TECHNICAL REPORT

Switched orthogonalization of fixed-codebook search in code-excited linear-predictive speech coder: Derivation of conditions for switching

Hiroyuki Ehara*, Takako Sanda and Toshiaki Sakurai

AVC Networks Company, Panasonic Corporation,
600 Saedo-cho, Tsuzuki-ku, Yokohama, 224-8539 Japan

(Received 10 February 2015, Accepted for publication 11 June 2015)

Abstract: This document presents an algorithm of switched orthogonalization of fixed-codebook (FCB) search in code-excited linear-predictive (CELP) speech coder and derivation of conditions for switching. Orthogonalization of FCB search is an early 1990's technology, and later efficient implementation was developed on some algebraic CELP (ACELP) speech coders standardized in 2000s. ITU-T Recommendation G.729.1 is such speech coder standardized in 2006. Orthogonalization of FCB search does not degrade CELP coder performance if ideal gain parameters are applied to codebook vectors. However, because of limited performance of gain quantization and overall optimization of CELP coding algorithm, the orthogonalization does not always give improved coder performance. This document presents a switched orthogonalization algorithm based on an estimated adaptive codebook (ACB) gain parameter, which is obtained through the orthogonalized FCB search. The algorithm was evaluated in G.729.1 coder. While the orthogonalization was switched off on 20% of voicing frames, segmental SNR was improved by around 0.1 dB in average.

Keywords: Code-excited linear-prediction, Speech coding, Orthogonalization

PACS number: 43.72.+q, 43.60.Ek [doi:10.1250/ast.37.30]

1. INTRODUCTION

Code-excited linear-predictive (CELP) speech coding [1] is based on a speech synthesis model. An excitation signal is passed through a linear predictive synthesis filter, and a synthesized speech signal is outputted from the synthesis filter. The excitation signal is generated by summing an adaptive codebook (ACB) vector scaled with an ACB gain and a fixed-codebook (FCB) vector scaled with a FCB gain. The ACB is a buffered sequence of the excitation signal generated in the past and used for representing periodic components in the excitation signal. The FCB contains a number of pre-fixed vectors. Such vectors can be represented in various ways. For example, a number of Gaussian noise sequences can be used for the pre-fixed vectors. In algebraic CELP (ACELP) [2], the pre-fixed vectors are generated by combining some unit pulses located at pre-determined positions. By using pulse-sequence vectors, clearness of the synthesized speech signal can be improved in comparison with using the Gaussian noise vectors. These days, ACELP is widely used in mobile telephony services [3].

How to select the best combination between the ACB vector and the FCB vector is critical for developing a high performance CELP algorithm. The CELP speech coding is performed to find the best combination of the four parameters, ACB and FCB vectors and respective gains, so that an error between a targeted speech vector and the synthesized speech vector is minimized. Typically, the ACB vector and the FCB vector are selected in a sequential manner, i.e. ACB vector search is followed by FCB search and quantization of corresponding gains. Therefore the combination of the two vectors and gains is a quasi-optimum. Orthogonalization of FCB vectors [4] can realize simultaneous optimization of the combination between the ACB vector and the FCB vector assuming ideal gains for corresponding gains. Geometric interpretation of the orthogonalized FCB search is studied closely in [5]. However, ideal gains are not always quantized well and may be out of range of gain quantization in some particular cases. Therefore proper switching mechanism between the orthogonalization and non-orthogonalization of FCB vector search can improve overall performance of the CELP coding. Conditional orthogonalized search based on the optimum pitch gain is proposed in [6]. However, theoretical background of the switching condition and its

*e-mail: ehara.hiroyuki@jp.panasonic.com

effectiveness on the ACELP coder have not been discussed closely and reported.

In this paper, switching based on expected ACB gain is derived from equation analysis and proposed to switch the orthogonalization of FCB vectors, and its effectiveness is evaluated using G.729.1 codec [7] which uses orthogonalized FCB vector search without any switching. The remainder of this paper is organized as follows. Section 2 reviews the orthogonalized FCB vector search principle. Switching criteria for the proposed switched orthogonalization is presented in Sect. 3, and Sect. 4 demonstrates its effectiveness. Section 5 discusses about the results of Sect. 4. Concluding summary of this report is given in Sect. 6.

2. ORTHOGONALIZED FCB SEARCH

Orthogonalization of a filtered FCB vector to a filtered ACB vector was proposed in 1990 [4]. It was further applied to ACELP [8], and its efficient implementation was adopted as a part of ITU-T Recommendation G.729.1 [9]. In this section, the principle of the orthogonalized FCB search algorithm is reviewed.

2.1. Principle of the Orthogonalized FCB Search

In CELP coding, a synthesized signal is composed of two elements, one is contribution from ACB, and the other is from FCB. Therefore the synthesized signal, X' , can be expressed by the following equation.

$$X' = ga \times Y + gf \times Z$$

Y , Z , ga and gf represent ACB contribution, FCB contribution, ACB gain and FCB gain, respectively.

In the case of non-orthogonalized FCB search, the process of selecting Y , Z , ga and gf can be explained as follows. Firstly, representing target signal as X , the error between the gain-scaled ACB contribution and the target signal is minimized. This can be achieved by maximizing the following term assuming the ideal ACB gain is applied to the ACB contribution.

$$(X^t Y)^2 / Y^t Y \quad (1)$$

Once the optimum Y is selected, ga is calculated by the following equation.

$$ga = X^t Y / Y^t Y \quad (2)$$

Using the selected Y and the calculated ga , the target signal is updated for selecting Z . The updated target W is expressed as follows.

$$W = X - ga \times Y \quad (3)$$

In the same procedure with the selection of Y , the error between the gain-scaled FCB contribution and W is minimized by maximizing the following term.

Fig. 1 Selection of the best combination between Y and Z .

$$(W^t Z)^2 / Z^t Z \quad (4)$$

By this way, the closest Z to W is selected. However, this is not always optimum because Z may be not on the X - Y - W plane, i.e. if there is another Z which is on the X - Y - W plane, such Z can minimize the error between X and X' with joint optimization of ga and gf . This is shown in Fig. 1, in which X , Y , Z , Z' and W are illustrated as 3-dimensional vectors. In Fig. 1, Z' is on the X - Y - W plane. Therefore, if ga and gf are optimized jointly, the target signal X can be expressed with the minimum error, while it is impossible by using Z which is not on the X - Y - W plane but closest to W . To avoid selecting Z (rather than Z') in the above-mentioned sequential selection of Y and Z , all candidates for Z are orthogonalized to Y , and the one, whose orthogonalized vector is the closest to the target vector X , is selected as the best candidate for Z . The orthogonalization of Z can be performed using the following equation [4].

$$Zo = Z - (Z^t Y / Y^t Y) \times Y \quad (5)$$

Therefore Eq. (4) can be rewritten as follows.

$$\frac{(W^t Zo)^2}{Zo^t Zo} = \frac{((W - (W^t Y / Y^t Y) Y)^t Z)^2}{Z^t Z - (Y^t Z)^2 / Y^t Y} \quad (6)$$

The vector which maximizes the term (6) is the optimum candidate for Z when ga and gf are jointly optimized. The optimum ga is given by the following equation.

$$\begin{aligned} ga &= \frac{(Z^t Z)(X^t Y) - (Z^t X)(Y^t Z)}{(Z^t Z)(Y^t Y) - (Z^t Y)(Y^t Z)} \\ &= \frac{X^t Y(Z^t Z - (Z^t X)(Y^t Z) / X^t Y)}{Y^t Y(Z^t Z - (Z^t Y)(Y^t Z) / Y^t Y)} \end{aligned} \quad (7)$$

2.2. Problem in the Orthogonalized FCB Search

The orthogonalized FCB search gives the best combination between Y and Z if the ideal ga and gf are given. However, there are some restrictions on ga . There are mainly two restrictions. Firstly, the quantized ga has the minimum value and the maximum value. The minimum value is usually zero, and the maximum value is typically

around 1.5, e.g. 1.2 in the case of AMR [3]. Therefore, if the optimum ga becomes out of this range, the orthogonalized FCB search could not give the best combination. Secondly, the ACB gain cannot be changed dramatically from the ACB gain calculated by Eq. (2). This is because the ACB gain is used as a “pitch gain” which is also used for pitch enhancement process for the FCB vector or reconstructed signal at the decoder side. Therefore, if the ACB gain values calculated by Eqs. (2) and (7) are different substantially, the orthogonalized FCB search could have undesirable effect on the codec performance.

3. SWITCHED ORTHOGONALIZATION

Conditional orthogonalized search is proposed in [6] in which the optimum pitch gain is used for switching between the two FCB search methods. However, there has been no equation analysis for the switching algorithm. In this section, switching criteria are derived from equation viewpoints.

3.1. Switching Criteria

As discussed as the first problem in Sect. 2.2, the quantized ga has the minimum value and the maximum value. In Eq. (6), the following term becomes zero if the ideal gain ga is used in Eq. (3). This can be confirmed using Eqs. (2) and (3) in Eq. (8).

$$d2 = W^t Y / Y^t Y \quad (8)$$

If the above term is not zero, this is because the ga was clipped and modified after the ACB search for adjusting the ga within the range of the ga , and such modified ga was used in Eq. (3) instead of using ga calculated by Eq. (2). For detecting such cases, threshold-based switching process can be introduced. That is, the orthogonalized FCB search is only used when $d2$ does not exceed a threshold. Segmental weighted signal-to-noise ratio (WSNRseg) defined in Eq. (13) is calculated for different threshold values between 0.0000001 and 1000 using 100 Japanese short sentences. Its result is shown in Fig. 2. Corresponding percentage of orthogonalized FCB search segments is shown in Fig. 3. Here the question is which threshold is appropriate for defining “ $d2 = 0$.” The value below 0.000001 is basically the error coming from computing precision. The range between 0.000001 and 0.01 is acceptable range for “ $d2$ is almost 0,” since there is no degradation in WSNRseg performance, and increase in the percentage shown in Fig. 3 is negligible. We choose 0.0001 as the threshold for $d2$.

As discussed as the second problem in Sect. 2.2, ga calculated in the orthogonalized search should not be far from ga calculated in the non-orthogonalized search. From the comparison between Eqs. (2) and (7), this can be interpreted as the following term being close to zero, which

Fig. 2 WSNRseg performance for $d2$ thresholds.

Fig. 3 Percentage of segments where $d2$ is under threshold.

makes Eq. (7) being approximated by Eq. (2) whichever FCB contribution is selected.

$$d1 = \sum_{i=1}^n (x_i / X^t Y - y_i / Y^t Y)^2 \quad (9)$$

If the above term becomes large, this could substantially decrease or increase the ACB gain, although its final influence depends on the FCB contribution. To reduce the possibility of such substantial change of the ACB gain, additional threshold is considered. That is, the orthogonalized FCB search is only used when $d1$ does not exceed a threshold. WSNRseg is calculated for different threshold values between 0.0000001 and 100000000 using the 100 Japanese short sentences. Its result is shown in Fig. 4. Corresponding percentage of orthogonalized FCB search

Fig. 4 WSNRseg performance for $d1$ thresholds.

Fig. 5 Percentage of segments where $d1$ is under threshold.

segments is shown in Fig. 5. In Fig. 4, 20.663388 dB is the WSNRseg for the case where the $d1$ criterion is not used. On the other hand, the $d2$ criterion is used, and its threshold is set to 0.0001. Similarly, in Fig. 5, 89.0325% is the percentage of orthogonalized FCB search segments for that case.

From Fig. 4, WSNRseg performance is saturated around 0.0001 to 0.0001 thresholds. We choose 0.0001 as the threshold for $d1$.

The above-mentioned two criteria, $d1$ and $d2$, are used for switching the orthogonalized FCB search and the non-orthogonalized FCB search. The flow of switching decision is shown in Fig. 6. In Fig. 6, $th1$ and $th2$ are the thresholds, and the both of them are set to 0.0001 as discussed in this section.

Fig. 6 Flow of orthogonalization switching decision.

3.2. FCB Search

The non-orthogonalized FCB search is performed by maximizing Eq. (4), which can be rewritten as the following equation.

$$(W^t H C)^2 / C^t H^t H C \quad (10)$$

where H is lower triangular matrix for convoluting the impulse response of a perceptually weighted synthesis filter, and C is a FCB vector. D and Φ are independent from C ; therefore they are calculated before searching the best C .

The orthogonalized FCB search is performed by maximizing Eq. (6), which can be rewritten as the following equation.

$$\begin{aligned} & \frac{((W - (W^t Y / Y^t Y) Y)^t H C)^2}{C^t H^t H C - (C^t H^t Y Y^t H C) / Y^t Y} \\ &= ((W - d2 \times Y)^t H C)^2 / C^t \left(H^t H - \frac{H^t Y Y^t H}{Y^t Y} \right) C \end{aligned} \quad (11)$$

The both Eqs. (10) and (11) can be written as the form of the following equation.

$$(D C)^2 / C^t \Phi C \quad (12)$$

Table 1 Objective test results.

	SNRseg [dB]	WSNRseg [dB]
Clean speech file (224 s)		
Non-orthogonalized	14.88	20.16
Orthogonalized	14.86	20.25
Switching	15.00	20.31
Speech with background noise (64 s)		
Non-orthogonalized	12.31	16.44
Orthogonalized	12.36	16.53
Switching	12.42	16.56

where $\mathbf{D} = \mathbf{W}^t \mathbf{H}$ and $\Phi = \mathbf{H}^t \mathbf{H}$ for the non-orthogonalized FCB search, and $\mathbf{D} = (\mathbf{W} - d2\mathbf{Y})^t \mathbf{H}$ and $\Phi = \mathbf{H}^t \mathbf{H} - \mathbf{H}^t \mathbf{Y} \mathbf{Y}^t \mathbf{H} / \mathbf{Y}^t \mathbf{Y}$ for the orthogonalized FCB search. Therefore, based on the decision of switching, the calculation $\mathbf{D}$ and Φ is switched as shown in Fig. 6. The process for maximizing Eq. (12) is common. Therefore the fast FCB search procedure used in ACELP can be used in the case of ACELP.

4. PERFORMANCE EVALUATION

The effectiveness of the presented switching method is demonstrated in this section. The switching mechanism has been tested in the ITU-T G.729.1 coder. Objective evaluation test has been performed using two speech files listed below.

- (1) clean speech (4 talkers $\times$ 14 short Japanese sentences, its duration is 224 s)
- (2) speech with a background noise (8 talkers $\times$ 2 short Japanese sentences, its duration is 64 s)

For objective measure, segmental SNR (SNRseg) and segmental WSNR (WSNRseg) are used. They are defined as follows.

$$\begin{aligned} \text{SNRseg} &= \frac{1}{N} \sum_{n=1}^N 10 \log \left(\frac{|S|^2}{|S - S'|^2} \right) \\ \text{WSNRseg} &= \frac{1}{N} \sum_{n=1}^N 10 \log \left(\frac{|S|^2}{|X - gaY - gfZ|^2} \right) \end{aligned} \quad (13)$$

where N is the total number of active segments, n is an active segment number, S is an input speech signal and S' is a synthesized speech signal at the 8 kbit/s layer of G.729.1.

Objective test results are shown in Table 1. For the switched orthogonalization, the percentages of the segments where non-orthogonalized FCB search are used are counted. They are 20.1% and 20.6% for the file (1) and (2) respectively. From these results, the presented switching process has effectively works to improve codec performance by around 0.1 dB segmental SNR. For the segments where the non-orthogonalized FCB search (FCBS) is performed, distribution of the cases $d1 > th1$, $d2 > th2$

Table 2 Distribution of each case of non-orthogonalized FCBS.

	segments	percentage
Clean speech file (224 s), active segments = 22,763		
a) $d1 > th1$ and $d2 \leq th2$	1,627	7.1%
b) $d2 > th2$ and $d1 \leq th1$	2,353	10.3%
c) $d1 > th1$ and $d2 > th2$	592	2.6%
a) + b) + c)	4,572	20.1%
Speech with background noise (64 s), active seg = 12,740		
a) $d1 > th1$ and $d2 \leq th2$	1,498	11.8%
b) $d2 > th2$ and $d1 \leq th1$	869	6.8%
c) $d1 > th1$ and $d2 > th2$	255	2.0%
a) + b) + c)	2,622	20.6%

has been examined. It is shown in Table 2. As shown in Table 2, the two criteria work separately.

5. DISCUSSION

As shown in the previous section, while the orthogonalized FCB search is off for more than 20% of active segments, the segmental SNRs are slightly improved by about 0.1 dB. Therefore the tested switching criteria practically work well for the presented switched orthogonalization method.

From Eq. (9) the criterion $d1$ can be understood as the similarity between the target vector and the synthesized adaptive codebook vector, and the criterion $d2$ means whether the adaptive codebook gain is within the range of the gain quantization. Therefore prediction gain or the like can be used as the $d1$ criterion, and whether the target vector $\mathbf{W}$ is calculated using non-ideal adaptive codebook gain can be used as the $d2$ criterion.

6. CONCLUDING SUMMARY

This paper presents the conditions where the orthogonalized FCB search is appropriate to be used. The criteria for switching between the orthogonalized FCB search and non-orthogonalized FCB search are derived from equation analysis on the FCB search algorithm in CELP coding. Objective evaluation test results show that the presented switching criteria successfully work and improve segmental SNR performance by about 0.1 dB.

REFERENCES

- [1] M. R. Schroeder and B. S. Atal, "Code-excited linear prediction (CELP): High-quality speech at very low bit rates," *Proc. IEEE ICASSP-85*, pp. 937–940 (1985).
- [2] J.-P. Adoul, P. Mabilieu, M. Delprat and S. Morissette, "Fast CELP coding based on algebraic codes," *Proc. IEEE ICASSP-87*, pp. 1957–1960 (1987).
- [3] 3GPP, "AMR Speech Codec; Transcoding functions (3G TS 26.090 version 3.1.0)" (1999).
- [4] M. Johnson and T. Taniguchi, "Pitch-orthogonal code-excited LPC," *Proc. IEEE GLOBECOM '90*, pp. 542–546 (1990).
- [5] T. Amada, M. Akamine and K. Maseki, "Fast excitation

- codebook search based on the orthogonalized target vector for CELP speech coding,” *IEICE Trans. Fundam.*, **J79-A**, 1092–1099 (1996).
- [6] A. Kataoka, S. Kurihara and S. Hayashi, “A 6.4-kbit/s extension to the G.729 speech coder,” *Proc. IEEE Workshop on Speech Coding for Telecommunications 1997*, pp. 69–70 (1997).
- [7] ITU-T, “G.729 based embedded variable bit-rate coder: An 8–32 kbit/s scalable wideband coder bitstream interoperable with G.729,” ITU-T Recommendation G.729.1 (2006).
- [8] J. Ikeda and A. Kataoka, “Low complexity CELP speech coder using orthogonalized search of algebraic code,” *Proc. IEICE-ISS Conf. 1995*, D-252, p. 255 (1995).
- [9] D. Massaloux, R. Trilling, C. Lamblin, S. Ragot, H. Ehara, M.-S. Lee, D.-Y. Kim and B. Bessette, “An 8–12 kbit/s embedded CELP coder interoperable with ITU-T G.729 coder: First stage of the new G.729.1 standard,” *Proc. IEEE ICASSP 2007*, pp. IV-1105–1108 (2007).

Hiroyuki Ehara received B.E. and M.E. degrees in electrical engineering and Ph.D. in applied physics and physico-informatics from Keio University in 1990, 1992 and 2008, respectively. Since joining Panasonic in 1992, he has been engaged in the research field of speech coding. He has participated in the international standardization of speech codecs since 1997. He served as the editor of ITU-T standards, G.729.1 Annex G and G.718 Annex C standardized in 2013. His recent involvement in the standardization is 3GPP EVS codec standardized in 2014. He received the Sato Prize Paper Award from ASJ in 2009. He is a member of ASJ, IEICE and IEEE.