Presented at the 1980 International Conference
on
Acoustics, Speech, and Signal Processing
April 9-11, Denver, Colorado

EXPERIMENTAL COMPARISON OF FORWARD AND BACKWARD ADAPTIVE PREDICTION IN DPCM

Jerry D. Gibson

Telecommunication and Control Systems Laboratory
Department of Electrical Engineering
Texas A&M University
College Station, Texas 77843

Louis C. Sauter

25 rue Georges Doublet
06100 Nice, France

ABSTRACT

Results of experimental comparisons of forward- and backward-adaptive prediction in differential pulse code modulation (DPCM) of speech are presented. Two different types of comparisons were conducted. In one comparison, both predictors were used with the same three/five-level pitch compensating quantizer (PCQ). For this comparison, forward prediction clearly outperforms backward prediction, but with the penalty of a 10% increase in data rate due to the need to transmit coefficients. In the second comparison, the forward-prediction DPCM system and the backward-prediction DPCM system are constrained to have the same data rate of 16 kbits/sec. The backward-adaptive predictor outperforms forward prediction for this latter comparison. The speech data base for the simulations is one sentence spoken by a male speaker in four different languages, English, French, German, and Arabic. The performance comparisons are based on signal-to-quantization noise ratio, signal-to-prediction error ratio, sound spectrograms, and formal subjective listening tests.

I. INTRODUCTION

Adaptive differential pulse code modulation (ADPCM) with adaptive prediction is one of several techniques under consideration for speech coding at data rates of 8 to 16 kilobits/sec. (kbps) [1]. The adaptive predictor in ADPCM can be forward-adaptive (FA) or backward-adaptive (BA). The term forward-adaptive indicates that the predictor is adapted according to information derived from the system input which is then transmitted to the receiver. The term backward adaptive means that the predictor is adapted based on the quantized error signal which is available at both the transmitter and receiver.

Previous experimental investigations of ADPCM with forward-adaptive prediction [1-3] and backward-adaptive prediction [4-8] have not performed comparative studies of forward- and backward-adaptive predictors, although some analytical results are available in [9]. It is the purpose of this paper to present some results which partially fill this void. The data base for the system simulations is the sentence, "A glass of milk is better than a piece of bread," spoken by one male speaker in four different languages, English, French, German, and Arabic. The sentences were low pass filtered to 3200 Hz. (3dB) using a seven-pole elliptic filter, sampled 8000 times/sec., and digitized to 12 bits.

II. THE ADPCM SYSTEM AND PERFORMANCE INDICATORS

The ADPCM system transmitter and receiver are shown in Figs. 1 and 2, respectively, where Q denotes the quantizer and P denotes the predictor. From Fig. 1, the prediction error is given by

$$e(k) = s(k) - \hat{s}(k|k-1) \tag{1}$$

where $s(k)$ is the input and $\hat{s}(k|k-1)$ is the predicted value at time k that is of the form

$$\hat{s}(k|k-1) = \sum_{i=1}^{N} \hat{a}_i(k)\hat{s}(k-i) \tag{2}$$

In Eq. (2), the $\{\hat{s}(k-i),\ i = 1, \ldots, N\}$ are past output values given by

$$\hat{s}(k) = \hat{s}(k|k-1) + e_q(k), \tag{3}$$

and the $\{\hat{a}_i(k),\ i = 1, \ldots, N\}$ are called feedback tap gains or predictor coefficients.

Since $e_q(k)$ is a quantized version of $e(k)$, it can be expressed as

$$e_q(k) = e(k) + n_q(k) \tag{4}$$

where $n_q(k)$ denotes quantization noise. Substituting Eq. (4) into Eq. (3) and using Eq. (1) yields

$$\hat{s}(k) = s(k) + n_q(k) \tag{5}$$

The sequence $\{\hat{s}(k)\}$ is the receiver output as well as the feedback signal in both the transmitter and receiver. It is important to note that the quantization noise in Eqs. (4) and (5) need not be white nor independent or uncorrelated with $s(k)$.

The principal quantitative performance indicator used here is the SNR defined by

$$\text{SNR} = \frac{\langle s^2(k)\rangle}{\langle n_q^2(k)\rangle} \tag{6}$$

where $\langle\cdot\rangle$ denotes time averaging. For characterization of adaptive prediction, another useful indicator is the signal-to-prediction error ratio (SPER),

$$\text{SPER} = \frac{\langle s^2(k)\rangle}{\langle e^2(k)\rangle} \tag{7}$$

The SNR can be written in terms of the SPER as

$$\text{SNR} = \text{SPER}\ \frac{\langle e^2(k)\rangle}{\langle n_q^2(k)\rangle} \tag{8}$$

where the quantity in brackets is the reciprocal of the normalized quantization noise power. Sound spectrograms and subjective listening tests are also used for performance comparisons.

III. FORWARD ADAPTIVE PREDICTION SYSTEM

The forward adaptive predictors used in this work calculate the set of predictor coefficients $\{\hat{a}_i(k),\ i = 1, 2, \ldots N\}$ from a rectangularly-windowed frame of input speech samples using the autocorrelation method of linear prediction [10]. In particular, N = 8 predictor coefficients are computed by solving the set of linear simultaneous equations represented by

$$RA = C \tag{9}$$

where A is the column vector of predictor coefficients R is an 8 by 8 Toeplitz autocorrelation matrix that has the first row $[r_0\ r_1\ \ldots\ r_7]$ and $C^T = [r_1\ r_2\ \ldots\ r_8]$. The components of R and C are given by

$$r_i = \sum_{k=0}^{M-i} s(k)s(k+i) \tag{10}$$

where $s(k) = 0$ for $k < 0$ and $k > M$. The predictor coefficients are updated every 25 or 17.5 msec. based on nonoverlapping blocks of input data. The coefficients are not quantized and coded for these simulations, although this would be necessary in practice. Further discussion of this point is presented in Sec. V.

Two different quantizers were used with the forward adaptive prediction system. One quantizer is the 3/5-level pitch compensating quantizer (PCQ) [5, 6] shown schematically in Fig. 3 and described in more detail in Sec. IV. When the PCQ is used, the forward adaptive predictor is updated every 25 msec. The other quantizer is a three-level, forward adaptive quantizer with step size, Δ, computed according to [2]

$$\Delta^2 = \alpha[r_0 - C^T A] \tag{11}$$

where α was selected experimentally to be 0.75. This quantizer is shown in Fig. 4. When the three-level quantizer is used, both the predictor and the quantizer are updated every 17.5 msec.

IV. BACKWARD ADAPTIVE PREDICTION SYSTEM

The backward adaptation of the predictor coefficients is accomplished using the same Kalman identification algorithm described in [7, 8]. The predictor coefficients are updated at each time instant according to the difference equation

$$\hat{A}(k+1) = \hat{A}(k) + K(k+1)\, e_q(k+1) \tag{12}$$

where $A^T(k) = [\hat{a}_1(k)\ \hat{a}_2(k)\ \hat{a}_3(k)\ \hat{a}_4(k)]$ is a vector of predictor coefficients with zero initial conditions. The Kalman gain vector in Eq. (12) is calculated at each time instant from

$$K(k+1) = V_{\tilde{a}}(k+1|k)\hat{S}(k)[\hat{S}^T(k)V_{\tilde{a}}(k+1|k)\hat{S}(k)+V_v]^{-1} \tag{13}$$

where $\hat{S}^T(k) = [\hat{s}(k)\ \hat{s}(k-1)\ \hat{s}(k-2)\ \hat{s}(k-3)]$ and where the 4 by 4 matrix $V_{\tilde{a}}(k+1|k)$ is given by

$$V_{\tilde{a}}(k+1|k) = [I-K(k)\hat{S}^T(k-1)]V_{\tilde{a}}(k|k-1) + V_w \tag{14}$$

The scalar V_v in Eq. (13) is chosen by experiment to be 100, and the 4 by 4 matrix $V_w = \text{diag}(10^{-7})$. With $V_{\tilde{a}}(0|0) = \text{diag}(0.01)$, Eq. (13) is processed first, followed by Eq. (12), and then Eq. (14). The procedure is then repeated for the next input sample. It is important to note that the bracketed quantity in Eq. (13) is a scalar, and hence no matrix inversion is required.

It is evident from the given equations that only a fourth order backward predictor is employed. Higher order predictors have not been investigated.

The quantizer used with the backward adaptive prediction system is the 3/5-level PCQ shown in Fig. 3. The quantity X(k) in Fig. 3 is given by

$$X(k) = 2^{G(k)} \tag{15}$$

where

$$G(k) = GP(k) + C(k) + 1 \tag{16}$$

and the quantities GP(k) and C(k) evolve according to the difference equations

$$GP(k+1) = \frac{63}{64}GP(k) + f_1(\ell_k) \tag{17}$$

and

$$C(k+1) = \frac{3}{4}C(k) + f_2(\ell_k) \tag{18}$$

The additive functions $f_1(\ell_k)$ and $f_2(\ell_k)$ depend on the quantizer output level at time instant k, denoted ℓ_k, (see [8]). The minimum value of X(k) was selected to be 2. Additional constraints are also placed on the X(k) adaptation as described in [8], which are not repeated here due to space limitations.

V. PERFORMANCE COMPARISONS

Two separate experimental comparisons of forward adaptive (FA) and backward adaptive (BA) prediction were performed. For one experiment, forward adaptive prediction combined with PCQ was compared to backward adaptive prediction combined with PCQ. Table I summarizes the quantitative results of this investigation. It is clear from this table that forward prediction (N=8) provides a uniformly higher SNR than backward prediction (N=4). In Table I, ΔSNR = FA SNR(dB) - BA SNR(dB) and ΔSPER = FA SPER(dB) - BA SPER(dB). Forward prediction provides an improvement in average SNR over backward prediction of 1.77 dB, with a minimum increase of 1.33 dB for French and a maximum improvement of 2.05 dB for German. Formal subjective listening tests using several untrained listeners indicate an almost unanimous preference for forward adaptive prediction and hence substantiate the SNR results.

Note from Eq. (8) that if $\langle e^2(k)\rangle/\langle n_q^2(k)\rangle$ remains constant, which is the usual assumption, then any change in SNR is due to a change in SPER alone. The results in Table I tend to reinforce this interpretation, since in general, ΔSNR ≈ ΔSPER.

The "average entropies" in Table I, that is, the averages at the bottom of the H(E) columns, are actually the entropies of the quantizer output average distributions, rather that the average of the entropies for each language. More explicitly, the "average" H(E) is obtained by accumulating the relative frequencies of the five quantizer output levels for all four sentences combined and then H(E) is found from $-\sum_{i=1}^{5} p(i)\log_2 p(i)$, where p(i) denotes the probability of the ith quantizer level. It is noteworthy that forward prediction also provides a reduction in entropy over backward prediction.

Although for these experiments no coding of the information to be transmitted to the receiver was actually performed, for these results to have any practical implications, reasonable coding schemes must be devised. For backward prediction, only the quantized error signal needs to be coded, while with forward prediction, the eight predictor coefficients must also be quantized and coded.

From the average entropies in Table I, backward prediction with perfect entropy coding would require a data rate of approximately 14.1 kbps at an 8000 samples/sec. sampling rate. For this same sampling rate, forward prediction with perfect entropy coding would require 14 kbps to send the quantized error signal. To compare the required data rates of the two methods, it is necessary to estimate the number of bits needed for the predictor coefficient transmission in forward prediction. For transmission to the receiver, the predictor coefficients would first be transformed into reflection coefficients because of their desirable stability and sensitivity properties [10]. To send these eight reflection coefficients to the receiver will require about 40 bits/frame or a "side information" data rate of 1.6 kbps. This estimate is consistent with previous work [1]. An example allocation of bits to the predictor coefficients is shown in Table II. The minimum total data rate required for the forward prediction method is thus 15.6 kbps as compared to 14.1 kbps for backward adaptive prediction.

Of course, perfect entropy coding cannot be easily achieved, hence it is necessary to devise practically acceptable source codes for the quantized error signal. A source code that achieves an average data rate of 2.0056 bits/sample for the backward adaptive prediction probabilities is shown in Table III. Thus, for this code and an 8000 sample/sec. input, the required data rate for backward prediction is 16.045 kbps. The source code in Table III achieves an average rate of 1.9966 bits/sample for the forward prediction quantized error signal, and therefore, the required data rate for forward prediction is 15.973 + 1.6 = 17.573 kbps. Thus, although DPCM-FA outperforms DPCM-BA, its data rate is substantially higher.

In order to compare the two types of prediction in a more realistic communication system structure, the total (coefficients + residual) transmitted data rate of each of the two systems was limited to 16 kbps. This limitation necessitated a redesign of the DPCM-FA system. After conducting experiments with several forward and backward adaptive quantizers, the three-level, forward adaptive quantizer described in Sec. III was selected for use with the forward adaptive predictor. To determine the transmitted data rate of this new DPCM-FA system, 7 bits are allocated to Δ, 40 bits to the coefficients, and these parameters are updated every 17.5 msec. to yield a data rate of 2.686 kbps. The quantizer output coding is accomplished by coding three ternary symbols as five binary symbols to yield a data rate of 13.336 kbps. The total transmitted data rate of the new DPCM-FA system is thus 16.022 kbps. The performance of the new DPCM-FA system is summarized in Table IV.

The transmitted data rate of the DPCM-BA is unchanged (16.045 kbps), and so is its' performance (see Table I). The DPCM-BA system thus outperforms the DPCM-FA system in terms of average SNR by approximately 2.2 dB. The results of formal, side-by-side subjective listening tests (with earphones) are summarized in Table V. Although a clear preference for DPCM-BA is evident, the two systems are perceptually very close.

VI. CONCLUSIONS

An experimental comparison of forward and backward adaptive prediction in APDCM speech coding has revealed that if the additional data rate required for side information is ignored, forward prediction is preferred over backward prediction. However, if the data rate required for side information is included, then backward prediction outperforms forward prediction for equal data rates. An important implication of these results is that comparisons of speech coders should not ignore the data rate required for side information since this can completely reverse experimental conclusions. As a final point, it is noted that before a choice can be made between the two adaptation methods, bit error rate results for the two systems are needed; such studies are in progress.

REFERENCES

[1] Tribolet, Noll, McDermott, and Crochiere, B.S.T.J., pp. 699-712, March 1979.

[2] Noll, B.S.T.J., pp. 1597-1614, Nov. 1975.

[3] Noll, Polytechnisch Tijdschrift, No. 19, pp. 623-629, Den Haag 1972.

[4] Gibson, Jones, and Melsa, IEEE Trans. Commun., pp. 1789-1797, Nov. 1974.

[5] Cohn and Melsa, IEEE Trans. Commun., pp. 935-941, Sept. 1975.

[6] Qureshi and Forney, Conf. Rec., 1975 ICC, pp. 30-31 through 30-34.

[7] Gibson, IEEE Trans. Commun., pp. 145-150, Jan. 1978.

[8] Gibson, Berglund, and Sauter, "Kalman backward adaptive predictor coefficient identification in ADPCM with PCQ," accepted for publication, IEEE Trans. Commun.

[9] Gibson, Conf. Rec., 1978 NTC, pp. 19.2.1-19.2.5.

[10] Makhoul, Proc. IEEE, pp. 561-580, Apr. 1975.

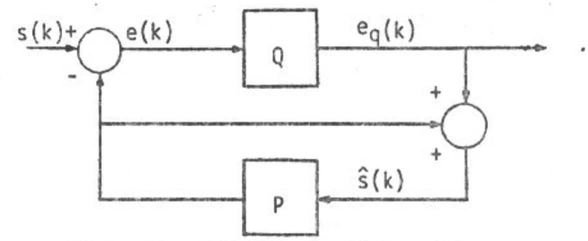

Figure 1. ADPCM System Transmitter

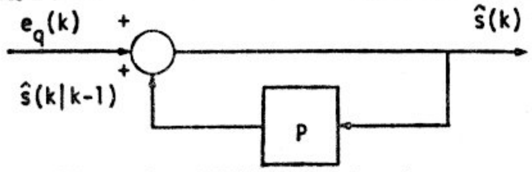

Figure 2. ADPCM System Receiver

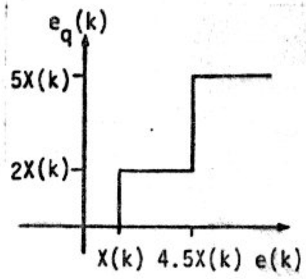

Fig. 3. 3/5 PCQ

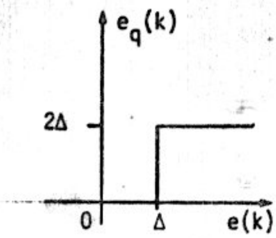

Fig. 4. 3-Level FA Quantizer

TABLE I. FORWARD AND BACKWARD ADAPTIVE PREDICTION WITH PCQ

Language	Forward SNR(dB)	Forward SPER(dB)	Forward H(E) (bits)	Backward SNR(dB)	Backward SPER(dB)	Backward H(E) (bits)	ΔSNR	ΔSPER
Arabic	17.94	9.33	1.745	16.29	7.71	1.775	1.65	1.62
English	16.31	7.70	1.750	14.51	6.13	1.787	1.80	1.57
French	16.23	7.69	1.704	14.90	6.31	1.747	1.33	1.38
German	19.90	11.43	1.704	17.85	9.65	1.747	2.05	1.78
Averages	17.87	9.32	1.750	16.10	7.69	1.762	1.77	1.63

TABLE II. ALLOCATION OF BITS TO REFLECTION COEFFICIENTS FOR FORWARD PREDICTION

Reflection Coefficients	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	Total
No. of Bits per frame	6	6	5	5	5	5	4	4	40

TABLE III. EXAMPLE SOURCE CODE FOR THE QUANTIZED ERROR SIGNAL

Message	Code Word	Message	Code Word
111	0000	4	1011
112	0001	5	1100
113	0010	14	1101
12	0011	15	1110
13	0100	24	1111 0000
21	0101	25	1111 0010
22	0110	34	1111 0100
23	0111	35	1111 1000
31	1000	114	1111 1001
32	1001	115	1111 1010
33	1010		

Note: 1 = 0 level, 2, 3 = inner levels, 4, 5 = outer levels

TABLE IV. DPCM-FA WITH A 3-LEVEL QUANTIZER

Sentence	A	E	F	G	Average
SNR(dB)	14.28	12.66	11.16	15.93	13.87

TABLE V. RESULTS OF LISTENING TESTS FOR KALMAN PREDICTOR WITH PCQ AND FORWARD PREDICTOR WITH 3-LEVEL QUANTIZER

Sentence	Number of listeners	Preferred backward	Preferred forward	Undecided
A	7	5(71%)	2(29%)	0
E	16	11(69%)	4(25%)	1(6%)
F	8	8(100%)	0	0
G	7	4(57%)	2(29%)	1(14%)
Total	38	28(74%)	8(21%)	2(5%)

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation under Grant No. ENG77-28527.