ISCA Archive
http://www.isca-speech.org/archive

EUROSPEECH '97
5th European Conference on Speech Communication and Technology
Rhodes, Greece, September 22-25, 1997

IMPROVED REGULAR PULSE VSELP CODING OF SPEECH AT LOW BIT-RATES

Yong-Soo Choi† *Hong-Goo Kang*‡ *Sang-Wook Park*† *Jae-Ha Yoo*∗ *Dae-Hee Youn*†

†ASSP Lab., Dept. of Electronic Eng., Yonsei University
Seoul 120-749, Korea
E-mail: cando@caas.yonsei.ac.kr
‡AT&T-Labs Research, Murray Hill, NJ07974, USA
∗LG Electronic Inc., Seoul Korea

10.21437/Eurospeech.1997-189

ABSTRACT

This paper describes an improved RP-VSELP (IRP-VSELP) speech coding. The RP-VSELP is classifed as a fast VSELP since it produces a comparable speech quality to the VSELP with much simplified system complexity. The new RP-VSELP coder proposed in this paper has additional new features, such as a fast codebook search obtained by employing backward filtering and pitch-adaptive regular pulse excitation. Due to new features added to the original RP-VSELP, the proposed method not only reduces the complexity of the original RP-VSELP but also provides an improved speech quality. Throughout objective and subjective tests, IRP-VSELP outperformed RP-VSELP as the reference coder. Simulation results are presented to verify the performance of the proposed method.

1. INTRODUCTION

Low bit-rate speech coding for the transmission over wireless channels requires the use of algorithms providing high quality speech signals with low system complexities. Code excited linear prediction (CELP) [1] coding schemes based on the analysis-by-synthesis technique are widely used in this application due to its high speech quality at the bit rate around 4.8 kbps. However, such quality was obtained at the expense of heavy computational power, which is the major drawback that makes it difficult to implement the CELP coding scheme using low cost and widespread digital signal processing (DSP) chips. Most of the complexity in realizing the CELP coders come from the codebook search. Many studies [2-5] have shown that the algebraic approach to the codebook design considerably reduces the computational load with a slight degradation of the speech quality, compared with stochastic codebooks composed of white gaussian noise sequences.

Among CELP coders based on the algebric approach, vector sum excited linear prediction (VSELP) [5] which was recently adopted for the GSM half-rate mobile radio is noticeable due to the robustness to channel errors and high quality speech with a reasonable complexity. However, further reduction of the complexity would be still desirable, if the implementation with present digital hardware is considered. In VSELP, since the basis vectors constructing the vector-sum codebook play an important role in enhancing speech quality and reducing computational complexity, the structure and the charateristic of the basis vectors are the keys in designing the VSELP coder. As an effort to reduce the VSELP complexity, we previously proposed a fast VSELP coding algorithm, called RP-VSELP [6], which reduced the complexity of the codebook search by a factor of 6 while producing comparable speech quality to VSELP. This advantage results from the use of an efficient vector-sum codebook in a form of mutually orthonormal regular pulses. In addition, its basis vectors are optimized to enhance the speech quality by an iterative closed-loop training process.

In this paper, we present an improved RP-VSELP coding method, referred to as *IRP-VSELP*. By introducing the backward filtering [3] to the codebook search, the IRP-VSELP reduces the complexity of the RP-VSELP. In addition, the IRP-VSELP enhances the speech quality by designing pitch-adaptive orthonormal RP basis vectors. In voiced frames, the innovation signal is adaptively converted to have pitch periodicity. Throughout objective and subjective tests, the IRP-VSELP outperformed the RP-VSELP as the reference coder at 4.8 kbps.

This paper is organized as follows. Section 2 describes the basic RP-VSELP algorithm. Improvements of the RP-VSELP focusing on the complexity reduction and the quality enhancement are presented and experimental results are discussed in Section 3. In Section 4, conclusion is made finally.

2. THE BASIC RP-VSELP ALGORITHM

In the RP-VSELP [6], a set of regular pulse (RP) basis vectors is employed to simplify the codebook search of VSELP. Therefore, the RP-VSELP is equivalent to VSELP except the excitation signal which consists of a linear combination of orthonormal basis vectors which are regularly spaced pulses shown in Fig. 1.

It is an important feature of the method that a RP basis matrix is expressed as a product of position and amplitude matrices:

$$\mathbf{V}_k = \mathbf{P}_k\mathbf{B}, \qquad 0 \le k \le K-1. \tag{1}$$

where $\mathbf{P}_k$, $\mathbf{B}$ and k represent a RP position matrix , a pulse amplitude matrix and a pulse shift index, respectively. A cost function for the RP vector-sum codebook is expressed as

$$\epsilon_k = \frac{((\mathbf{H}\mathbf{V}_k\mathbf{\Theta}_k)^T\mathbf{y})^2}{\mathbf{\Theta}_k^T\mathbf{B}^T(\mathbf{H}\mathbf{P}_k)^T(\mathbf{H}\mathbf{P}_k)\mathbf{B}\mathbf{\Theta}_k}. \tag{2}$$

In (2), $\mathbf{y}$ represents the target vector after long-term prediction, $\mathbf{H}$ is the impulse response matrix of the weighted synthesis filter $1/A(z/\lambda)$ and $\mathbf{\Theta}_k$ is the binary vector corresponding to the codebook index.

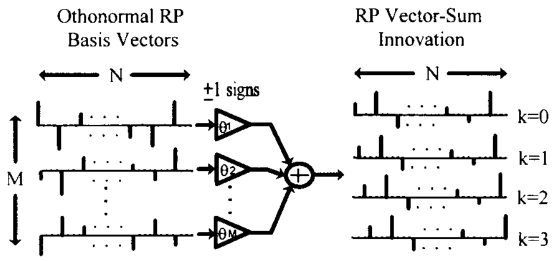

Fig. 1. The excitation structure of the RP-VSELP.

To simplify (2), the RP-VSELP utilizes three properties: (i) regular pulse property, i.e., $(\mathbf{H}\mathbf{P}_k)^T(\mathbf{H}\mathbf{P}_k) \approx r(0)\mathbf{I}$ where $r(0)$, $\mathbf{I}$ and D are the energy of the impulse response, the identity matrix and pulse spacing, respectively. (ii) mutually orthonormal property, i.e., $\mathbf{B}^T\mathbf{B} = \mathbf{I}$. (iii) binary vector property, i.e., $\mathbf{\Theta}_k^T\mathbf{\Theta}_k = M$ where M is the number of basis vectors. Using the three properties, (2) can be simplified as

$$\epsilon_k \approx \frac{1}{r(0)M}(\mathbf{\Theta}_k^T\mathbf{U}^T\mathbf{y})^2. \tag{3}$$

In RP-VSELP, the best codeword is selected by maximizing the numerator of (3). This is accomplished by taking the sign of $\mathbf{U}_k^T\mathbf{y}$ where $\mathbf{U}_k$ is computed by forward filtering the target vector $\mathbf{y}$. To find the optimal RP codeword, k and $\mathbf{\Theta}_k$ are determined sequentially. The optimal k maximizes an inner product of $\mathbf{\Theta}_k$ and $\mathbf{U}_k^T\mathbf{y}$. Here, $\mathbf{U}_k$ is recursively computed by right-shifting the forward filtered RP basis matrix $\mathbf{U}_0$. It is reported in [6] that the RP-VSELP algorithm reduces the VSELP complexity by a factor of 6 while maintaining the speech quality.

3. IMPROVED RP-VSELP

3.1. Complexity Reduction

A method to reduce the complexity of the RP-VSELP schematically shown in Fig. 2 is presented in this subsection.

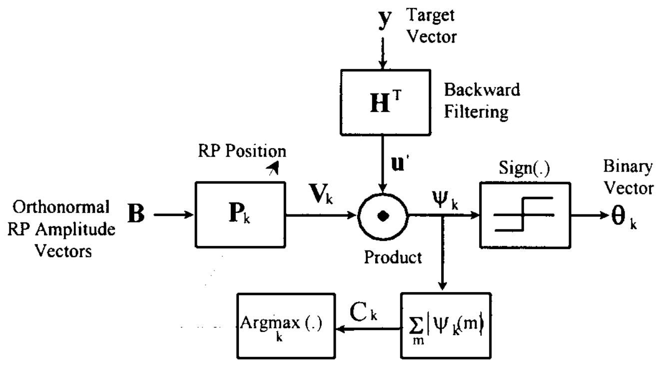

Fig. 2. A method to reduce the complexity of the RP-VSELP.

To further simplify the RP-VSELP algorithm, the numerator of (3) is reexpressed as

$$\epsilon_k' = (\mathbf{\Theta}_k^T\mathbf{V}_k^T\mathbf{H}^T\mathbf{y})^2. \tag{4}$$

Here, a new vector $\mathbf{u}' = \mathbf{H}^T\mathbf{y}$ is defined which is obtained by backward filtering [3] applied to the target vector $\mathbf{y}$ prior to the codebook search. It is straightforward to see that the backward filtering approach in (4) produces the same result that can be obtained using forward filtering as in (3). However, the reduction in computational complexity is significant. The optimal binary vector $\mathbf{\Theta}_k$ is then determined by taking the sign of $\mathbf{V}_k^T\mathbf{u}'$. That is

$$\mathbf{\Theta}_k = sign(\mathbf{\Psi}_k), \qquad \mathbf{\Psi}_k = \mathbf{V}_k^T\mathbf{u}'. \tag{5}$$

Consequently, the optimal regular pulse position index k is selected by maximizing C_k, defined as

$$C_k = \mathbf{\Theta}_k^T\mathbf{\Psi}_k = \sum_{m=1}^{M} |\Psi_k(m)|, \qquad 0 \le k \le K-1. \tag{6}$$

where the vector $\mathbf{\Psi}_k$ required to compute C_k can be formulated in a recursive form given by

$$\mathbf{\Psi}_k = \mathbf{V}_k^T\mathbf{u}' = (\mathbf{Z}\mathbf{V}_{k-1})^T\mathbf{u}', \qquad 1 \le k \le K-1 \tag{7}$$

where $\mathbf{Z}$ is a delay matrix and $\mathbf{V}_k = [\mathbf{v}_{1,k}\ \mathbf{v}_{2,k}\ \cdots\ \mathbf{v}_{M,k}]^T$. Since the mth regular pulse basis vector $\mathbf{v}_{m,k}$ for the kth pulse position is a D/K sample right-shifted version of $\mathbf{v}_{m,k-1}$ with zero input, the nth element of $\mathbf{v}_{m,k}$ can be obtained by

$$v_{m,k}(n) = \begin{cases} v_{m,k-1}(n - D/K), & D/K \le n \le N-1, \\ 0, & 0 \le n \le D/K - 1. \end{cases} \tag{8}$$

where D and N denote a pulse spacing and a subframe length. Using (8) and taking account of zeros between regular pulses, the mth element of Ψ_k is computed by

$$\Psi_k(m) = \sum_{j=1}^{Q} b_m(j) u'((j-1)D + kD/K), \ 1 \le m \le M \tag{9}$$

where $b_m(j)$ represents the jth element of the mth RP amplitude vector $\mathbf{b}_m$ and Q denotes the number of pulses.

For the codebook search, we can evaluate complexities of the RP-VSELP and the proposed method. The codebook search process consists of two parts: the weighted synthesis filtering and the codebook index search. It is assumed that the filtering means convoultion using the impulse response of length N. Approximate operations in terms of multiply-and-accumulate (MAC) are as follows taking account of zero entries between regular pulses.

- RP-VSELP Weighted synthesis (*forward*) filtering:

  $$\mathbf{U}_k = \mathbf{H}\mathbf{V}_k \ \Rightarrow \ Q(Q+1)MD/2$$

  Codebook index search:

  $$\mathbf{\Psi}_k = \mathbf{U}_k^T \mathbf{y} \ \Rightarrow \ NM$$

  $$C_k = \theta_k^T \mathbf{\Psi}_k \ \Rightarrow \ M$$

  $$\text{pulse shifts} \ \Rightarrow \ \times K$$

  $$\textit{Total} : Q(Q+1)MD/2 + (NM+M)K$$

- IRP-VSELP

  Weighted synthesis (*backward*) filtering:

  $$\mathbf{u}' = \mathbf{H}^T \mathbf{y} \ \Rightarrow \ N(N+1)/2$$

  Codebook index search:

  $$\mathbf{\Psi}_k = \mathbf{V}_k^T \mathbf{u}' \ \Rightarrow \ QM$$

  $$C_k = \theta_k^T \mathbf{\Psi}_k \ \Rightarrow \ M$$

  $$\text{pulse shifts} \ \Rightarrow \ \times K$$

  $$\textit{Total} : N(N+1)/2 + (QM+M)K$$

Taking $N = 60$, $M = 11, D = 4, K = 1$, operations for the RP-VSELP and the IRP-VSELP are 5940 and 1995, respectively. Fig. 3 shows the complexities of the two methods with respect to the number of the basis vectors. From Fig. 3, it can be observed that the IRP-VSELP algorithm redeuces significantly the computational load of the RP-VSELP method.

In this codebook search process, instead of the covariance method , the autocorrelation approach [7] (i.e. $h(n) \approx 0$, for $n \ge J$) can further reduce the complexity.

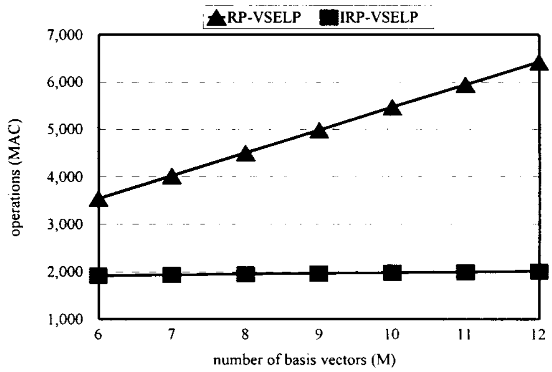

Fig. 3. The complexity with respect to the number of the basis vectors.

3.2. Performance Improvement

It is well known that accurate modeling of periodicity is of major importance in design of speech coders. In CELP coders, the periodicity is exploited by a pitch predictor or an adaptive codebook (ACB). In each subframes, the excitation for the synthesis linear prediction (LP) filter is the summation of an AC entry and a fixed codebook entry. The entries in the ACB are overlapping segments of previously constructed excitaiton. These entries are characterized by their delay relative to current subframe, called ACB delay which is usually close to the pitch period of the speech. However, this pitch periodicity still remains in the residual signal after the adaptive codebook search due to mismatching between pitch period and the ACB delay. Several methods to overcome this problem have been presented [8][9].

In this study, the method in [9] is exploited to enhance the speech quality of the RP-VSELP. It is accomplished by designing multiple pitch-adaptive RP vector-sum codebooks maintaining the advantage of the fast codebook search described in the previous subsection. The codebook design consists of three steps as follows.

- *First step:* to have more pitch-adaptive excitation structure, the RP basis vectors consisting of the vector-sum codebook is repeated according to the pitch period of the speech.
- *Second step:* pitch-adaptive RP basis vectors are a little modified as $D \ge 4$ so that the regular pulse property $((\mathbf{HP}_k)^T(\mathbf{HP}_k) \approx r(0)\mathbf{I})$ and the orthonormal property ($\mathbf{B}^T\mathbf{B} = \mathbf{I}$) is kept effective for the fast codebook search.
- *Third step:* modified pitch-adaptive RP basis vectors are orthonormalized by the Gram-Shmidt procedure.

In general, since the pitch-adaptation technique is applied to the basis vector only for $20 \le pitch \le N-1$, the number of multiple codebooks is $(N-20)/D$. For example, taking $N = 60$, $D = 4$, we require 10 pitch-adaptive RP vector-sum codebooks.

In the IRP-VSELP coder, multiple pitch-adaptive RP vector-sum codebooks are switched according to the ACB delay only for voiced frames. For the purpose of smooth pitch variation, the ACB delay is constrained to around a continuous pitch period trajectory obtained in an open-loop fashion. In unvoiced frames, random RP basis vectors are used. Fig. 4 shows a simplified block diagram of the synthesis process of the IRP-VSELP.

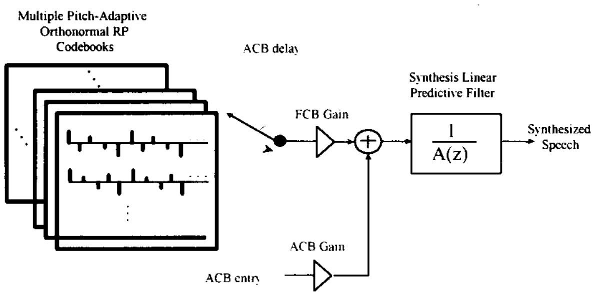

Fig. 4. A simplified block diagram of the synthesis process of the IRP-VSELP.

Performances of the RP-VSELP and the IRP-VSELP was evaluated under the same bit allocation in Table 1. From objective and subjective tests shown in Table 2, the IRP-VSELP demonstrated improvements of about 1.5 dB in the segmental signal-to-noise ratio (segSNR) and 0.4 in the mean opinion score (MOS), compared with the RP-VSELP.

Table 1. Bit allocation.

parameters	bits/7.5 ms (subframe)	bits/30 ms (frame)
LPC		38
frame energy		5
LTP	7(3)	20
RP CB-1 index	5	20
RP CB-1 positon	2	8
RP CB-1 index	5	20
RP CB-1 position	1	4
gain (GS,PO)	7	28
unused		1
total		144

Table 2. Performance evaluation.

	RP-VSELP	IRP-VSELP
segSNR (dB)	11.35	12.87
MOS	3.1	3.5

4. CONCLUSION

We have presented a improved RP-VSELP algorithm based on VSELP for low bit-rate speech coding. For an efficient codebook search, the method adopts the backward filtering and a pitch adaptive techniques. Advantages of the proposed IRP-VSELP method over the RP-VSELP are a significant reduction of the complexity and a performance improvement. Furthermore, the speech quality produced by the algorithm can be improved via codebook training process [6], which is another potential advantage of the algorithm.

A phase-adaptive technique can be adopted to further enhance quality at the expense of increasing complexity.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Young-Cheoul Park for his helpful discussion and valuable suggestions.

REFERENCES

[1] M. R. Schroeder and B. S. Atal, "Code-Excited Linear Prediction (CELP): high quality speech at very low bit rates," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp.25.1.1-25.1.4, 1985.

[2] J. P. Compbell, T. A. Tremain, and V. C. Welch, "DoD 4.8Kbps Standard (Proposed Federal Standard 1016)," *Advances in Speech Coding*, Kluwer Academic Publisher pp.121-133, 1991.

[3] J. P. Adoul, P. Mabilleau, M. Delprat and S. Morisette, "Fast CELP Coding Based on Algebraic Codes," *Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing*, pp.1957-1960, 1987.

[4] R. A. Salami, "Binary Pulse Excitation: A Novel Approach to Low Complexity CELP Coding," *Advances in Speech Coding*, Kluwer Academic Publisher pp.145-156, 1991.

[5] I. A. Gerson, and M. A. Jasuik, "Vector Sum Excited Linear Prediction (VSELP)," *Advances in Speech Coding*, Kluwer Academic Publisher pp.69-79, 1991.

[6] Y. S. Choi, H. G. Kang and D. H. Youn, "A Fast VSELP Speech Coder Based on Mutually Orthonormal Regular Pulse Vectors," *Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing*, pp.554-557, 1996.

[7] W. B. Kleijin, D. J. Kranski and R. H. Ketchum, "Fast Methods for the CELP Speech Coding Algorithm," *IEEE Trans. Speech and Audio Processing*, Vol. 38, No. 8, pp.1330-1341, 1990.

[8] P. Kroon and B. S. Atal, "Pitch predictors with high temporal resolution," *Proc. IEEE, Int. Conf. Acoust., Speech and Signal Processing*, pp.661-664, 1990.

[9] K. Mano, T. Moriya, S. Miki, H. Ohmouro, K. Ikeda and J. Ikedo, "Design of a Pitch Synchronous Innovation CELP Coders for Mobile Communications," *IEEE Journal on Selected Areas in Comm.*, Vol. 13, No. 1, pp.31-41. 1995.