Improved Excitation Coding for 13 kbps Variable Rate QCELP Coder

Sangwon Kang* and Dong-Ho Lee*

ABSTRACT

This paper reports on the optimal design of the excitation codebook in the 13 kbps variable rate QCELP coder of Korean speech. We present two optimal excitation codebooks which consist of 128 and 256 samples, respectively. For the design and test of the improved codebook, a data base of Korean spech is used. A quasi-Newton optimization algorithm was developed to design the codebook. The optimized codebook which remains sparse, can produce an average gain of 0.84 and 0.45 dB in SNR and SEGSNR respectively. Informal listening tests confirm the improvement in speech quality.

I. Introduction

8 kbps variable rate QCELP coder[1] is part of a CDMA digital cellular standard (IS-95), but it can not provide the toll speech quality which is required in the PCS and the FPLMTS system. Hence Qualcomm presents the 13 kbps variable rate QCELP speech coder [2], which provides the toll speech quality, as the service option of the CDMA digital cellular system and the part of the upbanded CDMA system. QCELP is a variable rate code-excited linear prediction (CELP) based coder. The QCELP excitation codebook consists of the 128 samples, arranged in a circular buffer, and adjacent vectors with dimension of L have $(L-1)$ samples overlapping. This kind of structure reduces the computation and the memory complexity, which are required in the excitation codebook search of the CELP structure[3], while preserving the reconstructed speech quality. However, current codebook is not optimized to the Korean language, and the improved codebook for the Korean speech can be designed by training the codebook sample by a data base of Korean speech.

This paper proposes the optimized codebooks with 128 samples and 256 samples for the 13 kbps QCELP coder. To retain compatibility with the existing 13 kbps QCELP coder, the optimized codebooks were required to have the similar structure as the original QCELP codebook[4].

The paper is organized as follows. The 13 kbps variable rate QCELP coding is discussed in section II. Section II presents design of the optimal codebook. In section IV, experimental results are given for the codebooks with 128 and 256 samples. A conclusion is provided in section V.

*Department of Control and Instrumentation Engineering, Hanyang University

Manuscript Received: July 7, 1997.

II. 13 kbps variable rate QCELP coder

Figure 1 shows the encoding part of the 13 kbps variable rate QCELP coder. The input speech is sampled at 8 khz and divided into 20 msec frames consisting of 160 samples. The 10^{th} order LPC coefficients are computed by the autocorrelation method, and tranformed to line spectrum pair (LSP) parameters[5], and then the split VQ[6], which consists of 5 vectors, is applied to quantize LSP parameters. The LSP parameters for pitch and codebook subframes is computed by interpolating the current LSP and the previous LSP. The QCELP coder uses an analysis-by-synthesis method to find optimal pitch and codebook parameters. That is, the encoding procedure determines the pitch and the codebook parameters to minimize the perceptual difference between the reconstructed speech and the original speech. The encoding procedure also includes quantizing the parameters and packing them into data packets for transmission. Bit allocations for each rate are listed in Fig. 2. Each frame is assigned one of the four different basic rates, namely 13.2, 6.2, 2.6, or 0.8 kbps, meaning either 264, 124, 52 or 16 bits, respectively. The coder first selects one of the four rates for each frame by a multiband energy thresholding scheme similar to the energy based decision algorithm used in IS-96-A over the entire speech band. Then seven features from speech signals are used optionally to determine the appropriate encoding mode.

Excitation sequence means the residual signal which subtracts the short term and the long term correlations from the input speech. Codebook is used to represent the excitation signal. The 13 kbps variable rate QCELP

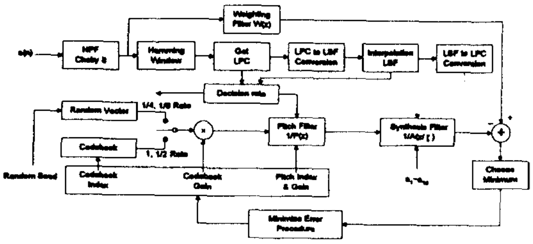

Figure 1. Encoding procedure of 13 kbps variable rate QCELP

Rate 1 (Total=264 bits)

LPC	32															
Pitch	11				11				11				11			
Codebook	12	12	12	11	12	12	12	11	12	12	12	11	12	12	12	11

Rate 1/2 (Total=124 bits)

LPC	32			
Pitch	11	11	11	11
Codebook	12	12	12	12

Rate 1/4 (Total = 52 bits)

LPC	32				
Pitch	0				
Codebook	4	4	4	4	4

Rate 1/8 (Total = 16 bits)

LPC	10
Pitch	0
Codebook	6

Figure 2. Bit allocations for the four different basic rates

coder uses two recursive codebooks with 128 samples to reduce the computation and the memory complexity, which are required in the excitation codebook search. Two codebooks are used; one for the full-rate encoding, and the other for use in the half-rate speech frame. Then 128 different vectors with dimension of L are generated as the window slides through the circular samples. Dimension L is 10 or 40, repectively, for rates of 13.2 or 6.2 kbps. Adjacent vectors have $L-1$ samples overlapping. Figure 3 shows analysis-by-synthesis procedure for finding the codebook index and gain parameters. MSE equation for the excitation codebook search is of the form

$$MSE=\sum_{n=1}^{L_c-1}\{x(n)-Gy_I(n)\}^2 \qquad (1)$$

The optimal gain G is then given by

$$G=\frac{E_{xyI}}{E_{yyI}} \qquad (2)$$

where,

$$E_{xyI}=\sum_{n=1}^{L_c-1}x(n)\,y_I(n)$$

$$E_{yyI}=\sum_{n=0}^{L_c-1}y_I^2(n),$$

L_c is the codebook subframe length, and x(n) corresponds to the weighted residual signal which substracts the zero input response of the weighted synthesis filter from the weighted input signal. Minimizing (1) is equivalent to minimizing the following error value.

$$Error=-2G\cdot E_{xyI}+G^2\cdot E_{yyI} \qquad (3)$$

For all index I and gain G, we find the optimal index I* and the optimal gain G* to minimize (3).

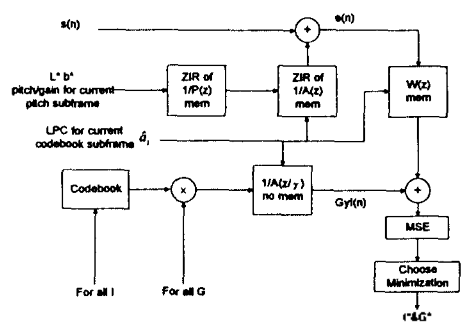

Figure 3. Analysis-by-synthesis for the codebook parameter search

Ⅲ. Design of the optimal codebook

We present how to design the optimal codebook. In this paper, we focus on an optimal codebook design for the full-rate encoding, since the half-rate speech encoding is rarely used(less than 1% of overall frames). Let the 128 scalar samples in the codebook be of the form:

$$C_i^T=(c(i),c(i+1 \bmod 128),\ldots,c(i+L-1 \bmod 128))\; i=0,1,\ldots,127 \qquad (4)$$

For a fixed training sequence of speech, say S, we can view the entire QCELP encoding algorithms as a mapping from R^{128} to R. R^{128} is the space of 128 entries $\{c(i), i=0,\ldots,127\}$, and R is the space corresponding to the distortion measure between the reconstructed and the original speech. Our goal of optimizing the circular codebook is a function minimization problem: find the values of c(0), c(1),..., c(127) which result in the minimum distortion for the training sequence of speech S.

Figure 4 shows the block diagram of the codebook

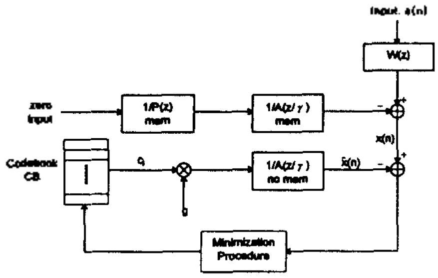

Figure 4. Codebook search structure

search in 13 kbps QCELP. Notice that the zero input response (due to the initial condition or filter memory) is subtracted to produce the weighted residual signal x(n). Lex $X=\{x(n): s(n)\in S\}$ be the corresponding residual training sequence. Let h(n) be the truncated impulse response of filter $1/A(z/\gamma)$. Then the filter output $\hat{x}(n)$ can be written in vector-matrix form as follows:

$$\hat{X}(n)=(\hat{x}(0),\ \hat{x}(1),\ldots,\hat{x}(L-1))^T \tag{5}$$

$$C_{i(X)}=(c(i(X)),\ c(i(X)+1 \bmod 128),\ldots,c(i(X)+L-1 \bmod 128))^T \tag{6}$$

$$H(X)=\begin{bmatrix} 0 & 0 & \cdots & 0 & h(0) \\ 0 & 0 & \cdots & h(0) & h(1) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ h(0) & h(1) & & h(L-2) & h(L-1) \end{bmatrix} \tag{7}$$

$$\hat{X}=g(X)\,H(X)\,C_{i(X)} \tag{8}$$

Where $H(X)$ is the impulse response matrix, which is function of input speech s(n) and hence implicitly dependent on X, g(X) is the gain factor, and $C_{i(X)}$ is the codeword selected for X. Let function $f: R^{128}\rightarrow R$ be the normalized weighted distortion defined as:

$$f=\sum_{X\in X}\|X-g(X)H(X)C_{i(X)}\|^2\Big/E_w \tag{9}$$

where E_w is the weighted signal energy which is a constant scalar for a fixed training sequence of speech S. Now f is a function of c(0), c(1),…, c(127). The problem is to minimize f over {c(0), c(1),…, c(127)} in R^{128}. This is a typical numerical optimization problem. Finding a global minimum in such a high-dimensional space is a very difficult task. However, many general numerical optimization algorithms can be efficiently applied to find a local minimum. Hence we can find many local minima by starting from widely varying initial points and then pick the one leading to the smallest function value. Among many numerical optimization algorithms, we pick the variable metric method, which is sometimes called a quasi-Newton method[4, 7].

Ⅳ. Experimental results

The training data is a sequence of 2,860,009 samples of Korean speech made up of phrases by 4 male and 4 female speakers. As seen in Table 1, the testing data is a sequence of 120,926 samples of Korean speech consisting of phrases by 3 male and 2 female speakers. Data is received from FM radio, and sampled at 8 khz. For finding the value close to a global minima, we obtain 15 local minimum values by starting from 15 initial points and then pick the one leading to the smallest value. The initial codebooks were produced from the memoryless Gaussian generator with unit varience and zero mean.

Table 1. Korean speech samples used for testing

Sent	Content(in Korean)	Samples	Sex
1	절대 세무조사는 없을것으로 알려졌습니다.	23,423	M
2	에어버스 310 여객기가 공중 납치되었습니다.	27,297	M
3	군 당국에 의해서 부사히 구조되었습니다.	21,509	M
4	지나친 흡연은 건강을 해칩니다.	25,679	F
5	성남쪽 차도가 일부 지장을 받고있습니다.	23,018	F

In this paper, two kinds of optimal codebooks are designed. First is a 128 sample codebook, similar to the original codebook, which adjacent vectors have L−1 samples overlapping. Second is a 256 sample codebook which adjacent vectors have L−2 samples overlapping. In the 256 sample codebook, 128 different vectors are generated as the window with length L slides by two samples. The 256 sample codebook produces better speech quality rather than the 128 sample codebook at the expense of little increase of the complexity as seen in Table 2.

Table 2. Additions and multiplies per sample(N = 128, L = 10)

Codebook	LPC filter	Pitch filter	Others	Total
Noncircular	704	132	160	996
Circular (128 samples)	133	132	160	425
Circular (256 samples)	260	132	160	552

Table 3 presents the simulation results for the improved 13 kbps QCELP coder with the optimized 128 and 256 sample codebook for the Korean speech. The objective performance advantage of the optimal 128 sample codebook is 0.53 and 0.32 dB, respectively, in the average SNR and SEGSNR values. For the optimal 256 sample codebook approach, the average SNR and SEGSNR performance is 0.84 and 0.45 dB better than those of the original codebook.

Table 3. Comparison of performance for different codebooks

Sent	Original codebook		Opimized codebook (128 samples)		Opimized codebook (256 samples)	
	SNR (dB)	SegSNR (dB)	SNR (dB)	SegSNR (dB)	SNR (dB)	SegSNR (dB)
1	16.57	12.54	16.96	12.84	17.32	12.94
2	16.54	13.57	16.83	13.98	17.54	14.06
3	17.10	14.01	17.87	14.28	17.92	14.32
4	21.57	16.74	22.24	17.26	22.56	17.40
5	17.74	14.61	18.27	14.71	18.37	14.98
Ave.	17.90	14.29	18.43	14.61	18.74	14.74

Informal listening tests confirm above objective performance. An informal listening test was performed with 8 listeners. The test consisted of five direct comparisons of the encoded sentences of the 13 kbps variable rate QCELP coder using the original codebook, the optimized codebook with 128 samples, and the optimized codebook with 256 samples. With the optimized codebooks, the improvement in speech quality is noticeable, especially with the 256 sample codebook.

V. Conclusion

In this paper, we perform the optimal design of the excitation codebook in the 13 kbps variable rate QCELP coder of Korean speech. It is suggested that two approaches be used; the optimized 128 sample codebook and the optimized 256 sample codebook. A quasi-Newton codebook design algorithm was used to design the optimized codebook. Objective performance evaluation demonstrated increases in SNR/SEGSNR of 0.53/0.32 dB for the opimized 128 sample codebook, and 0.84/0.45 dB for the optimized 256 sample codebook. Subjective evaluation demonstrated improved speech quality.

Reference

1. TIA/EIA/IS-96 Interim standard, *Speech service option for wideband spread spectrum digital cellular system*, 1994.
2. Qualcomm Inc., *High rate speech service option for wideband spread specturm communications system*, Feb. 1995.
3. B. S. Atal and M. R. Schroeder, "Stochastic coding of speech signals at very low bit rates," *Conf. Rec., Int. Conf. Commun.*, pp. 1610-1613, May 1984.
4. L. Yang, T. R. Fischer, S. W. Kang, I. Lee, "Codebook optimization in variable-rate CELP coders with sparse codebooks," *IEEE Workshop on speech coding*, pp. 85-86, Sept. 1995.
5. F. K. Soong and B. H. Juang, "Line spectrum pair(LSP) and speech data compression," *Proc. IEEE ICASSP*, pp. 1.10.1~1.10.4, Mar. 1984.
6. K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Trans. Speech and Audio Processing*, vol. 1, pp. 3-14, Jan. 1993.
7. W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes in C : the Art of Scientific Computing*, Second Edition, Cambridge University Press, 1992.

▲Sangwon Kang

Sangwon Kang received the B.S. degree in electronic engineering from Hanyang University, the M.S. degree in electronic engineering from Seoul national university, and the Ph. D. degree in electrical engineering from Texas A&M university in 1980, 1982, and 1990, respectively.

From December 1982 until February 1994, he was with Electronics and Telecommunications Research Institute. Since March 1994, he has been an assistant professor in the department of control and instr. engineering at Hanyang University.

▲Dong-Ho Lee

Dong-Ho Lee received the B. S. degree in electronic engineering from Hanyang University, and the M. S. and Ph. D. degrees in electrical and computer engineering from The Univ. of Texas at Austin in 1986, 1988, and 1991, respectively.

From December 1991 until February 1994, he was with LG Electronics, working in the area of Digital Image Compression. Since March 1994, he has been an assistant professor in the department of control and instr. engineering at Hanyang University.