GSM speech coding

Wolfgang Leister
Forelesning INF 5080
Vårsemester 2004

Norsk Regnesentral

Sources

This part contains material from:

- Web pages Universität Bremen, Arbeitsbereich Nachrichtentechnik (ANT): Prof. K.D. Kammeyer, Jörg Bitzer, Frank Jordan, Volker Kühn
  http://www.comm.uni-bremen.de/
  whomes/meyer/gsm_coder.html
- Sten Amundsen (INF-MKT presentation 2003)

Norsk Regnesentral
GSM

- Global System for Mobile communications,
  - originally developed for Europe
  - has now over 70% of the world market.
- Initially developed for operation in the 900MHz band and subsequently modified for the 850, 1800 and 1900MHz bands.
- GSM = Groupe Speciale Mobile (CEPT committee), 1982
- 1987: Memorandum to implement
- 1991: Start of operation

Best possible use of frequencies
- lower bandwidth than PSTN, LAN, WAN, etc: 13kBit/s
- mobile networks suffer from interferences and dropouts
  - but: not disturb end user
- GSM uses block coding
  - PCM and MPEG-1, MPEG-2 uses continuous coding
- Originally GSM is designed for:
  - One channel 16 kbit/s
  - Noise: BER of $10^{-2}$
  - Maximum coding delay: 30 ms
**GSM system architecture**

- radio access network and base net.
- Speech connection between GSM phone and PSTN or GSM

**Sender Functionality**

- Speech encoder receives PCM coded speech.
- Speech detector marks frames whether they contain speech.
- Background noise is sent in "silent suppression" mode.
### Receiver Functionality

- Speech decoder gets 13 kbit/s speech
- Speech frame interpolation replaces frames lost in transmission (using last received frame)
- Generate background noise using noise frames.

### Speech coding

- **Speech:**
  - bandwidth ca. 12 kHz
  - dynamics: 100 dB (whisper-shout)
  - 20 Bit resolution at 24kHz sampling rate: 480 kBit/s

- **Understandable speech:**
  - bandwidth 300Hz-3400Hz (telephone)
  - dynamics: 50 dB
  - 8 Bit resolution at 8kHz sampling rate: 64 kBit/s

- **Handy:** 13 kBit/s
  - coding of speech signal necessary (source coding)
What is speech?

- **Speech Generation:**
  - Lungs produce air stream
  - Vocal chord vibrates/opens
  - Mouth/Nose finalize sound

- **Artificial Speech Model:**
  - Source → All-Pol Filter → Synthetic Voice →

GSM speech encoder overview

Five Steps:
- AD conversion
- PCM sample blocks
- LPC = Linear Predictive Coding
- LTP = Long-Term Prediction
- RPE = Regular Pulse Excitation
GSM Speech encoder

- **Step 1:** produce digital audio/speech signal (microphone, low-pass filter, A/D converter)
- **Step 2:** Split up in blocks of 20 ms (160 samples) 
  statistical parameters do not change in short time intervals 
  (assume: short-time stationary behaviour)

---

**GSM Speech encoder**

- **Step 3:** Linear Predictive Coding
  Find filter coefficients with parametric estimation, e.g., Burg Algorithm
  Find reflection coefficients for lower data rate, quantize logarithmically
  Filter signal with quantized coefficients
  LPC uses 36 Bit/160 Samples transfer rate

<table>
<thead>
<tr>
<th>Parameter</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bits</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Werr</td>
<td>-0.932</td>
<td>0.735</td>
<td>0.145</td>
<td>0.69</td>
<td>-0.16</td>
<td>0.36</td>
<td>-0.2</td>
<td>-0.4</td>
</tr>
</tbody>
</table>

---

Norsk Regnesentral

Wolfgang Leister
GSM Speech encoder

- Step 4: Long-Term Prediction
  
  Split blocks into four parts of 40 samples each.
  
  There are still similarities in the encoding!
  
  For each block calculate difference to
  
  RPE from last step (cross-correlation)
  
  Calculate translation N0 and amplification b
  
  LPT uses (2+7)*4 Bit/160 samples transfer rate

- Step 5: Regular Pulse Excitation
  
  (a) Low pass filter with linear phase TP of degree 10
GSM Speech encoder

- Step 5: Regular Pulse Excitation
  - (b) Split rest signal into 3 polyphases
  - chose polyphase with largest energy
  - (c) Normation of max value of chose sequence (6bit)
  - quantize these 13 values with 3 bit linear

- RPE uses \((2+6+13*3)*4\) Bit/160 samples transfer rate

GSM transfer rate

For 160 samples we need:

<table>
<thead>
<tr>
<th></th>
<th>Bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPC filter coefficients</td>
<td>36</td>
</tr>
<tr>
<td>LTP filter coefficients</td>
<td>((7+2)*4) 36 Bit</td>
</tr>
<tr>
<td>Index and Norm Step 5</td>
<td>((6+2)*4) 32 Bit</td>
</tr>
<tr>
<td>Rest signal</td>
<td>(13<em>3</em>4) 156 Bit</td>
</tr>
<tr>
<td></td>
<td>260 Bit</td>
</tr>
</tbody>
</table>

Data rate: 260bit/20ms = 13kbit/s
GSM speech coding

- Encoder consists of following parts:
  - Short term LPC (Linear Prediction Coding) -analysis
  - Short term LPC-filter
  - Long term LTP (Long Term Prediction) -analysis
  - Long term LTP-filter
  - RPE (Regular Pulse Excitation) position and coding of rest signal

LPE-LPT speech decoder

- Decoder consists of three parts:
  - RPE decoding and position
  - LTP synthesis filter
  - LPC short time synthesis filter
GSM Sound examples

- Original
- GSM full-rate 13 kBit/s, error-free transmission
- Enhanced GSM full-rate 12.2 kBit/s, error-free transmission
- GSM half-rate 5.6 kBit/s, error-free transmission
- old Vocoder 5 kBit/s (no standard)

<table>
<thead>
<tr>
<th>Original</th>
<th>GSM full-rate, 13 kBit/s</th>
<th>Enhanced GSM full-rate, 12.2 kBit/s</th>
<th>GSM half-rate, 5.6 kBit/s</th>
<th>Old Vocoder (5 kBit/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male voice, English, 8 kHz, 16 Bit linear</td>
<td><img src="audio1_original.wav" alt="Audio" /></td>
<td><img src="audio1_enhanced.wav" alt="Audio" /></td>
<td><img src="audio1_halfrate.wav" alt="Audio" /></td>
<td><img src="audio1_vocoder.wav" alt="Audio" /></td>
</tr>
<tr>
<td>Male voice, German, 8 kHz, 16 Bit linear</td>
<td><img src="audio2_original.wav" alt="Audio" /></td>
<td><img src="audio2_enhanced.wav" alt="Audio" /></td>
<td><img src="audio2_halfrate.wav" alt="Audio" /></td>
<td><img src="audio2_vocoder.wav" alt="Audio" /></td>
</tr>
<tr>
<td>Music sample, 8 kHz, 16 Bit linear</td>
<td><img src="audio3_original.wav" alt="Audio" /></td>
<td><img src="audio3_enhanced.wav" alt="Audio" /></td>
<td><img src="audio3_halfrate.wav" alt="Audio" /></td>
<td><img src="audio3_vocoder.wav" alt="Audio" /></td>
</tr>
</tbody>
</table>

Future Development

- GSM is an ETSI standard
- RPE-LPT standardised for GSM by ETSI
- Three RPE-LTP speech coder standards:
  - FR: Full rate: 13 kbit/s
  - HR: Half rate: 5.6 kbit/s (frequency usage better, worse speech quality)
  - EFR: enhanced full rate: 12.2 kbit/s (better speech quality than FR)
- New specifications by ETSI
- Defines dynamic way to change speech encoders
End of Part

Thank you for your attention!