Proceeding 15-th IEEE in ISRAEL TEL-AVIV April 1987. pp. 1.5.6-1 - 1.5.6-4
IEEE Israel 1987

A Switched Mode ADPCM Coder for Voice and Voiceband Data Transmission at 16 kbps[1]

A. Zinaty and D. Malah
Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel

ABSTRACT

The standard ADPCM coder at 32 kbps is known to provide high quality digitization of telephone bandwidth speech, as well as voiceband data (VBD) signals. In this paper we present a modified ADPCM coder at 16 kbps based on a switched mode operation for speech and modems up to a data rate of 2400 bps. The coder's idle mode is speech coding. The coder is then ADPCM combined with an adaptive postfilter (PF), where the degree of postfiltering is matched to the instantaneous performance of the ADPCM coder. In the presence of VBD signals, the coder is automatically switched to APCM (i.e. without a predictor). The switching between modes is based on a specially devised algorithm for Speech/VBD discrimination. In computer simulations the proposed coder was found to provide good communication quality with telephone speech input, and results in a lower bit error rate (BER) then with conventional ADPCM, when coding VBD signals at data rate of 2400 bps.

1. INTRODUCTION

The 32 kbps ADPCM coder standardized by CCITT [1] offers a 2:1 reduction in the bit rate, as compared to 64 kbps log PCM, and is realizable in real time with current technology. The characteristics of this coder are summarized in section 2. In this work, we propose an algorithm for further reduction of the transmission bit rate to 16 kbps, with a minimal additional complexity. The proposed algorithm is based on using a switched mode ADPCM coder for speech and VBD signals. For speech signals a postfiltering technique [2], shown to provide a simple means for enhancing the quality of 16 kbps coded speech, is used. The PF proposed in [2] is an adaptively scaled version of the coder's predictor. As a result of the adaptation, the segments that are heavily filtered are those which have a large amount of quantization distortions. Noise spectral shaping techniques, including postfiltering are briefly discussed in section 3. Unfortunately, noise spectral shaping was found to be inappropriate for VBD signals. Furthermore, we found that for these signals a PCM coder with an appropriate quantizer, is superior over conventional ADPCM at the same bit rate. Therefore, the proposed coder automatically switches to APCM when coding VBD signals. In section 4 we describe an algorithm for Speech/VBD discrimination. A full description of the proposed coder is given in section 5, and section 6 summarizes the simulation results. Section 7 concludes the paper.

2. STANDARD 32 Kbps ADPCM CODER

Since the proposed coder is based on the standard 32 kbps ADPCM coder [1], we briefly describe here this coder and its major attributes. The standard algorithm provides digital transcoding from 64 kbps PCM (either for $A = 87.6$ or $\mu = 255$ law), to 32 kbps (8 k Hz sampling frequency and 4-bit per sample) and vice versa. In order to minimize transmission delay, and avoiding the need for side information, all adaptation processes are of the backward type. The algorithm is able to recover from digital transmission impairments and also includes a technique for eliminating cumulative distortions in synchronous tandem codings. Last but not least, adequate performance is maintained not only for speech but also for other signals such as VBD, up to a data rate of 4800 bps. A block diagram of the coder is shown in figure 1. (Not included in the diagram are the log-PCM / linear PCM converters and also not the synchronous adjustment unit at the decoder output).

Adaptive Quantizer: A non-uniform sixteen level quantizer is used. To accommodate, on one hand, a fast tracking capability for signals with large fluctuations, such as speech signals, and on the other hand, slow tracking of stationary signals, such as VBD signals, the coder applies a Bi-modal adaptive quantization scheme [3]. In this scheme the adaptation is realized by forming a linear combination of fast and slow scale factors, according to a weighting parameter which controls the adaptation speed. Thus the adaptation speed control is varied smoothly according to the ratio of the short-time to the long-time averages of the quantizer code words, avoiding mistracking problems between encoder and decoder following transmission errors.

Adaptive Predictor: The adaptive predictor is composed of a sixth order all-zero filter and a second order all-pole filter. The selection of this structure is the result of an investigation of several design factors: performance for narrowband and broadband signals, stability, tracking ability with transmission errors, and computational efficiency. A simplified gradient algorithm is used to update both sets of predictor coefficients. The problem of stability of the all-pole filter is overcome by employing a second order section only. Mistracking of predictor coefficients is solved by adapting the coefficients of an equivalent all-zero filter [4].

3. NOISE SHAPING

In an adaptive coder with backword adaptation, implying that no side information needs to be transmitted to the decoder, the quantizer is located inside the prediction loop, as shown in figure 1. In such a coder, if an optimal prediction filter (optimal in the minimum MSE sense), and sufficiently fine quantization are used, the resulting reconstruction error is white. However, human perception does not rank speech quality solely on the basis of MSE. The spectral shape of the noise, relative to the spectral shape of the speech signal, also comes into play. At medium bit rates, such as 16 kbps, flat noise masks the signal at high frequencies, and is perceived as a hissing sound. Shaping the noise spectrum to resemble the envelope of the speech spectrum itself cures the hissing problems, but unfortunately, on the expense of dramatically reducing the SNR at low frequencies, causing a certain amount of "roughness" and "rumble" [5]. An intermediate shaping, between the extreme cases of a flat and a speech shaped noise spectrum, can reduce the hissing effect without a significant decrease in speech quality and SNR at low frequencies. A block diagram of a system which adds noise spectral shaping to the conventional ADPCM coder, is shown in figure 2. In this system the quantization error is filtered and fed-back to the encoder's quantizer input. The reproduced signal $y(t)$ is given by the sum of the input signal $x(t)$ and the shaped quantization error, which in the z-transform domain can be written as:

[1]This work was supported by Tadiran - Communication Division (050-519).

$$Y(z) = X(z) + Q(z) \cdot F(z) \qquad (1)$$

where $Q(z)$ and $F(z)$ are the transform domain descriptions of the quantization error $q(z)$ and the noise feedback (NF) filter impulse response, respectively. For a pole-zero filter such as the one shown in figure 2, $F(z)$ is given by:

$$F(z) = [1 + B_F(z)]/[1 - A_F(z)] \qquad (2)$$

For sufficiently fine quantization, the reconstruction error has the spectral shape of $F(z)$.

Another method for enhancing the perceptual quality of encoded speech using noise spectral shaping is proposed in [2]. In this system an adaptive reconstruction filter (or "post-filter") $\mathrm{PF}(z)$ is added at the decoder output (see figure 3). While shaping the quantization error, the PF also distorts the speech signal itself. The reconstructed signal $Y_p(z)$ is given here by:

$$Y_P(z) = [X(z) + Q(z)] \cdot \mathrm{PF}(z) \qquad (3)$$

As seen from (3), frequencies that were masked by the quantization noise in a conventional ADPCM coder are masked here as well. Enhancement of speech quality is achieved by an overall reduction of output energy at those frequencies.

The above two techniques for noise spectral shaping were compared by informal listening tests. The ADPCM-PF speech quality was judged always to be perceptually superior to ADPCM-NF speech and to provide a significant improvement to the conventional ADPCM coder at 16 kbps. A detailed description of the PF used in our coder is given in section 5.

4. SPEECH / VBD DISCRIMINATION ALGORITHM

ADPCM coders with enhancement algorithms such as noise feedback or postfiltering discussed above, are adequate for coding speech signals, but none of them provides an acceptable output bit error rate (BER) for VBD signals. Furthermore, we obtained in our simulations a better output BER when coding VBD[2] signals at 16 or 32 kbps with a PCM coder using a fixed quantizer matched to the input signal, than with an ADPCM coder. These observations led us to the idea of using a dual mode coder which automatically switches between the two modes of speech coding and VBD coding, according to the type of the input signal. The proposed coder would switch therefore between an ADPCM-PF coder in the presence of speech signals and an APCM coder, in the presence of VBD signals. Switching between these operating modes must rely on some discrimination algorithm. We propose next an algorithm for the identification of VBD signals, suitable for real time implementation. It is a modified version of the algorithm in [6], based on zero-crossing and average signal magnitude.

Zero Crossing Count: For a Gaussian stationary ergodic process $\{X_t\}$ the mean number of zero-crossing per sampling interval is given by [7]:

$$\frac{E[N_L]}{L-1} = \frac{1}{\pi}\cos^{-1}\rho_1 = \frac{1}{\pi}\cos^{-1}\left[\int_o^\pi \cos\theta dF(\theta) / \int_o^\pi dF(\theta)\right] \qquad (4)$$

where N_L is the zero-crossing count in a block of L samples, ρ_1 is the correlation coefficient between two adjacent samples, and $F(\theta)$, $-\pi \le \theta \le \pi$, is the spectral distribution function of $\{X_t\}$. Thus, the number of zero-crossings is a weighted average of $dF(\theta)$.

[2]We used a PSK modem simulator for a data bit rate of 2400 bps. The output signal is a 1.8 kHz sinc-wave at one of four possible angles, multiplied by a smoothing tapered window without any further band limiting [8].

The energy spectrum of typical VBD signals is centered around 1700-1800 Hz - frequencies that are significantly higher than the frequencies of the first formant of voiced signals. This is also reflected in the zero-crossing rate of the signals. Discrimination of these VBD signals from speech is a relatively easy task, when the zero-crossing parameter is used. The zero-crossing rate of VBD signals is in the range spanned by the two tones used in FSK modems: 1.2 kHz and 2.4 kHz. Therefore, we use the values of zero-crossing rates corresponding to these frequencies as decision levels (denoted by Z_{low} and Z_{high}). A discrimination algorithm to be described later which relies only upon a zero-crossing count, was found to give good results for VBD signals. However, it suffers from some misclassifications when the input signal is speech, especially at the beginning and ending of an utterance. We did not get a significant improvement by adding a max-to-mean signal magnitude ratio estimator nor by an extremum count as suggested in [6]. Better results were obtained by adding measurements which relate to the non-stationarity characteristic of speech signals as explained below:

Adaptive Magnitude and Zero Crossing Thresholding: An obvious feature to be used for Speech/VBD discrimination is the stationarity of VBD signals as compared to syllabic-rate changes of the speech magnitude. A tail of an utterance is characterized also by a decrease of the signal magnitude. This could be easily detected by an adaptive threshold. The adaptive threshold can be based on a mean magnitude estimator with a fast rising time-constant and a slow decay time-constant. But, because of the higher max-to-mean magnitude ratio for most speech signals than for VBD signals, we use the maximum value rather than the mean magnitude value as the basis for setting of the adaptive threshold. A signal with a lower magnitude than the threshold is identified as speech. By freezing the last threshold value when entering into an idle channel state, the beginning of the next utterance is properly identified. The same principle is used for adaptive thresholding of the zero-crossing count. Here, the extremum count is used to establish the adaptive threshold.

5. DETAILS OF PROPOSED CODING SYSTEM

As explained earlier, our proposed coder is based on the standard 32 kbps ADPCM coder [1] with the postfilter proposed in [2]. We present here our modifications to these algorithms. The general form of adaptation of the coder's i-th parameter p_i is as follows:

$$p_i(t) = (1-\gamma_i)p_i(t-1) + \gamma_i \cdot \delta_i(t) \qquad (5)$$

where $\delta_i(t)$ is its correction term and γ_i is the recursion coefficient which determines the time constant.

Discriminator: The zero-crossing count - $N(t)$ is evaluated recursively as in (5) with a correction term - $Z(t)$ which is an indicator of a zero-crossing occurrence in the time interval between two successive output samples.

$$Z(t) = \begin{cases} 1, & sign[Y(t) \ne sign[Y(t-1)] \\ 0, & \text{else} \end{cases}$$

A scaled extremum count - $N^1(t)$, with a slower time constant, is used as the adaptive zero-crossing threshold.

The adaptive average magnitude threshold is given by the scaled maximum magnitude estimator - $M(t)$, which is evaluated recursively as in (5), except for having two time constants: a fast rising time constant, and a slow decaying time constant:

$$M(t)=\begin{cases} M(t-1)\,,\ Ave(t) \le \text{Silence threshold} \\ (1-\gamma_{n1})M(t-1)+\gamma_{n2}\{\max[M(t-1),Y(t)]-M(t-1)\}\,,\ \text{else}\end{cases}$$

where $Ave(t)$ is the average magnitude estimator. As one can see, $M(t)$ is frozen when the channel is observed to be in idle state.

A discrimination parameter $D(t)$ is used and takes the value 0 if the input is classified as VBD and the value 1 if it is classified as speech. A VBD decision is made only if the following three conditions are met:

(a) $Z_{\text{low}} \le N(t) \le Z_{\text{high}}$; (b) $N(t) \le N^1(t)/f_N$

(c) $Ave(t) \le M(t)/f_M$

where f_N and f_M are scaling factors.

Quantizer: The quantizer has a variable adaptation speed, similar to the one used in [1], but instead of a linear combination between two quantization scale factors we use a linear combination between two time constants, through its recursion coefficient $\gamma_Q(t)$:

$$\gamma_Q(t)=W_q(t)\gamma_u+[1-W_q(t)]\gamma_l\ ,\quad 0<\gamma_l \ll \gamma_u \ll 1$$

where $W_q(t)$ is the adaptation speed controler as defined in [1]. In the presence of stationary signals $W_q(t)=0$ and $\gamma_Q(t)=\gamma_l$, causing the quantizer to be in an almost locked state (slow adaptation), while in the presence of nonstationary signals $W_q(t)=1$ and $\gamma_Q(t)=\gamma_u$, resulting in fast adaptation.

Predictor: The predictor is composed of two poles and six zeroes with a simplified gradient adaptation algorithm as in [1]. It has two operation modes one is the active state, and the other is a degenerated mode in which all predictor coefficients are forced to zero. A smooth transition between operation modes is achieved simply by multiplying the correction terms in the adaptation scheme by the discrimination parameter $D(t)$.

Postfiltering: The desired transfer function of the PF is a flattened version of the speech short time spectrum. The PF suggested in [2] is a pole-zero filter similar to the reconstruction filter but with scaled coefficients. At the bit rate of 16 kbps it is proposed in [2] to have a variable degree of postfiltering. As a result of it, the segments that are heavily filtered are those which are reproduced very noisily by the ADPCM coder. An unconvenient side effect of the PF is that its gain is not unity. This can be compensated fairly effectively for speech signals by a fixed factor of about 1/3. However, in the switched coder, the PF becomes transparent in the presence of VBD signals, so that no gain compensation is needed.

A smooth transition of the gain compensation factor, from one state to another, with a minimal effect on the coder gain because of discrimination errors, is realized by low-pass filtering of the discrimination parameter with delay.

6. SIMULATION RESULTS

This section provides a summary of the proposed coder performance in coding speech and VBD signals in both a noisy and a free of error channel. For speech input we used three English sentences, spoken each by a male and a female, which were recorded from a high quality microphone and lowpass filtered at 3.4 kHz. For VBD input we used a data stream of 153,600 bits, modulated by a QPSK modem simulator[2].

The discrimination algorithm between speech and VBD signals was tested by coding the above signals and also by coding the two tones - 1.2 kHz and 2.4 kHz. With VBD signals no sample was misclassified, even when the signal was corrupted by additive white Gaussian noise at SNR values of down to 10 dB and channel BER of 10^{-3}. With the above two tones and the same transmission conditions, only a few samples were misclassified. For speech signals with additive white Gaussian noise at SNR values of down to 20 dB and channel BER of 10^{-3}, there were only unsignificant misclassifications, at the beginning and ending of utterances, which could not be perceptually noticeable.

When coding speech signals, the coder operates as an ADPCM-PF coder. Such a coder was reported in [2] to provide good communication quality with most telephone speech inputs. We also examined the coder's performance under noisy channel conditions. Because of the PF, input noise and channel errors are less detrimental to speech quality than other coders, such as ADPCM and PCM. This effect of input noise is also reflected in the SNR. Figure 4 compares the output segmental SNR of the proposed coder to that of an ADPCM coder at 16 kbps, as a function of input SNR. Because of the PF used in our coder we measure the SNR between the reconstructed signal to the filtered input signal by the same PF. However, this quality enhancement and robustness to channel impairment is obtained at the expense of a band limitation effect, which becomes unacceptable when several coders are connected in tandem.

With VBD signals, we measured the output BER as a function of the channel BER and the level of additive noise. The coder performance is compared against the standard 32 kbps ADPCM coder [1] and a 16 kbps ADPCM coder similar to our coder, but with an active predictor operating at the same conditions. The resulting output BER with the proposed coder was found to be lower at high SNR values. For lower SNR values there is a cross-over of the results, as seen in figure 5.

7. CONCLUSION

In this paper we introduced a switched mode 16 kbps ADPCM coder for speech and VBD signals.

With VBD signals the coder operates as an APCM coder. It and was found to be less sensitive to channel errors but more so to input noise as compared to the other ADPCM coders examined. The switching between the adaptive predictor, for speech signals, and no prediction for VBD signals, is realized by means of a Speech/VBD discrimination algorithm based on zero-crossing and average signal magnitude. In the presence of speech signals the coder operates as an ADPCM coder with a postfilter (PF). This coder provides good communication quality. Nevertheless, the band limiting effect and the non-unity gain of the PF becomes critical when several coders are connected in tandem.

The proposed coder is recommended, therefore, for coding speech and VBD signals for transmission at medium bit rates in applications where good communication speech quality is sufficient, and no tandeming is necessary.

REFERENCES:

[1] W.R. Daumer, X. Maitra, P. Mermelstein, I. Tokizawa, "Overview of the ADPCM Coding Algorithm", IEEE GLOBECOM '84 Conf. Proc., pp. 774-777, Nov. 1984.

[2] N.S. Jayant and V. Ramamoorthy, "Adaptive Postfiltering of 16 kb/s - ADPCM Speech", Proc. ICASSP, pp. 829-832, April 1986.

[3] D.W. Petr, "32 kb/s ADPCM-DLQ Coding for Network Applications", IEEE GLOBECOM '82 Conf. Proc., pp. 239-243, Nov. 29th - Dec. 2nd, 1982.

[4] D. Millar, P. Mermelstein, "Prevention of Predictor Mistracking in ADPCM Coders", IEEE ICC '84, Conf. Proc., pp. 1508-1512, 1984.

[5] J. Makhoul and M. Berouti, "Adaptive Noise Spectral Shaping and Entropy Coding in Predictive Coding of Speech", IEEE Trans. on ASSP, pp. 63-73, Feb. 1979.

[6] J.P. Adoul and D. Pradelles, "On Line Speech / Data-Modem Identifier for Telephone Network", Proc. ICASSP '77, pp. 332-335, May 1977.

[7] B. Kedem, "Search for Periodicities by Axis-Crossing of Filtered Time Series", Sig. Proc., pp. 129-144, March 1986.

[8] A. Zinaty, "Digital Coding of Speech and Voice-Band Data Signals", M.Sc. Thesis, Technion 1987.

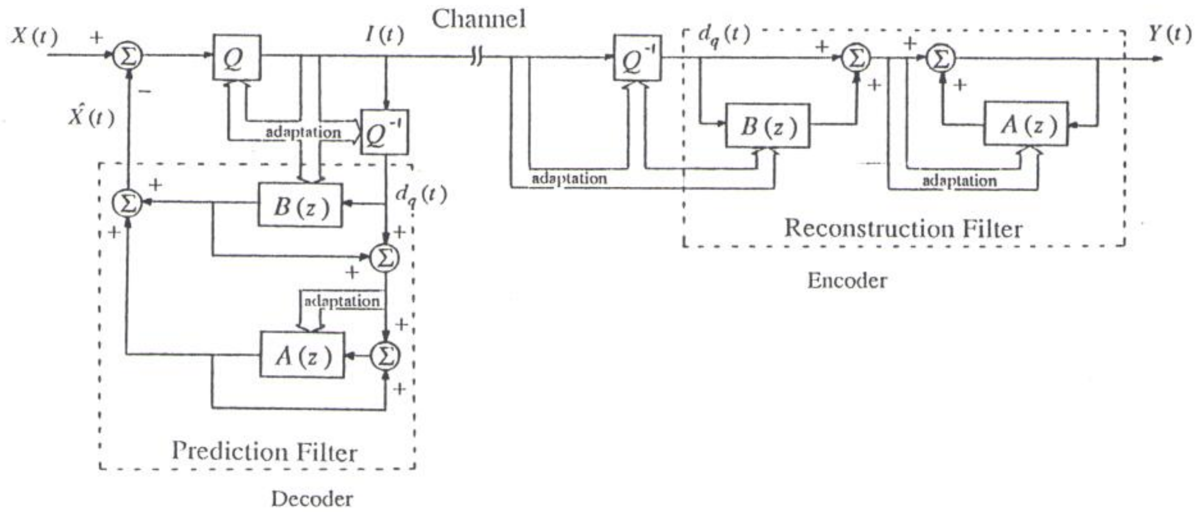

Figure 1: The standard 32 kbps ADPCM coder (without the linear/log PCM converters).

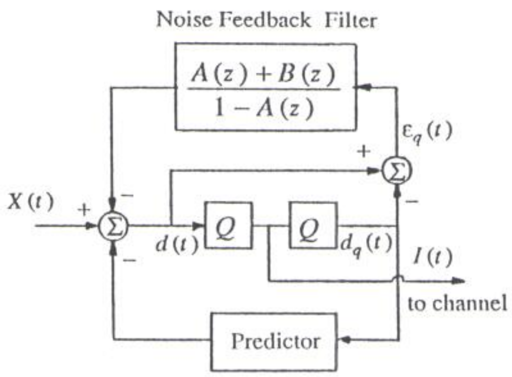

Figure 2: ADPCM encoder with a noise feedback filter.

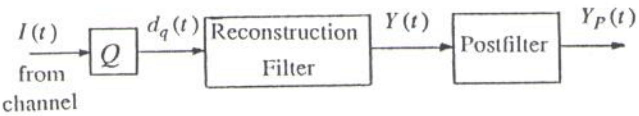

Figure 3: ADPCM decoder with a postfilter at the decoder output.

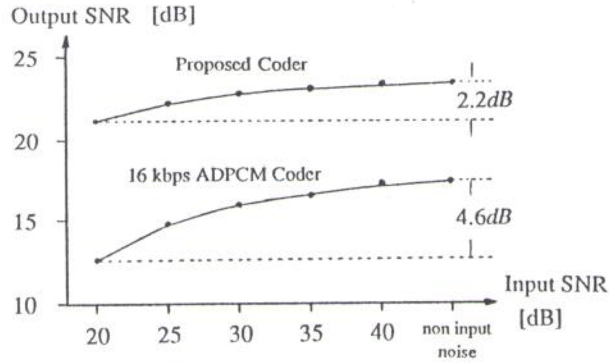

Figure 4: Output SNR as a function of input SNR for coded speech.

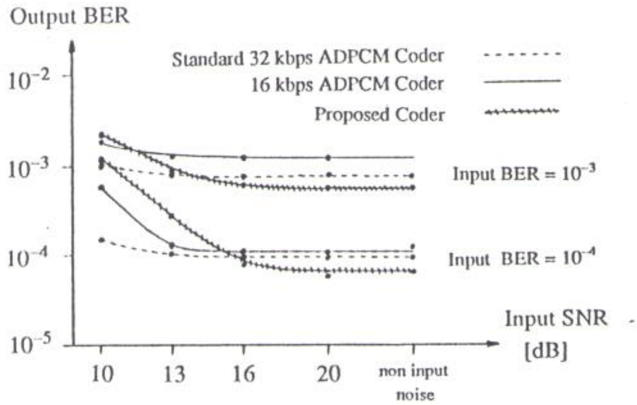

Figure 5: Output BER as a function of input SNR and channel errors