Proc. Int. Conf. Commun. Technol. (Beijing, China), pp. 26.03.1-26.03.4, Sept. 1992

BACKWARD ADAPTATION FOR SINGLE-PULSE EXCITATION CODER

Qian Yasheng[1,2] Yan Ming Cheng[2] Peter Kabal[2]

[1]Department of Electronic Engineering
Tsinghua University
Beijing, China
100084

[2]INRS-Telecommunications
Université du Québec
16 Place du Commerce
Verdun, Quebec
Canada H3E 1H6

ABSTRACT

Backward-adaptive linear prediction has been successfully used in medium rate speech coders with high quality and low delay(less than 2 ms) at 16 kb/s. The prediction gain of a forward-adaptive format predictor cascaded with a backward-adaptive format predictor has been first studied. We have found that if the analysis frame length of the backward predictor is larger than the pitch period, the backward prediction gain can reach that of a non-linear predictor or a cascaded forward format predictor. Results for several speech segments of male and female speakers, with different analysis window lengths, have been given and compared. The proposed cascaded adaptive filter configuration, the first forward-adaptive synthesizer followed by second backward-adaptive synthesizer, has been incorporated into a 3 kb/s Single-Pulse Excitation/ Code-Excited Linear Prediction (SPE/CELP) coder to improve the speech quality while maintaining almost the same bit-rate. Experimental results for the proposed SPE/CELP coder with backward adaptation show that the improvements of the segment SNR for voiced speech segments of several testing sentences can reach to 1.02–2.06 dB.

1. INTRODUCTION

Backward-adaptive linear prediction has been successfully used in medium rate speech coders with high quality and low delay, such as 32 kb/s ADPCM (G.721) coder and 16 kb/s LD-CELP coders (delay less than 2 ms) [1], [2]. Backward-adaptive linear prediction is based on the reconstructed signal rather than the original speech signal. Since the reconstructed signal is available to both the coder and decoder, both can adapt the prediction coefficients and gain factors. No side-information need be transmitted. The less the quantization error of the residual signal of the backward-adaptive predictor, the better the quality of the reconstructed signal. The backward-adaptive predictor has not been applied in low bit-rate speech coders, since the quantization noise is too high.

It has been recently shown that a non-linear predictor, based on non-linear dynamic system concepts, can obtain an additional prediction gain of 3 dB after a 12-th order forward format predictor [3]. It means that the 12-th predictor is not able to remove all redundancies from the speech signal. The non-linear predictor does not require the transmission of the side-information, while a second forward format predictor does need a large amount of the side-information to be transmitted to the decoder. However, the non-linear predictor is much more complicated than a linear predictor.

Since the backward-adaptive predictor possesses, to some degree, a non-linear predictor nature, a rapidly updated backward-adaptive predictor can play the same role as a non-linear predictor. Therefore, a backward-adaptive predictor followed by a forward format predictor is investigated to replace the cascaded non-linear predictor with no extra side-information required to be transmitted. We have applied the backward adaptive approach in a Single-Pulse and Code-Exited Linear Prediction speech coding system (SPE/CELP) which has been recently presented for coding at 3-4 kb/s [4]. The proposed cascaded adaptive configuration, the first forward adaptive synthesizer followed by second backward-adaptive synthesizer, has been incorporated into the 3–4 kb/s SPE/CELP coder to improve the speech quality. We allocate a small overhead to turning off the backward-adaptive synthesizer for frames in which it has negative impact.

In the following, we will first describe the backward-adaptive predictor, the optimized analysis window and then the proposed 3 kb/s SPE/CELP speech coder with the cascaded backward-adaptive synthesizer.

2. CASCADED BACKWARD-ADAPTIVE PREDICTORS

The block diagram of a first forward-adaptive format predictor followed by second backward-adaptive format predictor is shown in Fig. 1. The forward- adaptive format predictor updates its prediction coefficients based on analysis of the input speech. A model for calculating the predictor coefficients for a transversal implementation is illustrated in Fig. 2. The input signal $x(n)$ is multiplied by a data window $w_d(n)$ to give $x_w(n)$. The signal $x_w(n)$ is predicted from a set of its previous samples to form an error signal,

$$e(n) = x_w(n) - \sum_{k=1}^{N_p} c_k x_w(n-k) \tag{1}$$

The final step is to multiply the error signal by a error window $w_e(n)$ to obtain a windowed error signal $e_w(n)$ where $e_w(n) = w_e(n)e(n)$. The mean square error is defined by,

$$\varepsilon^2 = \sum_{n=-\infty}^{\infty} e_w^2(n) \tag{2}$$

The coefficients c_k are computed by minimizing ε^2. This leads to a linear system of equations which can be written in matrix form,

$$\Phi c = \alpha$$

$$\begin{bmatrix} \phi(1,1) & \cdots & \phi(1,N_p) \\ \phi(2,1) & \cdots & \phi(2,N_p) \\ \vdots & \vdots & \vdots \\ \phi(NP,1) & \cdots & \phi(N_p,N_p) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_{N_p} \end{bmatrix} = \begin{bmatrix} \phi(0,1) \\ \phi(0,2) \\ \vdots \\ \phi(0,N_p) \end{bmatrix} \tag{3}$$

where

$$\phi(i,j) = \sum_{n=-\infty}^{\infty} w_e^2(n) x_w(n-i) x_w(n-j) \tag{4}$$

There are two well-known methods to solve the linear system of equations : autocorrelation method and covariance method. The autocorrelation method results if $w_e(n) = 1$ for all n. The data window $w_d(n)$ is typically time-limited (rectangular, Hamming or others). The

This research was supported by a grant from the Canadian Institute for Telecommunications Research under the NCE program of the Government of Canada

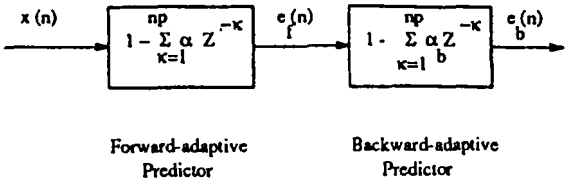

Fig. 1 Cascaded Backward-Adaptive Predictor

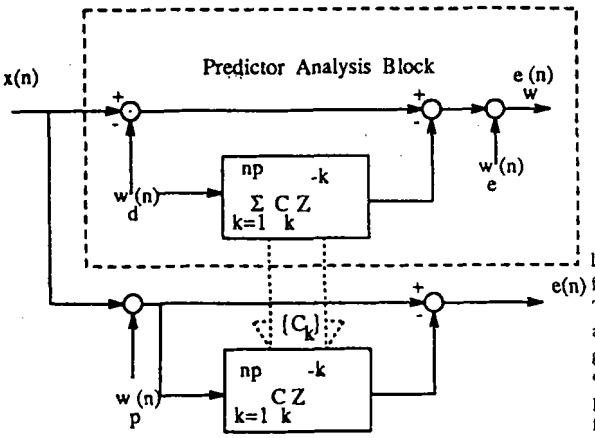

Fig. 2 Format Predictor

covariance method results if $w_d(n) = 1$ for all n and the error window is rectangular, $w_e(n) = 1$ for $0 \leq n \leq N-1$. The data window $w_d(n)$ of the forward-adaptive prediction coefficients analysis overlaps the predictor window $w_p(n)$. It means that the error signal samples are included in the data window. On the contrary, the data window $w_d(n)$ of the backward-adaptive prediction coefficients analysis does precede the predictor window $w_p(n)$, that is, the error signals are not included in the data analysis window. Therefore, the backward-adaptive predictor is not an optimum predictor in a sense of the Minimum Mean Square Error. There are a number of algorithms in the backward-adaptive predictor, such as the Least Mean Square (LMS) Method [5]. For LMS, the k-th prediction coefficients at $n+1$ instant, $c_k(n+1)$

$$c_k(n+1) = c_k(n) + \alpha_k(n) e_q(n) y(n-k); \qquad j = 1, 2, \cdots, N_p \qquad (5)$$

where

$c_k(n)$ —— k-th prediction coefficients at n instant

$e_q(n)$ —— quantized error signal

$y(n-k)$ —— reconstructed signal at $(n-k)$ instant

$\alpha_(n)$—— a normalized gradient coefficient which controls the rate of adaptation and stability.

For stationary signals in the steady state, the Mean Square Error may close to the minimum by using a small value of α or a function $\alpha(n)$ which decreases with time n. The backward adaptation algorithm used in our study is different from (5). The prediction coefficients are derived by autocorrelation and covariance algorithm with one sample lag of the predictor window from the data analysis window, as shown in Fig. 3. The prediction coefficients are updated sample-by-sample, that is, the backward-adaptive format predictor has a set of new prediction coefficients each time. This backward adaptation algorithm may employ different lengths for the data analysis and predictor window. In the LMS algorithm (5) the window lengths of them are not independent. Since the reconstructed signal is available to both the coder and decoder, no explicit transmission of prediction coefficients is needed. Therefore, the proposed cascaded backward-adaptive predictor does not require additional side-information to be transmitted while additional prediction gain obtained.

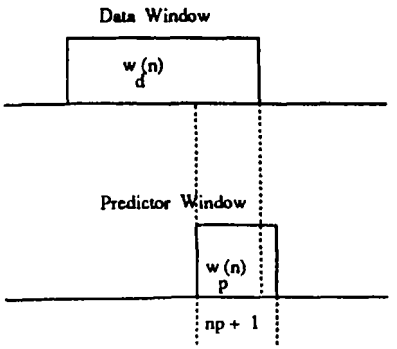

Fig. 3 Windows for Backward-adaptive Predictors

3. OPTIMIZED ANALYSIS WINDOW OF THE BACKWARD PREDICTORS

The first forward-adaptive predictor is a conventional 10-th order linear predictor using Durbin's recursive algorithm. The residuals of the first predictor is then fed to the second backward-adaptive predictor. The backward-adaptive predictor with one sample ahead of the data analysis window could not reach to an optimum forward prediction gain, unless the correlation matrix of the backward-adaptive predictor Φ in (4) is identical to the correlation matrix of the forward-adaptive predictor. Since there might be a big pulse in the residuals of the first forward format predictor. The difference between these two correlation matrixes could be quite large, as shown in Fig. 4. Therefore, the prediction gain of the backward predictor may fall down sharply. The larger the pulse component in the residuals, the less the prediction gain. It is not a serious problem for original speech with backward adaptation because of the slow variation for the speech characteristics in nature. For example, the backward prediction gain is 22.0 dB for a male speech segment with the data analysis window of 20 samples. However, there are many strong pulses, particularly, for male voiced segments. The backward prediction gain goes down to -10.7 dB for the residuals with the data analysis window of 12 samples.

As to a periodical signal, the correlation matrix is identical to the one-sample shifted version, if the data analysis window is larger than the period. Thus, the backward prediction gain can reach to the optimum forward prediction gain. Several speech segments of male and female speakers have been tested to calculate the backward prediction gains. The forward prediction gains are also counted for comparison. Both of the backward and forward prediction gains of different data analysis window length and predictor orders are listed in Table 1 and Table 2.

The pitch for the tested male voiced segments is 77 samples. The female's pitch is 36 samples. The backward prediction gain obtains as high as 6.35 dB, if the analysis window is 80 samples and the predictor is with 20th order. This backward prediction gain approximately approaches to the forward prediction gain of 8.28 dB for the same residuals and parameters. The backward prediction gain is 5.03 dB for 10th order backward predictor of 80-sample window length. The corresponding forward prediction gain is 5.99 dB. However, the backward prediction gains degrade rapidly, when the window length is shorter than the pitch period of 77 samples for male speech and 36 samples for female one. The corresponding residual waveforms of the first 10-th order forward predictor and second 20-th order backward predictor are shown in Fig. 4 to Fig. 5 for male speech. For these tests, the backward predictors use covariance algorithm to produce the prediction coefficients. With autocorrelation algorithm, as expected, the backward prediction gains are slightly lower than those listed in the tables.

4. A SPE/CELP CODER WITH BACKWARD ADAPTATION

The block diagram of the SPE/CELP coder with backward adaptation is given in Fig. 6. The reconstructed speech is synthesized

SECOND PREDICTOR GAIN (dB) CATM8 MALE				
			BACKWARD	FORWARD
Frame	Order	Framesize	PGain(dB)	PGain(dB)
5	20	80	6.350	8.281
12	10	110	5.542	5.313
9	10	100	5.013	5.013
8	10	90	4.585	5.657
6	10	80	5.033	5.989
4	10	70	3.23	5.929
5	10	60	2.76	6.635
5	10	50	1.666	7.238
4	10	40	0.391	7.865
4	10	30	-1.117	8.693
3	10	20	-3.902	9.526
3	10	12	-10.707	11.346
4	3	10	-2.769	4.436
3	1	5	-0.3383	2.221

Table 1 Prediction Gain of the Second Predictor for Male Speech. Each row shows the prediction gain for a voiced frame.

SECOND PREDICTOR GAIN (DB) FEMALE CATF8				
			BACKWARD	FORWARD
Frame	Order	Framesize	PGain (dB)	PGain (dB)
13	10	110	7.625	7.155
6	10	80	8.462	9.389
4	10	40	9.211	11.09
3	10	20	4.355	10.622
3	10	12	-0.636	12.75

Table 2 Prediction Gain of Second Predictor for Female Speech. Each row shows the prediction gain for a voiced frame.

by exciting a backward-adaptive format synthesizer cascaded with a forward-adaptive format synthesizer. Speech signal is subdivided into four subframes within the total coding frame of 200 samples. The excitation signal for the voiced subframes is modeled as a sequence of pulse, one pulse per pitch period. The location and amplitude of each pulse is determined to minimize a perceptually weighted error between the input and reconstructed speech using a dynamic programming strategy, as described in [4]. A Gaussian noise-like coded excitation is used for unvoiced subframes, as in a CELP coder. A fast pitch-detection algorithm is used to classify subframes into voiced or unvoiced ones. The speech signal is center-clipped to give a ternary output, +1, 0, −1. The autocorrelation function based on these ternary signal is used to determine the pitch period by peaks.

The forward-adaptive format parameters, Line Spectral Frequencies (LSF's) are analyzed and vector-quantized by 24 bits/frame. To keep the complexity within reasonable constraints, separate VQ codebooks with 12 bits each for the first 4 and the last 6 Line Spectral Frequencies are used. The spectral distortion of the LSF's quantization is less than 1 dB. Scalar quantization would require more than 30 bits with the same spectral distortion of VQ. For a pure unvoiced frame (all subframes are unvoiced), the excitation is obtained from an 8-bit codebook which is populated with zero-mean, unit-variance uncorrelated random Gaussian noise. 32 bits specify the codebook indices. The gain factor in each subframe is scalar-quantized with 4 bits. In a purely voiced frame, 34 bits encode all possible pulse locations. It assumes that the minimum pitch period is 2.5 ms or 20 samples. 10 bits specify the amplitude of the last pulse. The pulse amplitudes are linearly interpolated from the last pulse amplitudes between two consecutive frames. Therefore, only one pulse amplitude per frame is explicitly transmitted. For a transition frame from voiced to unvoiced or vice versa, assuming only one transition in a frame, 12 bits per subframe first are allocated to the unvoiced subframes. Then, the rest bits are used to the voiced subframes with appropriate interpolation. Total bits per frame of 200 samples at 3 kb/s are 75 bits.

The proposed SPE/CELP speech coder with the first backward-adaptive synthesizer followed by the forward-adaptive synthesizer has been tested by several male and female sentences. The segmental signal-to-noise ratio $SNR_{seg,bf}$ has been compared with SNR_{seg} of the original SPE/CELP coder without backward-adaptive synthesizer. The backward-adaptive format synthesizer has 20 orders with autocorrelation algorithm. The SNR_{seg} and $SNR_{seg,bf}$ are given in the Table 3. We have found that the improvements for the cascaded backward-adaptive synthesizer can reach to 1.02 – 2.60 dB for some voiced frames, such as 27–29 frames. However, for some voiced frames negative gains occur, as in 36–37, 41 frames. Since the single-pulse excitation model coarsely approximate the residuals of the cascaded format predictors, that is, the quantization noise is too high, it can not guarantee a positive gain in general. We have tried to turn off the backward-adaptive synthesizer, whenever the $SNR_{seg,bf}$ is lower than the SNR_{seg} by a decision block, which monitors the SNR_{seg} at the coder. Therefore, a decision bit per frame is required to send to the decoder. However, the additional bit-rate is only 50 bits/s. The proposed 3 kb/s SPE/CELP coder with cascaded format synthesizer operates at almost the same bit-rate.

Frame	SNR_{seg} (dB)	$SNR_{seg,bf}$
27	3.19	4.21
28	3.27	5.87
29	2.40	3.80
41	2.21	2.94
36	4.40	3.76
37	4.57	3.95

Table 3 Comparisons of SNR_{seg} and $SNR_{seg,cf}$

5. CONCLUSIONS

Since a forward-adaptive format predictor with reasonable side-information to be transmitted cannot remove the all redundancy of the speech signal, a second cascaded backward-adaptive predictor can obtain addition prediction gain with no extra side-information as a non-linear predictor. We have analyzed and shown that the analysis window of the backward-adaptive format predictor must be larger than the pitch period of the speech signal. The backward-adaptive format predictor used in this paper is updated based on sample-by-sample. The performance of the proposed backward-adaptive predictor is close to a non-linear predictor [3]. The proposed configuration can be applied to a low bit-rate speech coder to improve its quality. We have incorporated the backward- adaptive format synthesizer into a 3 kb/s SPE/CELP speech coder. It has shown that it can gain $SNR_{seg,bf}$ improvements of 1.02 to 2.60 dB for some voiced frames.

References

1. J. H. Chen, "A robust low-delay speech coder at 16 kbits/s", *IEEE Global Telecommun. Conf.* (Dallas, TX), Nov. 27-30, 1989, pp. 1237-1241
2. V. Iyengar and P.kabal,"A low delay 16 kbits/s speech coder", *IEEE Trans. Signal Processing*, Vol. 39, No. 5 May 1991, pp. 1049-1057
3. B. Townshend, "Non-linear prediction of speech", *Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Processing*, (Toronto, Ont.), May 1991, pp. 425-428
4. Wolfgang Granzow and Bishnu S. Atal, "High-quality digital speech at 4 kb/s", *Proc. IEEE Global Telecommun. Conf.* (San Diego, CA), Dec. 2-5, 1990, pp. 941-945
5. N.S.Jayant and Peter Noll, "Digital Coding of Waveforms", *Prentice-Hall, Inc.* (Englewood Cliff, NJ), 1984, pp. 303-306

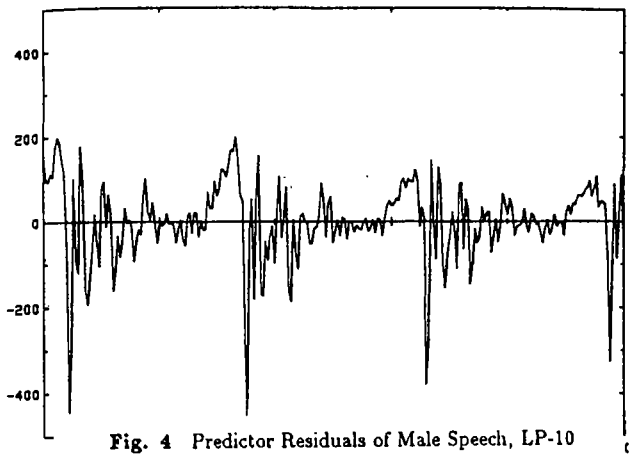

Fig. 4 Predictor Residuals of Male Speech, LP-10

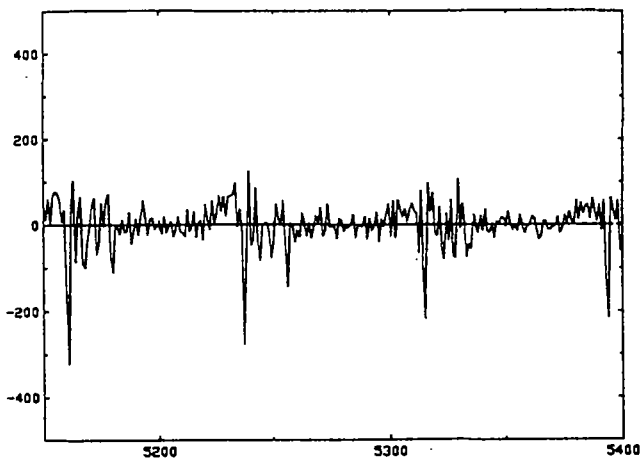

Fig. 5 Backward Predictor Residuals of Male Speech, $N - p = 20$, SNFRM=80

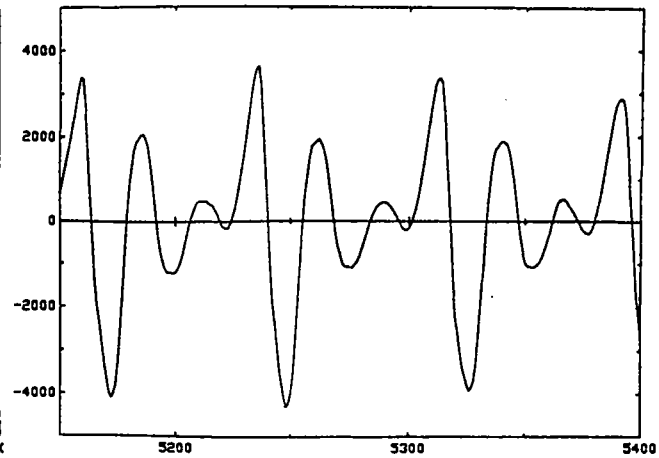

Fig. 7 Original Speech

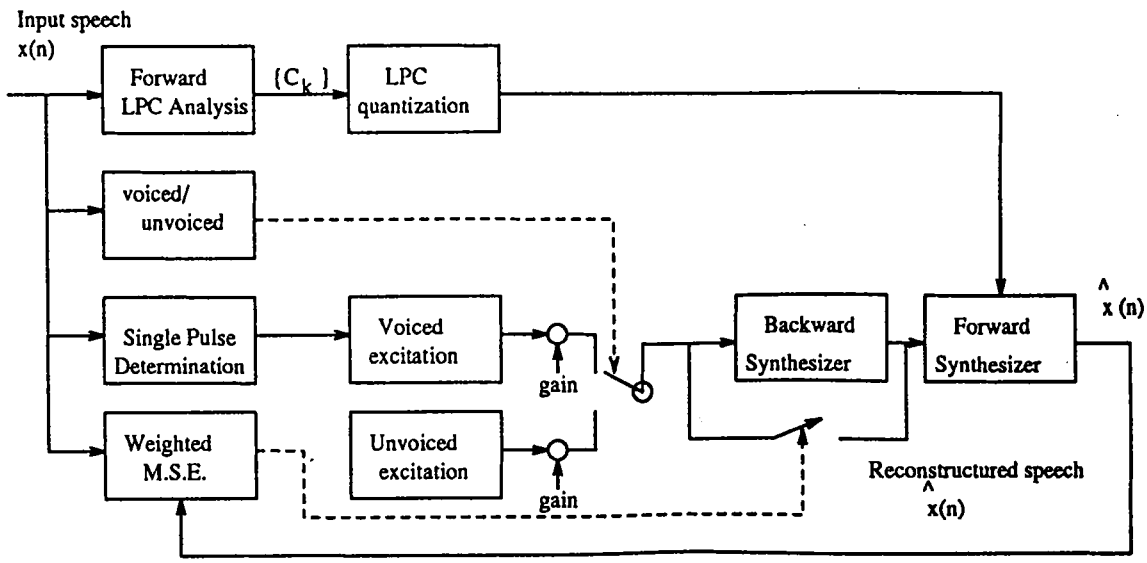

Fig. 6 The Block Diagram of Proposed SPE/CELP Coder