A 1200 bps SPEECH CODER WITH LSF MATRIX QUANTIZATION

Selma Ozaydin

Department of Electrical & Electronics Engineering,
Gazi University, Ankara,TURKEY
E-mail:ozaydin@mikasa.mmf.gazi.edu.tr

Buyurman Baykal

Department of Electrical & Electronics Engineering,
Middle East Technical University,Ankara,TURKEY
E-mail:buyurman@metu.edu.tr

ABSTRACT

A new 1200 bps speech coder designed with a tree searched multistage matrix quantization scheme is proposed. To improve speech quality and reduce the average bit rate, we have developed a new residual multistage matrix quantization method with the joint design technique. The new joint design algorithm reduces the codebook training complexity. Other new techniques for improving the performance include joint quantization of pitch and voiced/unvoiced/mixed decisions and gain interpolation. For the new matrix quantization based speech coder (MQBC), the listening tests have proven that an efficient and high quality coding has been achieved at bit rate 1200 bps. Test results are compared with the 2400 bps LPC10e coder and the new 2400 bps MELP coder which has been chosen as the new 2400 bps Federal Standard.

1. INTRODUCTION

Speech coding at very low bit rate below 2.4 kbps is useful for purposes such as voice communication over low capacity channels. For the linear prediction based speech coders, the main difficulty of reducing bit rate is due to the quantization of the LPC filter coefficients. Performance of the vector quantizer should be robust to speaker and channel errors. Several methods have been proposed for this problem. The multistage vector quantization (MSVQ) scheme presented in [1-3] has an efficient quantization performance at 22-24 bits per 20 ms frame. Furthermore, multistage structure has more flexibility in terms of search complexity, codebook storage and channel error protection. In the proposed quantization scheme, a residual multistage matrix quantization scheme (R-MSMQ) [4,5] is developed. The residual LSF vectors are obtained using a new backward prediction method involving past speech frames and then the resulting residual LSF vectors are combined and jointly quantized using consecutive frames. The proposed joint codebook design method, residual LSF vector quantization scheme, R-MSMQ scheme, pitch and voicing determination, interpolation of these parameters, test results and the conclusions are presented in this paper.

2. LSF QUANTIZATION

The major bit rate reduction in MQBC comes from the new residual LSF matrix quantization scheme, which gives a distortion result about 1 dB at 18bits/frame, with an acceptable storage and complexity.

2.1. MSVQ

In the MSVQ system [1-3], the parameter vector x consisting of p LSF parameters is approximated as a quantized parameter vector $\hat{x}$ using the minimum distortion rule. (All vectors are assumed to be column vectors)

$$\begin{aligned}\hat{x} &= y_0^{(l_0)} + y_1^{(l_1)} + ... + y_{K-1}^{(l_{K-1})},\\ &= B_0^{(l_0)} c_0 + B_1^{(l_1)} c_1 + ... + B_{K-1}^{(l_{K-1})} c_{K-1},\\ &= B.c,\end{aligned} \tag{1}$$

where K is the number of stages, $y_j^{(k)}$ (p by 1) is the k^{th} vector for the j^{th} stage and the vector c_j ($L_j p$ by 1) is created by stacking the codevectors,

$$c_j = [\, y_j^{(0)T} \; y_j^{(1)T} \; .... \; y_j^{(L_j - 1)T} \,] \tag{2}$$

where L_j is the size of codebook for the j^{th} stage. The column vector c (Lp by 1) is referred to as the stacked codebook where,

$$c = [\, c_0^{T} \; c_1^{T} \; .... \; c_{K-1}^{T} \,] \tag{3}$$

The selection matrix for the j^{th} stage $B_j^{(k)}$ is a sparse Toeplitz matrix (p by $L_j p$) constructed such that $B_j^{(k)} c_j = y_j^{(k)}$. The multistage selection matrix B (p by Lp) is,

$$B = [\, B_0^{(l_0)} \; B_1^{(l_1)} \; ... \; B_{K-1}^{(l_{K-1})} \,] \tag{4}$$

A weighted mean square error (WMSE) distortion criterion is used for training the codebooks and for the selection of the quantized vector in codebook [2]. The WMSE between the original and the quantized vector is defined as

$$d_r = \sum_n (x_n - \hat{x}_n)^T W_n (x_n - \hat{x}_n) \tag{5}$$

where W_n is a diagonal matrix which may depend on x.

2.2. Simultaneous Joint MSVQ

The goal of simultaneous joint codebook design is to jointly optimize all codevectors over all stages after each iteration, and therefore joint design can converge faster and produce a final distortion less than iterative sequential design [1]. In the joint codebook design procedure, the multistage codebook is considered as a single entity c and multistage selection matrix B is taken as in (4). To jointly optimize the stacked matrix codebook structure during training and testing, the average distortion d_r is minimized where $\hat{x}_n = B_n.c$. Hence,

$$
\begin{aligned}
d_r &= \sum_n (x_n - \hat{x}_n)^T W_n (x_n - \hat{x}_n) \\
&= \sum_n (x_n{}^T W_n x_n) - 2c^T \sum_n (B_n{}^T W_n x_n) + c^T \sum_n (x_n{}^T W_n B_n) c \\
&= d_0 - 2c^T Y + c^T Q c \ , \ Y : (L.\mathrm{p} \text{ by } 1) \ \ Q : (L.\mathrm{p} \text{ by } L.\mathrm{p})
\end{aligned} \tag{6}
$$

The minimizing solution should satisfy $Q.c = Y$ and a stacked codebook is then obtained which minimize d_r in the form $c = Q^{-1}Y$. Projection method is used to compute inverse Q matrix. If WMSE distortion criterion is used for training and selection of the best codevectors, W_n is a diagonal matrix and therefore the matrix Q has a smaller size.

2.3. The proposed simultaneous joint MSVQ

In the training procedure, in order to reduce Q matrix storage and to drop unnecessary zero multiplications, $c^T.Q.c$ is constructed as follows: The matrices c and Q are divided into small blocks. In these blocks, all-zero regions are dropped and the remaining parts are considered as vectors in order to simplify mathematical operations. Finally $c^T.Q.c$ is changed to a summation of vector multiplications. With the new arrangement, the diagonal matrix Q_{ii} (L_ip by L_ip) can be considered as a vector of size (L_ip by 1). Q_{ij} (L_ip by L_jp) is a matrix including p dimensional diagonal matrices that can be considered as p dimensional vectors and Q_{ij} is expressed as a vector in size of ($L_ip.L_j$ by 1).

$$
c^T.Q.c = \sum_{i=1}^{K} c_i{}^T.Q_{ii}.c_i + \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} c_i{}^T Q_{ij}.c_j + \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} c_j{}^T Q_{ji}.c_i \tag{7}
$$

where c_i (L_ip by 1) is a codebook for stage i. The vector Y is divided into Y_i (L_ip by 1) subvectors. Furthermore, with the new arrangement in d_r, it is not necessary to use multistage selection matrix B_n in d_r. During the training procedure, for a K stage MSVQ codebook design, if s_i is taken as a selected codevector index parameter for stage i and if the vectors lsf and wcb are current p dimensional input LSF training vector and weighting vector respectively, updated parameters for d_r are,

$$
\sum_{i=1}^{K} \sum_{j=1}^{p} Yi[s_i.p + j] \mathrel{+}= \sum_{j=1}^{p} wcb[j].lsf[j] \tag{8.a}
$$

$$
\sum_{i=1}^{K} \sum_{j=1}^{p} Q_{ii}[s_i.p + j] \mathrel{+}= \sum_{j=1}^{p} wcb[j] \tag{8.b}
$$

$$
\sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \sum_{n=1}^{p} Q_{ij}[s_i.p.L_i + L_j + n] \mathrel{+}= \sum_{n=1}^{p} wcb[n] \tag{8.c}
$$

$$
d_0 \mathrel{+}= \sum_{i=1}^{p} lsf[i].wcb[i].lsf[i] \tag{8.d}
$$

The above vectors are updated for all input training vectors and the vectors d_0, Q and Y are obtained for the computation of d_r. If any zero parameter occurs in the vector Q_{ii}, the solution for $c_i = Y/Q_{ii}$ is not possible. Therefore the related input codevector in the trained codebook must be rearranged. A scalar quantity between [0-1] is multiplied by this codevector. $\beta = 0.9$ is taken in our training procedure and the above procedure is repeated until a nonzero situation occurs in the vector Q_{ii}, i.e.,

$$
\sum_{i=1}^{K} \sum_{j=1}^{L_i p} (if \ \ (Q_{ii}[j] == 0) \quad c_i[j] = c_i[j].\beta\,) \tag{9}
$$

Taken as $C_{ij} = Q_{ij}.c_j$ and $C_{ji} = Q_{ji}.c_i$, the vectors C_{ij} and C_{ji} are calculated as follows:

$$
\sum_{a=1}^{Li.p} C_{ij}(a) = \sum_{a=1}^{Li} \sum_{b=1}^{Lj} \sum_{k=1}^{p} [C_{ij}(a.p+k) + Q_{ij}((a.p+k).L_j + b).c_j.(b.p+k)] \tag{10.a}
$$

$$
\sum_{a=1}^{Lj.p} C_{ji}(a) = \sum_{a=1}^{Li} \sum_{b=1}^{Lj} \sum_{k=1}^{p} [C_{ji}(a.p+k) + Q_{ji}((b.p+k).L_i + a).c_j.(b.p+k)] \tag{10.b}
$$

The new $c^TQ.c$ is applied to compute d_r in the training procedure and the calculation complexity is decreased approximately fifteen times for a 18 bits/frame vector quantization codebook (11). All other procedures for joint codebook design, is applied similar to [1-3]

$$
\sum_{i=1}^{K} L_i p.L_i p \Rightarrow \sum_{i=1}^{K} L_i p + \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} L_i p.L_j \tag{11}
$$

2.4. Residual LSF vector

The residual vector proposed in this paper is an extension of the residual vector in [4]. The LSF parameter vector is obtained by performing a 10^{th} order LPC analysis. Next, avarage LSF vector of the training set I_{DC} is subtracted from the LSF vector belonging to the i^{th} frame I_i to obtain a differantial LSF vector e_i given by,

$$
e_i = I_i - I_{DC} - \alpha.res_{i-1}, \quad res_0 = 0 \tag{12.a}
$$

$$
res_i = I_i{}^q - I_{DC} + \alpha.res_{i-1} \tag{12.b}
$$

$$
I_i{}^q = e_i{}^q + I_{DC} + \alpha.res_{i-1} \tag{12.c}
$$

where I_i^q and e_i^q are the quantized versions of I_i and e_i, respectively (figure 1). An experimentally determined scalar quantity $\alpha=0.325$ is used as the correlation coefficient for backward prediction of the residual LSF vector. A weighted Euclidean distance measure $d(e, \hat{e})$ between the input residual LSF vector e and the quantized residual LSF vector $\hat{e}$ is given by

$$
d(e,\hat{e}) = \sum_{j=1}^{p} w_j (e_j - \hat{e}_j)^2 \tag{13}
$$

where p (p=10 in our case) is the number of elements in the residual LSF vector and w_j is the weight assigned to the j^{th} LSF.

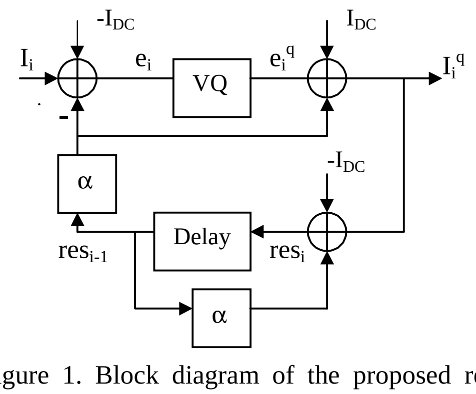

Figure 1. Block diagram of the proposed residual LSF coding scheme

Residual signal res_i is obtained by subtracting the current speech frame from the sum of the backward prediction of past speech frames. At the receiver, subtracted quantity is added again and quantized LSF vector $\hat{e}$ is obtained.

2.5. Residual-MSMQ

In a matrix quantization system, *M*-matrices of LPC vectors for *M* speech frames are grouped and the Generalized Lloyd Algorithm (GLA) is applied to this matrix sequence [5]. A codebook of reproduction matrices is referred to as a matrix codebook. Let x_i be a vector of LSF parameters for a speech frame, i.e.,. $x_i = [xx_0,xx_1,...,xx_{p-1}]^T$, $(i=1,2,..,M)$ where p is the order of the LPC filter. Then, if M is some integer ≥1, define the *pxM* matrix $X = [x_1,x_2,...,x_i,...,x_M]^T$ and reproduction matrix $Y = [y_1,y_2,...,y_i,...,y_M]^T$ where $y_i = [yy_0,yy_1,...,yy_{p-1}]^T$ and $(i = 1,2,..,M)$. A WMSE distortion criterion is used for training the codebooks and for the selection of the quantized vector in codebook [2]. The WMSE between the original and the quantized parameter vector is defined as,

$$D(X,Y) = \sum_{j=1}^{M} d(x_i, y_i) \tag{14}$$

$$d(x_i, y_i) = (x_i - y_i)^T w_i (x_i - y_i)$$

The training sequence of matrices is obtained by sliding the first matrix along the training sequence of speech frames to obtain the greatest number of matrices from the training sequence. According to the minimum distortion measure, a test sequence of M LPC frames $[t_1,t_2,...,t_j,...]$ is grouped into matrices $T_j = (t_{M(j-1)+1}, t_{M(j-1)+2},...,t_{Mj})$. Then, the reproduction matrix codebook Y is obtained according to the minimum distortion calculation and the transmitted codeword index m for the j^{th} matrix in the test sequence T_j is given by

$$m : D(T_j,Y_m) \le D(T_j,Y_k) \quad \text{for } Y_k \in [Y:Y_1,Y_2,...,Y_B] \tag{15}$$

where B is the size of codebook Y. In the proposed R-MSMQ scheme, we used an extension of the Multi Stage Vector Quantization (MSVQ) scheme presented in [1-3] in which we combined multiple frames for matrix quantization. Then, simultaneous joint codebook design method [1] was applied to R-MSMQ. The parameters that are quantized are residual LSF parameters of speech frames. The spectral distortion (SD) is calculated over the frequency band of 100-3800Hz for 8 kHz sampled speech. The training database (65,685 vectors) which consisted of English sentences were lowpass filtered and downsampled to 8 kHz. For the R-MSMQ codebook design, two consecutive speech frames are grouped into a superframe and jointly quantized.

Table 1. LSF quantizer performance for residual codebooks

Codebook	Used bits (bits/frame)	SD (dB)	%outlier [2-4dB]	%outlier [>4dB]
Joint MSVQ	[776]-20	1.167	3.693	0.029
Joint R- MSVQ	[776]-20	1.101	2.955	0.020
İerative R-MSMQ	[995544]-18	1.124	1.852	0.000
İterative R-MSMQ	[10-998]-18	1,102	1,325	0,000

The SD result and outlier performance for four stage and six stage R-MSMQ codebooks at 18 bits/frame, and three stage MSVQ and three stage R-MSVQ codebooks at 20 bits/frame are shown in table 1.

3. VOICED/UNVOICED/MIXED DECISION

Pitch and voicing information are coded with seven bits. For voiced frames, one of 63 pitch values [20-160] is selected and coded as shown in the table 2. The table 2 is used in decoding pitch and voicing information to determine if a frame is unvoiced (U), in voicing transition (mixed) or voiced (V). If voiced, the decoded value shall be used as the pitch period. The pitch determination algorithm relies on the original speech data and does not make use of the LP residual signal. Pitch analysis is performed in the spectral domain using the algorithm described in [6]. The voicing algorithm used in this paper adapts to various acoustic noise levels and features using a *K*-level adaptive linear discriminant classifier with *N* parameters,

$$\sum_{i=1}^{N} a_{ij} p_i + c_j > 0, \quad \text{where } j \in [0,.....,K-1] \tag{16}$$

where a_{ij} and c_j are appropriate weights and p_i are the measured signal parameters. Voicing decisions are made for each half frame of the windowed input speech using the following signal measurements: The zero crossing rate, energy measures, reflection coefficients and prediction gains. The weights a_{ij} and c_j adapt to the acoustic noise level by selecting j according to the signal-to-noise ratio (SNR). The input signal then classified as unvoiced (including silence) or voiced. This decision is made by a linear discriminant function. [7]

Table 2. Joint quantization of pitch and V/U decisions

V/U decisions	Quantization index for pitch and mixed excitation	
	Pitch index	Mixed excitation
UU	0	0
UV	0	1
VU	1->63	1
VV	1->63	0

For UU and UV situations, the pitch value in the decoder is taken as P=50 and mixed excitation is applied to the reproduced speech in UV and VU forms.

4. PARAMETER INTERPOLATION

Gain is coded using five bits and a new interpolation procedure is applied in synthesis to reduce discontinuities due to gain changes between frames. The improved interpolation procedure below assumes that if the gain difference between the gain of the old frame (G_p) and the gain of the current frame (G) is greater than 10dB, a nonlinear change occurs between the gains and a constant value=0.3 is used to reduce the complexity; if the gain difference is smaller than 10dB, the variation can be assumed as linear. The gain is linearly interpolated between the G_p and an average gain value G_x, if the starting point $f_0 = 0,1,....,frame$ of the new pitch period is less than *frame/2*, otherwise gain is interpolated between the G and G_x. The interpolation factor *fact* is based on the starting point of the new pitch period:

$fact = f_0 / frame$

if $(G < 10dB)$ $G_x = 10dB$
else if $(G > G_p)$ *then*
  if $(G-G_p) < 10dB$, $G_x = G_p+0.5*(G-G_p)$
  else if $(G-G_p) \geq 10dB$, $G_x = G - (G-G_p)*0.3$
else if $(G_p > G)$ *then*
  if $(G-G_p) < 10dB$, $G_x = G+0.5*(G_p-G)$
  else if $(G-G_p) \geq 10dB$ $G_x = G_p - (G_p-G)*0.3$

Figure 2. The new gain interpolation algorithm used in MQBC coder

Other parameters are linearly interpolated between the past and the current frames. The bit allocation scheme for 25 ms analysis frames is shown in table 3.

Table 3. Bit allocation of 1.2 kbps MQBC coder

Coding parameters	1.speech frame	2.speech frame
LSF parameters	36	
Pitch and U/V	6	6
Gain	5	5
Mixed excitation	1	1

5. SUBJECTIVE TEST RESULTS

A MOS (mean opinion score) experiment was done to assess the performance of the 1.2 kbps vocoder. For each condition, a set of twelve IRS-filtered speech was evaluated by ten non-experts. For comparison purposes, the 2.4 kpbs LPC10e vocoder [8] and the 2.4 kbps MELP standard coder [9] were used. The coders were tested on speech containing quiet background, office noise and car noise. For the speech containing car noise, DMOS (Degredation MOS) is used. All of the coders scored higher for male talkers than female talkers. The results are calculated by averaging the results of male and female scores as shown in table 4. In acoustic noise, such as office or car conditions, the speech retained its intelligibility and talker identity. The subjective quality of the proposed coder is found better than that of LPC10e and aproximately near the 2.4 kbps MELP standard.

Table 4. MOS testing results

Vocoder	Mean Opinion Score		
	Quiet	Office	Car Noise
2.4 kbps LPC10e	2.3	2.1	1.3
2.4 kbps MELP	3.5	3.0	2.7
1.2 kbps MQBC	3.1	2.7	2.2

6. CONCLUSION

A new matrix quantization based speech coder is presented which has been shown to produce good quality speech at a bit rate of 1.2 kbps. In the proposed coder, the transmitted parameters of consecutive frames are quantized together. The new MQBC coder uses new techniques for improving performance, such as joint quantization of pitch and V/U decisions, gain interpolation and residual LSF quantization. The new residual multistage matrix quantization scheme reduces the bit rate using residual LSF vectors obtained from the first-order backward prediction of LSF vectors. The new gain interpolation algorithm gives more correct and smooth gain estimations. The MOS test have indicated that the subjective quality of the proposed coder is found superior to that of LPC10e and approximately near the 2.4 kbps MELP standard.

REFERENCES

[1] W.P.LeBlanc, B.Bhattacharya, S.A. Mahmoud, “Efficient Search and Design Procedures for Robust Multi Stage Vector Quantization of LPC Parameters for 4 kbps Speech Coding”, *IEEE Trans. on Speech and Audio Processsing,* Vol.1, No:4, pp:373-385, October, 1993

[2] W.P.LeBlanc, S.A. Mahmoud, V. Cuperman, “Joint design of MSVQ Codebooks for LSP Quantization with Applications to 4 kbps Speech Coding”, *Speech and Audio processing for Wireless and Network Applications,* Kluwer Academics Publishers, 1993

[3] B.Bhattacharya, W.P.LeBlanc, S.A. Mahmoud, V. Cuperman, “Tree searched Multi Stage Vector Quantization of LPC Parameters for 4 kbps Speech Coding”, *IEEE Int. Conf. on Acoustics, Speech and Signal Processsing,* 1992

[4] Nandkumar, S., Swaminathan, K., Bhaskar, U., “Robust Speech Model based LSF Vector Quantization for Low Bit Rate Speech Coders”, *IEEE Int. Conf. on ASSP*, 1998

[5] Chieh Tsao, Robert M. Gray, “Matrix quantizer design for LPC speech using the generalized Lloyd algorithm”, *IEEE Trans. on Acoustics, Speech and Signal Processsing,* Vol.ASSP-33, No:3, June, 1985

[6] Yoav Medan, Eyal Yair, Dan Chazan, “Super resolution pitch determination of speech signals”, *IEEE Trans. on Signal Processsing*, Vol.39, No:1, pp:40-48, January, 1991

[7] Joseph P. Campbell, Jr. Thomas, E. Tremain, “Voiced/unvoiced classification of speech with applications to the U.S. Government LPC-10e algorithm”, *IEEE Int. Conf. on ASSP,pp:473-476,* 1986

[8] T.E. Tremain, "*The Government Standard Linear Predictive Coding Algorithm. LPC-10*", Speech Technology, pp. 40-49, April 1982.

[9] A.V. McCree and T.P. Barnwell III, “ *A Mixed Excitation LPC Vocoder Model for Low Bit Rate Speech Coding*”, IEEE Transactions on Speech and Audio Processing, Vol.3, No.4, pp. 242-250, July 1995