

CECS401

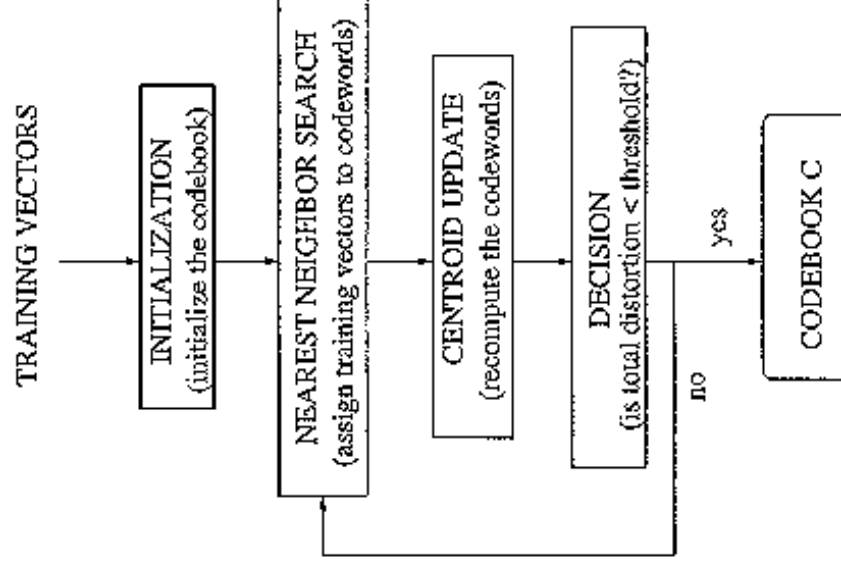
Fundamentals of Spoken Language Processing

Note-8

Thursday 9/23/99

D. Vector Quantization

Procedure of generating a codebook



Step-1. Initialization

Initialize codewords by choosing M vectors (randomly or informatively) from the set of L training vectors.

Example

$$y_1^{(0)} = v_1$$

$$y_2^{(0)} = v_{\lfloor L/M \rfloor}$$

$$y_2^{(0)} = v_{\lfloor 2L/M \rfloor}$$

⋮

$$y_M^{(0)} = v_L$$

$$C^{(0)} = \{ y_1^{(0)}, y_2^{(0)}, \dots, y_M^{(0)} \}$$

Assign $D^{(0)}$ a large number, and set the iteration count $n = 1$.

Step-2. Nearest neighbor search

Quantize each vector to its nearest codeword in the current codebook, i.e.,

$$\hat{v}_t^{(n)} = \arg \min_{y_i^{(n-1)} \in C^{(n-1)}} d(v_t, y_i^{(n-1)})$$
$$t = 1, 2, \dots, L$$

Assign v_t to the partition set $\Omega_i^{(n)}$ if $\hat{v}_t^{(n)} = y_i^{(n-1)}$.

Step-3. Centroid update

Update the codewords for each partition set $\Omega_i^{(n)}$, $i = 1, 2, \dots, M$:

$$y_i^{(n)} = \arg \min_{y_i} \sum_{v_t \in \Omega_i^{(n)}} d(v_t, y_i)$$

The actual updating formula for $y_i^{(n)}$ depends on the chosen distance measure.

Example

Euclidean distance measure:

$$d(v_t, y_i) = \|v_t - y_i\|^2$$

$$D_i = \sum_{v_t \in \Omega_i} \|v_t - y_i\|^2$$

Setting

$$\nabla_{y_i} D_i = \sum_{v_t \in \Omega_i} (-2v_t + 2y_i) = 0$$

leads to

$$y_i = \frac{1}{|\Omega_i|} \sum_{v_t \in \Omega_i} v_t$$

$|\Omega_i|$ denotes number of training vectors assigned to the partition.

Absolute distance measure:

$$d(v_t, y_i) = |v_t - y_i|$$

$$y_i = \text{median}\{v_t \in \Omega_i\}$$

where median is taken in each signal dimension.

Log spectrum distance measure:

$$d(v_t, y_i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\log V_t(\omega) - \log Y_i(\omega)|^2 d\omega$$

$$Y_i(\omega) = \left(\prod_{v_t \in \Omega_i} V_t(\omega) \right)^{\frac{1}{|\Omega_i|}}$$

Step-4. Decision

Compute the total distortion

$$D^{(n)} = \frac{1}{L} \sum_{t=1}^L d(v_t, \hat{v}_t^{(n)})$$

If $D^{(n-1)} - D^{(n)} < \delta$ (*threshold*)

then take

$$C = C^{(n)} = \{y_1^{(n)}, y_2^{(n)}, \dots, y_M^{(n)}\}$$

Stop.

Otherwise

$$n + 1 \rightarrow n$$

go back to step 2.

- The VQ training algorithm is local optimal. The resulting codebook C depends on the choice made on the initial codebook $C^{(0)}$.
- The total distortion D_{min} in general decreases with the increasing codebook size M .
- The codebook quality depends on the coverage of training vectors and the appropriateness of distance measure.

Binary split VQ algorithm

Design M-vector codebook in multiple stages:

- first design a 1-vector codebook
- then using a splitting technique on the codewords to initialize the search for a 2-vector codebook, e.g.,

$$y_i^+ = y_i(1 + \epsilon)$$

$$y_i^- = y_i(1 - \epsilon)$$

$$0.01 \leq \epsilon \leq 0.05$$

- continue the splitting process until the desired M-vector codebook is obtained.

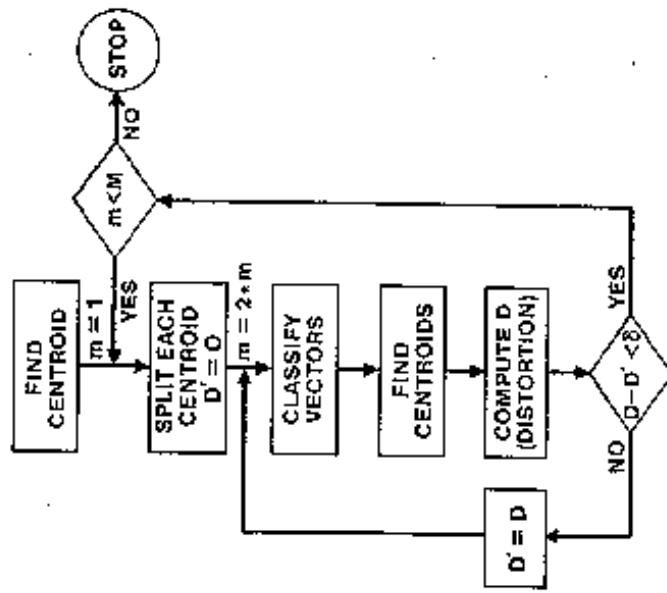


Figure 3.42 Flow diagram of binary split codebook generation algorithm.

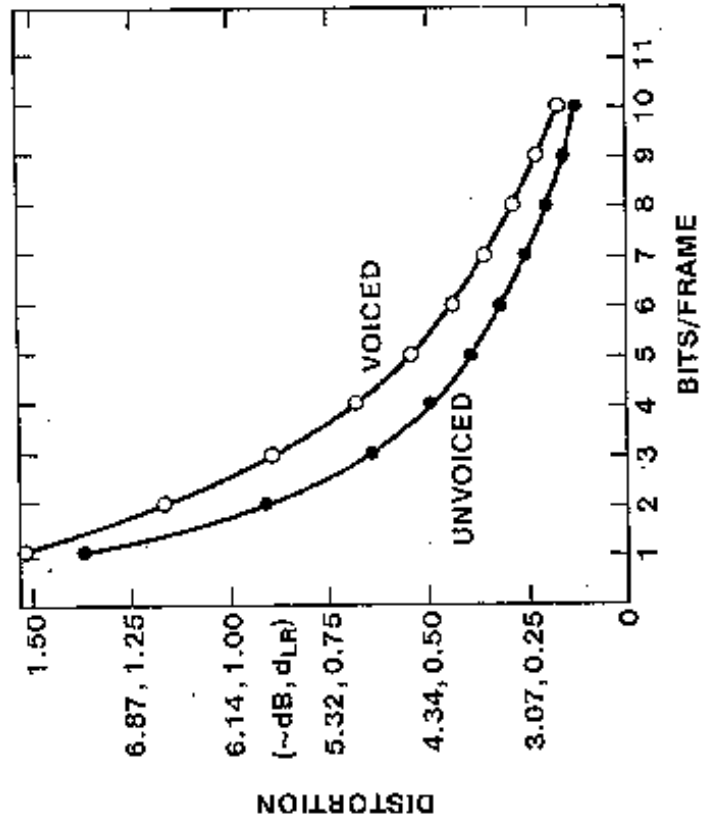


Figure 3.43 Codebook distortion versus codebook size (measured in bits per frame) for both voiced and unvoiced speech (after Juang et al. [12]).

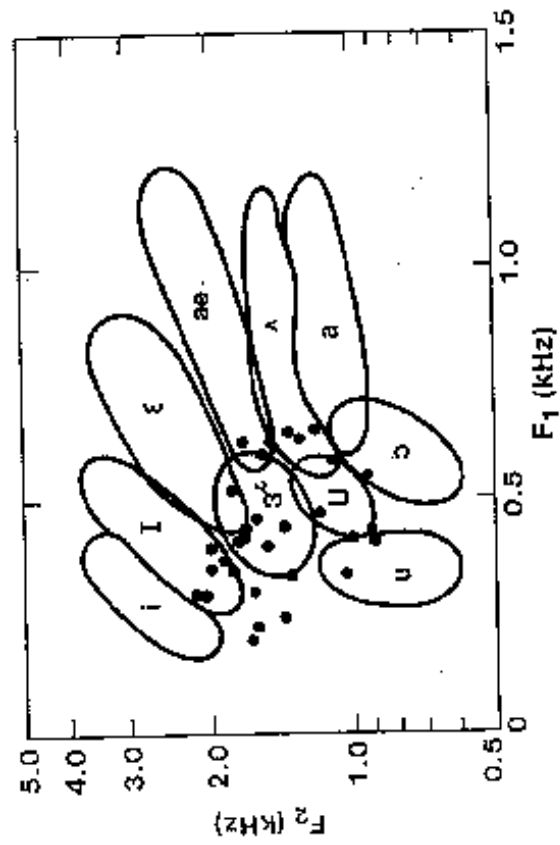


Figure 3.44 Codebook vector locations in the $F_1 - F_2$ plane (for a 32-vector codebook) superimposed on the vowel ellipses (after Juang et al. [12]).

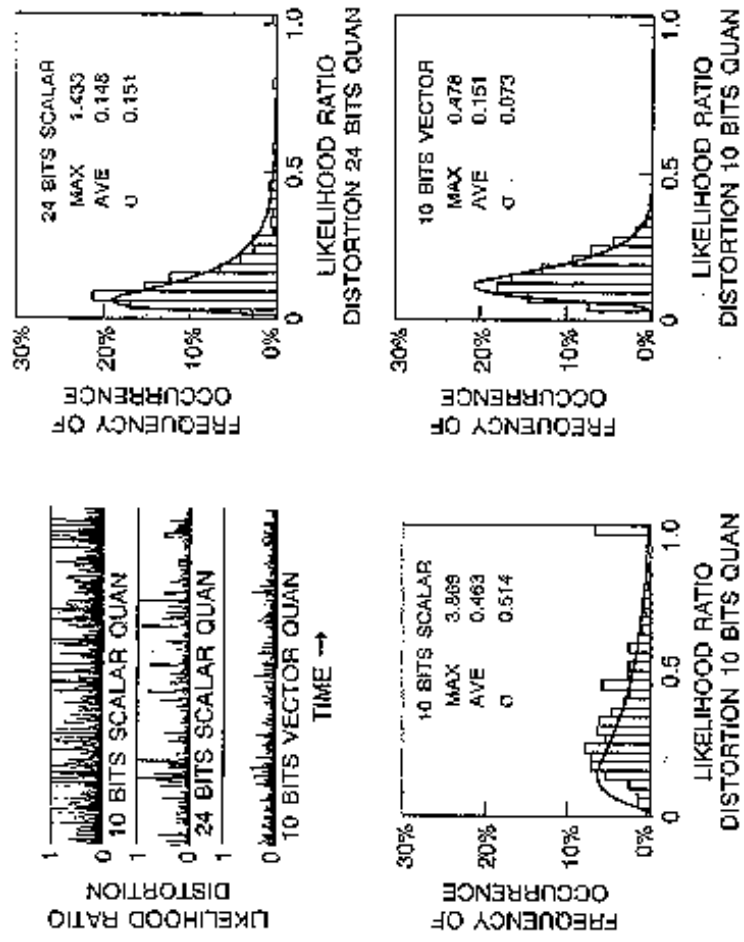


Figure 3.46 Plots and histograms of temporal distortion for scalar and vector quantizers (after Juang et al. [12]).

D. Application of LPC in Speech Coding

LPC is currently the most widely used technique for low bit rate speech coding.

Basic LPC coding system

Voiced/unvoiced	1 bit
pitch period (log)	6 bits
gain (log)	5 bits
reflection parameters	6 bits per coefficient
total bits per frame (10th order LPC)	72 bits

At frame rate of 33~66 frames/sec (frame shift of 30~15 ms), the resulting bit rate is in the range of 2,400~4,800 bps.

LPC-10 is such a codec produced by Texas Instrument in the 70s for the educational toy of “speaker-and-spell,” bit rate is 2,400 bps.

Multipulse LPC vocoder (Atal & Remede, 1982)

An analysis-by-synthesis method that produces multipulse excitation signals for LPC vocal filters.

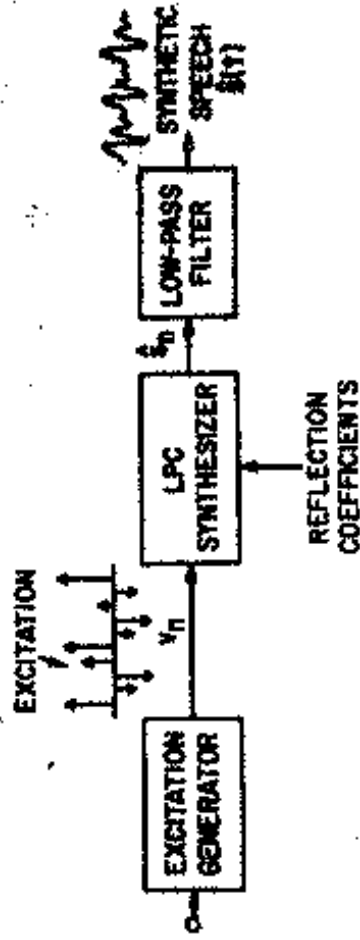


Fig. 3. Block diagram of an LPC speech synthesizer with multipulse excitation.

Motivation:



Fig. 2. An example of a short segment of the speech waveform.

- Certain regions of speech can be easily recognized as voiced and unvoiced regions, but there are regions where it is not clear whether the signal is voiced or unvoiced.
- The excitation for voiced speech should consist of several pulses in a pitch period rather than just one at the beginning of the period.

Multipulse excitation model:

- The excitation generator produces a sequence of pulses at times k_1, k_2, \dots, k_{n_e} with amplitudes a_1, a_2, \dots, a_{n_e} as input to LPC vocal filter (do not need voiced/unvoiced switch).
- The output of LPC vocal filter, after low pass filtering, becomes synthetic speech.