ALGORITHMIC ISSUES OF ADAPTIVE DIFFERENTIAL PULSE CODE MODULATION WITH REFERENCE TO CCITT RECOMMENDATION G.721

M.J. Flaherty

Telecom Australia Research Laboratories

ABSTRACT - A description of the CCITT 32 kbit/s adaptive differential pulse code modulation algorithm, G.721 (June, 1986 version) is presented with sufficient information to better appreciate its derivation.

INTRODUCTION

The ADPCM (adaptive differential pulse code modulation) algorithm, G.721, has been selected as an international standard for transcoding 64 kbit/s A/mu-law PCM to and from 32 kbit/s ADPCM. The algorithm's complexity stems partly from the requirements of coding both speech and non-speech services. Since its first recommendation, significant changes have been adopted to handle undesirable predictor behaviour in the presence of idle code inputs to the decoder and problems associated with unacceptably high error rates in character mode transmission using FSK modems (eg. V.23). Some authorities have also shown interest in changes to the algorithm or the adoption of a different ADPCM algorithm with superior voiceband data performance.

A full appreciation of these and other related issues requires a more detailed understanding of the algorithmic issues associated with this type of coder. A further reason for obtaining a more detailed understanding of the algorithm is the fact that many of the functional blocks contained within G.721 are likely to be embodied within other coding algorithms under study by the CCITT.

MAJOR FUNCTIONAL UNITS OF G.721.

The major functional units of the encoder and decoder are shown in Fig.1 and Fig.2 respectively. The processing functions of any differential coder generally fall under the two broad headings of quantization and prediction. Differential PCM removes signal redundancy by modelling signal statistics and subtracting a "predicted" estimate of the signal from the signal itself prior to quantization. If the predictor is a good model of the second order statistics, then the variance of the difference signal will be less than the variance of the original signal and fewer bits will be required to represent it for a given signal-to-noise ratio.

For the algorithm in question, the quantization process is handled by the following four functional blocks:

- Adaptive quantizer.
- Inverse adaptive quantizer.
- Quantizer scale factor adaptation.
- Adaptation speed control.

The tone and transition detector provides information to the adaptive predictor and the quantization functional units. The two remaining functional units are the format conversion and the synchronous coding adjustment. The paper will deal with each functional unit in turn.

Format conversion

The two format conversion functional blocks are the input PCM format conversion in the encoder and the output PCM format conversion in the decoder. The conversion follows CCITT Recommendation G.711 and converts input PCM into 13 bit linear signed magnitude for A-law and 14 bit linear signed magnitude for mu-law PCM. For output conversion a 16 bit twos complement number is converted into either an A-law or mu-law PCM representation following G.711.

Adaptive and Inverse adaptive quantizers

An adaptive quantizer can be viewed as having a time-varying stepsize which for optimum signal to noise ratio should be proportional to the standard deviation of the quantizer input signal. An equivalent operation to altering stepsize is to multiply the input to the quantizer by a time varying gain. Using the symbols as defined in G.721 and shown in Fig.1, the output of the adaptive quantizer can be written as

$$\tilde{I}(k) = Q[\ G(k).d(k)\ ] \qquad \ldots(1)$$

Replacing G(k) by 1/y(k) and defining the quantizer decision thresholds in the logarithmic domain we obtain

$$I(k) = Q_L[\ \log_2(d(k)) - y(k)\ ] \qquad \ldots(2)$$

The quantizer in the latest version of G.721 is a 15 level, mid tread quantizer. The decision thresholds as shown in Table 1 agree with those obtained by minimizing the expectation of the quantizer error assuming a quantizer input with a Gaussian probability density function (Max, 1960). This replaces the previous 16 level mid riser in order to alleviate undesirable idle code behaviour.

Quantizer scale factor adaptation

The purpose of the scale factor adaptation block is to provide the scale factor, y(k), to the adaptive quantizer. The scale factor is a measure of the difference signal variance. The approach adopted in G.721 in the fast or unlocked mode is a feedback adaptation with a one word memory (Jayant, 1973). When expressed in the logarithmic domain the adaptation stategy is given by

$$y(n) = \beta.y(n-1) + (1-\beta).W[I(n-1)] \qquad \ldots(3)$$

This will be recognised as a first order difference equation with a time constant of 4ms for the leakage factor $\beta=(1-2^{-5})$. This time constant is a measure of the propagation time of channel errors. The overall adaptation response time is also strongly determined by the choice of multipliers, W(.). For speech the adaptation response time needs to be high, however, many voiceband data signals are best modelled as having a constant or slowly changing variance. In order to cater for both speech and data, G.721 employs a dynamic locking quantizer with bi-modal adaptation (Petr, 1982) defined by the equations

$$y(k) = a_1(k).y_u(k-1) + (1-a_1(k)).y_1(k-1) \qquad \ldots(4a)$$

where a (k) is the speed control factor and determines the contribution to the overall scale factor from the locked and unlocked components and is in the range $0 \leq a_1(k) \leq 1$. The unlocked scale factor component in (4a) is given by

$$y_u(k) = (1 - 2^{-5}).y(k) + 2^{-5}.W[\ I(k)\ ] \qquad \ldots(4b)$$

where $y_u(k)$ is limited to $1.06 \leq y_u(k) \leq 10$. corresponding to a dynamic range of 1 to 491. The weights are given by

I(k)	7	6	5	4	3	2	1	0
W(I)	70.13	22.19	12.38	7.00	4.00	2.56	1.13	-0.75

The locked component is obtained by a first order filtering of y (k).

$$y_l(k) = (1 - 2^{-6}).y_l(k-1) + 2^{-6}.y_u(k) \qquad ...(4c)$$

It can be seen that in the special case where a (k)=1, (4a) reduces to (3). In the special case where a (k)=0 the above equations reduce to y(k)≃y(k-1) which corresponds to no adaptation.

Quantizer speed control

Scale factor estimation amounts to an estimate of signal variance while the determination of the speed control parameter amounts to an estimate of the rate of change of signal variance. The dynamics of the speed control parameter are governed by a first order difference equation with either a step or zero as input. The short-term average magnitude of the quantized difference signal, I(k), is given by

$$d_{ms}(k) = (1 - 2^{-5}).d_{ms}(k-1) + 2^{-5}.F\{ I(k) \} \qquad ...(5)$$

and the long-term average is given by

$$d_{ml}(k) = (1 - 2^{-7}).d_{ml}(k-1) + 2^{-7}.F\{ I(k) \} \qquad ...(6)$$

The time constants corresponding to (5) and (6) are 4ms and 16ms respectively and the driving function F[I(k)] is given by

|I(k)|	7	6	5	4	3	2	1	0
F[I(k)]	7	3	1	1	1	0	0	0

The driving function takes into account the non-uniform nature of the quantizer, squaring of the difference signal, and the rounding effects of a three bit representation. An unlimited speed control factor is given by

$$a_p(k) = \begin{cases} (1 - 2^{-4}).a_p(k-1) + 2^{-3}; & \text{if } |d_{ms}(k)-d_{ml}(k)| \geq 2^{-3}.d_{ml}(k) \\ ; & \text{or } y(k) < 3 \\ ; & \text{or } t_d(k) = 1 \\ 1 ; & \text{if } t_r(k) = 1 \\ (1 - 2^{-4}).a_p(k-1) ; & \text{otherwise} \end{cases} \qquad ...(7)$$

The first condition corresponds to a speech like signal with rapidly changing statistics. The second condition corresponds to an idle channel. The third condition corresponds to the detection of a partial band or tone like signal (eg. mark tone in FSK). When any of these three conditions are detected, the stepped input applies. The fourth condition corresponds to a transition from a partial band signal (eg. mark to space tone in FSK). For this condition the speed control is forced to 1. The final condition corresponds to a signal with slowly varying statistics (not partial band). Limiting is then applied to eliminate premature transitions for pulsed input signals such as switched carrier voiceband data. This limiting is defined by

$$a_l(k) = 1 \quad ; a_p(k-1) > 1 \qquad ...(8)$$
$$= a_p(k-1) \quad ; a_p(k-1) \leq 1$$

Tone and transition detector

This functional block is an addition to the latest revision of the algorithm and its purpose was to overcome serious problems which existed for V.23 FSK data transmission in the character mode. If the mark tone is transmitted for a sufficiently long period of time, the predictor adapts to it very well resulting in a relatively high prediction gain. This in turn causes the quantizer scale factor, $y(k)$, to decrease in order to scale a small difference signal. The problems arise in the transition from a mark to a space tone. The quantizer and predictor cannot adapt quickly enough resulting in quantizer overload. The tone and transition detector detects the presence of partial band signals (eg. FSK mark tone) by monitoring the second coefficient of the predictor's 2-pole component. For a second order predictor this coefficient gives a measure of the signal bandwidth. Specifically, tone detection is defined by

$$t_d(k) = \begin{cases} 1 & ;\ a_2(k) < -0.71875 \\ 0 & ;\ \text{otherwise.} \end{cases} \qquad \ldots(9)$$

A transition from a partial band signal is defined so that the predictor coefficients can be set to zero and the quantizer can be forced into the fast mode of adaptation. In addition to the condition given by (9) when the quantized difference signal exceeds a threshhold a transition is assumed to have occurred. Specifically,

$$t_r(k) = \begin{cases} 1 & ;\ a_2(k) < -0.71875\ \&\ |\ d_q(k)\ | > 24.2^{y_1(k)} \\ 0 & ;\ \text{otherwise.} \end{cases} \qquad \ldots(10)$$

Adaptive prediction

The signal estimate, s (k), is the sum of two pole plus six zero terms. Specifically,

$$s_e(k) = \sum_{i=1}^{2} a_i(k-1).s_r(k-i) + s_{ez}(k) \qquad \ldots(11)$$

where

$$s_{ez}(k) = \sum_{i=1}^{6} b_i(k-1).d_q(k-i) \qquad \ldots(12)$$

and the reconstruction signal is defined as

$$s_r(k-i) = s_e(k-i) + d_q(k-i) \qquad \ldots(13)$$

G.721 employs backward prediction coefficient adaptation using a simplified gradient technique as opposed to the well known block forward techniques. Justifications for adopting this procedure are that it has a computational advantage, requires no side information to be transmitted and most importantly introduces a negligible processing delay. The main disadvantage is that high order pole estimation cannot be realised with accuracy. The zero coefficients are updated according to the following equation (Nishitani et al, 1982)

$$b_i(k) = (1-2^{-8}).b_i(k-1) + 2^{-7}.\text{sgn}[d_q(k)].\text{sgn}[d_q(k-i)] \quad \text{for } i=1, 2, \ldots, 6 \qquad \ldots(14)$$

The pole update procedure is given by (Miller and Mermelstein, 1984)

$$a_1(k) = (1-2^{-8}).a_1(k-1) + (3.2^{-8}).\text{sgn}[p(k)].\text{sgn}[p(k-1)]$$
$$a_2(k) = (1-2^{-7}).a_2(k-1) + 2^{-7}.\{\text{sgn}[p(k)].\text{sgn}[p(k-2)] - f[a_1(k-1).\text{sgn}[p(k)].\text{sgn}[p(k-1)]\} \qquad \ldots(15)$$

where $p(k) = d_q(k) + s_{ez}(k)$

$$f(a_1) = 4.a_1 \quad ; \ |a_1| \le 1/2$$
$$= 2.sgn[a_1] \quad ; \ |a_1| > 1/2$$

Synchronous coding adjustment

Synchronous coding refers to the tandeming of transcoders with no intermediate digital/analog/digital conversion. The potential error accumulation is a consequence of having two different quantizers in the transcoding process. The final output stage of the decoder is a linear to PCM conversion which can be one PCM state removed from the previous encoder PCM input. When the PCM signal is linearised by the second encoder it can result in a difference signal not equal to the difference signal in the first encoder/decoder pair. The net result of this can be an accumulation of error in subsequent encoding and decoding operations. Synchronous coding adjustment provides a simple means of adjusting the decoder PCM output state to ensure that the quantized difference signal in tandem coders corresponds to the same adaptive ADPCM quantizer state. The decoder computes a difference signal, d.(k):

$$d_x(k) = s_{1x}(k) - s_e(k) \qquad ...(16)$$

where $s_{1x}(k)$ is the linearized version of the unadjusted PCM output. It is assumed that the states of all encoder/decoder pairs are the same and therefore the signal estimate in the first encoder/decoder pair is the same as in the second encoder/decoder pair. The difference signal given by (16) anticipates the corresponding calculation in the next encoder. The difference signal $d_x(k)$ is compared with the previous quantized difference signal, $d_q(k)$, as determined by I(k) and y(k). If $d_x(k)$ falls below the decision level corresponding to $d_q(k)$ then the output PCM state is moved to the next higher state. If $d_x(k)$ falls above or equal to the upper decision level then the output PCM state is altered to the next lower state. If $d_x(k)$ falls within the decision levels of $d_q(k)$ then the output PCM state is unchanged.

CONCLUDING REMARKS

The performance of a speech coding algorithm depends upon both fundamental knowledge and the cost of complexity associated with its hardware realisation. Since the initial recommendation of G.721 it has undergone a significant revision and still fails to meet the requirements of some authorities. The algorithm's complexity is well matched to current VLSI capabilities, however, improved algorithms exist and their cost effective implementation is likely to be realised in the near future. When used in conjunction with appropriate testing procedures, detailed knowledge of an algorithm and its derivation can enable potential users to determine its suitability for their application and the algorithm's "technological life".

ACKNOWLEDGEMENT

The permission of the Director Research, Telecom Australia, to publish this paper is hereby acknowledged.

REFERENCES

CCITT Recommendation G.721 (June, 1986) "32 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)", CCITT Study Group XVIII.

JAYANT, N.S.(1973) "Adaptive Quantization with a One Word Memory", Bell System Tech. J., Vol.52, No.7, pp. 1119-1144.

MAX, J. (1960) "Quantizing for Minimum Distortion", IRE Trans. Inform. Theory, Vol. IT-6, pp. 7-12.

NISHITANI, T., ET AL (1982) "A 32 kbit/s Toll Quality ADPCM Codec Using a Single Chip Signal Processor", ICASSP, pp. 960-963.

PETR, D.W. (1982) "32 kbit/s ADPCM-DLQ Coding for Network Applications", Globecom Conf., Miami FL., Vol. 1, pp. A8.3.1-A8.3.5.

Normalized quantizer input range $\log_2\lvert d(k)\rvert - y(k)$	$\lvert I(k)\rvert$	Normalized quantizer output $\log_2\lvert d_q(k)\rvert - y(k)$
[3.12, + ∞)	7	3.32
[2.72, 3.12)	6	2.91
[2.34, 2.72)	5	2.52
[1.91, 2.34)	4	2.13
[1.38, 1.91)	3	1.66
[0.62, 1.38)	2	1.05
[-0.98, 0.62)	1	0.031
(- ∞, -0.98)	0	- ∞

Table 1. Quantizer Decision Levels.

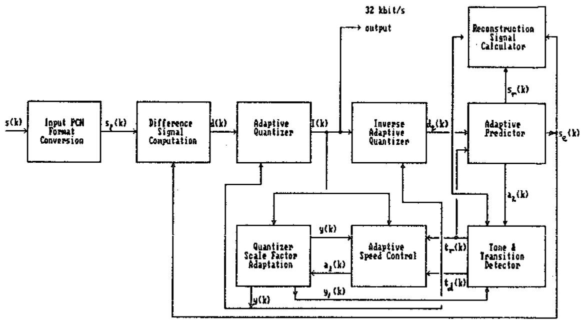

Figure 1. G.721 ADPCM Encoder

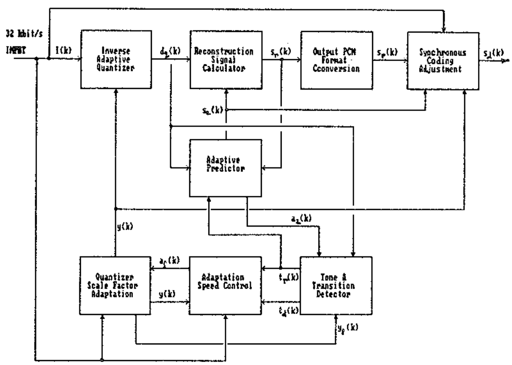

Figure 2. G.721 ADPCM Decoder.