University of California
Berkeley

College of Engineering
Department of Electrical Engineering
and Computer Sciences

Professors: N. Morgan / B. Gold
EE225D

Low Rate Coding

Lecture 25
## Digital Vocoder

<table>
<thead>
<tr>
<th>Date Rates</th>
<th>Applications &amp; Performance</th>
<th>Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>30,000 bps</td>
<td>Wide-band, High Fidelity Speech Transmission</td>
<td>Adoptive Differential</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pulse Code Modelation [Standard?]</td>
</tr>
<tr>
<td>10,000-20,000 bps</td>
<td>Medium Band, Good Quality Speech Transmission. Some Noise Immunity</td>
<td>Split Band Systems</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Part Modelling, Part Waveform Coding)</td>
</tr>
<tr>
<td>5,000-10,000 bps</td>
<td>Reasonable Quality Telephone Speech</td>
<td>Voice Excited Channel Vocoder</td>
</tr>
<tr>
<td></td>
<td></td>
<td>VELP &amp; RELP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adaptive Predictive Coding.</td>
</tr>
<tr>
<td>2,000-5,000 bps</td>
<td>More Vulnerable to Environment.</td>
<td>CELP, Multipolars, STC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speech Digitization Algorithms.</td>
</tr>
<tr>
<td>1,000-2,000 bps</td>
<td>Secrecy. Very Restricted Channels.</td>
<td>Frame-fill for Standard 2400bps Systems.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Transformation to Reduce # of Bits for Parameters</td>
</tr>
<tr>
<td>500-1000 bps</td>
<td>Extremely Restricted or Busy Channels.</td>
<td>Vector Quantization</td>
</tr>
<tr>
<td>100-500 bps</td>
<td>Underwater Transmission? Super-Restricted Channels.</td>
<td>Recognition-Synthesis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Phonetic Vocoder.</td>
</tr>
</tbody>
</table>
Frame - fill

* Start with a 2400 bps algorithm [Channel Vocoder].
* How do you get 2400bps?
  
  Assume 400bps for excitation.
  
  Assume 20ms frame or 50 frames/sec.

\[
\frac{2000}{5\varnothing} = 40 \text{ bits/channel for a channel vocoder.}
\]

Assume 16 channels  \[
\frac{40}{16} = 2.5 \text{ bits/channel, bits may be assigned on a perceptual basis. [Usually low channels need more bits.]}\]
**Question**

How to reduce bit rate below 2400bps?

**Algorithm**

- Transmit A&C
- Reconstruct B at Receiver.

Measurement are model at the transmitter.
Control bits tell Receiver how to handle the incoming data.

If $B \approx A$, reconstruct $B = A$.

- If $B \approx C$, reconstruct $B = C$.

- If $B \approx A \notin B \approx C$, reconstruct $B$ by averaging $A \notin C$.

So 40bits gets reduced to 22. Pitch can also be reduced.

$$40\text{bits} \times 50 = 2000$$  
$$42\text{bits} \times 25 = 1050$$
**Frame - fill for LPC**

Whereas the channel vocoder synthesizer is always the same structure (a filler bank with malulators for each b.p.filter) There are a few different LPC structures. For frame-fill, the trick is to find the best structure for finding a criterion of similarity.

- a-parameters — direct form
- k - parameters — lattice form
- area ration
- log area ration

**Experimental Result**

Lattice structure works best.

**DRT Results**

Table IV
**Pattern Matching on Vector Quantization**

C.P. Smith (1960’s)  
Buzo et al (1980’s ?)

Consider the 2400bps channel vocoder.

Assume 40 bits per frame.

* This corresponds to $2^{40}$ possible patterns.

Assume that $10^6$ patterns are perceptually distinguishable.

If all $10^6$ patterns were stored at both the receiver and transmitter.

**Algorithm**

- Compare incoming pattern with All $10^6$ stared patterns.
- Send the 20 bit address of the best stared pattern.
- Reconstruct this best pattern at receiver.

Thus, bit rate is decreasing by 2:1.

If only 1000 patterns were perceptually distinguishable, we gain from 40 to 10 bits per frame. [4:1 Reduction]
**Kany - Caulter**  600bps Vocoder

**Includes** LPC analysis

- clever LPC parameters to fewer formant parameters.

Vector quantization. - a form of frame fill.

**Viewgraphs**

- Fig. 32.1 - Complete block diagram of Kamy - Coulter.
- Fig. 32.2 - loci of poles is $K_{10} \rightarrow 1$
- Fig. 32.3 - Spectrum of all pole system as poles migrates towards unit circle.

**Vector Quantization** - 128 formant patterns were stored.

Figure overall rate, (including everything) (7bits) possible parameters to transmit.

Voiced formants PARCOR

Viewgraph of Lattice

Figure overall rate, to get 600 bps.
Frame - Fill vs. Merging

In frame-fill, alternate frames are not sent - instead, Control bits are sent to direct the receiver to “optimally” recreate the missing frames. The sent frames can still be vector quantized.

In merging, the analyzer compares adjacent frames and decides whether these frames can be merged into a single frame. Perhaps more than just two adjacent frames can be merged. This means that the transmission rate is Variable, and Control bits are needed. Merged framed can also be vector quantized.
Figure 29.8: Parallel formant synthesizer.
Pulse Source

Serial Structure for Vowels and Nasals

Parallel Structure for Most Consonants

Figure 29.12: Structure of Klatt Synthesizer.
Figure 29.2: OVE II Speech Synthesizer of Gunnar Fant. Form [20]
Figure 32.1: Kang-Coulter 600 bps Voice Digitizer.
Figure 32.1: Kang-Coulter 600bps Voice Digitizer

(b) Receiver
Figure 32.2: Loci of the Poles as $k_{10}$ Approaches Unity.
Figure 32.3: Spectra of All-Pole System as Poles Migrate Towards Unit

\[ Z^n - \alpha_1 Z^{n-1} - \alpha_2 Z^{n-2} \ldots - \alpha_n = \sum_{i=1}^{L} \prod (Z - Z_i) \]

as \( d_n \to 1 \), \( Z_i \)’s migrate to unit circle.
Predictor Coefficients

\[ y(n) = \frac{p}{k} k_{1} k_{2} k_{3} \ldots k_{p} \]

(a) Direct-Form Digital Filter with Variable “a” Coefficients

(b) Acoustic Tube with Variable Area Functions

Figure 29.5: Two configurations for all pole synthesizers based on LPC analysis. (cont.)
(c) All-Pole lattice Network with Variable “k” Parameters

Figure 29.5 : Two configurations for all pole synthesizers based on LPC analysis.

Synthesis Strategy for Voiced Sounds For unvoiced Sounds.

Vector quantization of the PARCOR Coefficients.
## Figure 32.4: Parameter Coding for 600bps Voice Digitizer

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Typical 2400-Bit-Per-Second Linear Predictive Encoder</th>
<th>600-Bit-Per-Second Voice Digitizer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame rate</td>
<td>44.444Hz</td>
<td>40Hz</td>
</tr>
<tr>
<td>Vocal-tract-filter parameters</td>
<td>40 bits/frame</td>
<td>7 bits/frame</td>
</tr>
<tr>
<td>Excitation parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voice/unvoice decision</td>
<td>1 bit/frame</td>
<td>1 bit/frame</td>
</tr>
<tr>
<td>Amplitude</td>
<td>6 bits/frame</td>
<td>4 bits/frame</td>
</tr>
<tr>
<td>Pitch</td>
<td>6 bits/frame</td>
<td>5 bits/double frame</td>
</tr>
<tr>
<td>Synthronization</td>
<td>1 bit/frame</td>
<td>1 bit/double frame</td>
</tr>
<tr>
<td>Total number of bits</td>
<td>54 bits/frame</td>
<td>30 bits/double frame</td>
</tr>
</tbody>
</table>
**V.O. Algorithm**

- Storage of the patterns.

  Compare incoming pattern with all previously stored patterns.

  \[ 2^N \] patterns stored.

- Receiver also has stored available.

  Send the address of the stored pattern nearest to the actual pattern.
**Kang-Coulter 600bps Vocoder**

- Started with 2400bps ---- LPC 10 Vocal Tract Parameter
- Reduce # of parameter to 4 or 5
- Vector quantization
- Frame-fill.

Speech → Recognizer  | Speech to Text  → Text to Speech
has no speaker I.D.

75bps
## Low Rate Vocoder

<table>
<thead>
<tr>
<th>Date Rates</th>
<th>Applications &amp; Performance</th>
<th>Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>30,000 bps</td>
<td>High Fidelity (Telephony)</td>
<td>ADPCM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DPCM</td>
</tr>
<tr>
<td>10,000-20,000 bps</td>
<td>Medium quality</td>
<td>Split Band Systems</td>
</tr>
<tr>
<td></td>
<td>Good Noise</td>
<td>(Part Modelling, Part Waveform Coding)</td>
</tr>
<tr>
<td></td>
<td>Some Noise Immunity</td>
<td></td>
</tr>
<tr>
<td>5,000-10,000 bps</td>
<td>Reasonable Quality [High Intelligibits]</td>
<td>Voice Excited Vocoders</td>
</tr>
<tr>
<td>2,400 bps (standard)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,000-2,000 bps</td>
<td>Secrecy. Very Restricted Channels.</td>
<td>Frame-Fill Transformation other Systems.</td>
</tr>
<tr>
<td>500-1000 bps</td>
<td>Extremely Restricted or Busy Channels.</td>
<td>Vector Quantization</td>
</tr>
<tr>
<td>100-500 bps</td>
<td>Underwater</td>
<td>Recognition-Synthesis Phanetic Vocoder.</td>
</tr>
</tbody>
</table>
**1000-2000bps Frame-fill**

start with a complete System ———— 2400bps

<table>
<thead>
<tr>
<th>400bps</th>
<th>Excitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000bps</td>
<td>Vocal Tract</td>
</tr>
</tbody>
</table>

Analyzer

| A | 0 | B | 20ms | C | 40ms |

not sent

Synthesizer

\[ n \quad F_1 \quad \frac{F_1 + F_2}{2} \quad n \quad F_2 \]

If B is close to A, \( B = A \).

If B is close to C, \( B = C \).

If neither is true, - Interpolate.

2400bps ———— 1200bps
Reduction of # of Parameters

500-10Mbps

C.P. Smith → Pattern Matching

2400bps Vocoder

2000bps Spectrum

50Hz Rate → 40 Bits Per Frame \[50 \times 40 = 2000\]

\[2^{40}\] Different Spectrum [Channel Vocoder]

\[2^{20}\] 1 million

Vector Quantization → LPC
Frame-fill for LPC?

Predictor coefficients $a'$s $\leftrightarrow$

K-Parameter [PARCOR]
Area Retios.
Log Area Ratios.

Sensitivity