Very Low Bit-Rate Speech Codec Using Temporal Decomposition

Ronen Krupnik
Erez Cohen

Supervisor : Guy Narkiss
Semester : Spring, 2005
In Cooperation With Tadiran Communication

5. 7.2006

Abstract

- The project deals with signal processing coding.
- Temporal Decomposition and VQ is used for compressing speech parameters.
- The project is a MATLAB simulation, which interfaces with MELP-2400 vocoder.
- The motivation: Decreasing Band-Width for transmitting, save space on disk etc…
- The Instruments:
  - Enforcing aggressive restrictions on the speech parameters
  - Applying better Vector Quantization methods
**LPC – Linear Predictive Coding**

- Basic speech signal representation
- Based on voiced/unvoiced excitation switching
- Not reliable enough

\[ H(z) = \frac{1}{1 - \sum_{j=1}^{I} a_jz^{-j}} \]

**Melp – Mixed Excitation Linear Prediction**

- The System is based on MELP-2400
- v/u analysis on 5 different band-passes
- LPC coefficients are represented in a frequency domain - LSF

- Speech Parameters data frame is transmitted every 22.5 ms
- Each data frame is 54 bit size:
  - 25 bits for LSF parameters
  - 7 bit for pitch
  - 8 bits for gain
  - 14 bits for band-pass voicing decision, Fourier magnitudes etc..
Temporal Decomposition

- Using speech signal’s correlation in time between adjacent frames.
- Buffering the signal to N length’s blocks consists of K speech events.

Speech event properties:
- Target Vector - $a(k), 1 \leq k \leq K$
- Event function - $\phi(k), 1 \leq k \leq N$ (target’s weight on every frame)
- Event location - $n_i, 1 \leq n_i \leq N$

Temporal Decomposition

$$\hat{y}(n) = \sum_{k=1}^{K} a(k) \cdot \phi(k) \Rightarrow \hat{Y}_{\phi_N} = \hat{A}_{\phi_N} \cdot \Phi_{\phi_N}$$

Iterative solution

- Restricted TD - only adjacent functions may overlap

This reduces the computation complexity
Additional Restrictions on Event Functions

- In most cases:
  \[ \varphi_n(k) + \varphi_n(k+1) \approx 1 \]
  No need to transmit both values
- In addition, in more than 95% of the cases:
  \[ 0 \leq \varphi_n(k), \varphi_n(k+1) \leq 1 \]

Sub-Optimal Algorithm

- Problem: Finding the combination of events locations that will yield the best \( \hat{Y} = A \cdot \Phi \) has very high complexity.
- Instead:
  - Initial uniform distribution of the events.
  - For each event, the best location between his two adjacent locations is found.
  

First BEST location found
Second BEST location found
Sub-Optimal Algorithm

[Diagram showing event locations and best locations]
TD on speech parameters

- Excitation Parameters: Pitch & Gain
- Much more singular than LSF
- Do not cover the same scale
  - Pitch is measured in [dB]
  - Gain is measured in percentages [%]

![Graph showing TD on speech parameters](image)

Vector Quantization

- Representing n-dimensional vector as an index of $2^k$ size given codebook ($k$ bits).

![Diagram of Vector Quantization](image)

- Using Generalized-Lloyd and LBG algorithms for creating optimal codebooks.
- Creating a codebook requires a training-set larger in few scales than the codebook's size ($>50$).
Vector Quantization - Algorithms

- LBG example on a 2-dimensional vectors \((x, y)\).
- Distortion function is euclidean distance

\[
D = 2.5
\]

Vector Quantization on TD Parameters

- Event functions quantization:
  - On each segment, only \(\phi_k\) is quantized, \(\phi_{k+1}\) is reconstructed using \(\phi_{k+1} = 1 - \phi_k\).
  - Each event function is quantized by itself - lengths are changing. There is a different codebook for each function’s length.

\[
\begin{pmatrix}
\phi_1(1) & \phi_2(1) & \phi_3(1) \\
\phi_1(2) & \phi_2(2) & \phi_3(2) \\
\phi_1(3) & \phi_2(3) & \phi_3(3) \\
\phi_1(4) & \phi_2(4) & \phi_3(4) \\
\phi_1(5) & \phi_2(5) & \phi_3(5)
\end{pmatrix}
\]

- In order to decrease number of codebooks, event functions lengths are limited to \(27\).
Vector Quantization on TD Parameters

- Target Vectors Quantization:
  - Target vectors refinement: After Event functions are quantized, the target vectors are calculated again.
  - ON LSF: 10-dim. vectors - Split VQ: $a(k)$ and $a'(k)$ quantized separately.

Transmission Vector

- Includes all the information for reconstructing $\hat{y}$.
- On LSF parameters:
  - On Pitch & Gain:
    - Bits allocation (LSF example):
      - Block size is 20 frames. In each block 6 events.
      - Transmission vector (every 0.45 seconds) consist of:
        - 15 bits for events location
        - 90 bits for target vectors
        - 25 bits for events function
      - In Total: 130 bits per block, 289 bits per second
Results - LSF

- Different bits allocation on quantization - LSD

\[ N \] - Number of frames in a block

\[ K \] - Number of events in a block

Results - LSF

- Different bits allocation on quantization – PESQ (Melp-2400 : 3.142)
Results – Pitch & Gain

- Different bits allocation on quantization – PESQ (Melp-2400 : 3.142)
- Separate analysis for males & females

Results – Quality Examples

- This is an example of two files – encoded with 545 bps (Total – 1167 bps)

<table>
<thead>
<tr>
<th></th>
<th>ORIGINAL</th>
<th>MELP</th>
<th>TD</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALES</td>
<td>4</td>
<td>4</td>
<td>2.574</td>
</tr>
<tr>
<td>FEMALES</td>
<td>4</td>
<td>4</td>
<td>1.732</td>
</tr>
</tbody>
</table>
Summary & Conclusions

- The aggressive compression of the event functions is the main factor in the bit rates reduction.
- The excitation parameters analysis is not satisfying. The gap between male & female voice should be explored.
- This project’s objective was to simulate the TD algorithm and applying the VQ on it, further enhancements may be applied, for example – Real Time implementation.