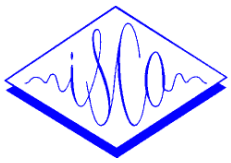

First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

# IMPROVEMENT ON 8 KB/S CELP USING LEARNED CODEBOOK: LCELP

Toshiki Miyano and Kazunori Ozawa

C&C Information Technology Research Labs., NEC Corporation
4-1-1 Miyazaki, Miyamae-ku, Kawasaki, Kanagawa 213, JAPAN

## ABSTRACT

This paper proposes 8kb/s LCELP (Learned Code Excited LPC Coding). In order to improve conventional CELP speech quality with relatively low computational complexity, LCELP uses a two-stage vector quantizer with learned and random codebooks. The advantages of using both the learned codebook and the random codebook are that synthetic quality is improved by the learned codebook and that robustness for any speech is enhanced by the random codebook. The learned codebook is designed using a speech database, and the random codebook is designed using white Gaussian signals. In order to allocate more bits to the learned and the random codebooks, LSP parameters are efficiently quantized by using a vector-scalar quantization (VQ-SQ) technique. Computer simulation results show that 8 kb/s LCELP achieves an average of 14.5 dB segmental SNR, which is 0.9 dB higher than that for the conventional CELP [2]. Informal listening tests show that LCELP speech quality is high and close to 56 kb/s $\mu$-law PCM.

## 1. INTRODUCTION

Demands for low bit rate speech coding have been rapidly increasing, especially in the areas of local area communications systems, mobile radio communications systems and mobile satellite communications systems.

CELP [1-4] is one of the promising low bit rate speech coding methods to produce good quality synthetic speech. In CELP, the excitation signal is quantized by a stochastic codebook. In order to obtain high-quality synthetic speech, the CELP requires a large size codebook, so that computation amount increases significantly. A two-stage stochastic vector quantizer can reduce the search complexity, but the synthetic speech quality is considerably lower than that for a single stage stochastic vector quantizer.

In order to cope with the above problem, this paper proposes an improved 8kb/s CELP(LCELP). LCELP has a two-stage vector quantizer with a learned codebook and a random codebook. The advantages of using both learned and random codebooks are that synthetic speech quality is improved by the learned codebook and that robustness for any speech is enhanced by the random codebook. Further, by employing two-stage vector quantizer structure [13], the computation amount for codebook search can be reduced, compared with that for a single codebook with the same codebook size. The learned codebook is designed by using a speech database, and the random codebook is designed by using white Gaussian signals. The distance measure, used in the learned codebook design, is the perceptually weighted error between original and synthetic vectors, while that used in the random codebook design is the perceptually weighted error between synthetic vectors.

In order to allocate more bits to the learned and the random codebooks, LSP parameters are efficiently quantized by using a vector-scalar quantization (VQ-SQ) technique.

## 2. LCELP ALGORITHM

Figure 1 shows a coder and decoder structure block diagram for the proposed LCELP algorithm [10]. In the coder side, LPC analysis is carried out using the autocorrelation method, and LPC coefficients are converted into LSP. LSP is quantized by a vector-scalar quantizer(VQ-SQ)[10]. In VQ-SQ, at first, LSP is quantized by VQ. Then, the VQ quantization error is quantized by SQ.

Input speech vector $x$ is perceptually weighted by

$$W(z) = \frac{A(z)}{A(z/\gamma)}, \tag{1}$$

where

$$A(z) = 1 - \sum_{i=1}^{p} \alpha_i z^{-i}, \tag{2}$$

and $p$ is LPC analysis order. The input vector $x'_w$ to the adaptive codebook search is computed by subtracting the zero input response of perceptually weighted synthesis filter, whose transfer function is shown in the following equation, from the weighted input speech vector $x_w$.

$$H_w(z) = \frac{W(z)}{A(z)} = \frac{1}{A(z/\gamma)}. \tag{3}$$

### 2-1. ADAPTIVE CODEBOOK SEARCH

The adaptive codevector $c_i^a$, with delay $i$, is produced from the past excitation signal. The optimal adaptive codevector is searched for, so as to minimize the squared Euclidean distance $d(x'_w, g_i^a H c_i^a)$ between $x'_w$ and $g_i^a H c_i^a$:

$$d(x'_w, g_i^a H c_i^a) = \| x'_w - g_i^a H c_i^a \|^2, \tag{4}$$

where $H$ is the lower triangular Toeplitz matrix, consisting of the impulse response of $H_w(z)$ and where $g_i^a$ is the optimal gain, which minimizes the distance $d(x'_w, g_i^a H c_i^a)$. The optimal gain is selected from the gain codebook for the adaptive codebook. The selected adaptive codevector and the optimal gain are denoted by $c^a$ and $g^a$.

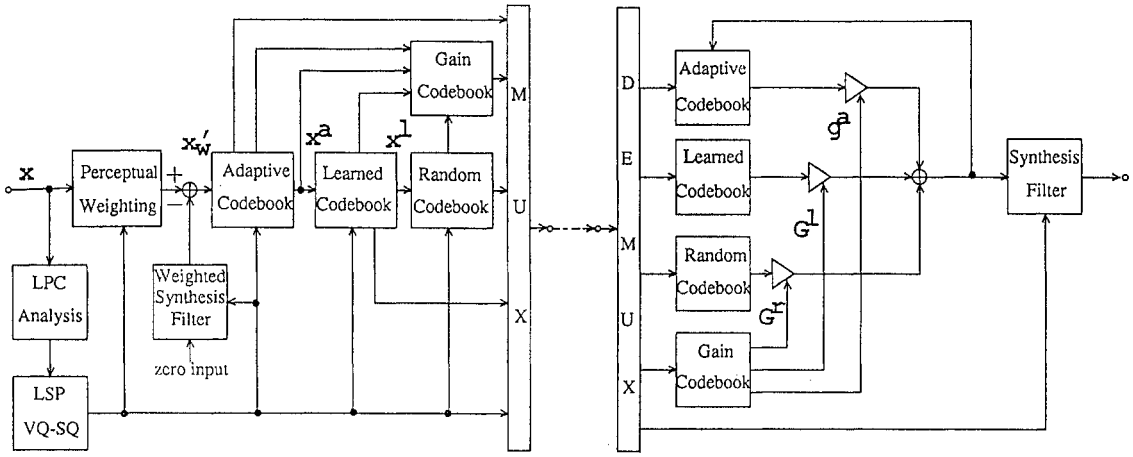


Fig.1 LCELP coder and decoder structure.

## 2-2. LEARNED CODEBOOK SEARCH

Next, learned codebook search is carried out to determine the learned codevector $c_i^l$, which minimizes the squared Euclidean distance $d(x^a, g_i^l H c_i^l)$ between $x^a$ and $g_i^l H c_i^l$:

$$d(x^a, g_i^l H c_i^l) = \| x^a - g_i^l H c_i^l \|^2, \quad (5)$$

$$x^a = x'_w - g^a H c^a, \quad (6)$$

where $g_i^l$ is the optimal gain which is determined by minimizing the distance $d(x^a, g_i^l H c_i^l)$:

$$g_i^l = \frac{< x^a, H c_i^l >}{< H c_i^l, H c_i^l >}. \quad (7)$$

$d(x^a, g_i^l H c_i^l)$ can be simplified to the following equation:

$$< x^a, x^a > - \frac{< x^a, H c_i^l >^2}{< H c_i^l, H c_i^l >} \quad (8)$$

$$= < x^a, x^a > - \frac{< x^a, H c_i^l >^2}{c_i^{l\,T} H^T H c_i^l}. \quad (9)$$

When Eq.(9) is directly carried out, a huge computation amount is required.

In order to drastically reduce the computation amount for codebook search, crosscorrelation $< x^a, H c_i^l >$ is calculated using the time-inversed filtering technique [7,8], and covariance matrix $H^T H$ is approximated by the autocorrelation matrix for the impulse response of the weighted synthesis filter,

$$R_h = \begin{pmatrix} R_h(0) & R_h(1) & \cdots & R_h(N-1) \\ R_h(1) & R_h(0) & \cdots & R_h(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_h(N-1) & R_h(N-2) & \cdots & R_h(0) \end{pmatrix} \quad (10)$$

which is a symmetric Toeplitz matrix. Thus, the Eq.(9) is modified as follows [7,8]:

$$< x^a, x^a > - \frac{(x^{a\,T} H c_i^l)^2}{R_h(0) R_l^i(0) + 2 \sum_{n=1}^{N-1} R_h(n) R_l^i(n)}, \quad (11)$$

where $R_l^i$ is the autocorrelation for the learned codevector $c_i^l$. The selected learned codevector and the optimal gain are denoted by $c^l$ and $g^l$.

## 2-3. RANDOM CODEBOOK SEARCH

The random codebook search is implemented in the same way as the learned codebook search. That is, it determines the random codevector $c_i^r$, which minimizes the squared Euclidean distance $d(x^l, g_i^r H c_i^r)$ between $x^l$ and $g_i^r H c_i^r$:

$$d(x^l, g_i^r H c_i^r) = \| x^l - g_i^r H c_i^r \|^2, \quad (12)$$

$$x^l = x^a - g^l H c^l, \quad (13)$$

where $g_i^r$ is the optimal gain for $c_i^r$, which is determined by minimizing the distance $d(x^l, g_i^r H c_i^r)$:

$$g_i^r = \frac{< x^l, H c_i^r >}{< H c_i^r, H c_i^r >}. \quad (14)$$

$d(x^l, g_i^r H c_i^r)$ can be simplified to the following equation:

$$< x^l, x^l > - \frac{(x^{l\,T} H c_i^r)^2}{R_h(0) R_r^i(0) + 2 \sum_{n=1}^{N-1} R_h(n) R_r^i(n)}, \quad (15)$$

where $R_r^i$ is the autocorrelation for the random codevector $c_i^r$.

## 3. CODEBOOK DESIGN

### 3-1. LEARNED CODEBOOK

In LCELP, the learned codebook is uesd to improve synthetic speech quality. The learned codebook is designed by the LBG method [5] Figure 2 is a block diagram, which shows how to produce training vectors for the learned codebook design(open-loop method). Training vector $x$ is produced by subtracting the zero input response for the weighted synthesis filter and the weighted synthetic vector for adaptive codevector from weighted input speech vector. Centroid vector $c$ for a cluster is calculated by minimizing the following distortion $D$:

$$D = \sum_{i=1}^{M} \| x_i - g_i H_i c \|^2, \quad (16)$$

where $x_i (1 \leq i \leq M)$ is the training vector in the cluster, $H_i$ is the impulse response matrix corresponding to $x_i$, and $g_i$ is the optimal gain. Taking the derivative of $D$ with respect to $c$ and setting the result equal to zero, we obtain

$$\sum_{i=1}^{M} g_i^2 H_i^T H_i c = \sum_{i=1}^{M} g_i H_i^T x_i. \tag{17}$$

By approximating $\sum_{i=1}^{M} g_i^2 H_i^T H_i$ to $\sum_{i=1}^{M} g_i^2 R_h^i$, which is a symmetric Toeplitz matrix, the above equation can be solved by the Levinson algorithm with reduced computation amount [11].

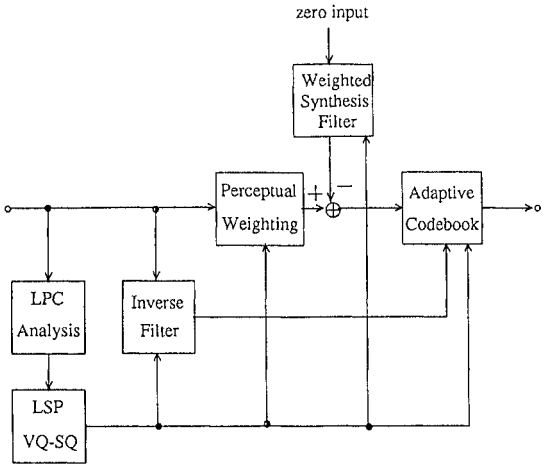


Fig.2 A blockdiagram showing training vector generation (open-loop method).

Moreover, in order to further improve the learned codebook performance, the closed-loop design procedure [9,11] is applied, until segmental SNR becomes sufficiently high. The improvement in SNR by the closed-loop design is shown in the next section.

### 3-2. RANDOM CODEBOOK

The quality of synthetic speech, which is produced by the learned codebook, is high for the training data. However, it varies with the speech signals, whose statistical characteristics are different from those for the training data. In order to maintain high speech quality for such speech, LCELP also uses a random codebook. It is assumed that the statistical property for a residual signal obtained from the output from the learned codebook, is approximated by the Gaussian probability density function. However, if the random codebook is designed by simply selecting samples from white Gaussian signals, it contains similar codevectors. Accordingly, the LBG method is invoked to design the random codebook efficiently [11]. Training vectors are white Gaussian signals. This method is similar to the method proposed in [12]. The different point is that not an average impulse response matrix, but many impulse response matrices $H_j (1 \leq j \leq T)$ are used in the design. Centroid $c$ is calculated by minimizing the following distortion:

$$D = \sum_{i=1}^{M} \sum_{j=1}^{T} \| H_j x_i - g_{ij} H_j c \|^2, \tag{18}$$

where $x_i (1 \leq i \leq M)$ is the Gaussian training vectors in the cluster, and where $g_{ij}$ is the optimal gain corresponding to $x_i$ and $H_j$. Taking the derivative of $D$ with respect to $c$ and setting the result equal to zero, we obtain

$$\sum_{i=1}^{T} \sum_{j=1}^{M} g_{ij}^2 H_j^T H_j c = \sum_{i=1}^{T} \sum_{j=1}^{M} g_{ij} H_j^T H_j x_i. \tag{19}$$

By approximating $\sum_{i=1}^{T} \sum_{j=1}^{M} g_{ij}^2 H_j^T H_j$ to $\sum_{i=1}^{T} \sum_{j=1}^{M} g_{ij}^2 R_h^j$, which is a symmetric Toeplitz matrix, the above equation can be efficiently solved by the Levinson algorithm.

## 4. EXPERIMENTS

Table 1 summarizes simulation conditions for the proposed method. The bit rate was 8 kb/s. Frame length was 20 ms and LPC analysis order was 10.

Table 1. 8 kb/s simulation condition

| | |
|---|---|
| Frame length | 20ms |
| LPC order | 10 |
| LSP VQ-SQ | 30bits |
| rms | 6bits |
| Adaptive codebook | 7bits |
| Learned codebook | 6bits |
| Random codebook | 7bits |
| Gain codebook 1 | 4bits |
| Gain codebook 2 | 7bits |

Gain codebook 1: Gain codebook for the adaptive codebook
Gain codebook 2: Two-dimensional gain codebook

### 4-1. VECTOR-SCALAR QUANTIZATION OF LSP

LSP coefficients are quantized by vector-scalar quantizer (VQ-SQ) [10]. The advantages of the VQ-SQ, compared with the conventional vector quantizer, are lower computational complexity and good performance for unknown data. In the proposed method, LSP coefficients are quantized by the vector quantizer designed by the LBG method. The error signal between the original LSP and the vector quantized LSP, is then quantized by a scalar quantizer. The structure for the VQ-SQ quantizer, used in the proposed method, is different from that for a conventional one [6]. The different point is that the quantization range for the scalar quantizer is determined by statistical analysis for the error signal. The training database contains about 7000 vectors. Table 2 shows the comparative performance between the VQ-SQ and the conventional scalar quantizer.

Table 2. LSP log-spectral distortion(dB)

| LSP quantization | bits | Log-spectral distortion |
|---|---|---|
| VQ-SQ | 30 | 0.95 |
| SQ | 32 | 1.17 |
| SQ | 34 | 1.05 |
| SQ | 36 | 0.914 |
| SQ | 38 | 0.784 |

The proposed VQ-SQ for 30 bits achieved about 0.95 dB log-spectral distortion and had a higher performance than the scalar quantizer for 34 bits.

### 4-2. RANDOM CODEBOOK

The training signals used in the random codebook design contain about 20000 vectors (about 100sec) of the white Gaussian signal. Table 3 shows segmental SNR (SNRseg) comparison of

two LCELP methods. The LBG method shows LCELP with the learned codebook and the random codebook designed by the LBG. The overlap method shows LCELP with the learned codebook and one sample shift overlap stochastic codebook. In both LCELPs, the learned codebook is designed by the open-loop method. The test data, used in the comparison, are eleven short Japanese sentences (about 33sec) uttered by 3 male and 4 female speakers. These data are different from the training database for the learned codebook design.

Table 3. SNRseg(dB) comparison

| | SNRseg |
|---|---|
| LBG Method | 14.2 |
| Overlap Method | 14.1 |

### 4-3. CLOSED-LOOP DESIGN FOR LEARNED CODEBOOK

In order to further improve the performance, the closed-loop design method [9] was applied to design the learned codebook [11]. The training database contains about 28000 vectors (about 140sec) uttered by 8 male and 9 female speakers. In this design, the random codebook designed by the LBG method was used for the second stage vector quantization. Table 4 shows SNRseg comparison between the open-loop design and the closed-loop design codebooks. The test data are eleven short Japanese sentences described above.

Table 4. SNRseg(dB) for the closed-loop and open-loop design codebooks

| | SNRseg |
|---|---|
| Closed-loop | 14.5 |
| Open-loop | 14.2 |

By using the closed-loop design codebook, LCELP achieved an average of 14.5 dB SNRseg, which was 0.3 dB higher than LCELP with the open-loop design codebook.

### 4-4. SEGMENTAL SNR COMPARISON

Table 5 shows averages of SNRseg, SNRseg for male speakers (SNRseg M) and SNRseg for female speakers (SNRseg F). In LCELP method, 6-bit learned codebook, designed by the closed-loop method, and 7-bit random codebook designed by the LBG method are used. Method 1 is CELP (frame length 20ms) with a 13-bit stochastic codebook. Method 2 is CELP (frame length 18ms) with an 11-bit stochastic codebook [2]. Method 3 is VSELP with two 7-bit codebooks [4]. All of the bit rates are 8 kb/s. No postfilter was used in any method. The test data were the eleven short Japanese sentences described above.

Table 5. SNRseg(dB) for LCELP, CELP and VSELP

| | SNRseg | SNRseg M | SNRseg F |
|---|---|---|---|
| LCELP | 14.5 | 13.2 | 15.5 |
| Method 1 | 13.9 | 12.9 | 14.6 |
| Method 2 | 13.6 | 12.6 | 14.3 |
| Method 3 | 13.5 | 13.1 | 13.8 |

Method 1: CELP with a 13-bit stochastic codebook
Method 2: CELP with an 11-bit stochastic codebook
Method 3: VSELP with two 7-bit codebooks

Experimental results showed that the 8 kb/s LCELP method achieved an average of 14.5 dB SNRseg, which was 0.6 dB, 0.9 dB and 1.0 dB higher than Method 1, Method 2 and Method 3, respectively.

Moreover, informal listening tests showed that LCELP produced high-quality synthetic speech, which was close to 56kb/s $\mu$-law PCM.

## 5. CONCLUSION

Improved 8 kb/s CELP (LCELP) was proposed. In LCELP, a two-stage vector quantizer with learned and random codebooks is used to obtain high-quality synthetic speech and robustness for any speech with relatively low complexity. In the first stage, a learned codebook, designed using a speech database, is used to improve speech quality. In the second stage, a random codebook, designed using white Gaussian signals, is used to enhance the robustness for any speech. In order to allocate more bits to the excitation codebooks, a vector-scalar quantizer is applied to efficiently quantize LSP parameters. Experimental results showed that 8 kb/s LCELP method achieved an average of 14.5 dB SNRseg, which was 0.9 dB higher than the 8 kb/s conventional CELP with an 11-bit stochastic codebook. Informal listening tests showed that LCELP produced high-quality synthetic speech, which was close to 56 kb/s $\mu$-law PCM.

## ACKNOWLEDGMENTS

The authors would like to thank Mr. Y. Unno, Miss M. Nakamura and Dr. M. Serizawa for carrying out this study.

## REFERENCES

[1] M. R. Schroeder and B. S. Atal, "Code-excited linear prediction (CELP): high-quality speech at very low bit rates," Proc. ICASSP, pp. 937-940, 1985.

[2] N. Jayant et al., "Speech coding with time-varying bit allocations to excitation and LPC parameters," Proc. ICASSP, pp .65-68, 1989.

[3] W. B. Kleijn, D. J. Krasinski and R. H. Ketchum, "Improved speech quality and efficient vector quantization in SELP," Proc. ICASSP, pp. 155-158, 1988.

[4] I. A. Gerson and M. A. Jasiuk, "Vector sum excited linear prediction (VSELP)," Proc. IEEE Workshop on Speech Coding, pp. 66-68, 1989.

[5] Y. Linde, A. Buzo and R. M. Gray, "An algorithm for vector quantizer design ," IEEE Trans. Commun.,vol.COM-28,pp. 84-95,Jan.1980.

[6] T. Moriya and M. Honda, "Transform coding of speech using a weighted vector quantizer," IEEE J. Sel. Areas, Commun., pp. 425-431, 1988.

[7] I. M. Trancoso and B. S. Atal, "Efficient procedures for finding the optimum innovation in stochastic coders," Proc. ICASSP, pp. 2375-2378, 1986.

[8] I. M. Trancoso and B. S. Atal, "Efficient search procedures for selecting the optimum innovation in stochastic coders," IEEE Trans. ASSP. vol.38, pp. 385-396, 1990.

[9] G. Davidson, M. Yong and A. Gersho, "Real-time vector excitation coding of speech at 4800 bps," Proc. ICASSP, pp. 2189-2192, 1987.

[10] T. Miyano and K. Ozawa, "Improvement on 8 kb/s CELP using learned codebook (LCELP)," Spring Nat. Convention Record of IEICE, pp.1-427-428, 1990 (in Japanese).

[11] T. Miyano and K. Ozawa, "Improvement on 8 kb/s LCELP," Proc. Autumn Meeting of ASJ, 1990 (in Japanese).

[12] T. Moriya and H. Suda, "An 8 kbit/s transform coder for noisy channels," Proc. ICASSP, pp. 196-199, 1989.

[13] B. H. Juang and A. H. Gray,Jr., "Multiple stage vector quantization for speech coding," Proc. ICASSP, pp. 597-600, 1982.