A FAST SEARCH METHOD OF ALGEBRAIC CODEBOOK BY REORDERING SEARCH SEQUENCE

Nam Kyu Ha

Human & Computer Interaction Lab.,
Information Processing Sector
Samsung Advanced Institute of Technology
San 14, Nongseo-Ri, Kiheung-Eup, Yongin City , Kyungki-Do, 449-712, Korea
nkha@green.sait.samsung.co.kr

ABSTRACT

This paper proposes a fast search method of algebraic codebook in CELP coders. In the proposed method, the sequence of codebook search is reordered according to the criterion of mean squared weighted error between target vector and filtered adaptive codebook vector, and the algebraic codebook is searched until a predetermined threshold is satisfied. This method reduces the computations considerably compared with G.729 at the expense of a slight degradation of speech quality. Moreover, it gives better speech quality with smaller average search space than G.729A.

1. INTRODUCTION

Code excited linear prediction (CELP) is widely used for a low-bit-rate speech coding [1]. However, it requires a large time complexity. Algebraic Code Excited Linear Prediction (ACELP) [2, 3] gives a good solution to this problem. Recently, algebraic codebook structure is widely adopted in standard speech coders such as G.729 [4], EVRC [5] and GSM EFR [6]. Also, with the increasing interest in the digital simultaneous voice and data (DSVD), computationally efficient G.729A was recently adopted as ITU-T Recommendation at the expense of a slight degradation of speech quality [7].

In this paper, a method which reduces the time complexity of algebraic codebook search is proposed. From the algebraic codebook which has 1 pulse, mean-squared weighted error between target vector and filtered adaptive codebook vector is caculated. The sequence of codebook search is reordered according to this MSE.

Section 2 describes the search method and section 3 analyzes the algebraic codebook search space of G.729, G.729A and the proposed method, respectively. Section 4 explains the results of the objective quality measurement. In section 5, the results is described.

2. THE SEARCH METHOD

The fixed codebook is searched by minimizing the mean squared weighted error defined by

$$E_k = \|\mathbf{x} - gH\mathbf{c}_k\|^2 \tag{1}$$

where $\mathbf{x}$ is the target vector given by the weighted input speech after subtracting the zero-input response of the weighted synthesis filter 1/A(z/γ), g is a scaling gain factor, and H is a lower triangular convolution matrix constructed from the impulse response of the weighted synthesis filter. Minimizing Eq. (1) gives

$$E_k = \mathbf{x}^T\mathbf{x} - \frac{(\mathbf{x}^T H\mathbf{c}_k)^2}{\mathbf{c}_k^T H^T H\mathbf{c}_k}. \tag{2}$$

From Eq.(2), the optimum codeword is determined by maximizing

$$\xi_k = \frac{(\mathbf{x}^T H\mathbf{c}_k)^2}{\mathbf{c}_k^T H^T H\mathbf{c}_k}. \tag{3}$$

In algebraic codebook structure, excitation code vector $\mathbf{c}_k$ is generated by $\mathbf{c}_k = F\mathbf{a}_k$, where F is a shaping matrix constructed from perceptual weighting filter, and $\mathbf{a}_k$ is an algebraic codeword.

In this paper, a codeword $\mathbf{a}_k$ has 4 nonzero pulses with sign ± 1 and, each pulse has 8 or 16 positions. The dimension of a codeword is 40 and codebook structure is shown in table 1.

A codeword $\mathbf{a}_k$ can be expressed as the sum of vector $\mathbf{a}_{k_0}$, $\mathbf{a}_{k_1}$, $\mathbf{a}_{k_2}$ and $\mathbf{a}_{k_3}$, i.e.,

$$\mathbf{a}_k = \sum_{i=0}^{3} \mathbf{a}_{k_i}, \tag{4}$$

Table 1: Structure of the algebraic codebook

Pulse No.	Track	positions
1	T_0	0 5 10 15 20 25 30 35
2	T_1	1 6 11 16 21 26 31 36
3	T_2	2 7 12 17 22 27 32 37
4	T_3	3 8 13 18 23 28 33 38
	T_4	4 9 14 19 24 29 34 39

Table 2: Structure of the reordered algebraic codebook

Pulse No.	Track	positions
1	T_0	20 15 5 10 25 30 35
2	T_1	36 26 6 11 16 21 31 1
3	T_2	2 7 12 17 22 27 37 32
4	T_3	8 3 13 18 23 38 33 28
	T_4	4 9 29 14 24 34 19 39

where $\mathbf{a}_{k_i} = (0, \ldots, \pm 1, \ldots, 0)$, and $\mathbf{a}_{k_i} \in P(\pm 1, 0^{39})$. $P(\pm 1, 0^{39})$ denotes the single pulse permutation codebook. Then, $\mathbf{c}_k$ is $F\mathbf{a}_k = \sum_{i=0}^{3} F\mathbf{a}_{k_i}$. Substituting $\mathbf{a}_{k_i}$ for $\mathbf{c}_k$ in Eq. (2) yields $E_{k_i} = \mathbf{x}^T\mathbf{x} - \frac{(\mathbf{x}^T HF\mathbf{a}_{k_i})^2}{\mathbf{a}_{k_i}^T F^T H^T HF\mathbf{a}_{k_i}}$. Dividing this equation by $\mathbf{x}^T\mathbf{x}$ means $1-(\cos\theta)^2$, where θ is the angle between target vector $\mathbf{x}$ and $\mathbf{a}_{k_i}$. The smaller the angle between $\mathbf{a}_{k_i}$ and $\mathbf{x}$, the more $\mathbf{a}_{k_i}$ expresses target vector $\mathbf{x}$ well. So, such $\mathbf{a}_{k_i}$ which has a small angle between $\mathbf{x}$ and $\mathbf{a}_{k_i}$ is more likely to form $\mathbf{a}_k$.

Now, for all 40 vectors $\mathbf{a}_{k_i}$, the error

$$\xi_{k_i} = \frac{(\mathbf{x}^T HF\mathbf{a}_{k_i})^2}{\mathbf{a}_{k_i}^T (HF)^T HF\mathbf{a}_{k_i}} \tag{5}$$

is calculated easily [4] by

$$\frac{E_{k_i}}{C_{k_i}^2} = \frac{s_i d(k_i)}{\phi(k_i, k_i)}, \tag{6}$$

where k_i is the position of i-th pulse and s_i is pulse sign. ξ_{k_i} represents the importance of pulse position. For each track, ξ_{k_i}s are sorted in the decreasing order. This order composes the search sequence of algebraic codebook. The pulse position which has higher ξ_{k_i} is tested at first to find $\mathbf{a}_k$. The reordered codebook structure ordered by ξ_{k_i} is shown in table 2, for example. $\mathbf{a}_{k_i}$ which has higher ξ_{k_i} is selected first from each track.

Table 3 shows that in each track's first four pulses of reordered codebook covers 95 percent hit ratio that one of them might be a member of codeword $\mathbf{a}_k$ which reconstructs target signal $\mathbf{x}$. This data was drawn empirically from 153,600 subframes of Korean speech in NATC CD-ROM.

For codebook search, a few pulse positions are selected out of 8 in each track. The number of candidate pulse positions are determined by a threshold or time limitation.

3. THE COMPARISON OF THE SEARCH SPACE

In this section, three search methods, G.729, G.729A and proposed method are compared. Search space is the number of possible position combinations. G.729 has 5 tracks and three pulses are selected from tarck 0, 1, 2 and a remaining pulse is selected from track 4 or track 5, respectively. Since each track has 8 pulse positions, possible position combinations are 8192. G.729 has a threshold to prune the codeword search, so in the worst case, 1440 possible position combinations are tested out of 2^{13} combinations(17.5 percent) [7].

G.729A uses depth-first tree search approach to reduce the search space and results in a smaller percentage of the possible position combinations. In total, 320 position combinations are constantly tested and about 50 percent reduction of the complexity in the coder part is attributed to the algebraic codebook search in G.729A [7].

This proposed method consists of three-stage search process. At the first stage explained in section 2, the pulse positions is selected from the first to n-th pulse per each track in reordered algebraic codebook by a threshold . The maximum possible position combinations are limited to $\frac{1}{16} * 8192 = 512$ position combinations, in this paper's experiment. That is, the number of positions to be tested are determined , satisfying $\prod_{i=0}^{4} N_{T_i} \leq 512$, where N_{T_i} is the number of pulse positions selected in track i (i=0, $\cdots$, 4). With these possible position combinations, search process is performed. and a position combination A = (i_0, i_1, i_2, i_3), which maximizes Eq. (3), is selected. At the second stage, pulse pair (i_0, i_1) in track T_0 and T_1 is tested with pulse pairs (i_2, i_3) in track T_2, T_3 and T_4 which are not tested at the first stage. This process is applied to the combination (i_0, i_2) (i_0, i_3) (i_1, i_2) (i_2, i_3) in A, like (i_0, i_1). This stage needs additional 128 possible position combinations. Lastly, for 4 pairs (i_0, i_1,i_2),(i_0, i_1,i_3), (i_0, i_2,i_3),(i_1, i_2,i_3), ξ_k is calculated with a remaining track. This also needs additional 32 position combinations. In total, $\prod_0^4 N_{T_i} + 160$ position combinations are tested.

Table 3: Pulse statistics

Pulse No.	Track	positions
1	T_0	(0.6716): (0.1728): (0.0674): (0.0350): (0.0213): (0.0140): (0.0103): (0.0076):
2	T_1	(0.6481): (0.1786): (0.0737): (0.0384): (0.0234): (0.0166): (0.0122): (0.0091):
3	T_2	(0.6628): (0.1742): (0.0709): (0.0359): (0.0219): (0.0154): (0.0111): (0.0078):
4	T_3	(0.3744): (0.0738): (0.0250): (0.0124): (0.0072): (0.0053): (0.0036): (0.0027):
	T_4	(0.3724): (0.0691): (0.0242): (0.0122): (0.0075): (0.0048): (0.0031): (0.0024):

4. EXPERIMENTS

The performance of proposed search method is assessed by comparing with those of G.729 and G.727A and with full search of algebraic codebook using G.729. Test materials contain Korean male and female speech and American's. The amount of analysed speech is 480 seconds. The performance of the algorithms is compared with respect to search space, SNR_{seg}.

Table 4 shows the search space of average, best and worst cases. The numbers in table 4 represent the number of position combinations tested. At the first stage in section 2, the threshold to determine the position combinations is determined by 1/16 * 8192. If this threshold becomes high, speed is down and speech quality becomes higher, vice versa. This table shows that in the average case this search method has about a half search space than that of G.729. Best speed of G.729 and this method is similar. Also, in the average case, the proposed method is slightly faster than G.729A.

Table 4: Search space comparison of two methods.

Language	Method	Worst	Best	Avg.
Korean	G.729	1440	176	510
	proposed	800	170	248
American	G.729	1440	192	516
	proposed	800	170	221

Codebook search error is also compared. G.729 has 0.048 percent search error and this method has 0.156 percent search error rate, comparing with the code word which full search version using G.729 generated. So, search correctness is needed to be upgraded to get a good speech quality.

In order to compare the speech quality, SNR_{seg} is used. SNR is defined by

$$SNR_i = 10 \log_{10} \frac{\sum_{i=0}^{n} x(i)^2}{\sum_{i=0}^{n} (x(i) - y(i))^2} \qquad (7)$$

Table 5: Segmental SNR (dB)

Language	$G.729^{fs}$	G.729	G.729A	proposed
Korean	9.577	9.588	7.791	9.166
American	9.509	9.499	7.744	8.985

and $SNR_{seg} = \frac{\sum_{i=0}^{N-1} SNR_i}{N}$, N is the number of total frames in a speech file. SNR_{seg} is gathered in table 5 which shows SNR_{seg} of each language of three search algorithm. $G.729^{fs}$ is the full search version using G.729. For two languages, SNR_{seg} of this search method are lower by about 0.5 dB. The proposed method has about 1.2dB higher quality than G.729A.

5. RESULTS

This paper proposes a fast search method of algebraic codebook. By focusing on some one-nonzero pulse positions which give low mean squared weighted error between target vector and filtered adaptive codebook vector, search time was reduced with a slight degradation of speech quality (about 0.5dB). However, the search time is about 2 times faster than G.729. In the average case, search space of proposed method is lower than G.729A but the speech quality of this method is slightly higher than that of G.729A. Codebook search error is higher than that of G.729 by about 4 times. More research is needed to develop a fast search algorithm with the equal speech quality to G.729.

6. ACKNOWLEDGEMENT

The author would like to thank Yong Duk Cho at SAIT for preparing many speech materials for test.

7. REFERENCES

[1] M. R. Schroeder and B. S. Atal, "Code Excited Linear Prediction (CELP) High Quality Speech at

Very Low Bit Rates," *IEEE Int. Conf. on Acoust. Speech Signal Process.*, pp. 937-940, 1985.

[2] J-P. Adoul, P. Mabilleau, M. Delprat and S. Morissette, "Fast CELP coding based on algebraic codes," *IEEE Int. Conf. on Acoust. Speech Signal Process.*, pp. 1957-1960, 1987.

[3] C. Laflamme, J-P. Adoul, R. Salami, S. Morissette, and P. Mabilleau, "16 kbps Wideband Speech Coding Technique Based on Algebraic CELP," *IEEE Int. Conf. on Acoust. Speech Signal Process.,* Toronto, CA, pp. 13-16, May 1991.

[4] ITU-T Draft Rec. G.729, "Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic Code-Excited Linear-Prediction (CS-ACELP)," 1995.

[5] TIA/EIA/IS-27, "Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems", *Telecommunications Industry Association*, 1997

[6] K. Jarvinen, et. al., "GSM Enhanced Full Rate Speech Codec," *IEEE Int. Conf. on Acoust. Speech Signal Process.*, 1997

[7] Redwan Salami, Claude Laflamme, Bruno Bessette, and Jean-Pierre Adoul, "ITU-T G.729 Annex A : Reduced Complexity 8 kb/s CS-ACELP Codec for Digital Simultaneous Voice and Data," *IEEE Communications Magazine,* vol. 35, no. 9, pp. 56-63, Sep. 1997.