Signal Processing 5 (1983) 257–265
North Holland Publishing Company

MULTILINGUAL COMPARISON OF FORWARD AND BACKWARD ADAPTIVE PREDICTION IN DPCM

Louis C. SAUTER
DEO/EP, Laboratoires de Marcoussis, Route de Nozay, 91460 Marcoussis, France

Jerry D. GIBSON
Department of Electrical Engineering, Texas A and M University, College Station, TX 77843, USA

Received 26 August 1982
Revised 4 November 1982 and 15 December 1982

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 are conducted. In one comparison, both predictors are 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.

Zusammenfassung. Dargestellt werden Ergebnisse einer vergleichenden experimentellen Untersuchung adaptiver Prädiktionsverfahren für DPCM-codierte Sprachsignale für sendeseitig angeordneten Prädiktor ('Vorwärtsprädiktor') und empfängerseitig angeordeneten Prädiktor ('Rückwärtsprädiktor'). Hierbei werden zwei verschiedene Vergleiche durchgeführt. Im ersten Fall benutzen beide Prädiktoren den gleichen drei, bzw. fünfstufigen adaptiven Quantisierer. Dabei verhält sich der Vorwärtsprädiktor eindeutig besser, allerdings auf Kosten einer um 10% getsigerten Datenübertragungsrate, da die Prädiktorkoeffizienten mit übertragen werden müssen. Im zweiten Fall wird verlangt, daβ beide Systeme mit der gleichen Datenübertragungsrate von 16 kbit/s auskommen müssen. In diesem Fall liefert der Rückwärtsprädiktor die besseren Ergebnisse. Als Sprachmaterial dient ein Satz, der von einem männlichen Sprecher in 4 Sprachen (Englisch, Deutsch, Französisch, Arabisch) gesprochen wurde. Als Vergleichskriterien dienten: der Signal-Rauschabstand, das Verhältnis der Energie von Signal und Prädiktionsfehler, Spektrogramme sowie Hörversuche.

Résumé. Les résultats expérimentaux de comparaison de prédiction adaptative en avant et en arrière dans la modulation par impulsions codée différentielle (MICD) de la parole sont présentés. Dans une comparaison, les deux prédicteurs sont utilisés avec le même quantificateur à compensation de fondamentale à 3/5 niveau. Pour cette comparaison, la prédiction en avant est nettement supérieure à la prédiction en arrière, avec un prix payé qui est une augmentation de 10% dans la cadence des données due à la nécessité de transmettre des coefficients. Dans la deuxième comparaison, le système de prédiction en avant MICD et le système de prédiction en arrière MICD sont soumis à la contrainte d'avoir la même cadence de données à 16 bits/sec. Le prédicteur en arrière adaptatif est nettement supérieur au prédicteur en avant dans cette comparaison. La banque de données de parole pur les simulations est une phrase prononcée par un locuteur mâle en quatre langues: anglais, français, allemand et arabe. Les comparaisons de performances sont basées sur les rapports signal sur bruit de quantification, singal sur l'erreur de prédiction, les sonogrammes et les tests d'écoute subjective.

Keywords. Adaptive prediction, ADPCM.

1. 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 updated according to information derived from the system input, and then this information is 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.

Except for a recent paper by Xydeas and Evci [12], 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. Some analytical comparisons are available in [9]. It is the purpose of this paper to present some experimental results on FA and BA prediction in ADPCM which complement the work in [12]. The comparisons presented here differ from those of Xydeas and Evci [12] in that:

(i) The Kalman backward predictor, which is not investigated in [12], is used exclusively here,

(ii) four different languages are considered rather than English alone, and

(iii) in our second comparison, the bit rate allocated to the quantized error signal is reduced by an amount needed to transmit the forward-adaptive predictor coefficients, whereas in [12], Noll's [2] expression for the loss in signal-to-noise ratio is used instead.

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 (3 dB) using a seven-pole elliptic filter, sampled 8000 times/sec, and digitized to 12 bits.

2. 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

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

where $s(k)$ is the input and $\hat{s}(k \mid k-1)$ is the predicted value at time k given past outputs through time instant $k-1$ that is of the form

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

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

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

and the $\{\hat{a}_i(k), i = 1, \ldots, N\}$ are called feedback tap gains or predictor coefficients. The hat, '^', on the coefficients indicates that they must be computed in some as yet unspecified manner.

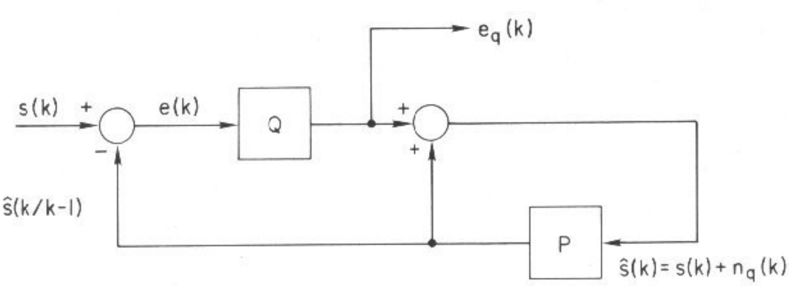

Fig. 1. ADPCM system transmitter.

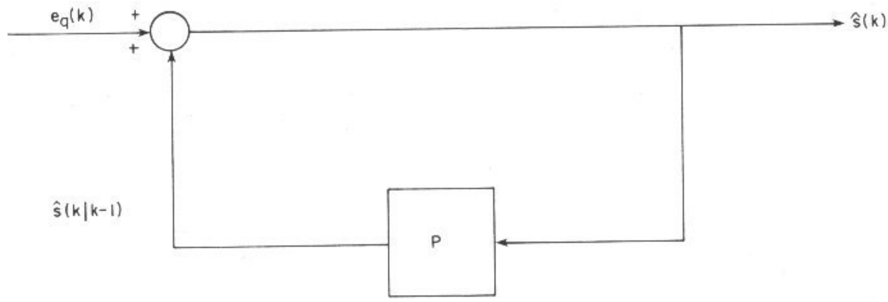

Fig. 2. ADPCM system receiver.

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, nor 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),

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

The SNR can be written in terms of the SPER as

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

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.

3. Forward adaptive prediction system

The forward adaptive predictors used in this work calculate the set of predictor coefficients $\{\hat{a}_i(k), i = 1, 2, \ldots 8\}$ from a rectangularly-windowed frame of input speech samples using the autocorrelation method of linear prediction [10]. The predictor coefficients are updated every 25 msec 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.

To compare the required data rates of the forward and backward adaptive systems, 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 be expressed as 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 for a 25 msec update rate. This estimate is consistent with previous work [1]. An example allocation of bits to the predictor coefficients is shown in Table 1.

Two different quantizers were used with the forward adaptive prediction system. One quantizer is the five-level pitch compensating quantizer (PCQ) [5, 6] shown schematically in Fig. 3 and described briefly in Section 4. When the PCQ is used, the forward adaptive predictor is updated every 25 msec. The other quantizer is a three-

Table 1

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

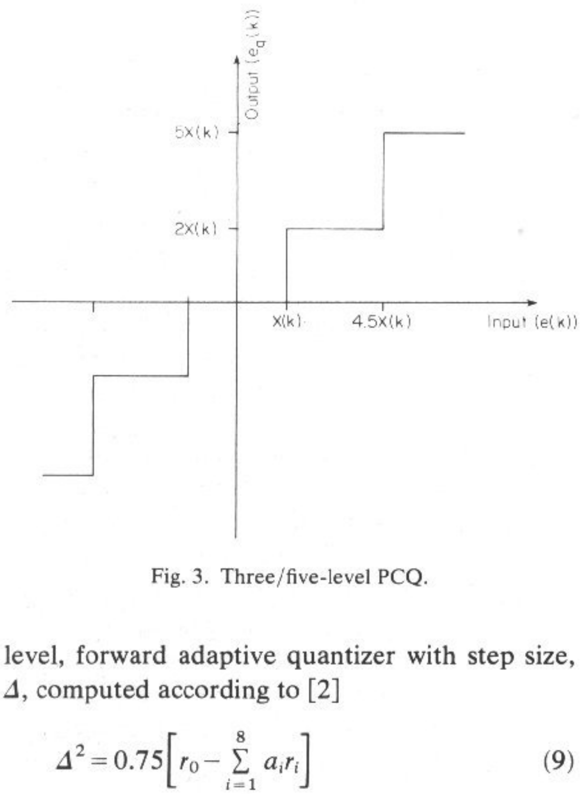

Fig. 3. Three/five-level PCQ.

level, forward adaptive quantizer with step size, Δ, computed according to [2]

$$\Delta^2 = 0.75\left[r_0 - \sum_{i=1}^{8} a_i r_i\right] \tag{9}$$

where

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

with $s(k)=0$ for $k<0$ and $k>M$. 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.

The selection of the predictor order, the quantizer, and coefficient and step size update rates are based on subjective experiments, and other forward-adaptive structures may be found which per-

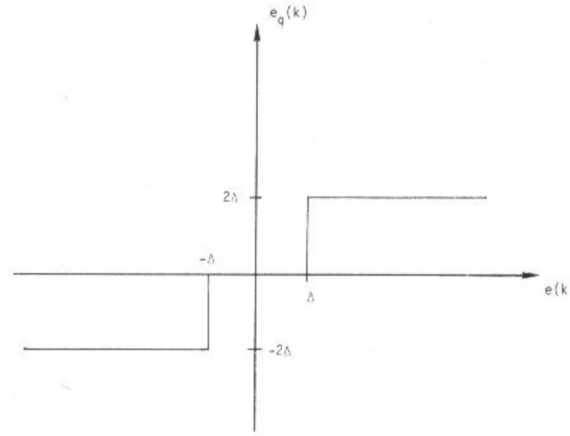

Fig. 4. Three-level forward adaptive quantizer.

form as well. We feel that the system designs used here are representative of the performance of such systems, however.

4. Backward adaptive prediction system

The backward adaptation of the predictor coefficients is accomplished using the same Kalman identification algorithm described in [7, 8]. The algorithms are not presented here due to space limitations. For the BA system, only a fourth order backward predictor is employed. Higher order predictors have not been investigated and may not yield a noticeable improvement. The quantizer used with the BA predictor is the five-level PCQ shown in Fig. 3. The step size adaptation algorithm for the PCQ is given in [8]. The minimum value of $X(k)$ is 2.0. Backward adaptive prediction with the PCQ is a completely backward-adaptive system which requires no side information to be transmitted to the receiver.

5. 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 2 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 2

$$\Delta\text{SNR} = \text{FA SNR(dB)} - \text{BA SNR(dB)}$$

and

$$\Delta\text{SPER} = \text{FA SPER(dB)} - \text{BA SPER(dB)}.$$

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 $\langle e^2(k)\rangle / n_q^2(k)\rangle$ remains constant, which is the usual assumption, then any change in SNR is due to a change in SPER alone. The results in Table 2 tend to reinforce this interpretation, since in general, $\Delta\text{SNR} \cong \Delta\text{SPER}$.

The 'average entropies' in Table 2, 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 2, 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. However, FA prediction requires that the predictor coefficients also be transmitted. Using the bit assignment in Table 1 and a 25 msec frame rate requires that 1.6 kbps be allocated to transmitting the coefficients. 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.

Table 2

Forward and backward adaptive prediction with PCQ

Language	Forward			Backward				
	SNR (dB)	SPER (dB)	$H(E)$ (bits)	SNR (dB)	SPER (dB)	$H(E)$ (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

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 3. Thus, for this code and an 8000 samples/sec. input, the required data rate for backward prediction is 16.045 kbps. The source code in Table 3 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.

Table 3
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.

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 Section 3 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 4.

Table 4
DPCM-FA with a 3-level quantizer

Sentence	Arabic	English	French	German	Average
SNR (dB)	14.28	12.66	11.16	15.93	13.87

The transmitted data rate of the DPCM-BA is unchanged (16.045 kbps), and so is its performance (see Table 2). 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 5.

Table 5
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
Arabic	7	5 (71%)	2 (29%)	0
English	16	11 (69%)	4 (25%)	1 (6%)
French	8	8 (100%)	0	0
German	7	4 (57%)	2 (29%)	1 (14%)
Total	38	28 (74%)	8 (21%)	2 (5%)

Listeners evaluated only those languages in which they were fluent. Although a clear preference for DPCM-BA is evident, the two systems are perceptually very close. Fig. 5 shows spectrograms of the sentence spoken in English. Fig. 5(a) is the spectrogram of the original sentence, Fig. 5(b) is the spectrogram of the output of the forward

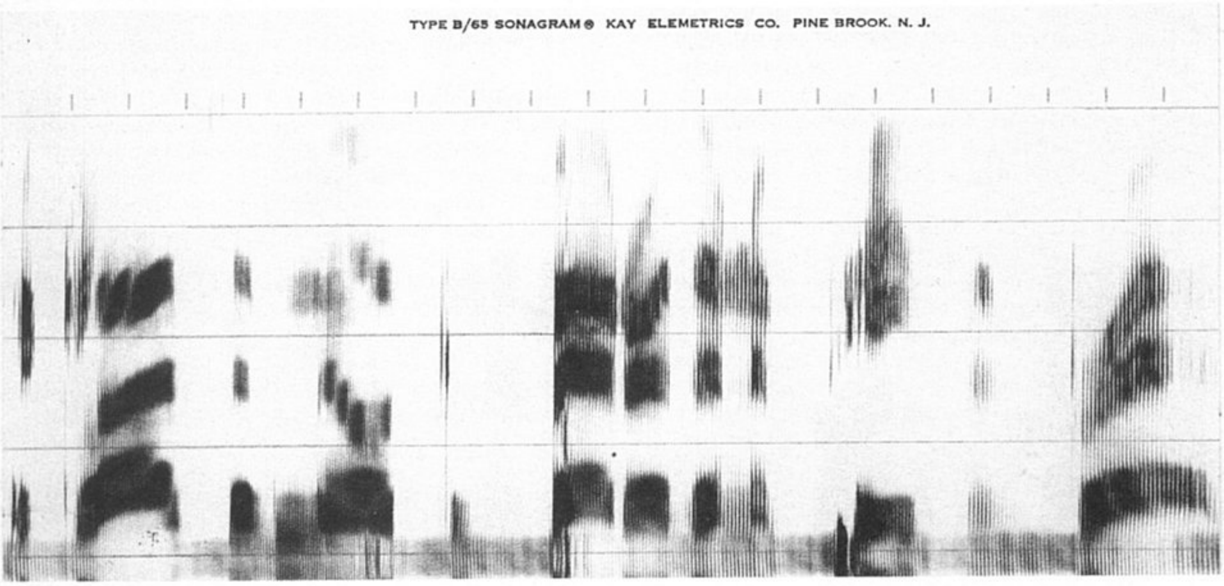

Fig. 5(a). Spectrogram of the original English sentence.

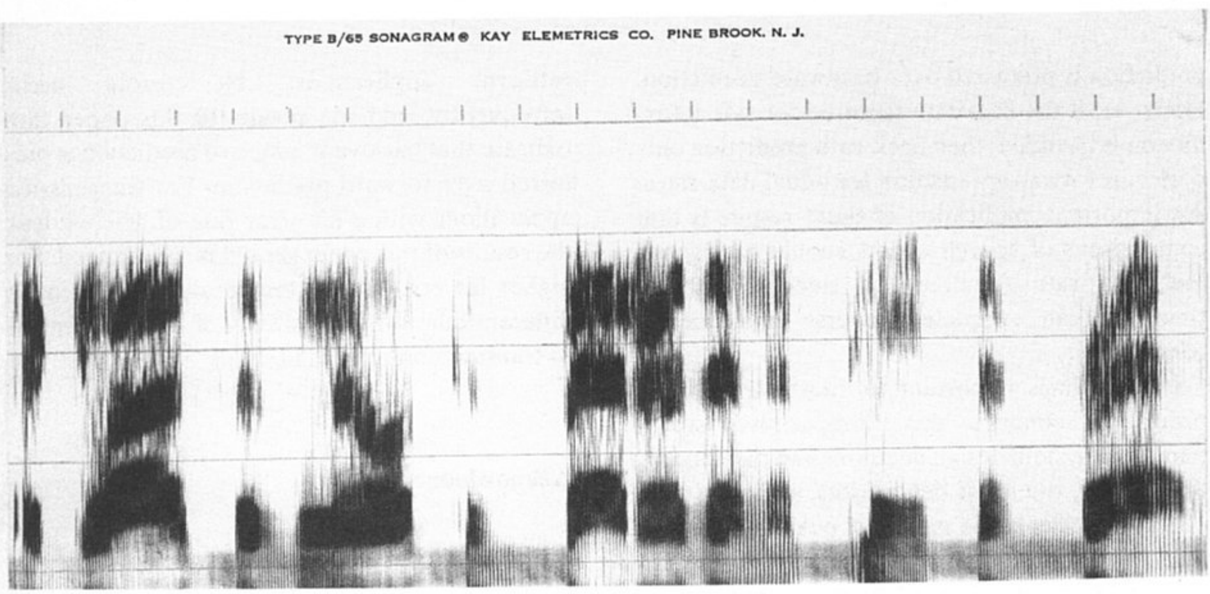

Fig. 5(b). Spectrogram of the forward adaptive system output (English).

adaptive system, and Fig. 5(c) is the spectrogram of the output of the backward adaptive system. Comparing Figs. 5(b) and 5(c), it is clear that the backward adaptive system output contains less interformant noise than does the forward adaptive system output.

6. Conclusions

An experimental comparison of forward and backward adaptive prediction in ADPCM speech coding has revealed that if the additional data rate required for side information is ignored, forward

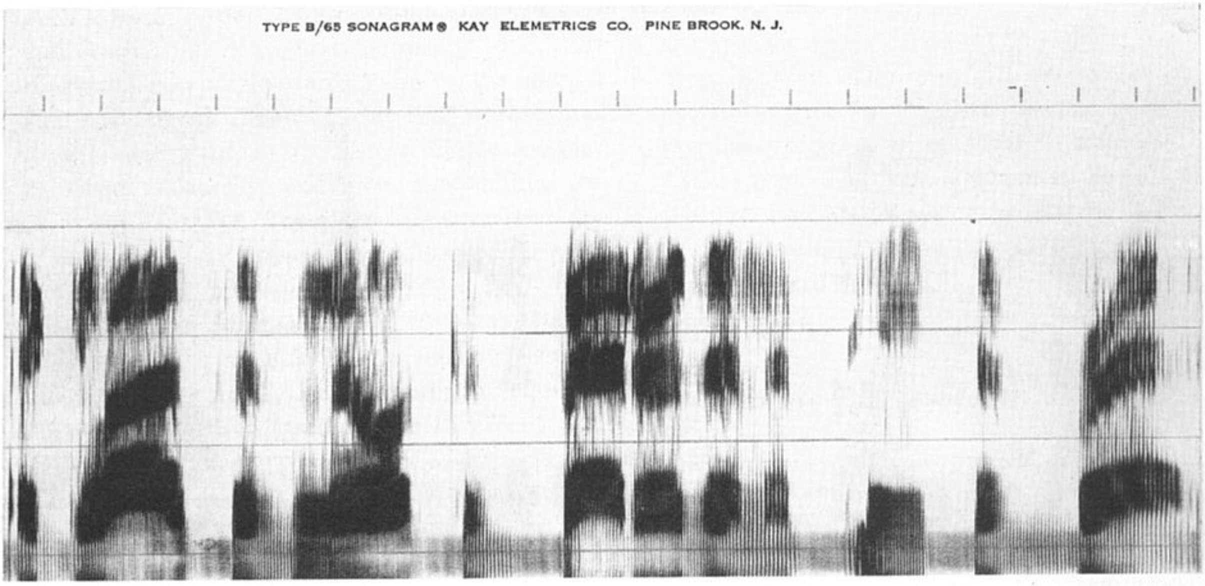

Fig. 5(c). Spectrogram of the backward adaptive system output (English).

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.

It is perhaps important to emphasize that in order to conduct this comparative study, numerous system design decisions had to be made. Specifically, the input bandwidths, sampling rate, coefficient calculation methods, quantizers, frame rates, and source code selections are representative of important system structures, but not all encompassing. Furthermore, only a single male speaker was used in the experiments. However, there is limited data comparing FA and BA systems in the literature, and this work yields interesting insights into the performance that can be expected.

The question of transmission errors has also not been addressed here. However, for storage and retrieval applications, bit errors occur infrequently, and the results of this paper thus indicate that backward adaptive prediction is preferred over forward prediction. For transmission applications with a bit error rate of 10^{-4} or less, the results of this paper should be unchanged. For higher bit error rates, the results may be quite different due to the sensitivity of the two systems to transmission errors [11].

Acknowledgement

This research was supported in part by the National Science Foundation under Grant Nos. ENG77-28527 and ECS-8020304.

References

[1] J.M. Tribolet, P. Noll, B.J. McDermott and R.E. Crochiere, "A comparison of the performance of four low-bit-rate speech waveform coders," *Bell Syst. Tech. J.*, Vol. 58, March 1979, pp. 699–712.

[2] P. Noll, "A comparative study of various quantization schemes for speech encoding," *Bell Syst. Tech. J.*, Vol. 54, November 1975, pp. 1597–1614.

[3] P. Noll, "Nonadaptive and adaptive differential pulse code modulation of speech signals," *Polytechnisch Tijdschrift*, No. 19, Den Haag, 1972, pp. 623–629.

[4] J.D. Gibson, S.K. Jones and J.L. Melsa, "Sequentially adaptive prediction and coding of speech signals," *IEEE Trans. Commun.*, Vol. COM-22, Nov. 1974, pp. 1789–1797.

[5] D.L. Cohn and J.L. Melsa, "The residual encoder-An improved ADPCM system for speech digitization," *IEEE Trans. Commun.*, Vol. COM-23, Sept. 1975, pp. 935–941.

[6] S.U.H. Qureshi and G.D. Forney, Jr., "A 9.6/16 Kb/s speech digitizer," *Conf. Rec., IEEE Int. Conf. on Commun.*, San Francisco, CA, June 1975, pp. 30-31 through 30-34.

[7] J.D. Gibson, "Sequentially adaptive backward prediction in ADPCM speech coders," *IEEE Trans. Commun.*, Vol. COM-26, Jan. 1978, pp. 145–150.

[8] J.D. Gibson, V.P. Berglund and L.C. Sauter, "Kalman backward adaptive predictor coefficient identification in ADPCM with PCQ," *IEEE Trans. Commun.*, Vol. COM-28, March 1980, pp. 361–371.

[9] J.D. Gibson, "Comparisons and analyses of forward and backward adaptive prediction in ADPCM," *Conf. Rec., Natl. Telecomm. Conf.*, Birmingham, AL, Dec. 1978, pp. 19.2.1–19.2.5.

[10] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, Vol. 63, April 1975, pp. 561–580.

[11] J.D. Gibson and R.C. Reininger, "Forward and Backward prediction in ADPCM over nonideal channels," *Conf. Rec., IEEE Int. Conf. on Commun.*, Seattle, WA, June 1980, pp. 42.5.1–42.5.5.

[12] C.S. Xydeas and C.C. Evci, "A comparative study of DPCM-AQF speech coders for bit rates of 16–32 kb/s," *Proc.* 1982 *Int. Conf. Acoust. Speech Signal Process.*, Paris, France, May 3–5, pp. 1680–1683.