Analysis-by-Synthesis Coding

Nimrod Peleg
Update: Jan 2004
A-by-S Coding

- RELP, APC, ATC and SBC have sufficient quality at rates of 9.6-16Kbps
- Two main reasons:
  - Coded speech is not analyzed to check efficiency (No distortion control)
  - Errors accumulated from previous frames are not taken into account in current frame (errors propagate since no "reset" in loop)
Analysis-by-Synthesis LP Coders

- Analysis-by-Synthesis coders use close-loop for the excitation sequence determination.
- An optimization process determines an excitation sequence which minimizes a measure of the difference between input and coded speech.
- A weighting function is chosen to optimize for human ear.
- A better quality achieved for 4.8-9.6Kbps.
General A-by-S Encoder

Input Speech → Time-Varying Filter

LPC Synthesis Filter + Error Signal → Pitch Synthesis Filter

Weighting Filter → Index Selection

Error Minimization → Selected Optimum Excitation Index
A-by-S Operation

• Initialize filters (low-level random noise)
• Determine LPC coeff. on a frame of samples
• Divide the frame to subframes, for each:
  – calculate pitch parameters (delay & scaling)
  – Pitch and LPC filters are grouped together. The cascaded filter is used to determine the best secondary excitation, to minimize error between original and synthesized speech
• Final speech is generated by passing optimum secondary excitation through cascaded filter
General A-by-S Decoder

- Selected Optimum Excitation Index
- Time-Varying Filter
  - Pitch Synthesis Filter
  - LPC Synthesis Filter
- Long-term
- Short-term
- Output Speech
Short-Term Predictor

• Models the short-term correlation in the speech: **Spectral envelope**
• It is time varying to reflect changes in speech spectrum
• Adaptation rates: 20-30mSec
• Filter order: 8-12
Long-Term Predictor

- Models the long-term correlation in the speech: **Pitch period**
- Higher adaptation rates: 5-10mSec
- LTP can be omitted (as in MPLPC)
- Both LTP and STP have buffer memory of previous frame that provide smoothing effect to distortion (caused by block oriented analysis)
Error Minimization

• Minimizing perceptual error by selecting filter(s) parameters such that \(\text{MSE}\) between original and reconstructed signal is minimized

• This results in a flat spectrum of quantization error, (better results can be obtained by exploiting the auditory masking of human hearing)

• All the process done in transmitter only!
Excitation Generator

• Includes all information missing in the model parameters (residual signal)
• Several methods for it:
  – CELP (Code Excited LP)
  – SELP (Self Excited LP)
  – MPLPC (Multi-Pulse LP)
  – RPE-LPC (Regular-Pulse LPC)
  – MELP (Mixed Excitation LP)
Excitation Types

- CELP
- SELP
- MPLPC
- RPE-LPC

Output Speech

Time-Varying Filter
CELP

• Excitation vector chosen from a set of $N$ possible stochastic vectors (+gain)
• The vector that produces lowest error is the desired excitation sequence
• Only the index needed to be sent to receiver
• Many schemes to codebook design, among them: unit variance white Gaussian random noise (Atal)
Speech Synthesis Model

White Gaussian Innovation

Long term Predictor

Fine structure
(Pitch)

Short term Predictor

Spectral Envelope
(Formants)

‘Speech’
CELP Basic Scheme

Original Speech

Gk

Stochastic CodeBook

Long term Predictor

Short term Predictor

Perceptual Weighting Filter (W(z))

Square

Average

Perceptual Error

e(n)
Optional Improvements

• Adaptive CodeBook (closed loop pitch prediction)
• Low-Delay (LD-CELP): Backward adaptive CELP Coder:
  – Linear prediction of order 50
  – Rate: 16Kbps (Toll quality)
Vector Sum Excitation LP (VSELP)

Orthogonal vectors create a basis for linear combinations of white Gaussian noise

$G_k + G_1 + G_2$
MPE-LPC

• The first A-by-S scheme (Multi Pulse Excitation)
• Excitation is specified by a small set of pulses with different amplitudes located at non-uniformly spaced intervals
• The only a-priory decision to make: The number of pulses required per block (usually about 1 per 10 samples, or 8 pulses for 10mSec)
• Locations and amplitudes should be sent
RPE-LPC

• regular Pulse Excitation LPC
• Pulses are equally spaced, and positions are specified by the first pulse position
• This prevents the need for position information
• Used in GSM
MPE Vs. RPE

MPE

RPE
GSM Encoder

LPC Analysis

+ Side Information

RPE

Transmitted

Inverse RPE

Long term predictor

S.I

Encoded Data

Similar Scheme to classic DPCM
GSM Decoder

Encoded Data → Inverse RPE → LTP → + → LPC Synthesis → Recon. speech

S.I
### Performance and Complexity (after Spanias, 1994)

<table>
<thead>
<tr>
<th>Technique</th>
<th>Rate (bps)</th>
<th>MOS</th>
<th>MIPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCM (G.711)</td>
<td>64K</td>
<td>4.3-4.4</td>
<td>0.01</td>
</tr>
<tr>
<td>ADPCM (G.726)</td>
<td>16K-40K</td>
<td>4.1-4.2</td>
<td>~2</td>
</tr>
<tr>
<td>LD-CELP (G.728)</td>
<td>16K</td>
<td>4.0-4.2</td>
<td>~19</td>
</tr>
<tr>
<td>RPE-LTP (GSM)</td>
<td>13k</td>
<td>3.47</td>
<td>6</td>
</tr>
<tr>
<td>VSELP (IS-54)</td>
<td>8k</td>
<td>3.45</td>
<td>13.5</td>
</tr>
<tr>
<td>CELP (DoD1016)</td>
<td>4.8k</td>
<td>3.2</td>
<td>16</td>
</tr>
<tr>
<td>LPC-10e (DoD1015)</td>
<td>2.4k</td>
<td>2.3</td>
<td>7</td>
</tr>
<tr>
<td>CS-ACELP (G.729/729a)</td>
<td>8k</td>
<td>4.2</td>
<td>17 / 10</td>
</tr>
<tr>
<td>CS-ACELP (G.723)</td>
<td>5.3k/6.3k</td>
<td>3.5/4.0</td>
<td>14.6 / 16</td>
</tr>
<tr>
<td>(Conjugate-Structure Algebraic CELP)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MELP</td>
<td>2.4Kbps</td>
<td>2.9*</td>
<td>20</td>
</tr>
</tbody>
</table>

* Better in Noisy environment!
Input Speech

Synthesis Filter

Weighting Filter

Excitation CodeBook

Index Selection

Spectral CodeBook
SELP

- Excitation signal is derived from the past history of the coded excitation function itself
- When the best vector is found, it is fed back with oldest samples discarded
- Its effectively a CELP coder with an adaptive codebook, containing no codebook...
MELP Principles

- Mixed Excitation
- Aperiodic Pulses
- Adaptive Spectral Enhancement
- Pulse Dispersion
- Fourier Magnitudes
MELP Encoder/Decoder

Encoder

Decoder

- LPC Analysis
- Voicing & Aperiodic Analysis
- Gain Analysis
- Pitch Analysis
- Fourier Magnitude Analysis

- bit
- 4 bit Quantization
- bit
- bit

- Aperiodic Flag
- P(n) pulse train
- W(n) white noise

- Adaptive Spectral Enhancement
- LPC Filter
- Gain
- Pulse Dispersion Filter

- Synthesized speech
MELP Coder Scheme