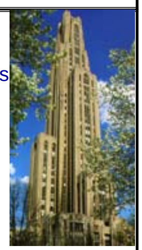
Digital Speech Processing

David Tipper
Associate Professor
Graduate Program of Telecommunications Networking
University of Pittsburgh
Telcom 2700/INFSCI 1072
Slides 7

http://www.sis.pitt.edu/~dtipper/tipper.html



Digital Speech Coding

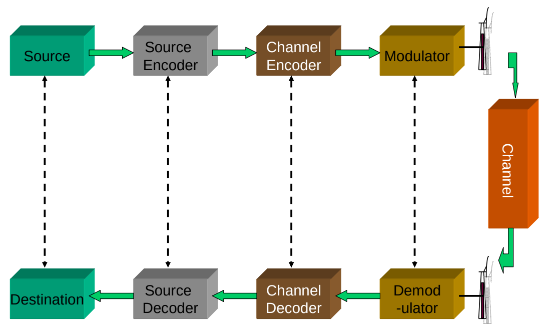
- Digital Speech
  – Convert analog speech to digital form and transmit digitally
- Applications
  – Telephony: (cellular, wired and Internet- VoIP)
  – Speech Storage (Automated call-centers)
  – High-Fidelity recordings/voice
  – Text-to-speech (machine generated speech)
- Issues
  – Efficient use of bandwidth
    - Compress to lower bit rate per user => more users
  – Speech Quality
    - Want tollgrade or better quality in a specific transmission environment
    - Environment (BER, packet lost, packet out of order, delay, etc.)
  – Hardware complexity
    - Speed (coding/decoding delay), computation requirement and power consumption

Digital Speech Processing

- Speech coding in *wireless* systems
  – All 1G systems have analog speech transmission
  – 2G and 3G systems have digital speech
  – Type of source coding

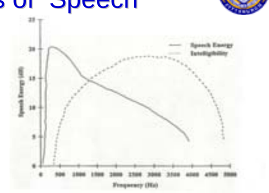
- Motivation for digital speech
  – Increase system capacity
    - Compression possible
    - Quality/bandwidth tradeoffs can be made
  – Improve quality of speech
    - Error control coding possible, equalization, etc.
  – Improve security as encryption possible for privacy
  – Reduce Cost and Operations and Maintenance (OAM)

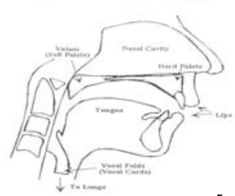
Typical Wireless Communication System



Characteristics of Speech

- Bandwidth
  - Most of energy between 20 Hz to about 7KHz ,
  - Human ear sensitive to energy between 50 Hz and 4KHz
- Time Signal
  - High correlation
  - Short term stationary
- Classified into four categories
  - Voiced : created by air passed through vocal cords *(e.g., ah, v)*
  - Unvoiced : created by air through mouth and lips *(e.g., s, f)*
  - Mixed or transitional
  - Silence

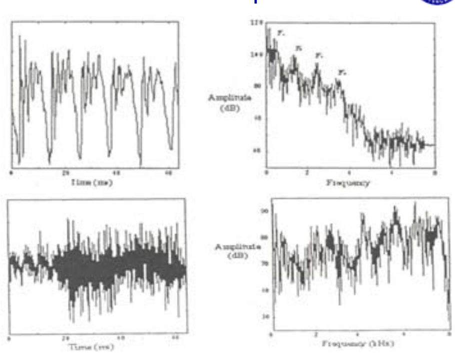




Characteristics of Speech

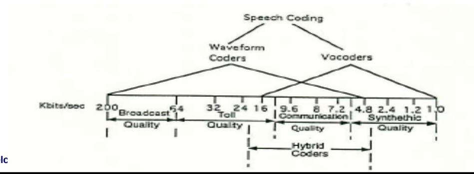
Typical Voiced speech *v*

Typical Unvoiced speech *s*



Digital Speech

- Speech Coder: device that converts speech to digital
- Types of speech coders
  - Waveform coders
    - Convert any analog signal to digital form
  - Vocoders (Parametric coders)
    - Try to exploit special properties of speech signal to reduce bit rate
    - Build model of speech – transmit parameters of model
  - Hybrid Coders
    - Combine features of waveform and vocoders



Speech Quality of Various Coders

Mean Opinion Score is a subjective measure of quality

Tradeoff in quality vs. data rate vs. complexity

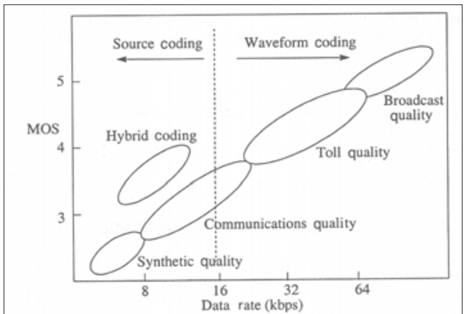
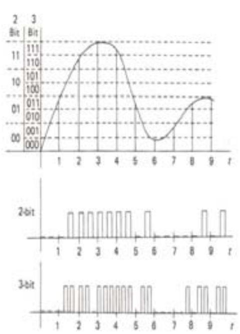


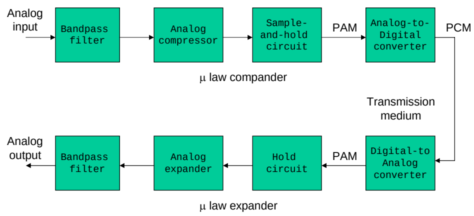
Figure 3.44 General speech quality versus transmission rate.

Waveform Coders (e.g.,PCM)

- **Waveform Coders**
  - Convert any analog signal to digital - basically A/D converter
  - Analog signal sampled > twice highest frequency- then quantized into ` *n* ' bit samples
  - Uniform quantization
  - Example Pulse Code Modulation
  - band limit speech < 4000 Hz
  - pass speech through μ–law compander
  - sample 8000 Hz, 8 bit samples
  - 64 Kbps DS0 rate
- **Characteristics**
  - Quality – High
  - Complexity – Low
  - Bit rate – High
  - Delay - Low
  - Robustness - High

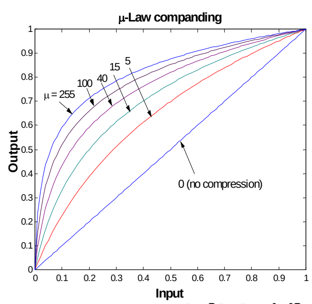


PCM Speech Coding



Pulse code modulation (PCM) system with analog companding then digital conversion
– ITU G.700 standard basis for speech coding In PSTN in 60's

Companding



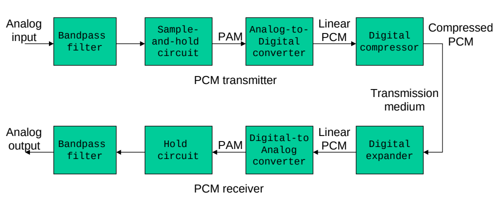
$$F(s) = \text{sgn}(s)\frac{\ln(1+\mu|s|)}{\ln(1+\mu)}$$

Analog Compander emphasizes small values, de-emphasizes large values in-order to equalize SNR across samples.

Reverse the mapping at the receiver with an expandor

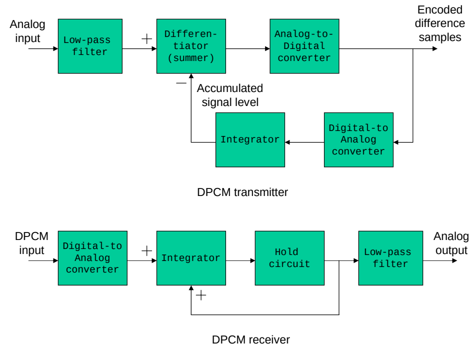
PCM Speech Coding

- Digitally companded PCM system – ITU G.711 standard
  - better quality speech than analog companding

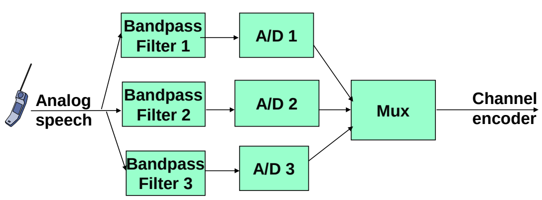


- Differential PCM (DPCM) : reduce bit rate from 64 Kbps to 32 Kbps)
  - since change is small between sample – transmit 1 sample
  - then on transmit difference between samples – use 4 bits to quantize
  - adaptively adjust range of quantizer – improves quality (ADPCM ITU G.726)

DPCM Speech Coding



Subband Speech Coding



Partition signal into non-overlapping frequency bands use different A/D quantizer for each band
Example: 3 subbands
5600+12000 + 13600 = 31.2 Kbps

band	Range	encoding
1	50- 700 Hz	4 bits
2	700-2000 Hz	3 bits
3	2000-3400Hz	2 bits

Vocoders

- Vocoders (Parametric Coders)
  - Models the vocalization of speech
  - Speech sampled and broken into frames (~25 msec)
  - Instead of transmitting digitized speech
    1. Build model of speech
    2. Transmit parameters of model
    3. Synthesize approximation of speech
- Linear Predictive Coders (LPC) basic Vocoder model
  - Models vocal tract as a filter
  - Filter excitation
    - periodic pulse (voiced speech) or noise (unvoiced speech)
  - Transmitted parameters:
    - gain, voiced/unvoiced decision, pitch (if voiced), LPC parameters

Vocoders

- Linear Predictive Coders (LPC)
  - Excitation
    - periodic pulse (voiced speech) or noise (unvoiced speech)
  - Transmitted parameters: gain, voiced/unvoiced decision, pitch (if voiced), LPC parameters

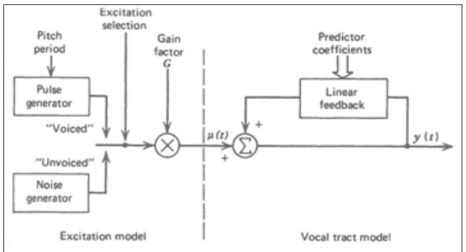


Figure 3.41 Speech generation model of linear predictive coding.

Vocoders

- Example Tenth Order Linear Predictive Coder
  - Samples Voice at 8000 Hz – buffer 240 samples => 30 msec
  - Filter Model
    - (M=10 is order, G is gain, z^{-1} unit delay, b_k are filter coefficients)

$$H(z) = \frac{G}{1 + \sum_{k=1}^{M} b_k z^{-k}}$$

- G = 5 bits, b_k = 8 bits each, voiced/unvoiced decision = 1 bit, pitch = 6 bits => 92 bits/30 msec = 3067 bps

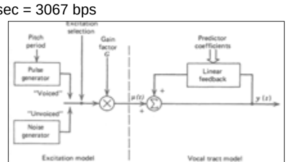


Figure 3.41 Speech generation model of linear predictive coding.

Vocoders

- LPC coders can achieve low bit rates 1.2 – 4.8 Kbps

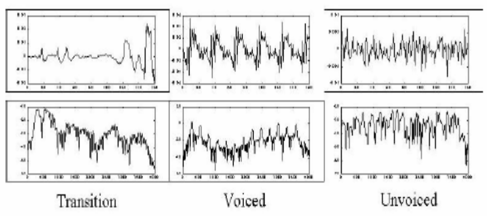
- Characteristics of LPC
  - Quality – Low
  - Complexity – Moderate
  - Bit Rate – Low
  - Delay – Moderate
  - Robustness – Low

- Quality of pure LPC vocoder to low for cellular telephony - try to improve quality by using hybrid coders

- Try to improve the quality by
  - refining model of speech,
  - improve accuracy of model
  - improve input to speech coder

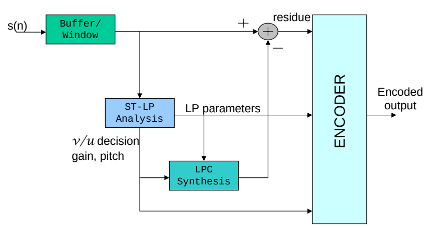
Vocoders

- Hybrid Coders
  - Combine Vocoder and Waveform Coder concept
    - Residual LPC (RELP)
    - Codebook excited LPC (CELP)



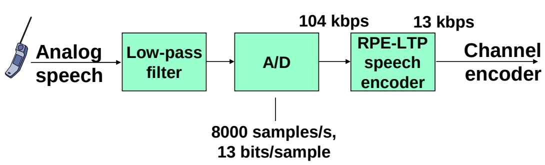
RELP Vocoder

- Residual Excited LPC
  - improve quality of LPC by transmitting error (residue) along with LPC parameters



Block diagram of a RELP encoder

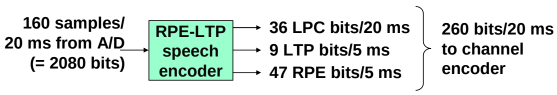
GSM Speech Coding



- GSM uses Regular Pulse Excited -- Linear Predictive Coder (RPE--LPC) for speech
  - –Basically combine DPCM concept with LPC
  - –Information from previous samples used to predict the current sample.
  - –The LPC coefficients, plus an encoded form of the residual (predicted - actual sample = error), represent the signal.

GSM Speech Coding (cont)

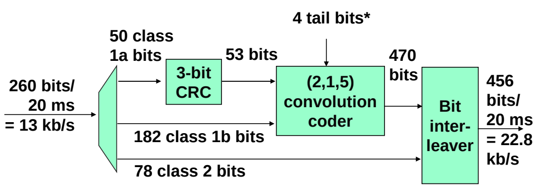
Regular pulse excited - long term prediction (RPE-LRP) speech encoder (RELP speech coder



LPC: linear prediction coding filter
LTP: long term prediction – pitch + input
RPE: Residual Prediction Error:

GSM Speech Coding (cont)

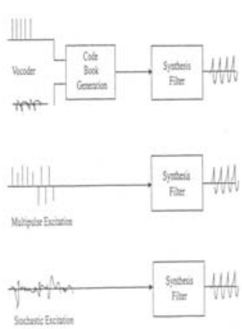
Channel encoder



Class 1a: CRC (3-bit error detection) and convolutional coding (error correction)
Class 1b: convolutional coding
Class 2: no error protection
*tail bits to periodically reset convolutional coder

Hybrid Vocoders

- **Codebook Excited LPC**
  - Problem with simple LPC is the voiced/unvoiced decision and pitch estimation doesn't model transitional speech well, and not always accurate
  - Codebook approach – pass speech through an analyzer to find closest match to a set of possible excitations (codebook)
  - Transmit codebook pointer + LPC parameters
  - NA-TDMA standard, IS-95, 3G, ITU G.729 standard



Typical CELP Encoder

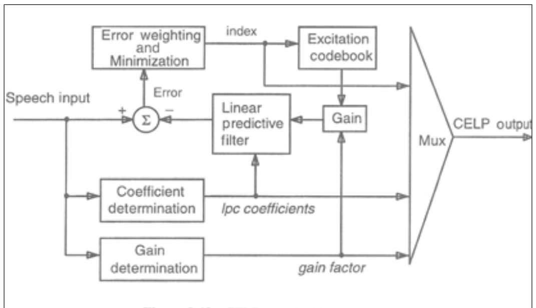
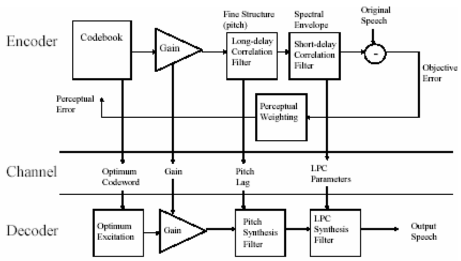


Figure 3.43 CELP encoder block diagram.

CELP Speech Coders

- General CELP architecture



Evaluating Speech Coders

- **Qualitative Comparison**
  - based on subjective procedures in ITU-T Rec. P. 830
- **Major Procedures**
- ***Absolute Category Rating***
  - Subjects listen to samples and rank them on an absolute scale - result is a mean opinion score (MOS)
- ***Comparison Category Rating***
  - Subjects listen to coded samples and original un-coded sample (PCM or analog), the two are compared on a relative scale – result is a comparison mean opinion score (CMOS)

Mean Opinion Score (MOS)

Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

Comparison MOS (CMOS)

Much Better	3
Better	2
Slightly Better	1
About the Same	0
Slightly Worse	-1
Worse	-2
Much Worse	-3

Evaluating Speech Coders

MOS for clear channel environment – no errors
Result vary a little with language and speaker gender

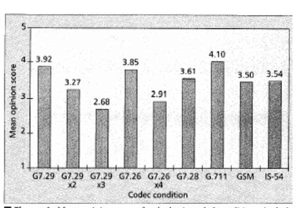
Standard	Speech coder	Bit rate	MOS
PCM	Waveform	64 Kbps	4.3
CT2	ADPCM	32 Kbps	4.1
DECT	ADPCM	32 Kbps	4.1
GSM	Hybrid RELPC	13 kbps	3.54
QCELP	Hybrid CELP	14.4 Kbps	3.4 – 4.0
QCELP	Hybrid CELP	9.6 Kbps	3.4
LPC	Vocoder	2.4 Kbps	2.5
ITU G.729	Hybrid CELP	8.Kbps	3.9

Qualcom Codebook Excited LP coder (cdmaone standard)

Evaluating Speech Coders

- Types of environments recommended for testing coder quality
  - Clean Channel no background noise
  - Vehicle : emulate car background noise
  - Street : emulate pedestrian environment
  - Hoth : emulate background noise in office environment (voice band interference)
- Consider environments above for cases of
  - Perfect Channel – no transmission errors
  - Random channel errors
  - Bursty channel errors
- May consider repeated encoding/decoding (e.g., mobile to mobile call)

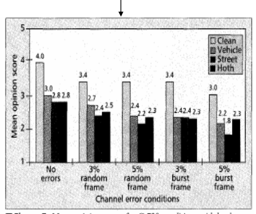
Evaluating Speech Coders



■ **Figure 1.** *Mean opinion scores for the basic coded conditions, including multiple incodings by a single codec.*

Repeated coding degrades quality

Background noise and errors degrade quality



■ **Figure 7.** *Mean opinion scores for G.729 conditions with background noise and random and burst frame erasure.*

Codec Selection

- For cellular need to consider Quality, Complexity, Delay, Compression Rate

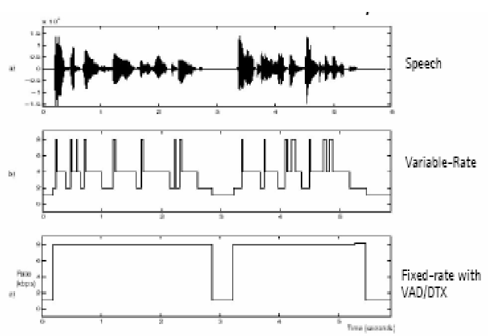
ITU Coder	Bit Rate	Coding Delay	Decoding Delay	Complexity
G.711	64 Kbps	0	0	Low
G.729	8 Kbps	15 ms	7.5 ms	Medium
G.723.a,b	6.4/5.3 Kbps	35.5 ms	18.75 ms	High

Silence Compression

- Much of a conversation is Silence (~40%)
  no need to transmit
- Voice Activity Detector (VAD)
  - Hardware to detect silence period quickly

1. Variable Bit Rate Coder Approach
   - reduce bit rate when silence detected – increase compression
   - Cdmaone and CDMA2000 codec use variable bit rate approach
2. Discontinuous transmission (DTX) Approach
   - Stop transmitting frames
     - Send minimal # of frames to keep connection up
   - Comfort Noise Generator (CNG)
     - Synthesize background noise - avoids: "Did you hang up?"
     - Random noise or reproduce speaker's ambient background
   - GSM, UMTS and popular VoIP G.723.1 codec use VAD/DTX/CNG

Silence Compression



Voice Coding

- Basic Voice Coding Approaches
  - Waveform
  - Vocoders
  - Hybrid Vocoders
- Evaluation of Vocoder Quality
- Codebook based vocoders use in new technology
- 3GPP and ITU recently standardized a
  - AMR wideband CELP
  - input 50-7000 HZ rather than 300-3400 Hz of current systems
  - more natural quality speech – slightly higher bit rate