PERFORMANCE COMPARISON OF THE SPLIT VECTOR QUANTIZATION OF SPEECH LSF PARAMETERS USING VARIABLE LENGTH CODEBOOK

E V KRISHNA RAO [1] Dr.P G KRISHNA MOHAN [2] Dr. P V SUBBAIAH [3]

[1] ASST. PROFESSOR OF ECE,
[3] PROFESSOR OF ECE,
V.R.SIDDHARTHA ENGINEERING COLLEGE,
VIJAYAWADA-520007, A.P., INDIA

[2] PROFESSOR OF ECE,
JNTU COLLEGE OF ENGINEERING,
KUKATPALLY, HYDERABAD, A.P., INDIA

ABSTRACT

This paper describes a Performance Comparison of Split Vector Quantization (SVQ) of Line Spectral Frequency (LSF) Parameters for Speech Coding using variable length codebook. This method is much better than the existing algorithms like LBG (LINDE, BUZO, GRAY) algorithm. The Linear Predictive Coding (LPC) parameters consisting of 10 dimensional LSF's are splitting into sub-vectors and each sub-vector is quantized separately using vector quantization. In the LSF quantization, performance of the Split Vector Quantizer is studied for different splittings such as (3,7), (4,6), (5,5) and (3,3,4). The variable length optimum codebook is prepared by using rawcodebook and hitbook (Raw-Hit) procedure. By removing the unvoiced frames, the Split LSF vectors are quantized using optimum codebook and then the index is transmitted to the receiver. The index of the removed unvoiced frames is maintained and the same is introduced at synthesis to reconstruct the original signal. It is shown that the SVQ can quantize LSF vector in 21 bits with an average Spectral Distortion (SD) is less than 0.5 dB in two-part splitting and it is slightly higher in three-part splitting. It can be seen that average SD is very much less in the proposed SVQ algorithm and best performance for the split (4,6).

Key words: Linear Predictive Coding, Line Spectral Frequencies, Split Vector Quantization, Speech Coding, Spectral Distortion.

1. Introduction

Linear Prediction (LP) is among the most widely used methods of Speech Processing. Especially in low bit rate speech coding applications, the LPC coefficients are very important. In order to model the envelope of speech spectrum accurately enough with LP, the prediction order is typically adjusted to equal the frequency in kHz added by a small integer. A p^{th} order LPC analysis results in an all-pole filter with p poles whose transfer function is denoted by [1],[2].

$$H(z) = 1/A(z) \qquad (1)$$

Where Linear Predictor (LP)

$$A(z) = 1 + a_1 z^{-1} + \ldots + a_p z^{-p} \qquad (2)$$

and $[a_1, a_2, \ldots a_p]$ are the LPC coefficients. The LSF parameters are represented as roots of the LPC polynomial denoted by

$$l = [l_1, l_2, \ldots l_p]^T \qquad (3)$$

The ordering property of the LSF parameters states that parameters are ordered and bounded within a range. i.e. $0 < l_1 < l_2 < \ldots < l_p < 0.5$, that the reconstructed LPC filter will be stable.

Line Spectral Pair (LSP) decomposition defines two polynomials of order p+1, the symmetric polynomial U(z) and anti-symmetric polynomial V(z), as follows [3],[4].

$$U(z) = A(z) + z^{-p-1} A(z^{-1})$$
$$V(z) = A(z) - z^{-p-1} A(z^{-1}) \qquad (4)$$

The LSP decomposition has the following properties

1. The zeros of U(z) and V(z) are always on the unit circle
2. When p is even, U(z) has a trivial root located at z = -1 and V(z) has a trivial root located at z=1. When p is odd, U(z) has trivial roots at z=1 and z= -1, while V(z) has no trivial roots.
3. When A(z) is minimum-phase, zeros of U(z) and V(z) are interlaced. This property is called the intramodel interlacing theorem
4. The roots of U(z) computed from an LP predictor of order p are interlaced with the roots of U(z) computed from an LP-predictor of order p-1. Similarly, roots of V(z) computed from a pth order LP-predictor interlace with those defined from an LP-predictor of order p-1. This property is called the intermodel interlacing theorem.

Predictor A(z) can be obtained from the LSP polynomials as $A(z) = \frac{1}{2} [U(z) + V(z)]$.

For low bit rate speech coding applications, it is important to quantize these parameters using as few bits as possible. Efficient LSF quantization using Vector Quantization is achieved by encoding the LSF Parameters in a vector form. [5]-[9].

Based on *new* Raw-Hit VQ algorithm developed by EVKR for the design of variable length optimum codebook [10],[11] which is more efficient and less complex, using Split Vector Quantization (SVQ) the performance of the splitting (3,7), (4,6), (5,5), (3,3,4) is observed in this paper. In two-part splitting, two optimum codebooks are prepared by splitting each 10^{th} order LSF vector into two parts using above optimum codebook design procedure [12],[13]. In three-part splitting, three optimum codebooks are prepared by splitting each 10^{th} order LSF vector into three parts [14]. The length of the codebook is not fixed. The LSF vector is splitting into sub-vectors and each sub-vector is quantized to the nearest vector of the corresponding optimum codebook. The unvoiced frames are removed before quantization and added at the receiver to get original signal. It is known that split vector quantizer reduces the complexity at the cost of degraded performance. Varying the number of bits per frame the average Spectral Distortion (SD) is calculated in speech coding without considering unvoiced frames to study the performance of SVQ.

2. Design Of Optimum Codebook

In the design of Optimum Codebook , first it has to make rawcodebook from all possible LSF vectors and then hitbook is prepared based on the number of times the LSF vector is accessed from the rawcodebook. The hitsum is calculated summing all the hits generated from different number of speech files. Rearranging the hitsum in descending order codebook is prepared depending on the length of the codebook required [11].

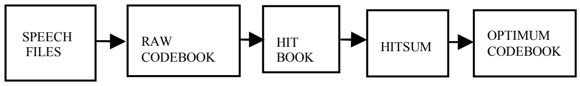

Figure 1. Optimum codebook

2.1 Design Of Rawcodebook

For each sub-vector a rawcodebook is to be made. The purpose of this codebook is to satisfy the error criterion. To save the memory in processing of long speech file, it is considered N small files from different speakers, which is expected to contain all possible LSF parameters as vectors. For a particular file the LSF parameters are calculated assuming Q number of vectors represented by x_a, a= 1,2,3…. Q

Each LSF vector split into two parts in two-part splitting, each part is considered as vector for the preparation of codebook. The first LSF vector x_1 is taken as the 1^{st} codebook element in r_i from one file. Then the second LSF vector x_2 is taken and compared with the first x_1. If it is at a distance greater than a minimum specified distance (S) from all existing codebook elements, it is considered in the rawcodebook r_i, else it is discarded. This procedure is repeated for each incoming LSF vector.
The rawcodebook sub-elements of each file having dimension q and k_i number of elements and there are N rawcodebooks corresponding to N speech files as

$$r_i = \{ r_{n,1}, r_{n,2}, \ldots r_{n,q} \} \quad n = 1,2, \ldots k_i \quad i= 1,2, \ldots N \qquad (5)$$

The nearest neighbour condition is

$$\|x_a - r_i\|^2 \geq \in \qquad (6)$$

where error selected $\in = 0.0002$.
Now there exists sub-elements in one rawcodebook corresponding to one file of length k_1. This procedure is repeated for all N files, preparing N rawcodebooks r_i, i = 1,2,..N corresponding to the lengths $k_1, k_2, \ldots k_N$. Appending all these N rawcodebook elements gives one long rawcodebook R_N called appended rawcodebook.

$$R_N = \{r_1 \, r_2 \ldots\ldots r_N\} \qquad (7)$$

Using the above procedure, the length of the rawcodebook will eventually saturate, say to length

$$K_N = k_1 + k_2 + \ldots + k_N. \qquad (8)$$

This rawcodebook is not an optimal codebook as there are entries rarely accessed. These entries are called less probability cells. This leads to wastage of memory space and increased processing time. Similar rawcodebook is prepared for second part of sub-vectors

2.2 Calculating The Accessibility Of Each Of The Rawcodebook Vector

A 'hit' is said to have occurred if a match is found between the input block and an entry of the rawcodebook. A number of input sequences are to be given and hits are calculated for each element of the rawcodebook. To prepare hitsums h_i, i=1,2,…N, corresponding to the rawcodebook r_i of length k_i, hitbook program is run for every rawcodebook r_i , i=1,2,..N with M number of speech files where

$$h_i = \sum_{j=1}^{M} h_{ij} \qquad i=1,2, \ldots N \qquad (9)$$

and $h_{i1}, h_{i2}, \ldots h_{iM}$ are the hitbooks of M speech files for i^{th} rawcodebook r_i . Appending all these hitsums to give one appended hitbook

$$H_N=\{h_1\ h_2 \ldots h_N\} \qquad (10)$$

of length K_N .

The length of the appended rawcodebook and appended hitbook is equal

2.3 Elimination Of Redundant Vectors From Appended Rawcodebook

As the rawcodebooks and hitbooks are prepared for individual files and appended to get appended rawcodebook and appended hitbook, there is a possibility of occurrence of redundant vectors in the appended rawcodebook. Again the nearest neighbor condition is checked to remove the redundant vectors. Removal of the redundant vectors from the appended rawcodebook (R_N) results in Modified Rawcodebook (MR_N). Here the procedure for obtaining Modified Rawcodebook, rawcodebook procedure is repeated with the appended rawcodebook vectors as training vectors.

If a vector is found to have a nearest neighbour in the Modified Rawcodebook it has to be discarded from the appended rawcodebook. Its hits are added to the hits of the nearest neighbour and discarded from the appended hitbook resulting in the Modified Hitbook (MH_N).

2.4 Optimizing The Modified Rawcodebook

In order to obtain the optimum codebook, the modified rawcodebook elements are to be arranged in the descending order of hits.
The *advantages* of this *new SVQ* algorithm over the LBG algorithm are given below.

1. In the LBG algorithm the length of the codebook is fixed. In this *new* algorithm the number of hits of all the entries of the codebook is checked and those entries with less number of hits are neglected. The length of the optimum codebook can be varied depends upon requirement
2. In LBG algorithm the codebook should be trained several times until mean square error falls below a certain predetermined value which requires lot of time. In the present algorithm training of the codebook is not required. Preparation of each rawcodebook is done from individual files and appended which requires *less memory*.
3. In SVQ, the input vector is accessed at *higher speed* because the length of the codebook is less compared with direct LSF Vector codebook.

3. Spectral Distortion

The Spectral Distortion is employed to measure the objective quality of the distortion introduced in the power spectral density of speech in each particular frame. The Spectral Distortion in the *n*th frame is given by [15]-[17]

$$SD_n = [(1/F_s) \int [10 \log_{10}(P_n(f) / \hat{P}_n(f))]\, df]^{1/2} \quad (11)$$

Where

$$P_n(f) = 1/|A_n(\exp(j2\pi f/F_s))|^2 \qquad (12)$$

and

$$\hat{P}_n(f) = 1/| \hat{A}_n(\exp(j2\pi f/F_s))^2 \qquad (13)$$

are the original and quantized power spectral densities of the *n*th frame, respectively. The terms A_n (z) and $\hat{A}_n(z)$ are the corresponding original and quantized LPC filters and F_s is the sampling rate of the signal. It is considered that transparent quality is achieved when the average Spectral Distortion is about 1 dB, and the fraction of 2 dB outliers is less than 2%. Here the SD is much lower than 1 dB.

4. Results

The Speech Signal is sampled at 8 KHz and 8 bits per sample, 180 samples per frame are taken with an overlapping of 60 samples. A database of about 1,06,208 frames of speech is used to quantize the LSF parameters as well as residual parameters. The test data of about 4068 frames are taken. Generation of Optimum codebook is very simple and less complex. The length of the Optimum codebook can be varied depending upon the number of bits/frame (power of 2). After finding the LSF Parameters, the LSF Parameters and residual parameters are quantized using proposed SVQ. Average Spectral Distortion (SD) is used as performance measure. The unvoiced frames are removed before quantization and added at the receiver to get original signal. The compression ratio is very much increased by removing these unvoiced frames. The performance of the proposed SVQ is best for the split (4,6).

Table 1 show the average SD performance of the 24 bits/frame SVQ using different splittings of the LSF vector into (3,7), (4,6), (5,5) and (3,3,4). Table 2 show the results for the average SD without considering unvoiced frames in the proposed algorithm compared with LBG algorithm at different bit rates for the split (4,6). In table 3 show the results for the average SD without considering unvoiced frames in the proposed algorithm at different bit rates for the three-part splitting (3,3,4). As number of splittings are increased the SD is slightly increased due to the length of the each codebook is reduced. Fig. 2 shows Bits/frame Vs. average SD for LBG algorithm [17] and our proposed SVQ algorithm.It is observed that the reconstructed signal is very much recognizable with least Spectral Distortion. The Fig.3 shows the original and reconstructed signal quantized at 21 bits/frame with removed unvoiced frames. There is no outlier voiced frame having SD larger than 2 dB.

Table1

Spectral Distortion (SD) performance of the 24 bits/frame SVQ using different Splittings of the LSF Vector

Splitting	Average SD (dB)
(3,7)	0.4435
(4,6)	0.4309
(5,5)	0.4484
(3,3,4)	0.6341

Table 2

Bits/frame vs. Spectral Distortion (SD) by Splitting LSF Vector into (4,6)

Bits / frame	Avg.SD with LBG algorithm (dB)	Avg.SD in our proposed two-part SVQ algorithm (dB)
21	1.27	0.4631
22	1.17	0.4514
23	1.1	0.4433
24	1.03	0.4309
25	0.96	0.4238
26	0.9	0.4154

Table 3

Bits/frame vs. Spectral Distortion (SD) by Splitting LSF Vector into (3,3,4)

Bits / frame	Avg.SD after removing unvoiced frames in our proposed 3 part SVQ algorithm (dB)
21	0.6714
22	0.6579
23	0.6517
24	0.6340
25	0.6301
26	0.5712

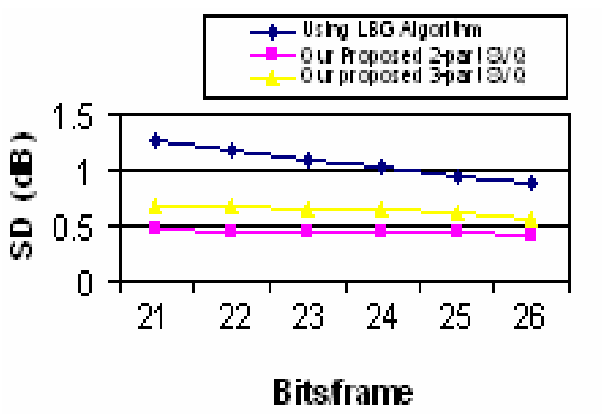

Figure 2. Bits/frame Vs. SD for the split (4,6)

5. Conclusions

In this algorithm training of the codebook is not required. The length of the codebook can be varied depending upon the number of bits required to encode the coefficients. This proposed SVQ algorithm is found to be more efficient because less average SD is obtained when compared to LBG. It is observed that average Spectral Distortion in quantizing LSF vector using two-part SVQ with 21 bits/frame is less than 0.5 dB . Complexity of this quantizer is reduced by splitting the LSF vector into three parts, but this results the degradation in performance. The limitation of this algorithm is that the highest probability LSF vectors are considered in the optimum codebook.

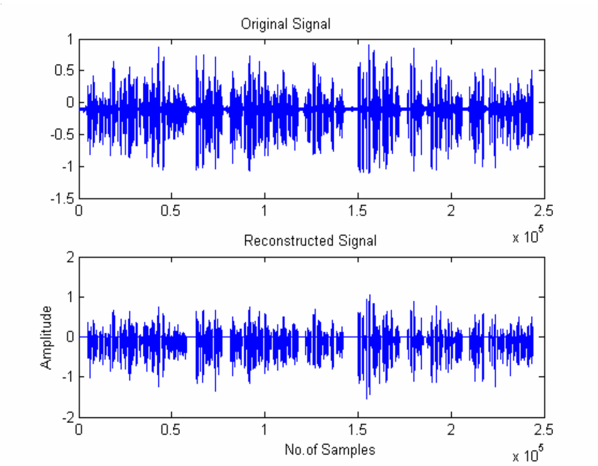

Figure 3. Original and Reconstructed signal at 21 bits/frame with removed unvoiced frames using SVQ for the split (4,6).

References

1. L.R.Rabiner and R.W.Schafer, '*Digital Processing of Speech Signals'*, Prentice-Hall, Englewood Cliffs, NJ, 1978.
2. J Makhoul, 'Linear Prediction: A Tutorial Review', *Proc. IEEE*, April 1975, pp. 561-580
3. F Itakura, 'Line Spectrum Representation of Linear Predictive Coefficients of Speech Signals', *J. Acoust. Soc. Amer.*, vol.57, 1975, pp.S35.
4. H k Kim and H S Lee, 'Interlacing Properties of Line Spectrum Pair Frequencies', *IEEE Transactions on Speech and Audio Processing,* January 1999, pp.87-91.
5. J Pan and Thomas R Fischer, 'Vector Quantization of Speech Line Spectrum Pair Parameters and Reflection Coefficients', *IEEE Transactions on Speech and Audio Processing,* March 1998, pp.106-115.
6. John Makhoul, Salim Roucos and Herbert Gish, 'Vector Quantization in Speech Coding', *Proceedings of the IEEE*, Nov. 1985, pp.1551-1588.
7. R M Gray, 'Vector Quantization', *IEEE ASSP Magazine*, April 1984.
8. Y.Linde, A.Buzo, and R.M.Gray, 'An Algorithm for Vector Quantizer Design', *IEEE Transactions on Communications, COM-28*, Jan.1980, pp. 84-95.
9. A Gersho and R.M.Gray, '*Vector Quantization and Signal Compression'*, Kluwer Academic Publishers, Norwell, MA,1991.
10. E V Krishna Rao, P G Krishna Mohan, P V Subbaiah and N V N Prathyusha, 'Efficient Vector Quantization of LSF Parameters', *International Conf. Of IASTED On Communications, Internet and Information Technology (CIIT-2004),* St.Thomas, US Virgin Islands, USA, Nov. 2004.
11. E V Krishna Rao, P G Krishna Mohan, P V Subbaiah and N V N Prathyusha, 'Efficient Vector Quantization of LSF Parameters', Accepted in *WSEAS Transactions on Systems* to be published in June 2005.
12. E V Krishna Rao, P G Krishna Mohan and P V Subbaiah, 'Efficient Split Vector Quantization of LSF Parameters', *International Conf. Of Bio-Medical Electronics and Telecommunications (BET-2004)* at Andhra University, Visakhapatnam, India, Dec. 2004, pp. 260-266.
13. E V Krishna Rao, P G Krishna Mohan and P V Subbaiah, 'Performance of Split Vector Quantization of Speech LSF Parameters', *International Conf. Of Hunan Machine Interfaces(ICHMI-2004)* at Indian Institute of Sciences (IISc), Bangalore, India, Dec. 2004, pp.341-349.
14. E V Krishna Rao, P G Krishna Mohan and P V Subbaiah, 'Performance of Split Vector Quantization of Speech Line Spectral Frequencies' *International Conf. on Systemics, Cybernetics and Informatics ICSCI 2005, Hyderabad, India*, Jan. 2005, pp. 362-365
15. H Saito, Isao Umoto, A Sasou, S Nakamura, Y Horio and T Kubota, 'Subadaptive Piecewise Linear Quantization for Speech Signal (64 kbit/s) Compression', *IEEE Transactions on Speech and Audio Processing*, Sept. 1996, pp. 379-382.
16. E V Krishna Rao, P G Krishna Mohan and P V Subbaiah, 'Multi-Channel Filter Bank for Speech Signal Compression Using LPC Parameters', *International Conf. on Systemics, Cybernetics and Informatics ICSCI 2004, Hyderabad, India*, Feb. 2004, pp.411-414.
17. K.K.Paliwal and B.S.Atal, 'Efficient Vector Quantization of LPC Parameters at 24 bits/frame', *IEEE Transactions on Speech and Audio Processing,* Jan. 1993, pp. 3-14.