MAXIMUM-TAKE-PRECEDENCE ACELP: A LOW COMPLEXITY SEARCH METHOD

Fu-Kun Chen and Jar-Ferr Yang

Department of Electrical Engineering, National Cheng Kung University
1 University Road, Tainan, Taiwan 70101
Email: jfyang@ee.ncku.edu.tw

ABSTRACT[+]

The ACELP method makes use of multipulse structure to represent the excitation pulses of residual signal. With the purpose of computational complexity reduction, this paper provides the Maximum-Take-Precedence ACELP (MTP-ACELP) search method under the acceptable degradation in performance. Because the maximum of target signal is preferentially compensated, the degradation of performance would be diminished. By predicting the locations of pulses, the computational complexity would be reduced. We not only reduce the possible pulse combinations in search procedure but also avoid the computation of useless correlation functions before the search procedure. Furthermore, the proposed method is compatible to any ACELP type vocoder, e.g. the G.723.1, G.729, GSM- EFR standards.

1. INTRODUCTION

With short-term prediction, long-term prediction and stochastic code excited search, the linear predictive coding (LPC) speech coders that are realized in the Analysis-by-Synthesis (AbS) optimization can achieve very good speech quality in low rate but with high computational complexity [1]. The stochastic code excited vector provides a good compensated information to the LTP residue and the LPC model error in speech coding especially for unvoiced regions. In this stochastic compensation, the Algebraic Code Excited LPC (ACELP) coding scheme has been adopted in many speech codecs, such as the ITU-T G.723.1 low rate 5.3kbps [2] and the G.729 [3, 4] coders and the ETSI GSM Enhanced Full Rate (EFR) speech transcoder [5].

The ACELP only searches for the best combination of pulses and signs to obtain the best-matched synthesis signal in minimum mean square error senses. However, the global optimization of pulse and sign combinations still requires a huge amount of computation load for the ACELP search procedure. For computational complexity reduction, the efficient ACELP search algorithms have been proposed from the aspect in reduction of pulse combinations [6-8].

The next section of this paper presents a briefly review of the standardized ACELP search. In Section 3, we proposed the maximum-take-precedence scheme to reduce the computation in the ACELP optimization. Without loss of generality, the G.729 and G.729A coders are chosen as examples to analyze the computational complexity in Section 4. Simulation results for the traditional and proposed ACELP coder are then presented. Finally, the conclusion is addressed in Section 5.

2. ACELP CODEBOOK SEARCH

In the ACELP coders, the multiple pulses (4 pulses for G.723.1, G.729 and G.729A, 10 pulses for GSM-EFR and G.729E [9]) which include their positions and amplitudes need to be found to synthesize the best-matched portion of the target signal. Table 1, for example, illustrates the structure of the G.729 ACELP codebook.

The original ACELP search involves the determination of pulse positions and pulse amplitudes, which produces the minimum mean square error E_ξ:

$$E_\xi = \left\| \mathbf{r} - G\mathbf{H}\mathbf{v}_\xi \right\|^2, \tag{1}$$

where $\mathbf{r}$ is the target signal, $\mathbf{v}_\xi$ is the algebraic code vector noted with the index ξ, and G is the codebook gain. In (1), the matrix $\mathbf{H}=h^{\mathrm{T}}h$ is defined as the lower triangular Toepliz convolution matrix, where $h[n]$ is the impulse response of vocal tract model. It can be shown that the optimum codevector is the one, which maximizes the term,

$$\tau_\xi = \frac{\left(\mathbf{r}^T\mathbf{H}\mathbf{v}_\xi\right)^2}{\mathbf{v}_\xi^T\mathbf{H}^T\mathbf{H}\mathbf{v}_\xi} = \frac{C_\xi{}^2}{\varepsilon_\xi} = \frac{\left(\mathbf{d}^T\mathbf{v}_\xi\right)^2}{\mathbf{v}_\xi^T\Phi\mathbf{v}_\xi}, \tag{2}$$

where $\mathbf{d}=\mathbf{H}^T\mathbf{r}$ is defined as the correlation function between the target signal $r[n]$ and impulse response $h[n]$, and $\mathbf{\Phi}= \mathbf{H}^T\mathbf{H}$ is the covariance matrix of the impulse response. Of course, the correlation function $\mathbf{d}$ and the covariance matrix $\mathbf{\Phi}$ should be computed before the codebook search. The code vector $\mathbf{v}_\xi$ is sparse and with amplitudes of ±1 only at certain positions $\{m_0, m_1, m_2, m_3\}$. The correlation C_ξ in the numerator of (2) can be simplified as

$$C_\xi = \left|d(m_0)\right| + \left|d(m_1)\right| + \left|d(m_2)\right| + \left|d(m_3)\right|, \tag{3}$$

since the signs of nonzero amplitudes are chosen as $v_\xi(i) = \mathrm{sign}(d(i))$, for $i = m_0, m_1, m_2$ and m_3. Similarly, the energy ε_ξ in the denominator of (2) can be expressed by

[+] This research was partially supported by National Science Council under Contract #NSC87-2215-E006-008, Taiwan, Republic of China.

$$
\begin{aligned}
\varepsilon_\xi = {} & \Phi'(m_0, m_0) \\
& + \Phi'(m_1, m_1) + 2\Phi'(m_0, m_1) \\
& + \Phi'(m_2, m_2) + 2[\Phi'(m_0, m_2) + \Phi'(m_1, m_2)] \\
& + \Phi'(m_3, m_3) + 2[\Phi'(m_0, m_3) + \Phi'(m_1, m_3) + \Phi'(m_2, m_3)],
\end{aligned} \tag{4}
$$

where $\Phi(i,j)$=sign($d(i)$)sign($d(i)$)$\Phi'(i,j)$. With pre-computed $d(m)$ for m = 0, 1, ..., (L–1) and $\Phi(i, j)$ for i, j = 0, 1, ..., (L–1), we can perform the ACELP search through the computation of (3) and (4) to maximize (2) to find the best set of indices. Therefore, for each search loop, there is a need of 12 additions, 3 shift operators, 1 multiplication, and 1 division to decide one set of positions, $\{m_0, m_1, m_2, m_3\}$, once the values of $d(m)$ and $\Phi'(i, j)$ have been pre-computed.

Finally, the nest-loop search scheme is the optimal method to find the solution of pulse locations which the criterion (2) would be maximized. Each loop corresponds to one pulse position, and the contribution of a new pulse is added in the loop. The total combinations are huge; for example, the total possible pulse combinations are 8192 in the G.729. To reduce the nest-loop search method, there will be simplified methods that set some additional restrictions to decrease the possible combinations of pulses. For real application, the ITU-T and the ETSI committees suggest the focused search and depth-first tree search methods.

3. MTP-ACELP SEARCH METHOD

To maximize τ_ξ in (2), it is noted that the pre-computation of **Φ** and **d** are needed for both focused search and depth-first tree search procedures. By observing (3), for each $d(m)$, we need (L–m) multiplications and (L–m–1) additions. Totally, the computation of **d** needs (L+1)L/2 multiplications and (L–1)L/2 additions. And, there is a need of (L–max(i,j)) multiplications and (L–max(i,j)–1) additions for computation of $\Phi(i,j)$.

If the possible region of the desired pulses can be predicted, then the number of candidate positions can be obviously condensed. Additionally, the computational complexity of **Φ** and **d** can be also reduced accordingly. In the proposed scheme, the table of pulse positions is partitioned into eight regions in a subframe. Then, we introduce a pilot function, $E(k)$ k=0, …, 7, to indicate the possibility of location which the desired pulse existed in that region. Table 1 illustrates the partition assignment of the proposed method. The column of the ACELP position table is collected into a region. The value of $E(k)$ is equivalent to the probability of the pulse located in the region R_k, k=0, …, 7. It is noted that the effective factors of pilot function include the target signal and the impulse response of synthesis filter.

First, consider the information of target signal, it can be observed that the synthesis signal is reconstructed by adding shifted impulse response after finding the pulse positions and pulse signs. Therefore, the position of maximum synthesis waveform must be same as the position of excitation pulse in unit pulse case. In order to forecast the possibility of pulse positions, we offer a simple solution to utilize the information of target signal under the consideration of computation load. The pilot function $E(k)$ is defined as the maximum amplitude of the target signal in each region to involve the information of target signal.

$$
E(k) = p_k \operatorname*{Max}_{n \in R_k} |r(n)| \; , k = 0, 1, 2, \; ..., \; K-1 , \tag{5}
$$

where p_k represents the polarity of maximum amplitude in the k-th region. It can be observed that the larger value of $E(k)$ the larger possibility of pulse be allocated in that region. When M regions that had large value of $E(k)$ were chosen, the pulse positions are reduced to 5M candidate locations as listed in the table 2. The total combination of pulse position is reduced from 2^{13} to $2M^4$.

It is noted that the selected codebook vector should be filtered through the LTP pre-filter that enhances harmonic components to improve the quality of the reconstructed speech. When the pitch delay, T is less than the subframe size, L, the true codebook $v(n)$ is modified as

$$
v'[n] = \begin{cases} v[n] & n = 0, ..., T-1 \\ v[n] + \beta v[n-T] & n = T, ..., L-1 \end{cases} \tag{6}
$$

where β is pitch gain and bounded by $0.2 \le \beta \le 0.8$ in the G.729, and $0 \le \beta \le 1$ in the GSM-EFR. Equivalently, the above modification is performed in the fixed codebook search by modifying the impulse response $h[n]$ of the vocal model. Consequently, the pilot function should be modified with the information of $h[n]$ which would disturb the location of pulse positions. Due to the repetition of impulse response, the synthesis waveform not only arises from the position of the searched excitation pulse, but also has another pick at the position of lag T from the selected pulse position. The pulse should not be located in the position of the duplication caused by the modification of $h[n]$ in actually. For the modification of the pilot function, two matters must be considered. One is the sign of the pulse, and the other is pitch gain and lag. For example, for the G.729 coder, the modified pilot function is described as follows:

$$
\tilde{E}(k) = \begin{cases} E(k) + \beta E(k+\tilde{T}) & k = 0, ..., (K-1-\tilde{T}) \\ E(k) & k = (K-\tilde{T}), ..., (\tilde{T}-1) \\ (1 - \beta p_k p_{k-\tilde{T}}) E(k) & k = \tilde{T}, \; ..., \; (K-1) \end{cases} , \tag{7}
$$

where $\tilde{T} = \left\lfloor \frac{T}{5} \right\rfloor$, and the symbol $\lfloor \cdot \rfloor$ means Gauss operator. Finally, the desired number of pulse positions and their corresponding values through (2) should be founded in the selected candidate pulse positions. It is quite obvious that only little extra operations of addition is needed to predict the possibility of pulse positions before the ACELP search and related computation. The search process does not perform the entire position combinations. Concurrently, the needed elements of $d(m_i)$ and $\Phi(m_i, m_j)$ are greatly reduced before the search process is performed. Furthermore, the proposed method can easy apply to the focused search method or the depth-first tree search method in ACELP structure.

4. COMPUTATIONAL COMPLEXITY AND SIMULATION RESULTS

The computation reduction of the proposed scheme includes minimizing the computation of required $d(m_i)$ and $\Phi(m_i, m_j)$ and decreasing the total number of search loops. With prediction of pulse positions, it is unnecessary to compute the useless elements of $d(m_i)$ and $\Phi(m_i, m_j)$. Only the $d(m_i)$ of candidate position needs to be computed. It is noted that the computational complexity of $\Phi(m_i, m_j)$ depends on the site of the chosen regions. Because the computation of $\Phi(m_i, m_j)$ can be recursively obtained, such that more elements of $\Phi(m_i, m_j)$ are required to be computed if the selected regions are closer to the front of subframe. Oppositely, a little elements of $\Phi(m_i, m_j)$ need to be computed if the chose regions are closer and converge on the end of subframe. Generally, the computation load depends on terminal positions of candidate pulses table that is chosen by the condensing function.

Table 3 shows the numbers of multiplications and additions needed in the calculation of $d(m_i)$ and $\Phi(m_i, m_j)$ for both the worst and the best cases. The reduction of computational complexity is about 27%~48% off for $d(m_i)$ and $\Phi(m_i, m_j)$ if we preset the number of regions as 5, i.e. $M = 5$. For $M = 5$, it is noted that the numbers of combinations are changed from 1440 to 220 for the focused search method and from 320 to 140 for the depth-first tree search method. The pulse combinations would be significantly reduced in the search procedure. Table 4 shows the number of combinations in searching procedure for different methods.

In the following, we performed the MTP-ACELP method on the G.729 and the G.729A codec, and compared the degradation of performance. In order to have better speech quality or lower computation, the selected number of regions can be increased or decreased. With M varying, Table 5 shows the performances of focused search and depth-first tree search methods by using MTP-ACELP scheme. From Table 7, we can find that the critical case is $M = 4$ because there are four pulses be chosen in the G.729 standard. In the worst case, the pulses straggle over the different regions. The case M=5 is a better trade-off between the quality of reconstructed signal and the computation load in the G.729. However, the number of pulse combinations is only 140 for the proposed method combined with the depth-first tree search for M=5. In listening tests, we almost can not tell any speech degradation for M= 5. Even with M=2, there still does not seem to be any serious or injurious speech for listening. Because the maximum target signal is preferentially compensated. Accompanying with the above subjective tests, we provide the readers the decoded speech files on

http://netcity.hinet.net/chenfk/listening.htm

for objective listening.

It is noted that the proposed candidate scheme (with M=5) saves about 27%~48% computation for $d(m_i)$ and $\Phi(m_i, m_j)$ and reduces 84% of search loops for the focused search method and 56% for the depth-first tree search method.

5. CONCLUSION

The proposed maximum-take-precedence scheme with selected regions greatly reduces the computational complexity for standardized ACELP search procedure with slightly performance degradation. The reduction is not only in the possible pulse combinations but also in the computational load of correlation functions. Using the MTP-ACELP scheme, we can setup the desired quality of speech to control the computational complexity for ACELP codevector search procedure. In average, about 50% ~ 80% computational load can be reduced with perceptually intangible degradation in performance. Furthermore, the proposed method can be applied to any ACELP type vocoder, for example, the G.723.1, the G.729, or the GSM-EFR to greatly reduce the computation load. The reduction would be especially helpful to save the power consumption for the mobile-machine in wireless application.

6. REFERENCES

[1] A. M. Kondoz, “Digital Speech-Coding for Low Bit Rate Communications Systems”, by John Wiley & Sons Ltd., 1994.

[2] ITU-T Recommendation G.723.1, Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s, March 1996.

[3] ITU-T Recommendation G.729, Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP), March 1996.

[4] ITU-T Recommendation G.729 - Annex A: Reduced complexity 8 kbit/s CS-ACELP speech codec, March 1996.

[5] ETSI, Digital cellular telecommunications system (Phase 2); Enhanced Full Rate (EFR) speech transcoding (GSM 06.60 version 6.0.0), ETSI EN 300 726, 1997.

[6] M. A. Ramirez and M. Gerken, “A multistage search of algebraic CELP codebooks”, Acoustics, Speech, and Signal Processing, Proceedings of IEEE International Conference, Vol.1, pp.17-20, 1999.

[7] M. A. Ramirez and M. Gerken, “Efficient algebraic multipulse search”, ITS '98 Proceedings, SBT/IEEE International Telecommunications Symposium, vol.1, pp.231-236, 1998

[8] N. K. Ha, “A fast search method of algebraic codebook by reordering search sequence”, Acoustics, Speech, and Signal Processing, Proceedings of IEEE International Conference, Vol.1, pp.21-24, 1999

[9] ITU-T Recommendation G.729 - Annex E: 11.8 kbit/s CS-ACELP speech coding algorithm, September 1998.

Table 1. Region assignment of the proposed scheme

	R0	*R1*	*R2*	*R3*	*R4*	*R5*	*R6*	*R7*
m_0	0	5	10	15	20	25	30	35
m_1	1	6	11	16	21	26	31	36
m_2	2	7	12	17	22	27	32	37
m_3	3 4	8 9	13 14	18 19	23 24	28 29	31 32	38 39

Table 2. The index of position with *M* selected regions

	Rk_0	Rk_1	…	Rk_{M-1}
m_0	$5k_0$	$5k_1$	…	$5k_{M-1}$
m_1	$5k_0+1$	$5k_1+1$	…	$5k_{M-1}+1$
m_2	$5k_0+2$	$5k_1+2$	…	$5k_{M-1}+2$
m_3	$5k_0+3$ $5k_0+4$	$5k_1+3$ $5k_1+4$	… …	$5k_{M-1}+3$ $5k_{M-1}+4$

Table 3. The computation comparison of *d* and Φ

		d	Φ
Multiplication	original	820	680
	Worst case	$\frac{(81-5M)\times 5M}{2}$	$40+160M-10M^2$
	Best case	$\frac{(1+5M)\times 5M}{2}$	$10M^2+5M$
Addition	original	780	647
	Worst case	$\frac{(79-5M)\times 5M}{2}$	$39+156M-10M^2$
	Best case	$\frac{(5M-1)\times 5M}{2}$	$10M^2+M-1$

Table 4. The number of combinations in ACELP search methods in G.729

	Exhaustive search		Focused search (maximum)		Depth-First tree search	
	Original	MTP-ACELP	Original	MTP-ACELP	Original	MTP-ACELP
No. of search loops	8192	$2M^4$	< 1440	$< 2M\left\lceil \frac{90}{(8/M)^3} \right\rceil$	320	$4(2M+M^2)$
M=4	8193	512	< 1440	< 96	320	96
M=5	8193	1250	< 1440	< 220	320	140

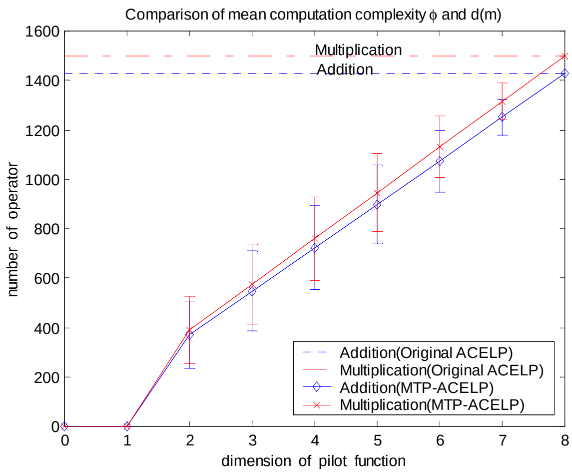

Figure 1. The comparison of total computational complexity of the proposed method

Table 5. Performance comparison of the proposed method (M regions selected)

Total no. of region		*M*=2	*M*=3	*M*=4	*M*=5	*M*=6	*M*=7	*M*=8 original
Proposed method combined with focused search in G.729								
Male (speaker 1)	SNR	8.872	10.070	10.868	11.444	11.523	11.431	11.795
	SEGSNR	10.118	11.151	11.780	12.259	12.405	12.537	12.626
Female (speaker 2)	SNR	11.514	12.830	13.690	14.151	14.485	14.628	14.654
	SEGSNR	11.455	12.378	13.021	13.299	13.600	13.697	13.727
Proposed method combined with depth-first tree search in G.729A								
Male (speaker 1)	SNR	9.547	10.500	11.157	11.487	11.783	11.871	12.054
	SEGSNR	10.352	11.250	11.883	12.143	12.382	12.473	12.593
Female (speaker 2)	SNR	12.162	13.155	13.669	13.958	14.032	14.150	14.259
	SEGSNR	11.463	12.256	12.653	12.977	13.163	13.244	13.374