Technion-
Israel Institute of Technology
Department of Eelctrical
Engineering

Signal and Image Processing lab

Low Bit-Rate Speech Coding Using Joint Segmentation and Vector Quantization

Ronen Mayrench

Supervisor: Prof. David Malah

Supported by: TADIRAN COM.

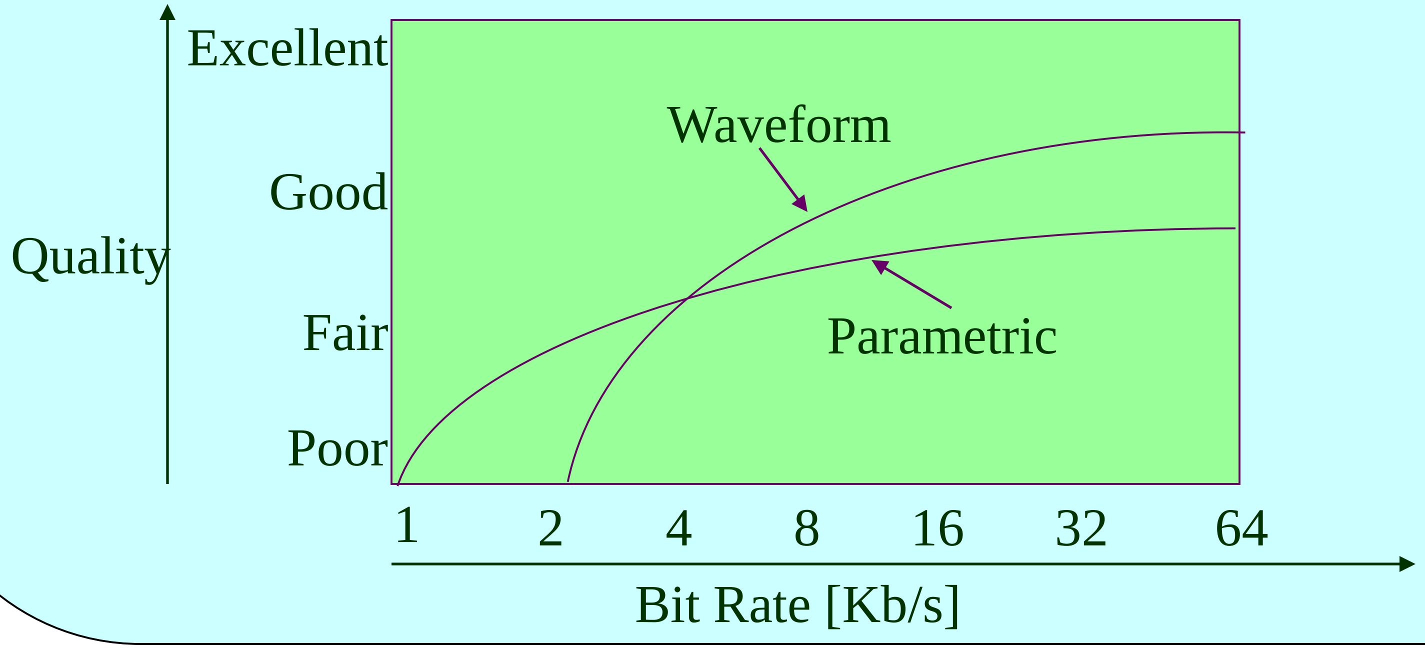
July 1999

Motivation

- PCM bit rate is 64kb/s.
- Low bit-rate speech coding is essential for:
  - Cellular and satellite links.
  - Voice mail.
  - Secure communication.
- Lossy speech coding is accepted in many applications, using:
  - Speech redundancy.
  - Limitation of human perception.

Classification of Speech Coding Procedures

- **Waveform** - reconstructed signal approximates original speech signal.
- **Parametric** -reconstructed signal sounds like speech (<u>does not</u> converge towards the original signal with decreasing quantization error).



Objective

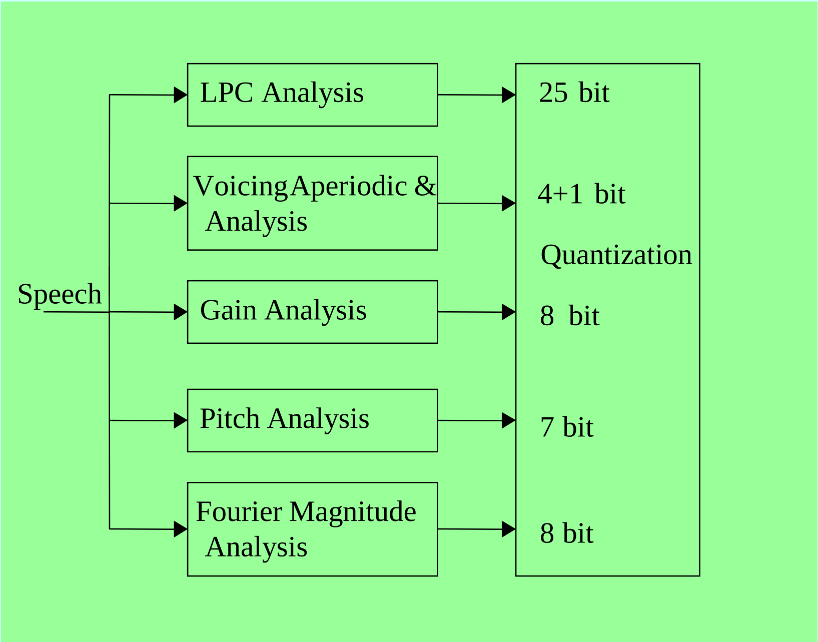
- Design a low bit-rate (1200bps) speech coder, based on the Mixed Excitation Linear Prediction (MELP) coder which is the new 2400bps DoD standard.

Outline

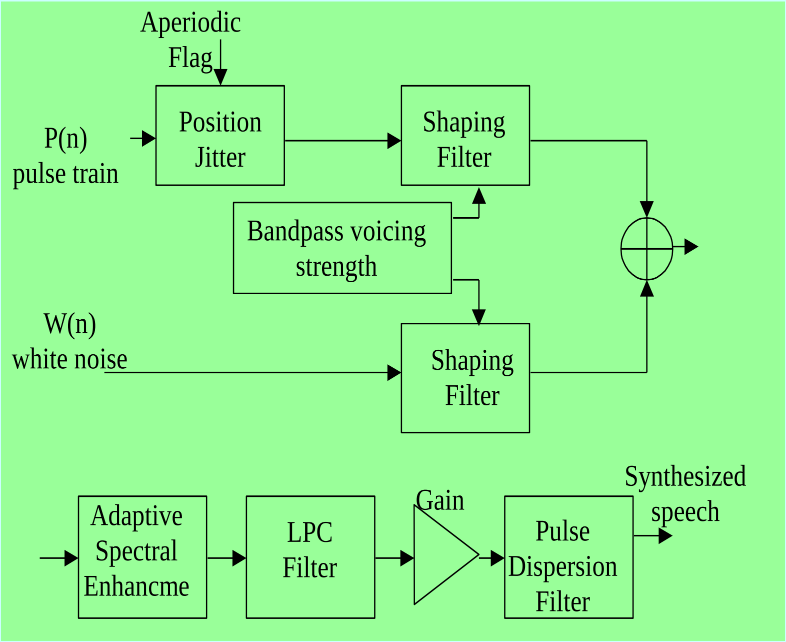
- Introduction
  - MELP analysis and synthesis
  - Low bit-rate schemes
- Trellis Segmentation-Quantization (TSQ)
  - Variable length segment quantization
  - Trellis segmentation
  - LSF quantization distortion function
  - Codebook design
- Channel error control
- Low bit-rate scheme based on TSQ
- Simulation results

Mixed Excitation Linear Prediction-MELP

- Improvement of the traditional LPC speech coder.
- The coder contains four additional features:
  - Mixed pulse and noise excitation.
  - Periodic or aperiodic pulses.
  - Adaptive spectral enhancement.
  - Pulse dispersion filter.



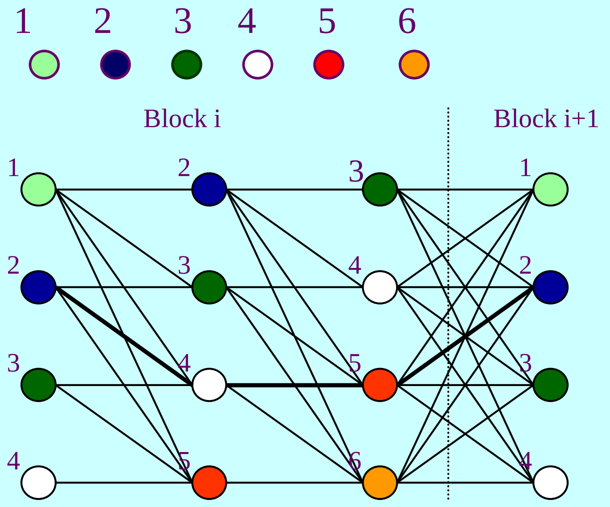
Encoder



Decoder

Low Bit-Rate Approaches

- Alternate frame transmission- AF [schwartz & roucos, 1983]
- Matrix quantization- MQ [tsao & gray, 1985]
- Adaptive frame selection-TQ [george, 1996]
  - Objective- select M frames out of a block of N frames such that the total block distortion is minimized.

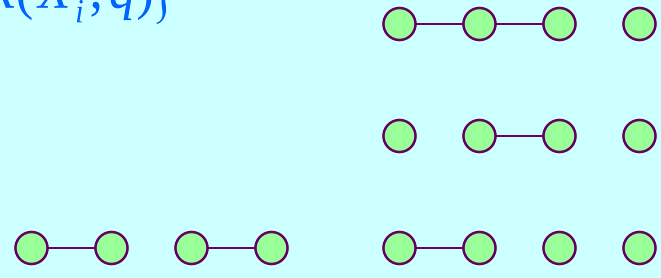


Low Bit-rate Approaches (Cont'd)

- Variable-length segment quantization [shiraki & honda, 1983]
  - Representing a block of N LSF vectors with M disjoint quantized segments
  - The algorithm determines simultaneously both the sequence of quantized segments and their lengths
- Optimal time-segmentation and resource allocation for signal modeling and compression [vetterli, 1997]

$$L(\lambda) = \min_{s \in S} \sum_{i=1}^{N_s} \min_{q \in C_s} \left\{ D(X_i, q) + \lambda R(X_i, q) \right\}$$

  - Example for a segment set (N=4):

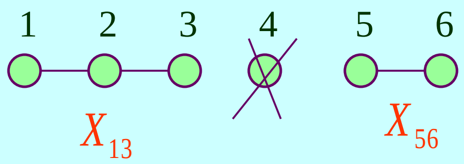


Proposed Scheme
Trellis Segmentation-Quantization (TSQ)

- Extension of the frame selection approach (TQ).
  - Modeling the input speech as a sequence of variable-length segments.
- The algorithm has a richer partition set.
  - The segmentation allow frames skipping.
- Interpolate frames in skipped segments.
- MQ and TQ are specific cases of TSQ.
- Fixed bit rate.

TSQ (cont'd)

- TSQ Objective: select a fixed number of M *segments* from a block of N (N>M) LSF vectors with minimum quantization error.
- Missing frames are linearly interpolated.
- Example for N=6 and M=2:



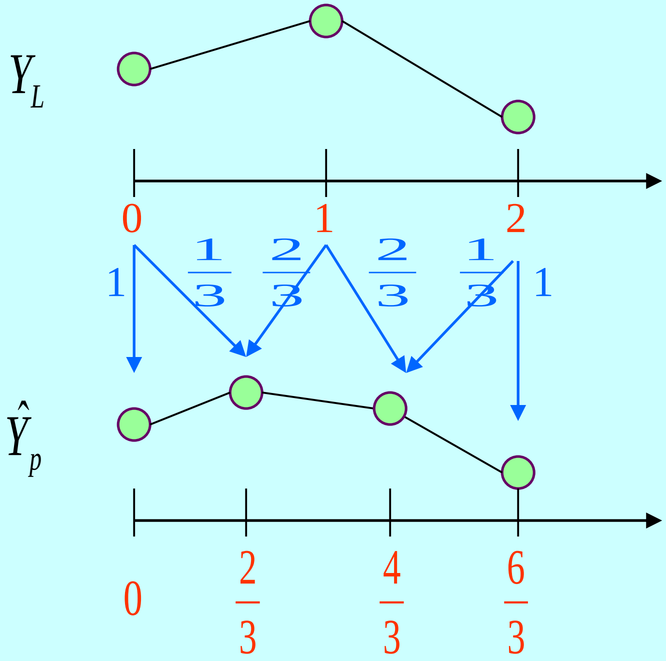
- The quantization process represents variable-length segments X_{ij} with a fixed-length codebook.
- The process of choosing segments is done using a trellis diagram.

TSQ: Variable Length Segment Quantization

- Let $C = \{Y_L\}_{k=1}^{K}$ denote the codebook where L is the *codeword* length and K is the *codebook* length

- The code vectors are linearly interpolated and re-sampled at *p*=*j*-*i*+1 equi-spaced points to receive a stacked row LSF vector.

- Example for L=3, p=4 and using the following ratio: $\beta = \frac{L-1}{p-1} = \frac{2}{3}$



- The transformation is specified only by the original segment length.

TSQ: Variable Length Segment Quantization (cont'd)

- The transformation can be performed by matrix multiplication: $\hat{Y}_p = Y_L H_p$, [Shiraki, 1988]

$$H_p = \begin{bmatrix} h_{11}[I] & h_{12}[I] & \cdots & h_{1p}[I] \\ h_{21}[I] & h_{22}[I] & \cdots & h_{2p}[I] \\ \vdots & \vdots & \ddots & \vdots \\ h_{L1}[I] & h_{L2}[I] & \cdots & h_{Lp}[I] \end{bmatrix}$$

$$H_p : \nu L \times \nu p$$

$$Y_L : 1 \times \nu L$$

$$\hat{Y}_p : 1 \times \nu p$$

$$h_{ij} = \begin{cases} 1-\alpha_j, & i = \lfloor \beta_j \rfloor + 1 \\ \alpha_j, & i = \lfloor \beta_j \rfloor + 2 \\ 0, & else \end{cases}$$

$$j = 1, 2, \ldots, p$$

$$\alpha_j = \beta_j - \lfloor \beta_j \rfloor$$

$$\beta_j = \frac{(j-1)(L-1)}{(p-1)}$$

TSQ: LSF Distortion Function

- The distortion function is important for codebook design.
- Log Spectral Distance (LSD) is highly correlated with human perception, but is complicated for practical design.

$$d_{LSD}(A,\hat{A}) = \sqrt{\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(10\log_{10}\left|\frac{1}{A(\omega)}\right|^2 - 10\log_{10}\left|\frac{1}{\hat{A}(\omega)}\right|^2\right)^2 d\omega}$$

- Usually, WMSE is used in practical designs.

$$d_{WMSE}(a,\hat{a}) = (a-\hat{a})W_a(a-\hat{a})^T$$

TSQ: LSF Distortion Function (cont'd)

<u>Atal & Paliwal's Weighting</u> [1993]

- W is a diagonal matrix with elements proportional to the synthesis filter spectrum.

$$w_i = [P(f_i)]^r, \quad P(f) = \frac{1}{\left|A(e^{j2\pi f/F_s})\right|^2}$$

$$r = 0.15$$

<u>Gardner's Weighting</u> [1994]

- Approximate LSD using WMSE.

$$d(a,\hat{a}) \cong \frac{1}{2}(a-\hat{a})W(a-\hat{a})^T, \qquad W = \left.\frac{\partial^2 d_{LSD}(a,\overline{a})}{\partial \hat{a}_k \partial \hat{a}_l}\right|_{a=\hat{a}} = 4\beta R_A(k-l)$$

$$R_A(k) = \sum_{n=0}^{\infty} h(n)h(n+k), \qquad h(n) = F^{-1}\left\{\frac{1}{A(z)}\right\}$$

β=constant

TSQ: Segment Distortion

- Segment distortion is defined as accumulated wighted distortion between original and quantized LSF vectors.

$$d(X_{ij}, Y_L) = \sum_{k=1}^{p} d_1(X_k, \tilde{Y}_k)$$

$$\tilde{Y}_k = (Y_L H_p)_k$$

$$d_1(X_k, \tilde{Y}_k) = (X_k - \tilde{Y}_k) W_{X_k} (X_k - \tilde{Y}_k)^T$$

$$d(X_{ij}, Y_L) = (X_{ij} - Y_L H_p) W_X (X_{ij} - Y_L H_p)^T$$

- W_X is a diagonal matrix with W_{X_k} on its main diagonal.

TSQ: Codebook Design

- Define the following maping:

$$Q:\Re^{\nu} \rightarrow C = \{Y_L(k)\}_{k=1}^{K}$$

- The Codebook design problem is: Given segmented data, find the code vectors that minimize:

$$D = E\{d(X_{ij}, Y_L) \mid X_{ij} \in S\}$$

- Necessary conditions are *Nearest Neighbor* and *Centroid* conditions.

TSQ: Centroid Derivation

- For the k'th cluster, the centroid is obtained by minimizing D_k.

$$D_k = \sum_{X_{ij} \in R_k} d(X_{ij}, Y_L(k)) =$$

$$\sum_{X_{ij} \in R_k} (X_{ij} - Y_L(k) H_p) W_X (X_{ij} - Y_L(k) H_p)^T$$

- By differentiating D_k with respect to $Y_L(k)$ and equating to zero, we obtain:

$$Y_L^T(k) = \left[\sum_{X_{ij} \in R_k} H_p W_X H_p^T \right]^{-1} \sum_{X_{ij} \in R_k} H_p W_X X_{ij}^T$$

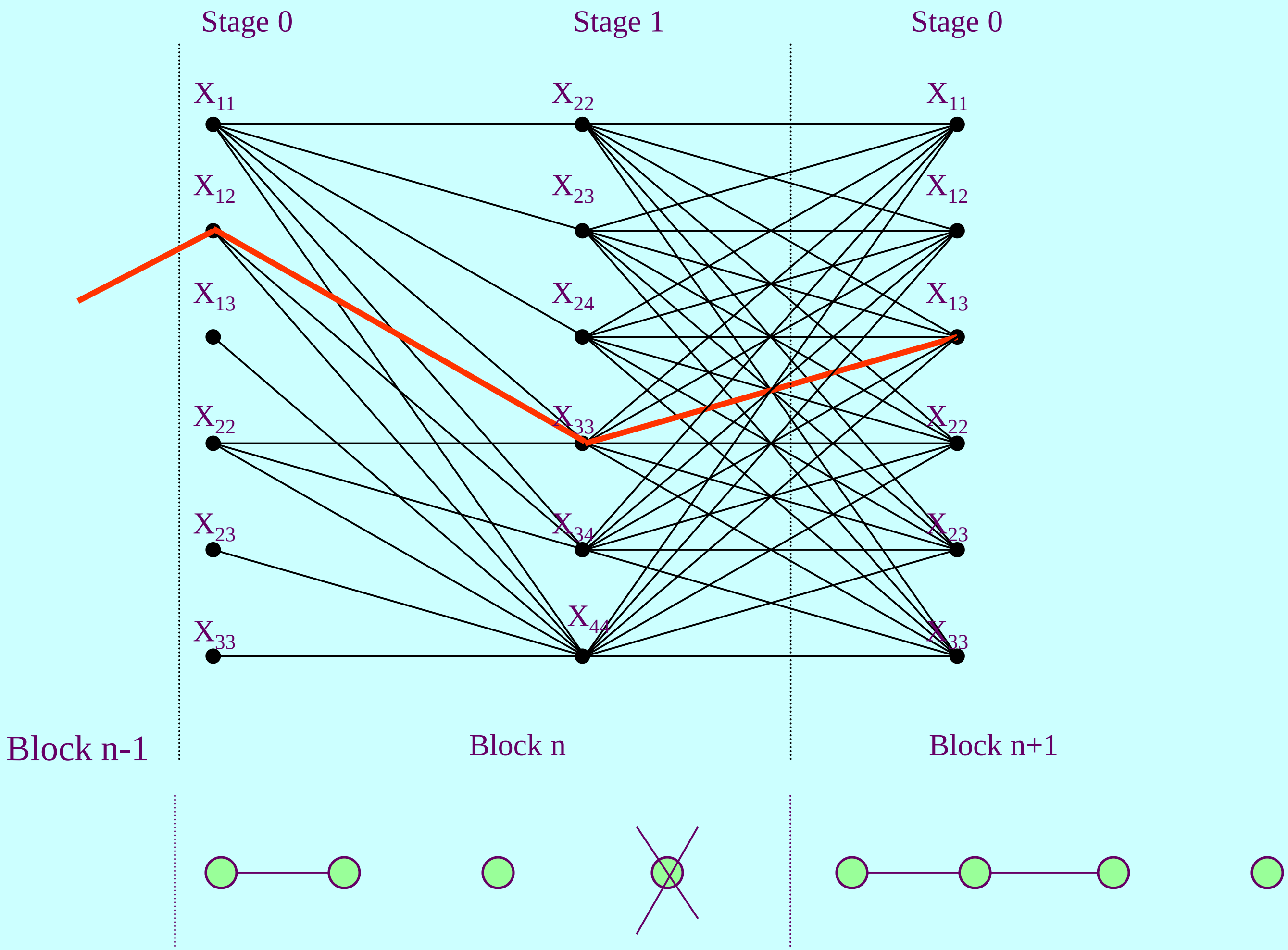
TSQ: Trellis Segmentation

- The process of choosing M segments from a block of N frames is done using a trellis diagram.

- The transmitted segments should be selected to minimize the total path cost with 2 constraints:
  - Segments should be in ascending order.
  - Maximum segment length should be specified (no longer than N-M+1).

TSQ: Trellis Segmentation (cont'd)

- Example for M=2, N=4:



TSQ: Trellis Segmentation (cont'd)

- The path cost function is defined by:

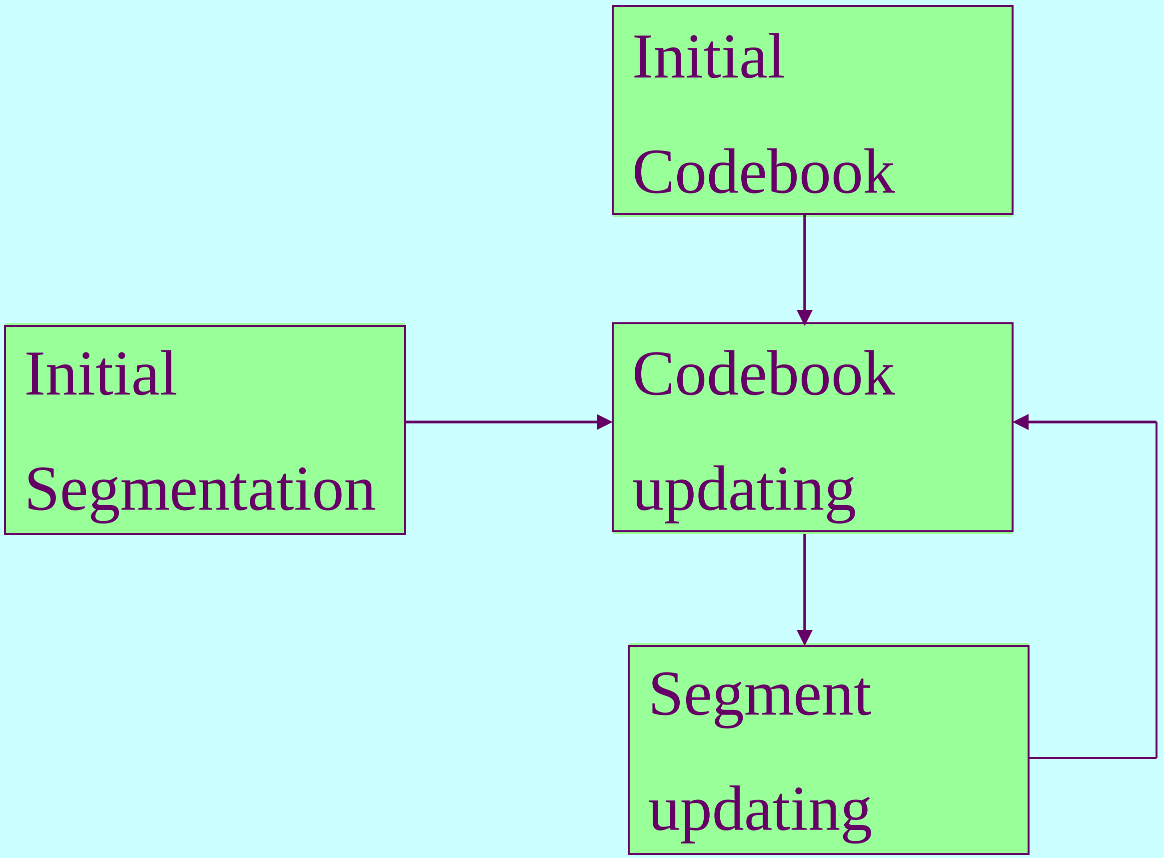
$$D_{path}(k) = \frac{\sum_{l=k_0}^{k} g_l d(X_l, \hat{X}_l)}{\sum_{l=k_0}^{k} g_l}$$

g_l - gain-dependent weight

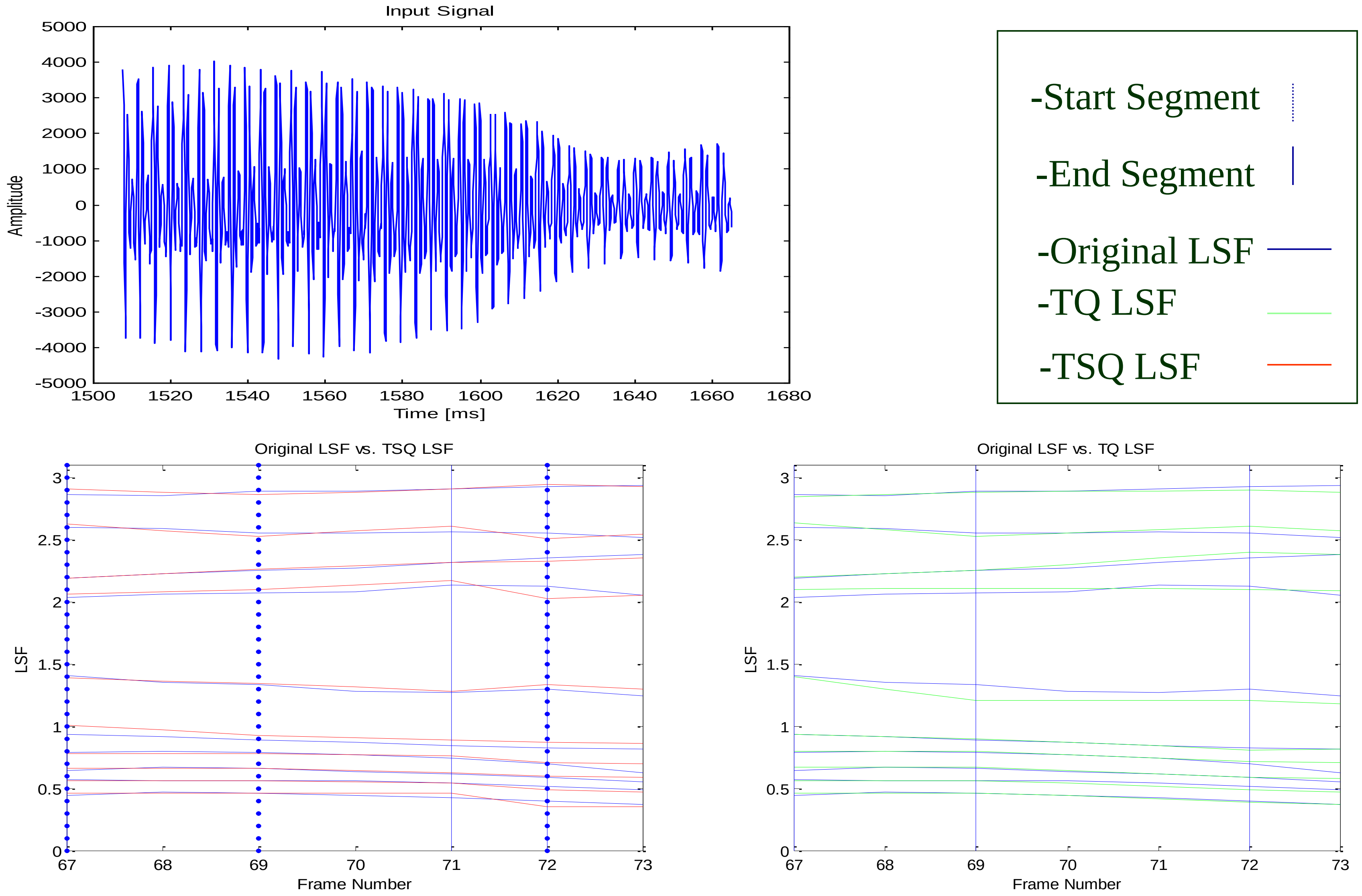
- High gain segments are more likely to be choosen than low gain segments.
- The purpose of the trellis search is to minimize path cost, $D_{path}(k)$.
- The search for the best segmentation-interpolation path is done efficiently with DP using the Viterbi algorithm.

Codebook Design Algorithm For TSQ

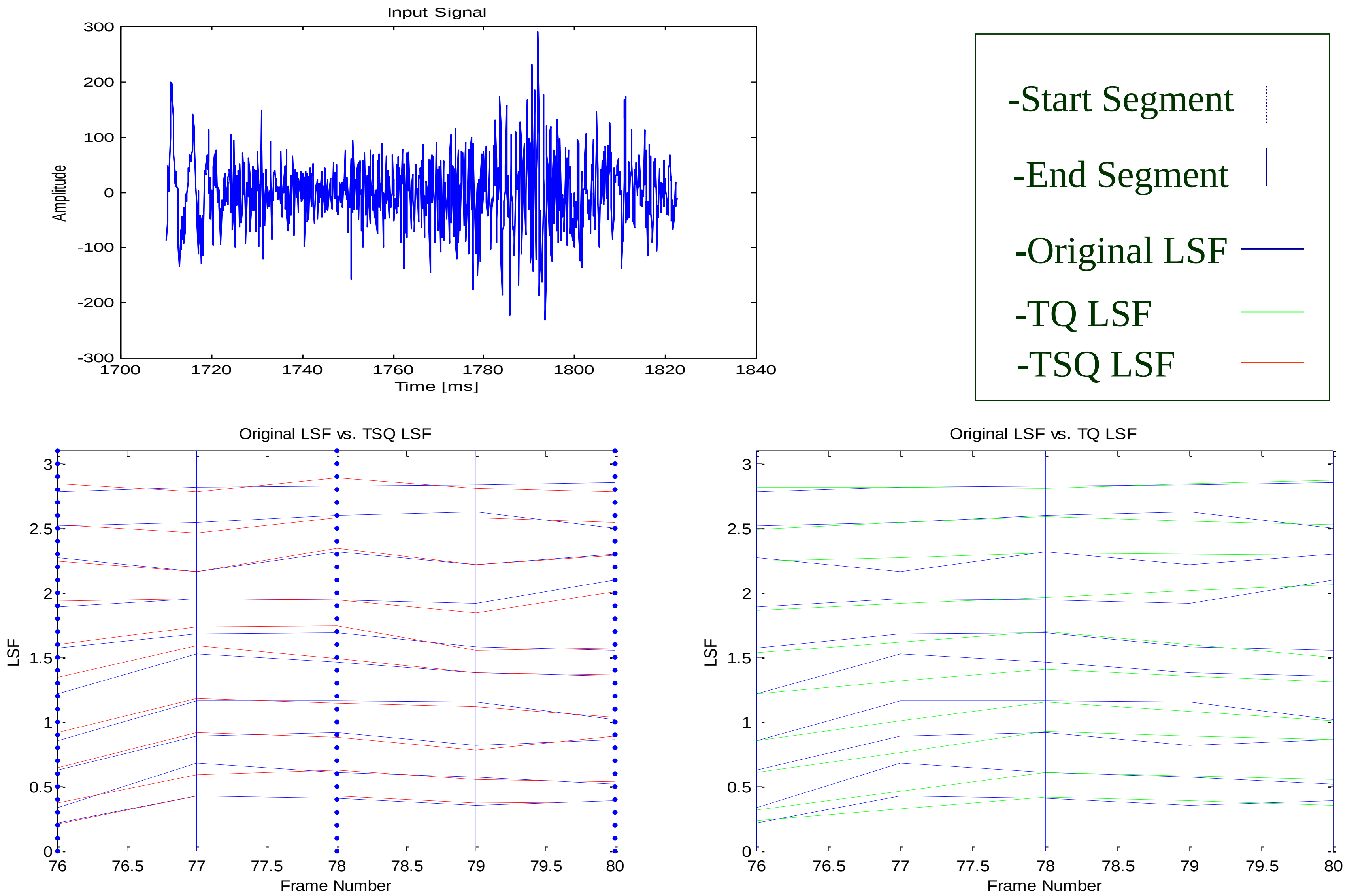
- Iterative approach for TSQ codebook design:



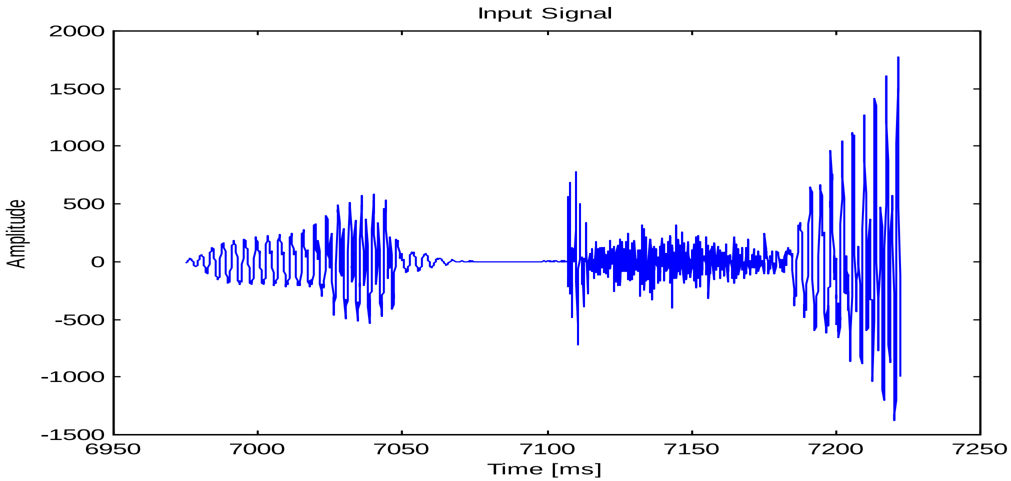
TSQ - Simulation Results

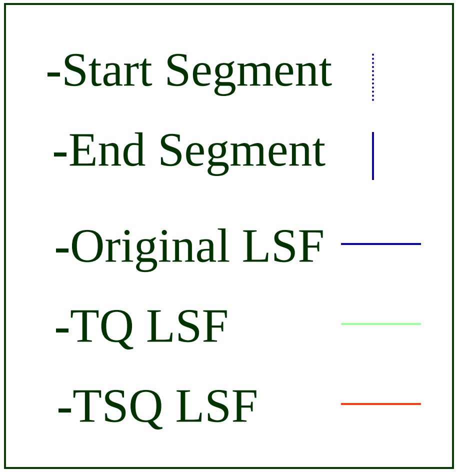


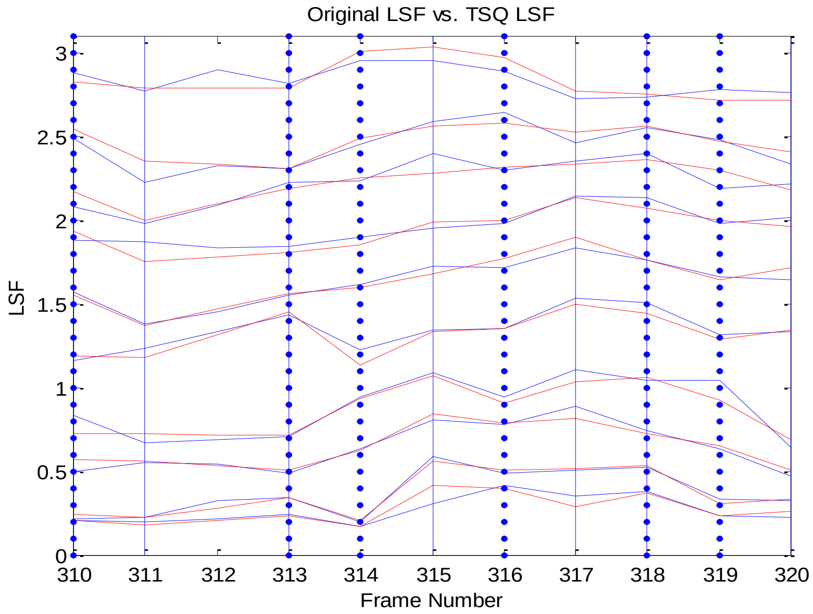
TSQ - Simulation Results (cont'd)

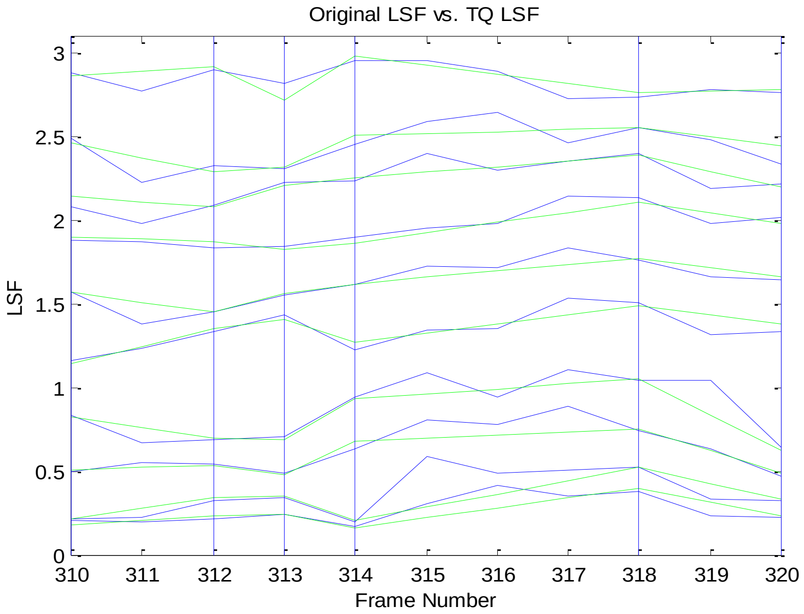


TSQ - Simulation Results (cont'd)









TSQ: Simulation Results (cont'd)

- Log Spectral Distortion (LSD [dB]) - Atal's vs. Gardner's weighting (Split-VQ):

	Atal	Gardner
20 bit	2.17	1.75
22 bit	1.98	1.67

- Log Spectral Distortion (LSD) - TSQ vs. TQ, MQ, AF with Gardner's wighting (Split-VQ):

Split VQ (22bit)	AF (11 bit)	MQ (11 bit)	TQ (11 bit)	**TSQ** (11 bit)
1.67	2.43	2.41	2.23	2.01

Channel Error Control

- Channel codes
- Channel Optimized VQ (COVQ) [Favardin, 1990]
- Codebook scaling [Ben-David & Malah, 1995]
- Index Assignment (IA)
  - Full search complexity is of the order of N!
  - Pseudo-Gray algorithm [Zeger & Gersho, 1990]
  - Codebook design with LBG splitting method
- Frame-erasure concealment

Index Assigment

- Channel distortion increases the average error:

$$D = E\{d(X, Y_j)\} \geq E\{d(X, Y_i)\}$$

- Upper bound on D: $D \leq E\{d(X, Y_i)\} + E\{d(Y_i, Y_j)\} = D_Q + D_c$

- Channel distortion for VQ (fixed word-length):

$$D_c = \sum_{i=1}^{N} \sum_{j=1}^{N} p_{ij} d(Y_i, Y_j) q_i$$

$q_i = \Pr(Y_i)$

$p_{ij} = \Pr(Y_j | Y_i)$

- Channel distortion for VQ (variable word-length):

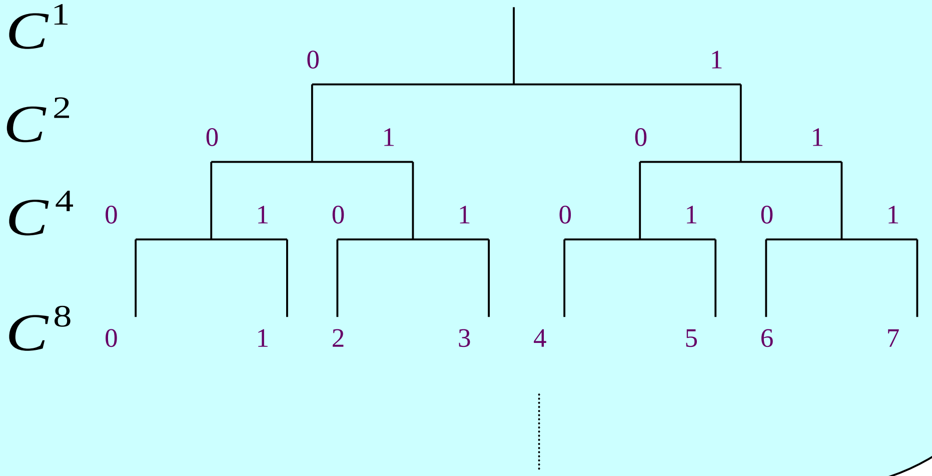
$$D_c = \sum_{l=1}^{L} \sum_{i=1}^{N} \sum_{j=1}^{N} p_{ij} \mu_l d(Y_i^{(l)}, Y_j^{(l)}) q_i$$

μ_l = Probability of segment length

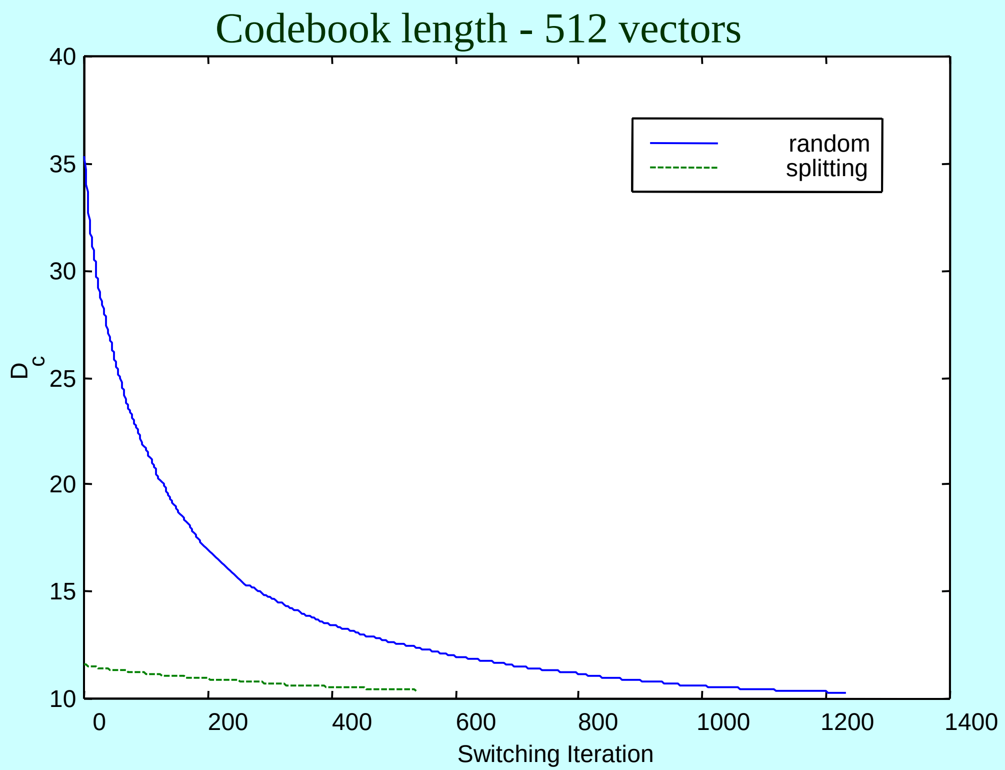
(Assuming independence between segment lengths and codeword indices)

Index Assigment (cont'd)

- Index Switching Algorithm (assuming BSC):
  - Calculate distortion matrix for the codebook
  - Assign cost for each codeword
  - Sort all costs with decreasing order
  - Find an index switch that maximize the distortion reduction (searching from highest cost codeword to lowest cost codeword)
- Splitting algorithm:

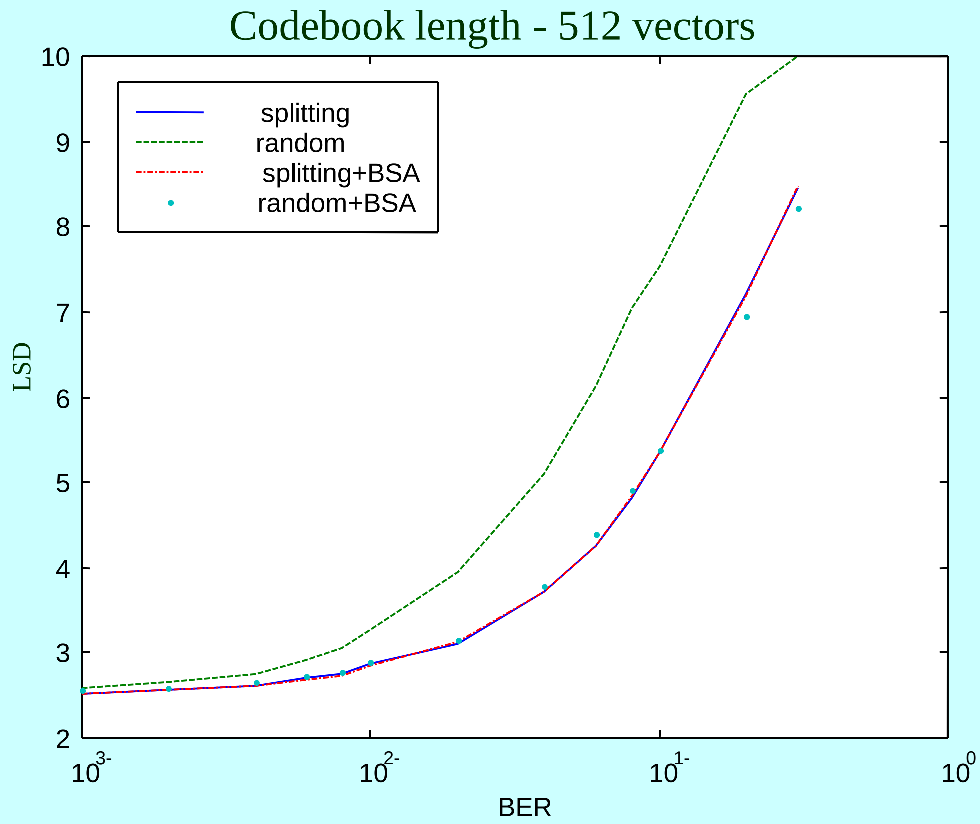


Index Assigment: Simulation Results



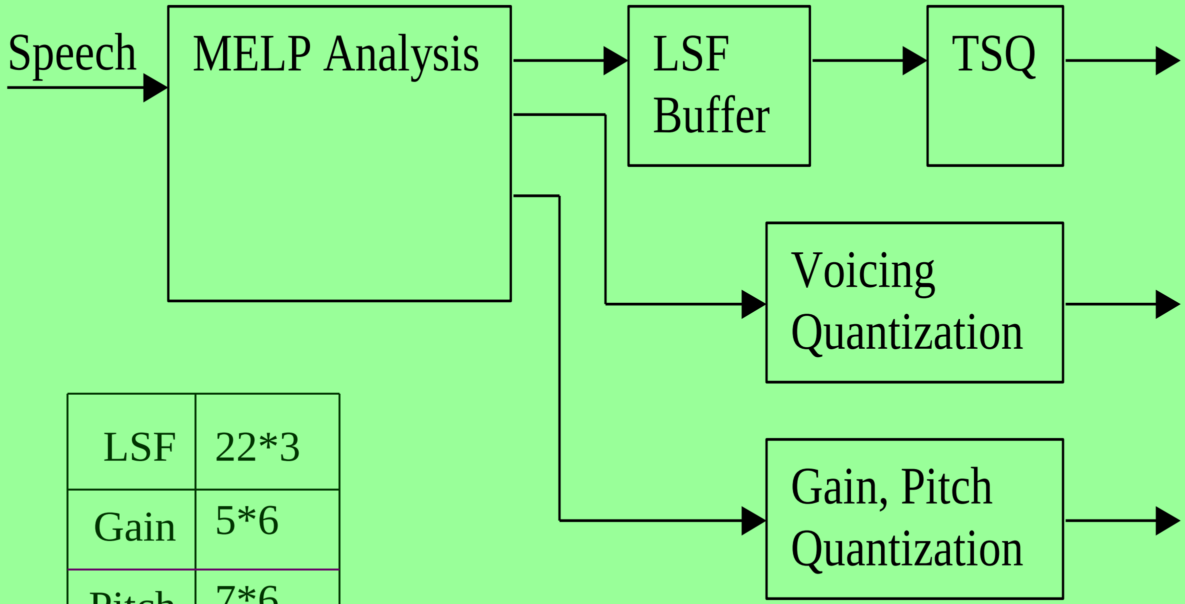
Typical behavior of the channel distortion (D_c) in the Index Switching Algorithm.

Index Assigment: Simulation Results (cont'd)



It has been observed that the splitting method achieves almost all of the IA gain as compared to the Index Switching Algorithm.

1200bps Speech Coder



LSF	22*3
Gain	5*6
Pitch	7*6
V/UV	2*6
Path	9

Codebook design using splitting algorithm

159 bits per 6 frames (26.5 bits per frame)
N=6 frames, M=3 segments, frame rate=44.44 f/s,
Total Rate:1178bps (unused bits can synchronize frames)

Coding Examples

	Male	Female
• Original		
• LPC10 (2400 bps)		
• MELP (2400 bps)		
• MELP-AF (1200 bps)		
• MELP-TQ (1200 bps)		
• MELP-TSQ (1200 bps)		

Summary

- Trellis Segmentation-Quantization
  - TSQ is an extension of the AF, TQ and MQ schemes.
  - Better results (lower LSD) than AF, TQ and MQ.
  - Gardner's weighting function is better (lower LSD in VQ design) than Atal's weighting function.
- Index Assigment
  - Good results using the splitting scheme compared to the index switching algorithm and random IA.
  - The splitting scheme enables the use of Unequal Error Protection (UEP) to better protect sensitive bits.

Summary (cont'd)

- The TSQ algorithm enables halving the bit-rate of the MELP coder with only a slight degradation.

Suggestions For Further Research

- TSQ with a non-linear interpolation schemes.
- Using the LSD in the index switching algorithm (Needed to calculate the LSD from partial LSF vector representations).
- What kind of IA strategy is needed for transmission over channels other than BSC?
- Is it possible to gain by combining IA and channel codes (joint IA and channel codes)?
- Frame-erasure and parameter-error concealment.