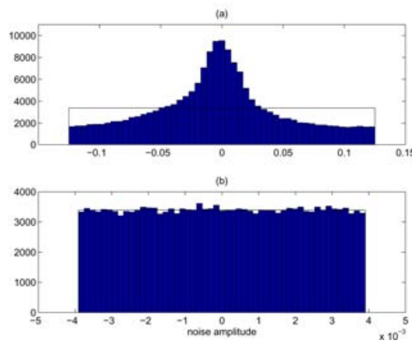


Input to quantizer

5-bit quantization error

42

Histograms of Quantization Noise

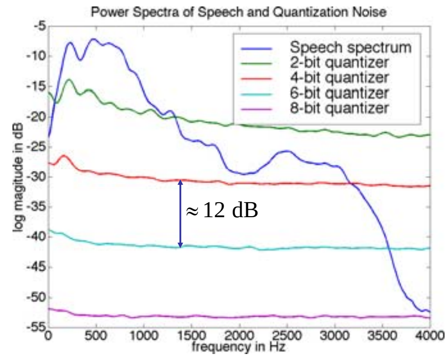


3-bit quantization histogram

8-bit quantization histogram

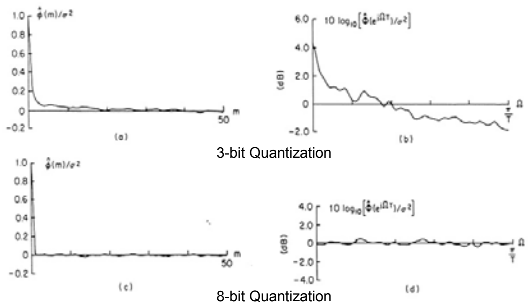
43

Spectra of Quantization Noise



44

Correlation of Quantization Error



assumptions look good at 8-bit quantization, not as good at 3-bit levels (however <u>only 6 db variation</u> in power spectrum level)

Sound Demo

- "Original" speech sampled at 16kHz, 16 bits/sample
- Quantized to 10 bits/sample
- Quantization error (x32) for 10 bits/sample
- Quantized to 5 bits/sample
- Quantization error for 5 bits/sample

46

SNR for Quantization

- can determine SNR for quantized speech as

$$SNR = \frac{\sigma_x^2}{\sigma_e^2} = \frac{E\left(x^2(n)\right)}{E\left(e^2(n)\right)} = \frac{\sum_n x^2(n)}{\sum_n e^2(n)}$$

$$\Delta = \frac{2X_{max}}{2^B} \text{ (uniform quantizer step size)}$$

- assume $p(e) = \frac{1}{\Delta} \quad -\frac{\Delta}{2} \le e \le \frac{\Delta}{2}$ (uniform distribution)

$\quad = 0$ *otherwise*

$$\sigma_e^2 = \frac{\Delta^2}{12} = \frac{X_{max}^2}{(3)2^{2B}}$$

47

SNR for Quantization

- can determine SNR for quantized speech as

$$SNR = \frac{\sigma_x^2}{\sigma_e^2} = \frac{E\left(x^2(n)\right)}{E\left(e^2(n)\right)} = \frac{\sum_n x^2(n)}{\sum_n e^2(n)}$$

$$\sigma_e^2 = \frac{\Delta^2}{12} = \frac{X_{max}^2}{(3)2^{2B}}$$

$$SNR = \frac{(3)2^{2B}}{\left[\frac{X_{max}}{\sigma_x}\right]^2}; \quad SNR(dB) = 10\log_{10}\left[\frac{\sigma_x^2}{\sigma_e^2}\right] = 6B + 4.77 - 20\log_{10}\left[\frac{X_{max}}{\sigma_x}\right]$$

- if we choose $X_{max} = 4\sigma_x$, then $SNR = 6B - 7.2$

$B = 16, \quad SNR = 88.8\ dB$

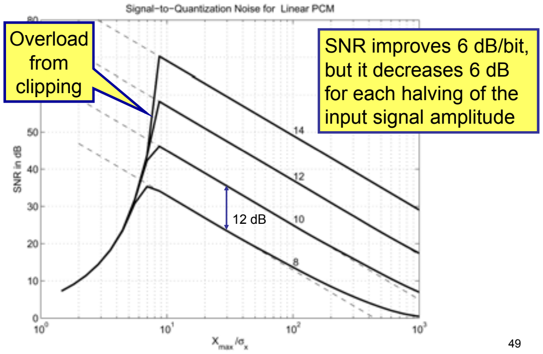
$B = 8, \quad SNR = 40.8\ dB$

$B = 3, \quad SNR = 10.8\ dB$

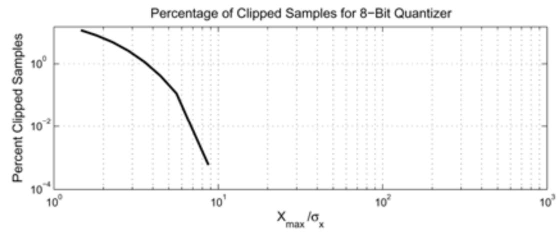
The term X_{max}/σ_x tells how big a signal can be accurately represented

48

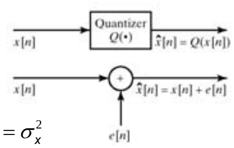
Variation of SNR with Signal Level



Clipping Statistics



Review--Linear Noise Model



$E\{(x[n])^2\}=\sigma_x^2$

- assume speech is stationary random signal.
- error is uncorrelated with the input.

$$E\{x[n]e[n]\}=E\{x[n]\}E\{e[n]\}=0$$

- error is uniformly distributed over the interval

$$-(\Delta/2)<e[n]\le(\Delta/2).$$

- error is stationary white noise, (i.e. flat spectrum)

$$P_e(\omega)=\sigma_e^2=\frac{\Delta^2}{12}, \quad |\omega|\le\pi$$

Review of Quantization Assumptions

1. input signal fluctuates in a complicated manner so a statistical model is valid
2. quantization step size is small enough to remove any signal correlated patterns in quantization error
3. range of quantizer matches peak-to-peak range of signal, utilizing full quantizer range with essentially no clipping
4. for a uniform quantizer with a peak-to-peak range of $\pm 4\,\sigma_x$, the resulting *SNR(dB)* is *6B-7.2*

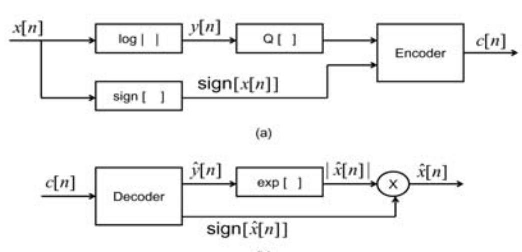
Uniform Quantizer SNR Issues

SNR=6B-7.2

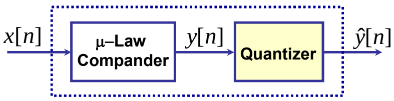
- to get an *SNR* of at least 30 dB, need at least $B\geq6$ bits (assuming $X_{max}=4\,\sigma_x$)
  - this assumption is weak across speakers and different transmission environments since σ_x varies so much (order of 40 dB) across sounds, speakers, and input conditions
  - *SNR(dB)* predictions can be off by significant amounts if full quantizer range is not used; e.g., for unvoiced segments => need more than 6 bits for real communication systems, more like 11-12 bits
  - need a quantizing system where the *SNR* is independent of the signal level => constant percentage error rather than constant variance error => need non-uniform quantization

Instantaneous Companding

- to order to get constant percentage error (rather than constant variance error), need logarithmically spaced quantization levels
  - quantize logarithm of input signal rather than input signal itself



μ-Law Companding



$$y(n) = \ln|x(n)|$$

$$x(n) = \exp[y(n)] \cdot sign[x(n)]$$

- where $sign[x(n)] = +1 \quad x(n) \geq 0$

  $= -1 \quad x(n) < 0$
- the quantized log magnitude is

$$\hat{y}(n) = Q[\log|x(n)|]$$

$$= \log|x(n)| + \varepsilon(n) \quad \text{new error signal}$$

μ-Law Companding

- assume that $\varepsilon(n)$ is independent of $\log|x(n)|$. The inverse is

$$\hat{x}(n) = \exp[\hat{y}(n)] \cdot sign[x(n)]$$
$$= |x(n)| \cdot sign[x(n)] \exp[\varepsilon(n)]$$
$$= x(n) \cdot \exp[\varepsilon(n)]$$

- assume $\varepsilon(n)$ is small, then $\exp[\varepsilon(n)] \approx 1 + \varepsilon(n) + ...$

$$\hat{x}(n) = x(n)[1 + \varepsilon(n)] = x(n) + \varepsilon(n)x(n) = x(n) + f(n)$$

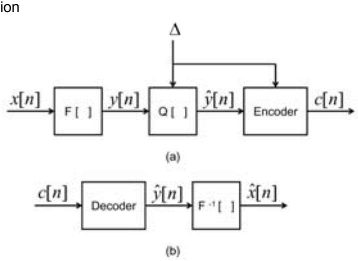
- since we assume $x(n)$ and $\varepsilon(n)$ are independent, then

$$\sigma_f^2 = \sigma_x^2 \cdot \sigma_\varepsilon^2$$
$$SNR = \frac{\sigma_x^2}{\sigma_f^2} = \frac{1}{\sigma_\varepsilon^2}$$

- SNR is independent of σ_x^2 --it depends only on step size

Pseudo-Logarithmic Compression

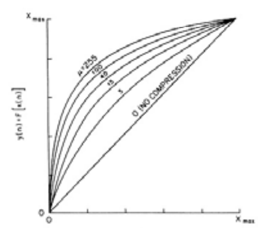
- unfortunately true logarithmic compression is not practical, since the dynamic range (ratio between the largest and smallest values) is infinite => need an infinite number of quantization levels
- need an approximation to logarithmic compression => μ-law/A-law compression



μ-law Compression

$$y(n) = F[x(n)]$$

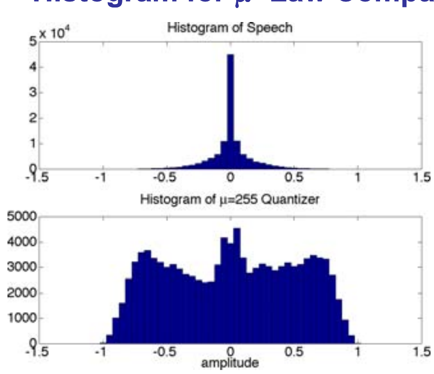
$$= X_{max} \frac{\log\left[1 + \mu \frac{|x(n)|}{X_{max}}\right]}{\log(1+\mu)} \cdot sign[x(n)]$$



- when $x(n) = 0 \Rightarrow y(n) = 0$
- when $\mu = 0$, $y(n) = x(n) \Rightarrow$ linear compression
- when μ is large, and for large $|x(n)|$

$$|y(n)| \approx \frac{X_{max}}{\log \mu} \cdot \log\left[\frac{\mu|x(n)|}{X_{max}}\right]$$

Histogram for μ -Law Companding

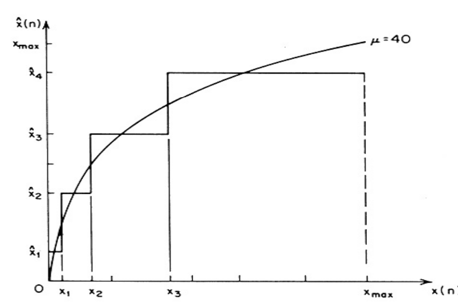


Speech waveform

Output of μ-Law compander

μ-law approximation to log

- μ – law encoding gives a good approximation to constant percentage error ($|y(n)| \propto \log|x(n)|$)



SNR for μ-law Quantizer

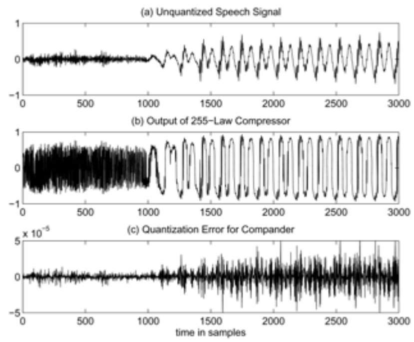
$$SNR(dB) = 6B + 4.77 - 20\log_{10}\left[\ln(1+\mu)\right] - 10\log_{10}\left[1 + \left(\frac{X_{\max}}{\mu\sigma_x}\right)^2 + \sqrt{2}\left(\frac{X_{\max}}{\mu\sigma_x}\right)\right]$$

- $6B$ dependence on $B \Rightarrow$ ***good***
- much less dependence on $\frac{X_{\max}}{\sigma_x} \Rightarrow$ ***good***
- for large μ, *SNR* is less sensitive to changes in $\frac{X_{\max}}{\sigma_x} \Rightarrow$ ***good***

  - μ-law quantizer used in wireline telephony for more than 40 years

61

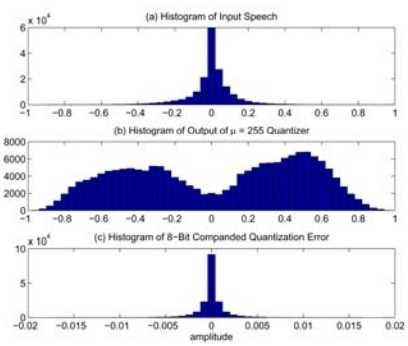
μ-Law Companding



Mu-law compressed signal utilizes almost the full dynamic range (±1) much more effectively than the original speech signal

62

μ-Law Quantization Error



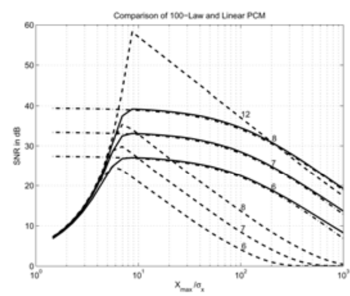
(a) Input speech signal

(b) Histogram of μ=255 compressed speech

(c) Histogram of 8-bit companded quantization error

63

Comparison of Linear and μ-law Quantizers

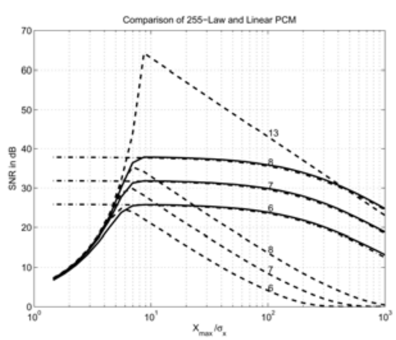


Dashed line – linear (uniform) quantizers with 6, 7, 8 and 12 bit quantizers

Solid line – μ-law quantizers with 6, 7 and 8 bit quantizers (μ=100)

• can see in these plots that X_{max} characterizes the quantizer (it specifies the 'overload' amplitude), and σ_x characterizes the signal (it specifies the signal amplitude), with the ratio (X_{max}/σ_x) showing how the signal is matched to the quantizer

Comparison of Linear and μ-law Quantizers



Dashed line – linear (uniform) quantizers with 6, 7, 8 and 13 bit quantizers

Solid line – μ-law quantizers with 6, 7 and 8 bit quantizers (μ=255)

65

Analysis of μ-Law Performance

- curves show that μ-law quantization can maintain roughly the same *SNR* over a wide range of X_{max}/σ_x, for reasonably large values of μ
  - for $\mu=100$, *SNR* stays within 2 dB for $8 < X_{max}/\sigma_x < 30$
  - for $\mu=500$, *SNR* stays within 2 dB for $8 < X_{max}/\sigma_x < 150$
  - loss in *SNR* in going from $\mu=100$ to $\mu=500$ is about 2.6 dB => rather small sacrifice for much greater dynamic range
  - $B=7$ gives *SNR*=34 dB for $\mu=100$ => this is 7-bit μ-law PCM-the standard for toll quality speech coding in the PSTN => would need about 11 bits to achieve this dynamic range and *SNR* using a linear quantizer

66

CCITT G.711 Standard

- **Mu-law** characteristic is approximated by 15 linear segments with uniform quantization within a segment.
  - uses a mid-tread quantizer. +0 and –0 are defined.
  - decision and reconstruction levels defined to be integers
- **A-law** characteristic is approximated by 13 linear segments
  - uses a mid-riser quantizer

67

Summary of Uniform and μ-Law PCM

- quantization of sample values is unavoidable in DSP applications and in digital transmission and storage of speech
- we can analyze quantization error using a random noise model
- the more bits in the number representation, the lower the noise. The ‘fundamental theorem’ of uniform quantization is that “the signal-to-noise ratio increases 6 dB with each added bit in the quantizer”; however, if the signal level decreases while keeping the quantizer step-size the same, it is equivalent to throwing away one bit for each halving of the input signal amplitude
- *μ*-law compression can maintain constant SNR over a wide dynamic range, thereby reducing the dependency on signal level remaining constant

68

Quantization for Optimum SNR (MMSE)

- goal is to match quantizer to actual signal density to achieve optimum *SNR*
  - *μ*-law tries to achieve constant *SNR* over wide range of signal variances => some sacrifice over *SNR* performance when quantizer step size is matched to signal variance
  - if σ_x is known, you can choose quantizer levels to minimize quantization error variance and maximize *SNR*

69

Quantization for MMSE

- if we know the size of the signal (i.e., we know the signal variance, σ_x^2) then we can design a quantizer to minimize the mean-squared quantization error.
  - the basic idea is to quantize the most probable samples with low error and least probable with higher error.
  - this would maximize the *SNR*
  - general quantizer is defined by defining *M* reconstruction levels and a set of *M* decision levels defining the quantization “slots”.
  - this problem was first studied by J. Max, “Quantizing for Minimum Distortion,” *IRE Trans. Info. Theory*, March,1960.

70

Quantizer Levels for Maximum SNR

- variance of quantization noise is:

$$\sigma_e^2 = E\left[e^2(n)\right] = E\left[\left(\hat{x}(n) - x(n)\right)^2\right]$$

- with $\hat{x}(n) = Q\left[x(n)\right]$. Assume M quantization levels

$$\left[\hat{x}_{-(M/2)}, \hat{x}_{-(M/2)+1}, \ldots, \hat{x}_{-1}, \hat{x}_1, \ldots, \hat{x}_{(M/2)}\right]$$

- associating quantization level with signal intervals as:

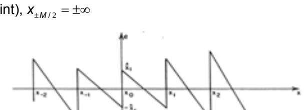
$\hat{x}_j =$ quantization level for interval $\left[x_{j-1}, x_j\right]$

- for symmetric, zero-mean distributions, with large amplitudes (∞) it makes sense to define the boundary points:

$x_0 = 0$ (central boundary point), $x_{\pm M/2} = \pm\infty$

- the error variance is thus

$$\sigma_e^2 = \int_{-\infty}^{\infty} e^2 p_e(e)de$$



71

Optimum Quantization Levels

- by definition, $e = \hat{x} - x$; thus we can make a change of variables to give:

$$p_e(e) = p_e(\hat{x} - x) = p_{x/\hat{x}}(x/\hat{x}) = p_x(x)$$

- giving

$$\sigma_e^2 = \sum_{i=-M/2+1}^{M/2} \int_{x_{i-1}}^{x_i} (\hat{x}_i - x)^2 p_x(x)dx$$

- assuming $p(x) = p(-x)$ so that the optimum quantizer is antisymmetric, then

$\hat{x}_i = -\hat{x}_{-i}$ and $x_i = -x_{-i}$

- thus we can write the error variance as

$$\sigma_e^2 = 2\sum_{i=1}^{M/2} \int_{x_{i-1}}^{x_i} (\hat{x}_i - x)^2 p_x(x)dx$$

- goal is to minimize σ_e^2 through choices of $\{x_i\}$ and $\{\hat{x}_i\}$

72

Solution for Optimum Levels

- to solve for optimum values for $\{\hat{x}_i\}$ and $\{x_i\}$, we differentiate σ_e^2 wrt the parameters, set the derivative to 0, and solve numerically:

$$\int_{x_{i-1}}^{x_i} (\hat{x}_i - x)^2 p_x(x)dx = 0 \qquad i = 1,2,\ldots,M/2 \qquad (1)$$

$$x_i = \frac{1}{2}(\hat{x}_i + \hat{x}_{i+1}) \qquad i = 1,2,\ldots,M/2-1 \qquad (2)$$

- with boundary conditions of $x_0 = 0,\ x_{\pm M/2} = \pm\infty \qquad (3)$
- can also constrain quantizer to be uniform and solve for value of Δ that maximizes SNR
  - optimum boundary points lie halfway between $M/2$ quantizer levels
  - optimum location of quantization level $\hat{x}_i$ is at the centroid of the probability density over the interval x_{i-1} to x_i
  - solve the above set of equations (1,2,3) iteratively for $\{\hat{x}_i\},\{x_i\}$

Optimum Quantizers for Laplace Density

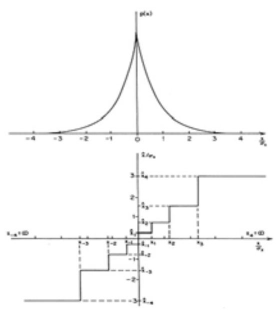
Assumes

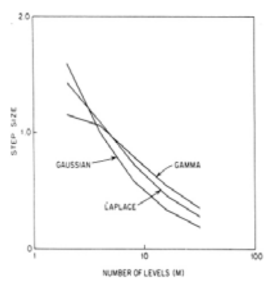
$\sigma_x = 1$

N	2		4		8		16		32	
i	x_i	$\hat{x}_i$	x_i	$\hat{x}_i$	x_i	$\hat{x}_i$	x_i	$\hat{x}_i$	x_i	$\hat{x}_i$
1	∞	0.707	1.102	0.395	0.504	0.222	0.266	0.126	0.147	0.072
2			∞	1.810	1.181	0.785	0.566	0.407	0.302	0.222
3					2.285	1.576	0.910	0.726	0.467	0.382
4					∞	2.994	1.317	1.095	0.642	0.551
5							1.821	1.540	0.829	0.732
6							2.499	2.103	1.031	0.926
7							3.605	2.895	1.250	1.136
8							∞	4.316	1.490	1.365
9									1.756	1.616
10									2.055	1.896
11									2.398	2.214
12									2.804	2.583
13									3.305	3.025
14									3.978	3.586
15									5.069	4.371
16									∞	5.768
MSE	0.5		0.1765		0.0548		0.0154		0.00414	
SNR dB	3.01		7.53		12.61		18.12		23.83	

M. D. Paez and T. H. Glisson, "Minimum Mean-Squared Error Quantization In Speech," *IEEE Trans. Comm.*, April 1972.

Optimum Quantizer for 3-bit Laplace Density; Uniform Case





quantization levels get further apart as the probability density decreases

step size decreases roughly exponentially with increasing number of bits

Performance of Optimum Quantizers

Table 5.3 Signal-to-Noise Ratios for 3-bit Quantizers. (After Noll [12]).

Nonuniform Quantizers	*SNR (dB)*	*Smallest Level* (σ_x=1)
μ-law ($x_{max} = 8\sigma_x$, μ=100)	9.5	0.062
Gaussian	14.6	0.245
Laplace	12.6	0.222
Gamma	11.5	0.149
Speech	12.1	0.124

Uniform Quantizers	*SNR*	*Smallest Level* (σ_x=1)
Gaussian	14.3	0.293
Laplace	11.4	0.366
Gamma	11.5	0.398
Speech	8.4	0.398

Summary

- examined a statistical model of speech showing probability densities, autocorrelations, and power spectra
- studied instantaneous uniform quantization model and derived *SNR* as a function of the number of bits in the quantizer, and the ratio of signal peak to signal variance
- studied a companded model of speech that approximated logarithmic compression and showed that the resulting *SNR* was a weaker function of the ratio of signal peak to signal variance
- examined a model for deriving quantization levels that were optimum in the sense of matching the quantizer to the actual signal density, thereby achieving optimum *SNR* for a given number of bits in the quantizer