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

1

Abstract

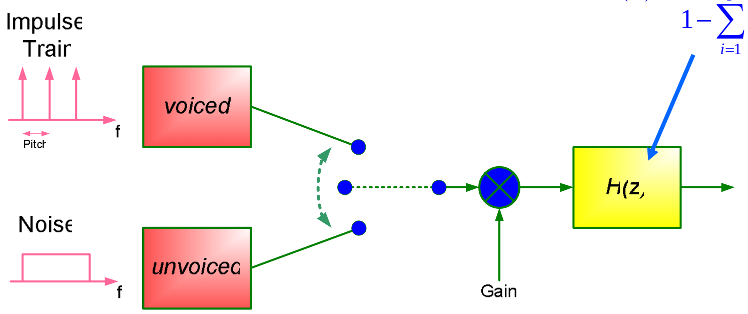
- The project deals with signal processing coding.
- Temporal Decomposition and VQ is used for compressing speech parameters.
- The project is a MATLAB simulation, which interfaces with MELP-2400 vocoder.
- The motivation : Decreasing Band-Width for transmitting, save space on disk etc…
- The Instruments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

2

LPC – Linear Predictive Coding

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

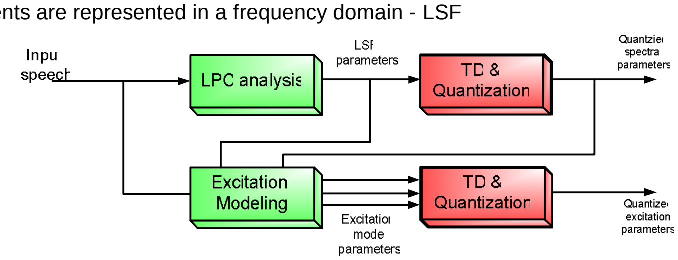
$$H(z) = \frac{1}{1 - \sum_{i=1}^{I} a_i z^{-i}}$$



3

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

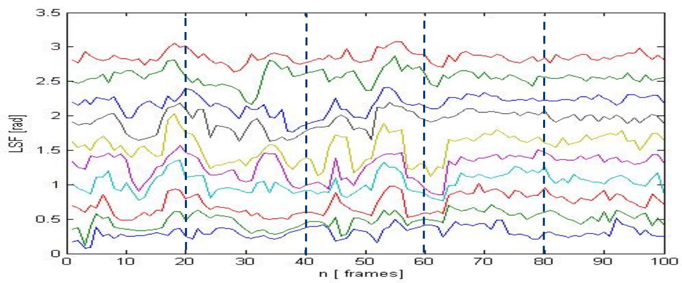


- Speech Parameters data frame is transmitted every 22.5 ms
- Each data frame is 54 bit size :
  - 25 bits for LSF parameters
  - 7 bit for pitch
  - 8 bits for gain
  - 14 bits for band-pass voicing decision, Fourier magnitudes etc..

4

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 - $\underline{a(k)}\,,\ 1 \le k \le K$
  - Event function - $\phi_n(k),\ 1 \le n \le N$ (target's weight on every frame)
  - Event location - $n_k,\ 1 \le n_k \le N$

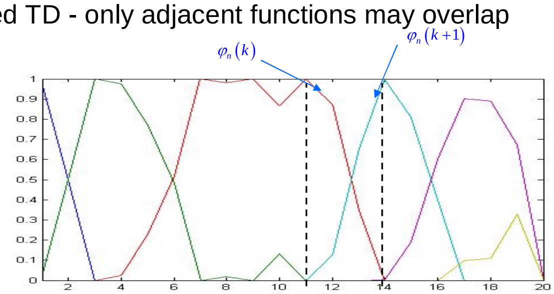
5

Temporal Decomposition

$$\underline{\hat{y}(n)} = \sum_{k=1}^{K} \underline{a(k)} \cdot \phi_n(k) \Rightarrow \underline{\underline{\hat{Y}}}_{I\times N} = \underline{\underline{A}}_{I\times K} \cdot \underline{\underline{\Phi}}_{K\times N}$$ ← *Iterative solution*

$$\begin{pmatrix} \vdots & \vdots & & \vdots \\ \hat{y}(1) & \hat{y}(2) & \cdots & \hat{y}(N) \\ \vdots & \vdots & & \vdots \end{pmatrix} = \begin{pmatrix} \vdots & \vdots & & \vdots \\ a(1) & a(2) & \cdots & a(K) \\ \vdots & \vdots & & \vdots \end{pmatrix} \bullet \begin{pmatrix} \cdots & \varphi(1) & \cdots \\ \cdots & \varphi(2) & \cdots \\ & \vdots & \\ \cdots & \varphi(K) & \cdots \end{pmatrix}$$

- Restricted TD - only adjacent functions may overlap



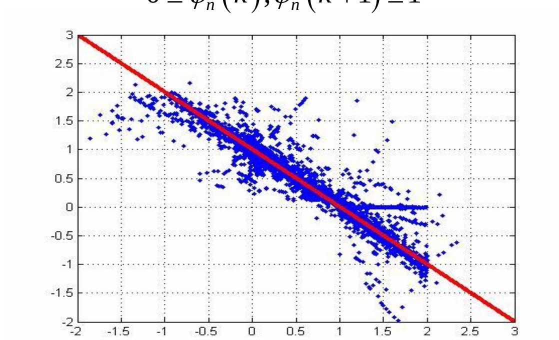
This reduces the computation complexity

6

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 :

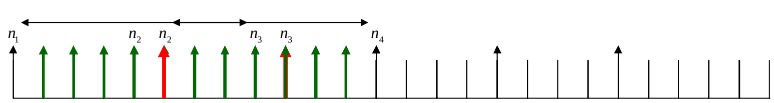
$$0\le\varphi_n(k),\varphi_n(k+1)\le 1$$



Sub-Optimal Algorithm

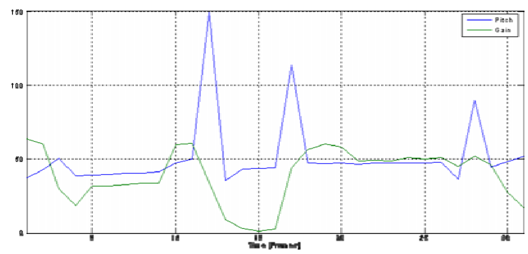
- Problem : Finding the combination of events locations that will yield the best $\underline{\underline{\hat{Y}}} = \underline{\underline{A}} \bullet \underline{\underline{\Phi}}$ has very high complexity.
- Instead :
  - Initial uniform distribution of the events.
  - For each event, the best location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

First BEST location found / Second BEST location found



TD on speech parameters

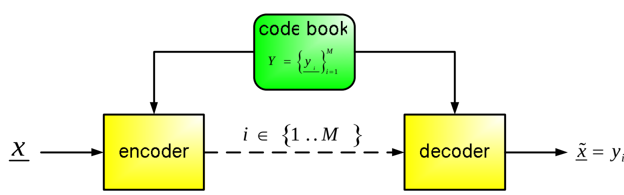
- Excitation Parameters : Pitch & Gain
- Much more singular than LSF
- Do not cover the same scale
  - Pitch is measured in [dB]
  - Gain is measured in percentages [%]



9

Vector Quantization

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

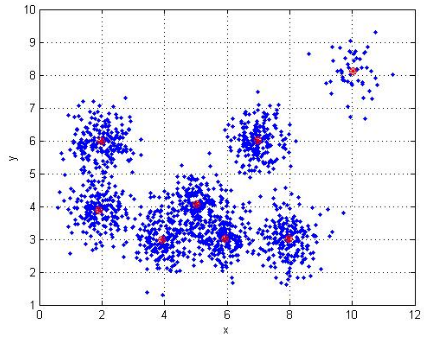


- Using Generalized-Lloyd and LBG algorithms for creating optimal codebooks.
- Creating a codebook requires a training-set larger in few scales than the codebook's size (>50).

10

Vector Quantization - Algorithms

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



$D = 0.89$

11

Vector Quantization on TD Parameters

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

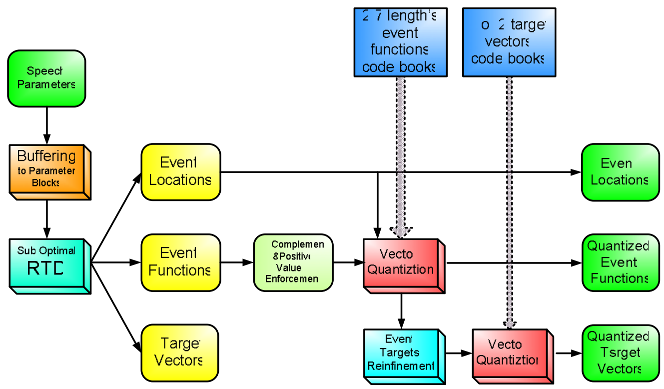
$$\begin{pmatrix} \varphi_1(1) & \varphi_2(1) & \varphi_3(1) & . & . & . & . & . & . & . & . \\ \varphi_1(2) & \varphi_2(2) & \varphi_4(2) & \varphi_4(2) & \varphi_5(2) & \varphi_6(2) & . & . & . & . & . \\ . & . & . & \varphi_4(3) & \varphi_5(3) & \varphi_6(3) & \varphi_7(3) & \varphi_8(3) & . & . & . \\ . & . & . & . & . & . & \varphi_7(4) & \varphi_8(4) & \varphi_9(4) & \varphi_{10}(4) & \varphi_{11}(4) \\ . & . & . & . & . & . & . & . & \varphi_9(5) & \varphi_{10}(5) & \varphi_{11}(5) \end{pmatrix}$$

  - In order to decrease number of codebooks, event functions lengths are limited to $2 \div 7$.

12

Vector Quantization on TD Parameters

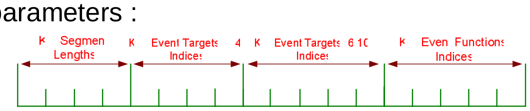
- Target Vectors Quantization :
  - Target vectors refinement : After Event functions are quantized, the target vectors are calculated again.
  - ON LSF : 10-dim. vectors - Split VQ : $\underline{a(k)}_{1:4}$ and $\underline{a(k)}_{5:10}$ quantized separately.



13

Transmission Vector

- Includes all the information for reconstructing $\underline{\hat{Y}}$.
- On LSF parameters :



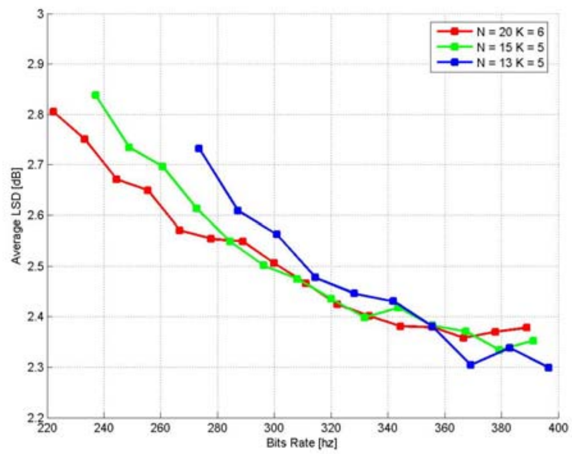
- On Pitch & Gain :

- 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

14

Results - LSF

- Different bits allocation on quantization - LSD



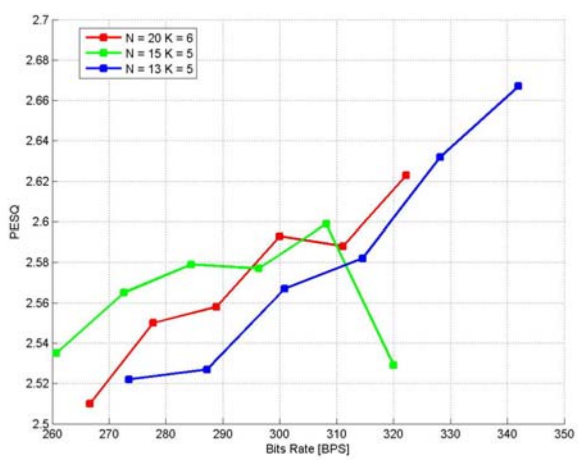
N - Number of frames in a block

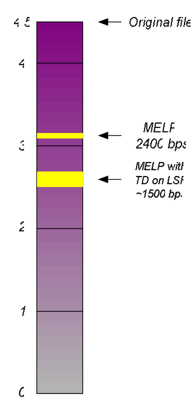
K - Number of events in a block

15

Results - LSF

- Different bits allocation on quantization – PESQ (Melp-2400 : 3.142)

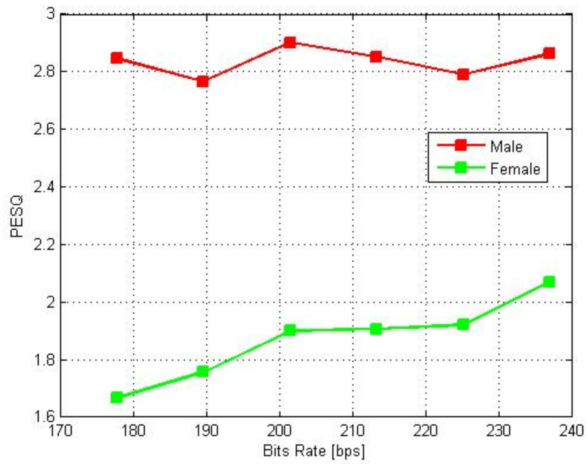


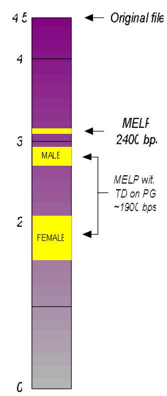


16

Results – Pitch & Gain

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





17

Results – Quality Examples

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

	ORIGINAL	MELP	TD
MALES			2.574
FEMALES			1.732

18

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 .

19