Digital Speech Processing

David Tipper
Associate Professor
Graduate Program of Telecommunications Networking
University of Pittsburgh
Telcom 2720
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

![Typical Wireless Communication System Diagram](image-url)
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

### Typical Voiced speech

### Typical Unvoiced speech
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 \( \mu \)-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

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

Reverse the mapping at the receiver with an expander.

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

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

DPCM transmitter

```
Analog input
| Low-pass filter + Differentiator (summer) + Analog-to-Digital converter |
| + Accumulated signal level + Digital-to-Analog converter |
| + Integrator + Hold circuit + Low-pass filter |
```

DPCM receiver

```
DPCM input
| Digital-to Analog converter + Integrator + Hold circuit + Low-pass filter |
```

Subband Speech Coding

```
Analog speech
| Bandpass Filter 1 + A/D 1 + Mux |
| Bandpass Filter 2 + A/D 2 + Channel encoder |
| Bandpass Filter 3 + A/D 3 |
```

Partition signal into non-overlapping frequency bands use different A/D quantizer for each band
Example: 3 subbands

<table>
<thead>
<tr>
<th>band</th>
<th>Range</th>
<th>encoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50-700 Hz</td>
<td>4 bits</td>
</tr>
<tr>
<td>2</td>
<td>700-2000 Hz</td>
<td>3 bits</td>
</tr>
<tr>
<td>3</td>
<td>2000-3400Hz</td>
<td>2 bits</td>
</tr>
</tbody>
</table>

5600 + 12000 + 13600 = 31.2 Kbps
**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

---

**Linear Predictive Coders (LPC)**

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

![Speech generation model of linear predictive coding.](image)
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

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
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

- 260 bits/20 ms = 13 kb/s
- 50 class 1a bits
- 3-bit CRC
- 182 class 1b bits
- 53 bits
- (2,1,5) convolution coder
- 4 tail bits*
- 78 class 2 bits
- 470 bits
- 456 bits/20 ms = 22.8 kb/s

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 Vcoders**

- **Codebook Excited LPC**
  - Problem with simple LPC is U/V 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**

![CELP Encoder Block Diagram](image)

*Figure 3.43 CELP encoder block diagram.*

**CELP Speech Coders**

- **General CELP architecture**

![CELP Speech Coders Diagram](image)
CELP Speech Coders

Block diagram of the NA-TDMA (IS-54) speech coder – subband codebook approach – termed vector sum excited LPC (VSELPC)

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 uncoded sample (PCM or analog), the two are compared on a relative scale - result is a comparison mean opinion score (CMOS)

---

**Mean Opinion Score (MOS)**

<table>
<thead>
<tr>
<th>MOS</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>5</td>
</tr>
<tr>
<td>Good</td>
<td>4</td>
</tr>
<tr>
<td>Fair</td>
<td>3</td>
</tr>
<tr>
<td>Poor</td>
<td>2</td>
</tr>
<tr>
<td>Bad</td>
<td>1</td>
</tr>
</tbody>
</table>

**Comparison MOS (CMOS)**

<table>
<thead>
<tr>
<th>CMOS</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much Better</td>
<td>3</td>
</tr>
<tr>
<td>Better</td>
<td>2</td>
</tr>
<tr>
<td>Slightly Better</td>
<td>1</td>
</tr>
<tr>
<td>About the Same</td>
<td>0</td>
</tr>
<tr>
<td>Slightly Worse</td>
<td>-1</td>
</tr>
<tr>
<td>Worse</td>
<td>-2</td>
</tr>
<tr>
<td>Much Worse</td>
<td>-3</td>
</tr>
</tbody>
</table>
Evaluating Speech Coders

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

<table>
<thead>
<tr>
<th>Standard</th>
<th>Speech coder</th>
<th>Bit rate</th>
<th>MOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCM</td>
<td>Waveform</td>
<td>64 Kbps</td>
<td>4.3</td>
</tr>
<tr>
<td>CT2</td>
<td>ADPCM</td>
<td>32 Kbps</td>
<td>4.1</td>
</tr>
<tr>
<td>DECT</td>
<td>ADPCM</td>
<td>32 Kbps</td>
<td>4.1</td>
</tr>
<tr>
<td>NA-TDMA</td>
<td>Hybrid VSELP</td>
<td>8Kbps</td>
<td>3.0</td>
</tr>
<tr>
<td>GSM</td>
<td>Hybrid RELPC</td>
<td>13 kbps</td>
<td>3.54</td>
</tr>
<tr>
<td>QCELP</td>
<td>Hybrid CELP</td>
<td>14.4 Kbps</td>
<td>3.4 – 4.0</td>
</tr>
<tr>
<td>QCELP</td>
<td>Hybrid CELP</td>
<td>9.6 Kbps</td>
<td>3.4</td>
</tr>
<tr>
<td>LPC</td>
<td>Vocoder</td>
<td>2.4 Kbps</td>
<td>2.5</td>
</tr>
<tr>
<td>ITU G.729</td>
<td>Hybrid CELP</td>
<td>8 Kbps</td>
<td>3.9</td>
</tr>
</tbody>
</table>

• Types of environments recommended for testing coder quality
  – Clean Channel no background noise
  – Vehicle: emulate car background noise
  – Street: emulate pedestrian environment
  – Hoith: 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

Repeated coding degrades quality

Background noise and errors degrade quality

Codec Selection

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

<table>
<thead>
<tr>
<th>ITU Coder</th>
<th>Bit Rate</th>
<th>Coding Delay</th>
<th>Decoding Delay</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>G.711</td>
<td>64 Kbps</td>
<td>0</td>
<td>0</td>
<td>Low</td>
</tr>
<tr>
<td>G.729</td>
<td>8 Kbps</td>
<td>15 ms</td>
<td>7.5 ms</td>
<td>Medium</td>
</tr>
<tr>
<td>G.723.a,b</td>
<td>6.4/5.3 Kbps</td>
<td>35.5 ms</td>
<td>18.75 ms</td>
<td>High</td>
</tr>
</tbody>
</table>
3G Standards

- Two competing 3G standards
- Both standards use multi-mode CELP vocoders

1. 3GPP/cdma2000
   - (SMV – Multimode rate set 1)
   - Variable bit rate vocoder
   - Source Control of bit rate
   - Channel coding treats all bits equally

2. 3GPP/UMTS
   - (AMR-NB Multi-rate)
   - Fixed rate vocoder
   - Voice Activity Detection
   - Discontinuous Transmission
     - network control of coder rate
   - Tailors Channel coding to speech coder

Silence Compression

- Much of a conversation is Silence (~40%)
  - no need to transmit
- Voice Activity Detector (VAD)
  - Hardware to detect silence period quickly
- Variable Bit Rate coders – reduce bit rate when silence
- Discontinuous transmission (DTX)
  - 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
- For example GSM codec and popular VoIP G.723.1 codec has VAD/DTX/CNG
- Cdmaone and CDMA2000 codec use variable bit rate approach
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