APPROXIMATED CORRELATION MATRIX AND PULSE PREDICTION FOR FAST ALGEBRAIC CODE-EXCITED LPC SPEECH CODERS

Shu-Min Tsai and Jar-Ferr Yang

Institute of Computer and Communication Engineering,
Department of Electrical Engineering, National Cheng Kung University
1 University Road, Tainan, Taiwan, ROC
Tel: +886-6-2757575 ext 62362 Fax: +886-6-2345482
Email:jfyang@ee.ncku.edu.tw

ABSTRACT

With reducing computational complexity, an approximated correlation matrix of the vocal impulse response is proposed in algebraic-code-excited linear prediction (ACELP) coders. By exploring statistical characteristics, we only need to calculate a small portion of correlation coefficients before ACELP search procedure. If we further combine a pulse position prediction algorithm, we can reduce the arithmetic complexity in pre-computing autocorrelation matrix and the number of pulse position combinations with imperceptible degradation in speech quality performance. The proposed scheme can be applied to all ACELP coders such as ITU G.729 and G.723.1.

1. INTRODUCTION

For efficient speech communication, the code-excited linear prediction (CELP) structure is widely adapted by low-bit-rate speech coders [1]. Most of the complexity in the CELP coders comes from the search procedure, which needs to select the optimal excitation by feeding all possible excitations into a vocal tract synthesized filter to find the best-fitted synthesized speech. Several structured designs of excitation developed for effective determination of the optimum parameters include the multi-pulse maximum likelihood quantization (MP-MLQ) adopted by International Telecommunication Union (ITU) G.723.1 high bit rate coder. Specially, the algebraic-code-excited linear prediction (ACELP) structure has been adopted by G.729 coders [2, 3], GSM enhanced full rate (EFR) coder, G.723.1 low bit rate speech coder, and enhanced variable rate coding (EVRC) in recent years. The ACELP coding structure is popular due to the embedded efficient search for the optimal solution and no actual storage of the codebooks.

Generally, the ACELP structure needs to find the best combination of pulses and their corresponding signs from several fixed tracks to characterize the optimal speech excitation for minimizing the weighted mean square error. However, the ACELP search procedure still requires plenty of computational loads with the full search method to obtain a globally optimized excitation vector. The focused search and depth first tree search algorithms are known as very efficient search approaches to reduce the search computational complexity. In addition to the above methods, various fast methods are proposed to reduce computational complexity by decrease of pulse combinations [4-6].

In this paper, we suggested an approximated correlation matrix to reduce the computational complexity in the ACELP codebook search. Associated with a pulse prediction method, the proposed method can achieve even lower complexity and higher quality before starting the search procedure. In Section 2, we briefly introduce the search proceedings of algebraic codebook in CELP coder. Then, we propose a modified correlation matrix to form a fast search method to reduce the computation of algebraic codebook search in Section 3. Experimental results of the proposed and existed methods adopted into G.729 speech coder are depicted in Section 4. Finally, the conclusions about this paper are present in Section 5.

2. ALGEBRAIC CODEBOOK SEARCH

Algebraic codebooks are deterministic codebooks in which the codebook vectors are determined from the transmitted indices by using simple algebraic combinations rather than truly table lookup. This structure has advantages in term of table storage and search complexity. The random excitation vector is denoted by P pulse positions with P corresponding amplitudes. The algebraic codebooks are divided into P tracks; each pulse is selected from each track, which is a set of fixed positions. In G.729 coders, for example, each codevector with 40 samples contains 4 nonzero pulses, which are characterized by 4 position indices and 4 magnitudes with +1 or −1, i.e., P = 4. Table 1 shows pulse positions of 4 tracks and their bit allocations specified in the G.729 coders.

The optimal codevector is searched by minimizing the mean square error ε_ξ between the weighted target speech $\mathbf{u}$ and the reconstructed speech, $\tilde{\mathbf{u}}_\xi = \gamma\mathbf{H}\mathbf{c}_\xi$ as

$$\varepsilon_\xi = \left\|\mathbf{u} - \tilde{\mathbf{u}}_\xi\right\|^2 = \left\|\mathbf{u} - \gamma\mathbf{H}\mathbf{c}_\xi\right\|^2, \quad (1)$$

where ξ is the algebraic codevector index, γ denotes the codebook gain, and $\boldsymbol{c}_\xi$ represents the ξ^{th} codevector with subframe length L. In (1), the $L \times L$ convolution matrix $\mathbf{H}$ is a lower triangular Toeplitz matrix. In the ACELP optimization, we need to search all possible codevectors $\boldsymbol{c}_\xi$ to minimize (1). By taking the partial derivative of (1) with respect to γ and letting to be zero, we can find the optimal of γ for each fixed $\boldsymbol{c}_\xi$. Substituting the optimal γ into (1) and eliminating irrelevant terms, the optimum codevector can be achieved by maximizing

$$\tau_\xi = \frac{C_\xi^{\,2}}{E_\xi} = \frac{\left(\mathbf{d}^T \mathbf{c}_\xi\right)^2}{\mathbf{c}_\xi^T \mathbf{\Phi} \mathbf{c}_\xi}, \tag{2}$$

where the superscript T denotes the transpose operator. The inverse filtered target signal $\mathbf{d} = \mathbf{H}^T\mathbf{u}$, which can be treated as the target excitation, can be computed by

$$d(n) = \sum_{i=n}^{L-1} u(i)h(i-n) \quad n = 0,\ldots,L-1. \tag{3}$$

The correlation matrix of the impulse response, $\mathbf{\Phi} = \mathbf{H}^T\mathbf{H}$, can be given by

$$\phi(i,j) = \sum_{n=j}^{L-1} h(n-i)h(n-j) \; i = 0,\ldots,L-1, \; j = i,\ldots,L-1. \tag{4}$$

In (2), C_ξ is expressed by $C_\xi = \mathbf{d}^T\mathbf{c}_\xi$, which can be treated as the cross correlation of target excitation vector $\mathbf{d}$ and the ξ^{th} codevector $\boldsymbol{c}_\xi$. E_ξ is stated as $E_\xi = \mathbf{c}_\xi^T\mathbf{\Phi}\mathbf{c}_\xi$, which denotes the energy of the filtered ξ^{th} codevector, i.e., $\mathbf{Hc}_\xi$.

Before the ACELP search, all target excitation $d(n)$ and all correlation function $\phi(i,j)$ should be calculated in advance. Since the codevector $\boldsymbol{c}_\xi$ contains only P non-zero pulses with magnitudes of +1 or −1, the search of the optimal solution can be performed in P nested loops by selecting a pulse position from each track. During the search, the contribution of the pulse in the i^{th} track in each loop is algebraically added together. The cross correlation of target excitation and codevector, C_ξ can be simply expressed by the summation of P selected target excitations as

$$C_\xi = \sum_{i=0}^{P-1} s_i d(m_i), \tag{5}$$

where m_i denotes the selected position in the i^{th} track and s_i represents its corresponding amplitude (or sign) with +1 or −1. The energy of the filtered codevector is expressed as

$$E_\xi = \mathbf{c}_\xi^{\mathbf{T}}\mathbf{\Phi}\mathbf{c}_\xi = \sum_{i=0}^{P-1} \phi(m_i, m_i) + 2\sum_{i=0}^{P-2}\sum_{j=i+1}^{P-1} s_i s_j \phi(m_i, m_j). \tag{6}$$

To exhaustive search all pulse combinations for 4 tracks, the G.729 coder totally need to test all $2 \times 8^4 = 8192$ combinations. For each combination, we need 9 additions and 1 shift operator to compute E_ξ as stated in (6), 3 additions to compute C_ξ as depicted in (5), and 1 multiplication and 1 division to compute τ_ξ as specified in (2), if we have pre-calculated all $d(m)$ and $\phi(i,j)$. To avoid the full search of ACELP codebooks, the algorithms of pre-selected pulses are proposed to reduce the computational complexity of pulse selection by [4-6].

3. EFFICIENT SEARCH

The ACELP codebook search even with focus search strategy takes up about 20.3% computational load in the G.729 encoder. Among the search computation, it is noted that the search loops consume 74.9% computation while the inverse filtered target vector $d(n)$ and the correlation matrix $\mathbf{\Phi}$ take the remaining 25.1% computation. In this section, we will propose an approximated computation of correlation matrix to reduce arithmetic operations. If we further apply the approximated correlation matrix to a pulse position prediction algorithm, we not only decrease arithmetic complexity in pre-computing autocorrelation matrix but also reduce the number of pulse position combinations.

3.1. Approximated correlation matrix

Before computation reduction of the correlation matrix, we first analyze its mathematical and statistical behavior. The correlation matrix $\mathbf{\Phi}$ consists of $\phi(i,j)$, which is the autocorrelation of $h(n)$. Hence, the matrix $\mathbf{\Phi}$ is symmetric, i.e., $\phi(i,j) = \phi(j,i)$. Thus, the computation of correlations can only limit to right upper triangular, i.e., $j \geq i$. For simplicity, we define the r^{th} diagonal reverse-ordered correlation function $\varphi_r(k)$ with $r = j - i$ as

$$\varphi_r(k) = \phi(L-1-k-r, L-1-k), \tag{7}$$

where $k = 0,1,\ldots,(L-1-r)$. For $r = 0$, i.e., the main diagonal of $\mathbf{\Phi}$ can be expressed by $\varphi_0(k) = \phi(L-1-k, L-1-k)$, which recursively gives $\varphi_0(0) = \phi(L-1, L-1)$,…, and $\varphi_0(L-1) = \phi(0,0)$. Along the r^{th} diagonal direction, Figure 1 shows the computation of correlation matrix of impulse response for L = 40.

Figure 2 shows the averaged magnitudes of impulse responses $h(n)$ obtained from 79643 speech subframes when their pitches are greater than 39. The results exhibit that the amplitudes of $h(n)$ are very small for index n greater than six. With this fact, the autocorrelation becomes very small and near the same for $|i-j| \geq 5$. Besides, the reverse-order diagonal elements $\varphi_r(k)$ will increase slowly for high index k. From using these two observations, the most autocorrelations $\phi(i,j)$ actually are not needed to be computed. Fig. 1 also conceptually exhibits that the correlations only concentrate along a few principle diagonals. Along the diagonal (upper left) direction, i.e., $\varphi_r(k)$ with increasing k, the correlations are gradually saturated. Along the orthogonal to diagonal (upper right) direction, $\varphi_r(k)$ with increasing r, the correlations are rapidly decayed.

In order to decide elements which need not to be computed, we made two criterions. As shown in Fig. 1, if the difference of neighbor correlations along the diagonal direction is less than α as

$$\left|\varphi_r(K+1) - \varphi_r(K)\right| < \alpha, \tag{8}$$

the correlations corresponding to gray points can be directly set as

$$\varphi_r(K+1) = \varphi_r(K). \tag{9}$$

Thus, we don't have to compute the correlations for $k > K$. If the amplitude of first element of the R^{th} diagonal is less than η,

$$|\varphi_R(0)| < \eta, \tag{10}$$

the others are set to zeros as

$$\varphi_r(k) = 0,\ r \geq R, \text{ and for all } k. \tag{11}$$

To avoid extra computation, in this paper, the number of maximum elements of each diagonal, K and the number of main diagonals, R are not dynamically calculated. The determination of K and R is pre-determined by trade-off between speech quality and computational complexity. To find applicable thresholds α and η, which are respectively used to determine K and R stated in (8) and (10), we analyze 79643 subframes in advance. From experiences, we found that the thresholds with α=0.005, η=0.01 achieve better quality performances, and only need to compute about 50 elements for G.729 coders. The proposed simplification scheme saves about 92% computation for calculating elements of the correlation matrix. As shown in Fig. 1, we now only need to compute partial elements, which are marked with black points, the remaining terms, which are marked with gray (saturate values) and white (zero values) points can be approximated by (9) and (11), respectively.

3.2. Combined approximated correlation matrix with prediction

The most ACELP fast search algorithms [4-6] adopt prediction scheme to limit the search range instead of exhaustive search. It is physically meaningful that the pulse position has higher probability occurred at the larger target excitation $d(n)$. So we use the target excitation $d(n)$ to predict the possible pulse positions to reduce the search range. Before the search procedure, the first D pulses in order of preceding power are picked up from the excitation $d(n)$ per track and then exhaustively are searched to obtain the optimal pulse combination instead of the original full search. The numbers of pulse position combinations by prediction approach are depicted as Table 2.

To achieve effective ACELP search, we could get an even faster algorithm if the proposed approximated correlation matrix (ACM) method in conjunction with prediction of pulse positions. Thus, the ACM with pulse prediction combined method not only decreases arithmetic complexity in pre-computing autocorrelation matrix but also reduces the numbers of pulse position combinations.

4. PERFORMANCE COMPARISONS

For each frame, Table 3 exhibits the numbers of arithmetic operations acquired to compute the correlation matrix by various approaches. When choosing α=0.005, η=0.01, it is observed that the proposed ACM algorithm only needs 41 additions and 50 multiplications. We can reduce 93.7% additions and 92.6% multiplications in computation of correlation matrix, $\mathbf{\Phi}$ as shown in Table 4, before the backward-filtered residual target signal enters searching loop. Comparing the other methods, the ACM can save more computation complexity than the MCP [5].

The number of pre-selected pulses for per track is specified to be 4, i.e., D = 4. Tables 5 and 6 respectively show the SNR and SEGSNR performances of the original G.729 speech coder and the proposed fast schemes. Six sentences selected from the TIMIT database with different dialect regions are obtained by three males and three females, where two sentences are in Chinese language generated by one male and one female. Simulations show that the proposed method, comparing to the MCP, is with slight performance degradation but reduces arithmetic complexity in autocorrelation matrix. For the subjective evaluation, we also provide the decoded speech files on the web for readers to subjective listen test. The web site is http://www.ce.npu.edu.tw/member/faculty/anny/listen.htm. The test results indicate that the listeners cannot distinguish the speech quality of the original methods and proposed approaches.

5. CONCLUSIONS

This paper proposes an efficient algorithm to reduce the computational complexity of searching pulses for random excitation. Using the proposed ACM scheme, there will be about 93.7% off for additions and 92.6% off for multiplications in computation of correlation matrix, $\mathbf{\Phi}$ with slightly performance degradation. With the combination of both ACM and prediction concepts, the scheme can further reduce the computation of the ACELP. Not limited to G.729 coder, the proposed algorithm can be compatible to the other speech coders, such as the GSM-EFR coder, G.723.1 low bit rate coder and IS-641 speech coders, once their speech coders are with the ACELP structures.

6. REFERENCES

[1] A. M. Kondoz, "Digital Speech," *Wiley Publisher*, pp. 135, 1994.

[2] "Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP)," *ITU-T Recommendation G.729*, 1996.

[3] "Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP), Annex A: Reduced Complexity 8kbit/s CS-ACELP Speech Codec," *ITU-T Recommendation G.729-Annex A*, 1996.

[4] Shu-Min Tsai, Jia-Ching Wang, Jar-Ferr Yang, and Jhing-Fa Wang, "Efficient Coding Translation of GSM and G.729 Speech Coders across Mobile and IP Networks," *IEICE Trans. Information and Systems,* vol. E87-D, no.2, pp.444-452, February 2004.

[5] M. L. Wang, and J. F. Yang, "A Generalized Candidate Scheme of Stochastic Codebook Search for Scalable CELP Coders," *IEE Proceedings: Vision, Image and Signal Processing*, 2004.

[6] F. K. Chen, J. F. Yang and Y. L. Yan," Candidate Scheme for Fast ACELP Search," *IEE Pro.-Vis. Image and Signal Process.*, vol. 149, no. 1, February, pp. 10-16, 2002.

Table 1. Structure of algebraic codebook

Pulse	Sign	Positions	Bits
p_0	± 1	m_0 : 0, 5, 10, 15, 20, 25, 30, 35	1+3
p_1	± 1	m_1 : 1, 6, 11, 16, 21, 26, 31, 36	1+3
p_2	± 1	m_2 : 2, 7, 12, 17, 22, 27, 32, 37	1+3
p_3	± 1	m_3 : 3, 8, 13, 18, 23, 28, 33, 38 4, 9, 14, 19, 24, 29, 34, 39	1+4

Table 2. Comparison of searching loop in codebook

	Original	Prediction (D=4)
Full search	8192	512
Focus search	<1440	<96
Depth-first tree search	320	96

Table 3. Comparison of arithmetic operations

Methods	Additions	Multiplications
G.729	647	680
MCP [5]	647	680
ACM*	41	50

(ACM*: thresholds with $\alpha = 0.005, \eta = 0.01$)

Table 4. Comparison of complexity reduction rate

Methods	Additions	Multiplications
G.729	N/A	N/A
MCP	0%	0%
ACM	93.7%	92.6%

Table 5. Comparison of SNR performance

Methods	Female				Male				Average
	#1	#2	#3	#4	#1	#2	#3	#4	
G.729	13.874	14.432	13.434	13.553	13.513	13.416	14.292	10.284	13.349
MCP	13.873	14.313	13.421	13.486	13.523	13.183	13.868	10.181	13.231
Combined*	13.59	14.233	13.224	13.43	13.376	13.248	13.622	9.947	13.084

(Combined*: ACM + Prediction)

Table 6. Comparison of SegSNR performance

Methods	Female				Male				Average
	#1	#2	#3	#4	#1	#2	#3	#4	
G.729	11.494	10.847	10.129	11.402	10.142	10.272	10.115	8.446	10.356
MCP	11.472	10.67	10.047	11.205	10.124	10.055	9.879	8.332	10.223
Combined	11.292	10.604	9.967	11.175	10.013	9.961	9.75	8.147	10.114

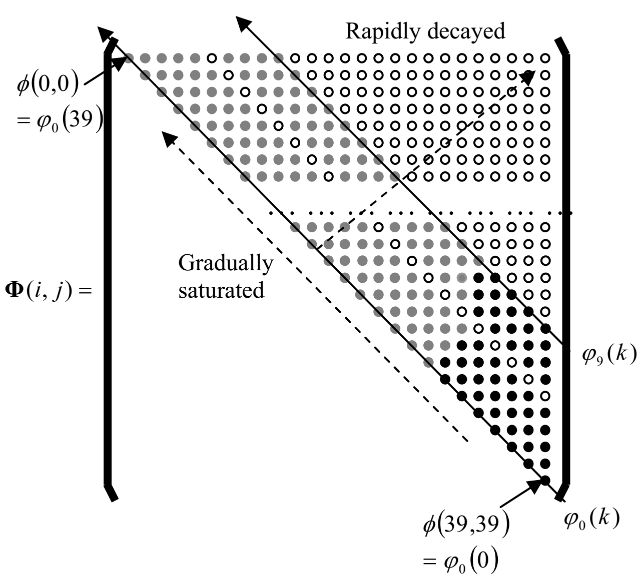

Fig. 1 The computation of the correlation function $\varphi_r(k)$ of impulse response along the r^{th} diagonal direction from low ($k = 0$) to top ($k = 39$-r) elements.

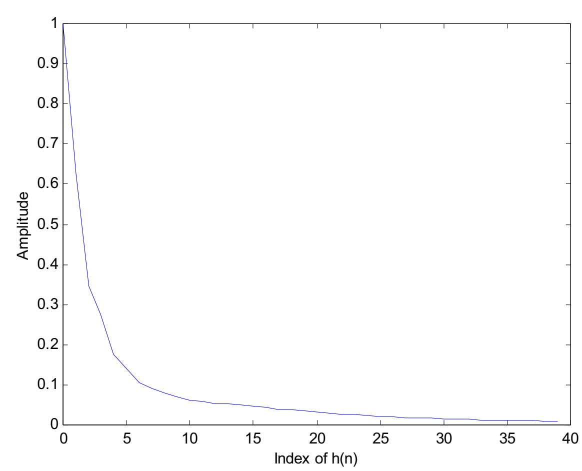

Fig. 2 Average amplitudes of impulse responses