A Packet-Based CAPDM Speech Coder for PCN Applications

Chia-Horng Liu and Chia-Chi Huang

***Abstract*—In this paper, we present a median-rate speech coder, the controlled adaptive prediction delta modulation coder (CAPDM), which operates at 16 kb/s with good speech quality and low algorithm complexity [15]. The coder is dedicated to personal communication network (PCN) applications and transmits speech samples on the basis of packets. It combines the features of one-step looking forward decision, syllabic companding, instantaneous companding, and adaptive prediction. In addition to the use of a short-term prediction filter, CAPDM also exploits the pitch property to predict speech waveform explicitly. With the aid of a pitch prediction filter, the performance of a CAPDM codec improves about 3 dB in segmental signal-to-noise ratio (SEGSNR). The average SEGSNR of CAPDM.FF is about 21 dB, which is 7 dB over traditional CVSD at 16 kb/s. We also utilize an adaptive postfilter (APF) to enhance the perceptual quality of the decoded speech. The mean opinion score (MOS) listening test of CAPDM.FF with APF shows that its average score achieves 4.19, which is as good as G.728 16-kb/s LD-CELP and is comparable with CCITT G.721 32-kb/s ADPCM. The complexity of CAPDM.FF is evaluated to be 8 MIPS, which is much lower than that of LD-CELP and could be further reduced by adopting a smaller correlation window for pitch detection.**

To solve the problem of packet loss, we developed a packet-based waveform substitution method by reinitializing the codec parameters at the beginning of each packet. The simulation results show that CAPDM.FF could tolerate 5% of packet loss and still keep an SEGSNR at 10 dB and an MOS at about 3.0.

***Index Terms*—Adaptive prediction, instantaneous companding, packet recovery, pitch detection, speech coder, syllabic companding, waveform substitution.**

I. Introduction

A PERSONAL communication network (PCN) has to provide telephone services at any time, anywhere with a satisfactory quality of service (QoS) measure [1]. Two most important requirements of this kind of service are low cost and high speech quality. Accordingly, the 32-kb/s adaptive differential pulse code modulation (ADPCM) coder recommended as G.721 by the International Telephone and Telegraph Communication Committee (CCITT) [2] is adopted by several personal communication systems. For example, it is adopted by the second-generation cordless telephone (CT2 system), the digital European cordless telecommunication (DECT) system, the personal access communication system (PACS), and the personal handiphone system (PHS). All of these systems use the G.721 ADPCM codec operating at 32 kb/s without considering voice activity detection (VAD). In this paper, we propose a 16-kb/s speech codec with appropriate speech quality for PCN applications. The adoption of this speech codec will double the PCN system capacity as compared with using the G.721 ADPCM codec. Furthermore, if VAD mechanism is employed with this codec, much higher system capacity gain can be achieved.

The structure of a controlled adaptive prediction delta modulation coder (CAPDM) codec consists mainly of four parts: a logic unit, a stepsize estimation unit, a pitch predictor, and an adaptive prediction unit. The logic unit is used for one-step look forward decision. The stepsize estimation unit is used to estimate both instantaneous and syllabic stepsizes. The pitch predictor is used to find out vowel pitch periods. The adaptive prediction unit is used to predict the current speech sample based on both a short- and a long-term predictor.

For PCN applications, digital transmission schemes are used to carry voice or data packets over a shared radio medium [26]. When a network is in a heavy traffic, a voice packet might be held in a queue. When a voice packet is delayed over a certain time limit, it must be dropped. Another problem of packet voice transmission comes from packet loss. In order to sustain speech codec performance in an error prompt environment, people use different waveform substitution techniques to recover lost voice packets. In this scenario, we have surveyed the zero substitution technique [3], the previous packet repetition technique [3], the pattern matching technique [3], the pitch-based replication technique [4], etc. [5]. Besides waveform substitution, the continuity of codec parameters must be maintained in order to bridge over the gaps of the lost packets. In the paper, we show that the isolation of each transmission packet will avoid the problem of divergence in the reconstructed speech waveform and maintain good codec performance when voice packets are lost.

This paper is composed of six parts. In Section II, we describe the basic structure of a CAPDM codec. In Section III, we describe two pitch prediction methods which can be used in a CAPDM coder. In Section IV, we present computer simulation results of CAPDM codec in an ideal channel in which both subjective and objective performance are evaluated. In Section V, we describe CAPDM performance evaluation in a noisy channel. Finally, we draw some conclusions in Section VI.

II. Basic Structure of CAPDM

Primarily, CAPDM [16] has four features including one-step look forward decision [18], syllabic companding, instantaneous companding, and adaptive prediction [21].

1) One-step look forward decision: CAPDM looks one sample ahead before making any decision. That is,

Manuscript received January 19, 1998; revised November 16, 1998. This paper was presented in part at the Symposium of Personal Indoor Mobile Radio Communication, Taipei, Taiwan, R.O.C., October 1996.

The authors are with the Department of Communication Engineering, National Chiao Tung University, Hsinchu, Taiwan, R.O.C.

Publisher Item Identifier S 0018-9545(00)03696-3.

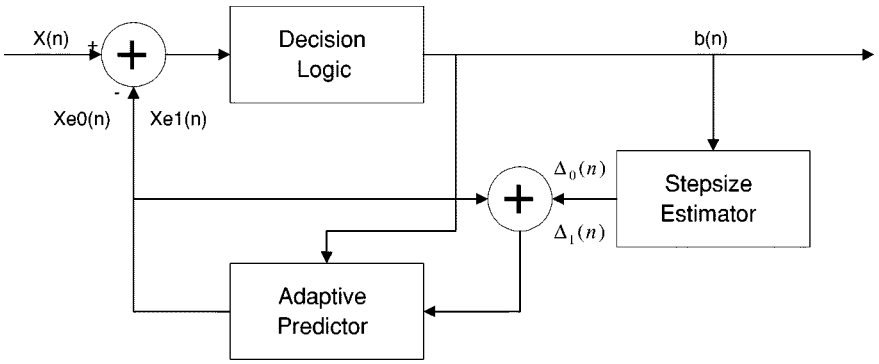


Fig. 1. CAPDM encoder.

it generates simultaneously two estimated samples, $Xe1(n)$ for bit 1 and $Xe0(n)$ for bit 0. The one that is closer to the current speech sample, $x(n)$, is chosen as a valid prediction, $Xe(n)$. This mechanism results in less prediction error and closer waveform matching.

2) Syllabic companding: This is to estimate long-term stepsize from averaged speech waveform slope over a short period of time (about 15 ms).
3) Instantaneous companding: This is to adapt stepsize sample-by-sample in order to track the dynamic range of speech waveform.
4) Adaptive prediction: The prediction is based on the short-term correlation in speech samples with predictor coefficients updated by a simplied stochastic approximation of the gradient method [6].

The basic structure of a CAPDM codec is simply a decision logic, a stepsize estimation unit, and an adaptive prediction unit [16]. The block diagram of a CAPDM encoder is shown in Fig. 1. Each individual unit is described in the following.

A. Decision Logic

The function of this unit is to compare the distance between the current speech sample and two estimated samples from the adaptive prediction unit, i.e., comparing $|X(n)-Xe0(n)|$ with $|X(n)-Xe1(n)|$. It decides a zero to transmit if the former value is smaller than the latter. Otherwise, it transmits a one. This logic unit thus implements the feature of one-step look forward decision.

B. Stepsize Estimation Unit

This unit produces both syllabic and instantaneous stepsize estimation and combines them to generate the current stepsize estimation. The block diagram of this unit is shown in Fig. 2. It consists of a 3-bit memory, a last stepsize logic, a basic stepsize logic, and an instantaneous stepsize estimator unit. The inputs to this unit are the previously transmitted bits and its outputs are two stepsize estimations. Before we explain the operation of this unit, two states must be defined first. If the last two transmitted bits are the same, the codec is in state "O." Otherwise, it is in state "U." Intuitively, "O" represents slope overload and "U" represents slope underload.

1) 3-bit memory: The outputs of the decision logic are recorded in this memory for 3 bits long.

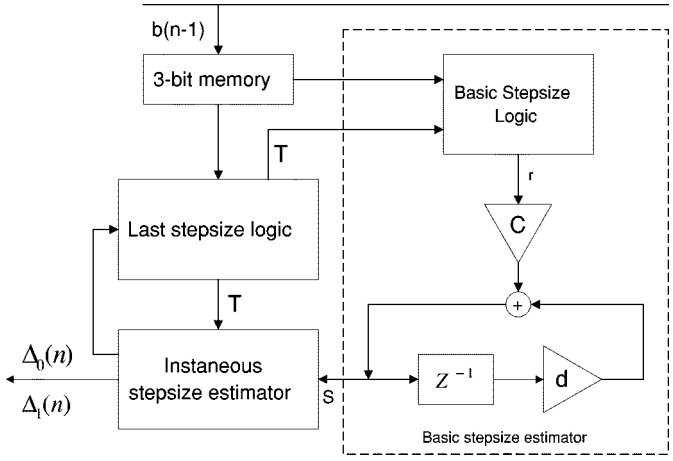


Fig. 2. Stepsize estimation unit of CAPDM.

2) Last stepsize logic: If the last transmitted bit equals to zero, a reference stepsize T is set to the absolute value of $\Delta_0(n-1)$, which is the previously estimated stepsize for bit 0. Otherwise, T is set to the absolute value of $\Delta_1(n-1)$, which is the previously estimated stepsize for bit 1

$$T=|\Delta_0(n-1)|, \quad \text{if } b(n-1)=0 \tag{1}$$

$$T=|\Delta_1(n-1)|, \quad \text{if } b(n-1)=1. \tag{2}$$

3) Basic stepsize logic: If the last three transmitted bits are the same, the reference level T is amplified by a factor C and fed into a low-pass filter (LPF). This T could be viewed as an input to drive the LPF. If the last three transmitted bits are not the same, the basic stepsize S is decreased

$$r=T, \quad \text{if } b(n-1)=b(n-2)=b(n-3) \tag{3}$$

$$r=0, \quad \text{otherwise.} \tag{4}$$

4) Instantaneous stepsize estimator: This unit produces two current stepsize estimates, $\Delta_0(n)$ and $\Delta_1(n)$, for both bits 0 and 1 according to the last two transmitted bits and the current preassumed bits, as shown in Fig. 3.
   a) If the state combination is O + O, the codec is in the state of slope overload and the stepsize has to

b(n-2)	b(n-1)	$\Delta_0(n)$	$\Delta_1(n)$
0	0	-T*P	S*h
0	1	-T/P	S
1	0	-S	T/P
1	1	-S*h	T*P

Fig. 3. Instantaneous stepsize table of CAPDM.

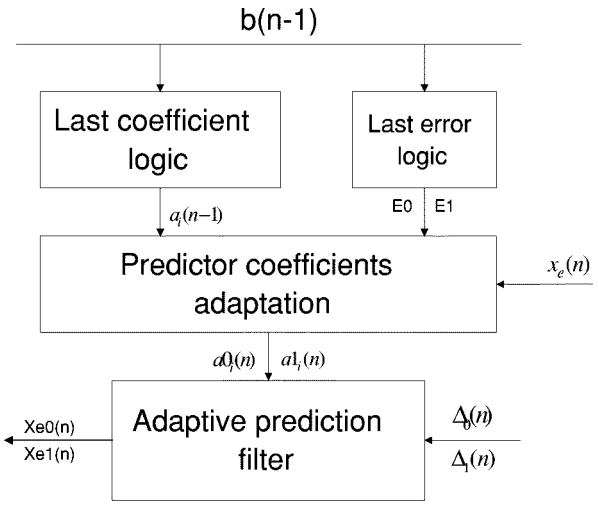


Fig. 4. Adaptive prediction unit of CAPDM.

be increased to $T*P$. Here, P is a constant slightly greater than one.

b) If the state combination is U + U, the codec is in the state of slope underload and the stepsize has to be decreased to T/P.

c) The state combination U + O occurs when the codec state changes from underload to overload. In this case, the stepsize immediately returns to the basic stepsize S, which is the average slope for the last 15 ms.

d) The state combination O + U occurs when codec state switches from overload to underload. In this case, the stepsize is set to $S*h$, where h is a value smaller than one. This means that the stepsize has become too large and has to be decreased.

The reason for the asymmetry between case c) and case d) (also observed in Fig. 3) is that the coder needs to adapt more quickly to the rising edge of speech waveform.

C. Adaptive Prediction Unit

This unit consists of a last coefficient logic, a last error logic, a predictor coefficients adaptation block, and an adaptive prediction filter. The block diagram of this unit is shown in Fig. 4. Assuming that $\{Xe(n-1), \cdots, Xe(n-N)\}$ are the N previously estimated speech samples, the two estimates $Xe0(n)$ and $Xe1(n)$ of the current speech sample are given by an Nth-order linear prediction filter as

$$Xe0(n) = \sum_{i=1}^{N} a0i(n) * Xe(n-i) + \Delta_0(n) \tag{5}$$

TABLE I
LAST ERROR LOGIC

b(n-1)	E0	E1
0	-1	0
1	0	1

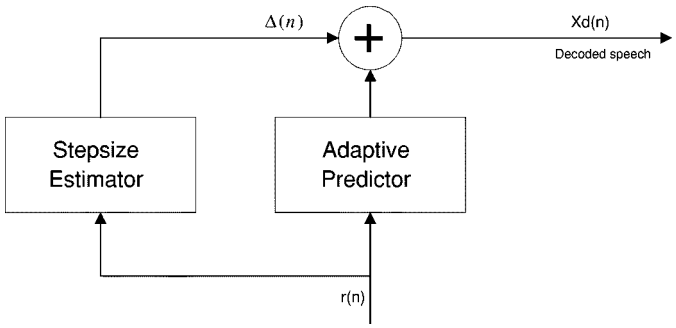


Fig. 5. CAPDM decoder.

$$Xe1(n) = \sum_{i=1}^{N} a1i(n) * Xe(n-i) + \Delta_1(n) \tag{6}$$

assuming a bit 0 or 1 is to be transmitted, respectively. The two stepsize estimates $\Delta_0(n)$ and $\Delta_1(n)$, coming from the stepsize estimation unit, are used to drive the two linear prediction equations, separately.

The two sets of filter coefficients $a0_i(n)$ and $a1_i(n)$ are adapted recursively using a simplified stochastic approximation of the gradient method [6] in the predictor coefficients adaptation block. By the method, the prediction filter coefficients are calculated as

$$a0_i(n) = (1-\alpha) * a_i(n-1) + \alpha * b_i + \beta * E0 * \text{SIGN}[Xe(n-i)] \tag{7}$$

$$a1_i(n) = (1-\alpha) * a_i(n-1) + \alpha * b_i + \beta * E1 * \text{SIGN}[Xe(n-i)] \tag{8}$$

where only the signs of the estimated speech samples are used. $(1-\alpha)$, which we set to 0.999 in our computer simulation, is the leaky factor of a first-order low-pass filter. At a sampling rate of 16 KHz, this leaky factor corresponds to an average time constant of 62.5 ms. b_1 is set to 0.95 and all other b_i's are set to zero. β is a parameter which controls the adaptation speed and is set to 0.0022 in our simulation. The previous filter coefficients, $a_i(n-1)$'s, are updated in the last coefficient logic according to the following:

$$a_i(n-1) = a0_i(n-1), \quad \text{if } b(n-1) = 0 \tag{9}$$

$$a_i(n-1) = a1_i(n-1), \quad \text{if } b(n-1) = 1 \tag{10}$$

i.e., they are updated according to the last transmitted bit.

It is noted that the error samples in the equations are replaced by E0 and E1 for bit 0 and bit 1, respectively, according to Table I.

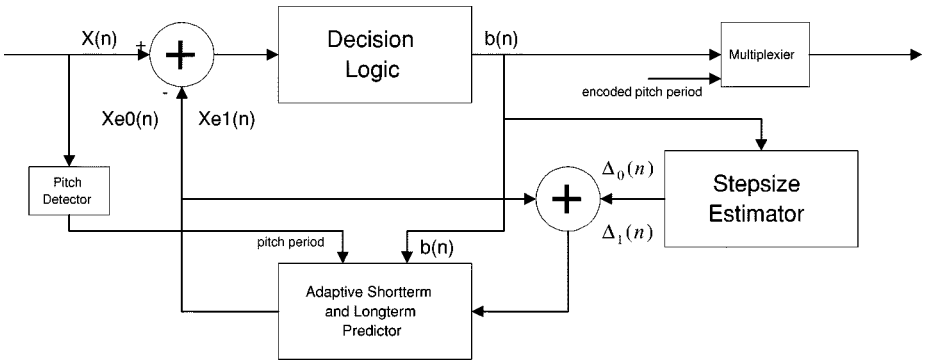


Fig. 6. CAPDM encoder with feedforward pitch detection.

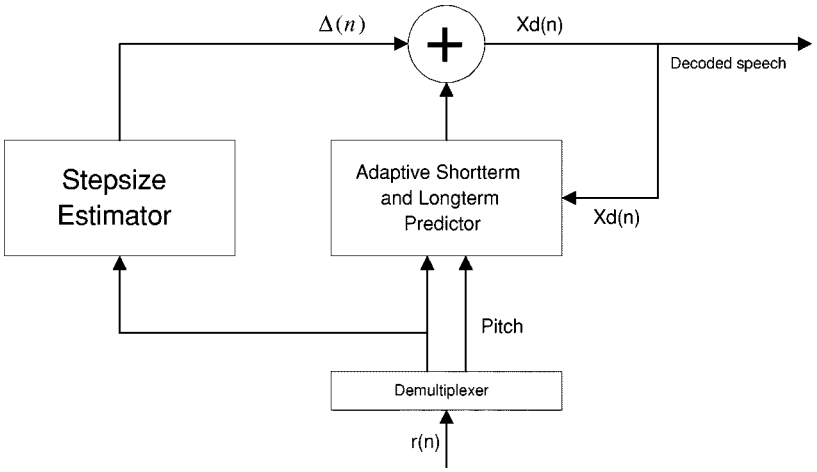


Fig. 7. CAPDM decoder with feedforward pitch detection.

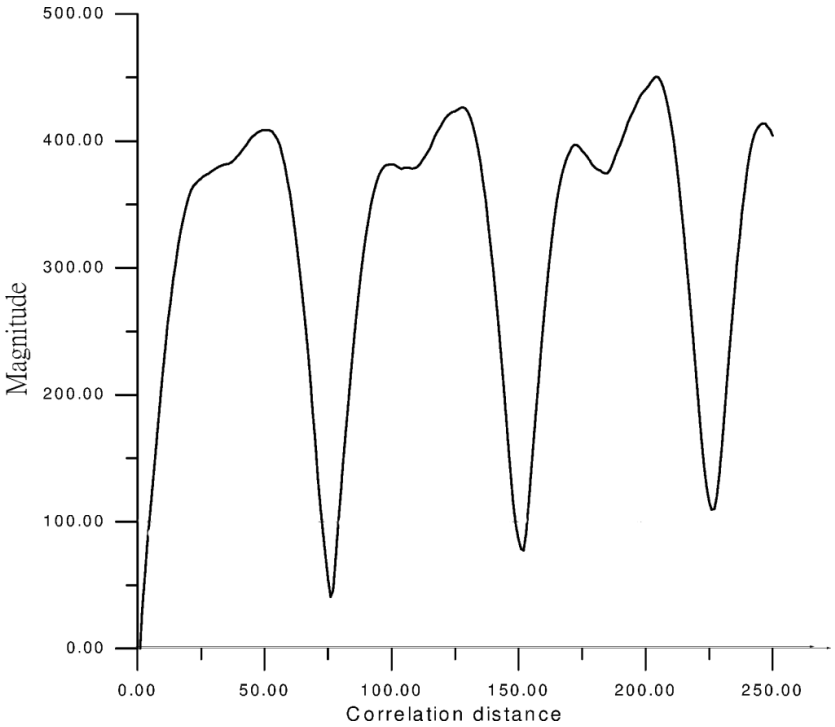


Fig. 8. The rest of AMDF analysis.

From the table, we observed that the prediction filter coefficients are adapted only when consecutive ones or zeros occur in the encoded bit stream, which is the situation when a voiced signal is being transmitted. The adaptive predictor will go back

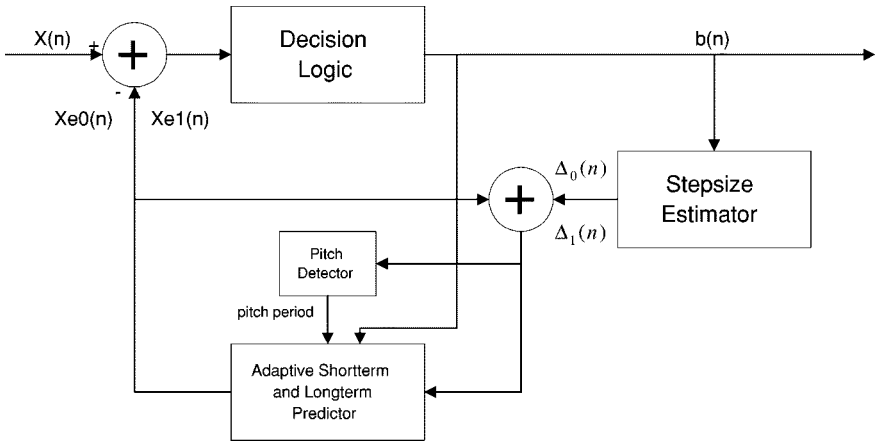


Fig. 9. CAPDM encoder with feedback pitch detection.

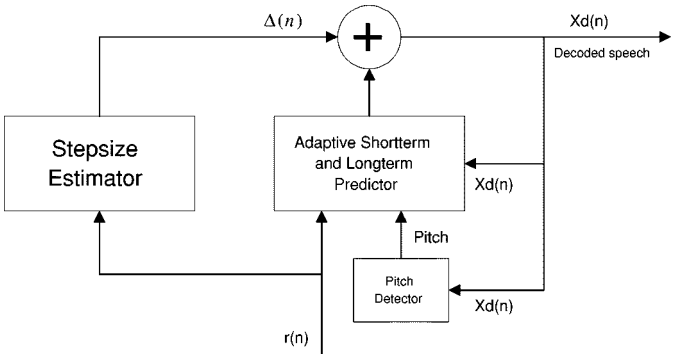


Fig. 10. CAPDM decoder with feedback pitch detection.

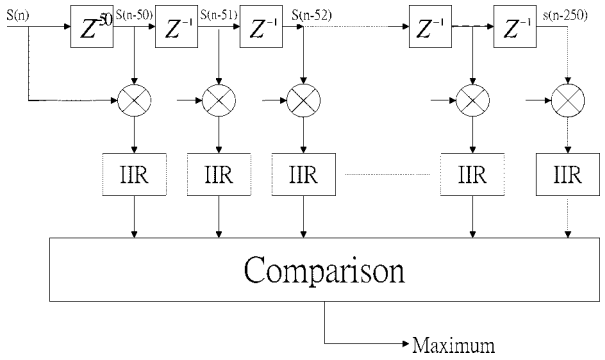


Fig. 11. Feedback pitch detection method.

to a fast-leaky first-order predictor when an unvoiced sound is transmitted or a receiver encounters a high channel error rate.

D. Decoder

The CAPDM decoder consists of a stepsize estimation unit and an adaptive prediction filter unit, as shown in Fig. 5. Here, we assume that the encoded bit stream $\{b(n), b(n-1), \cdots\}$ is packetized, digitally transmitted through a channel, demodulated, and depacketized into received bit stream $\{r(n), r(n-1), \cdots\}$. The decoded speech samples are denoted as $X_d(n)$.

The received bit stream $r(n)$ is sent to both a stepsize estimation unit and an adaptive prediction unit to determine the current decoded speech sample $Xd(n)$. $Xd(n)$ is fed back to the adaptive prediction unit. Comparing Fig. 5 with the feedback path in Fig. 1, we observe that these two structures are similar. The main difference is that we do not need to produce two estimates for bits 0 and 1 in the decoder as proceeded in the encoder. In our simulation, we found that overestimated stepsizes caused by channel errors are very disturbing in codec performance. In order to enhance the speech quality, we set an upper limit to the basic stepsize S at 200 and a lower limit at one, where the dynamic range of the speech samples is ± 1024.

III. CAPDM with Pitch Detection

The CAPDM codec described in Section II removes the short-term redundancy in speech waveform and the residual signal still contains pitch periods, which carry the long-term redundancy. If this pitch-related redundancy had been removed, the residual signal would have been like a random noise signal and easier to be encoded due to its smaller dynamic range [17].

Both feedforward and feedback pitch detection algorithm (PDA's) [8] are investigated and simulated for our codec. In the feedforward method, pitch periods are detected from the original speech samples. These detected pitch periods are transmitted to the receiver as side information. In the feedback method, pitch periods are detected from the predicted speech samples, which are available at both the transmitter and the receiver sides. Therefore, no side information is required. For convenience, we denote the CAPDM codec with feedforward pitch detection as "CAPDM.FF" and the one with feedback pitch detection as "CAPDM.FB."

A. Feedforward Pitch Detection

The block diagrams of CAPDM.FF encoder and decoder are shown in Figs. 6 and 7. Pitch periods are detected from original speech samples by performing the average magnitude difference function (AMDF) [7] analysis on the current and the previous speech waveform segments. The length of a segment is 250 samples, which is the maximum allowable pitch period at a 16-kb/s sampling rate. An example of AMDF analysis result is shown in Fig. 8, in which the distance between two local minima corresponds to the pitch period. After a pitch period is

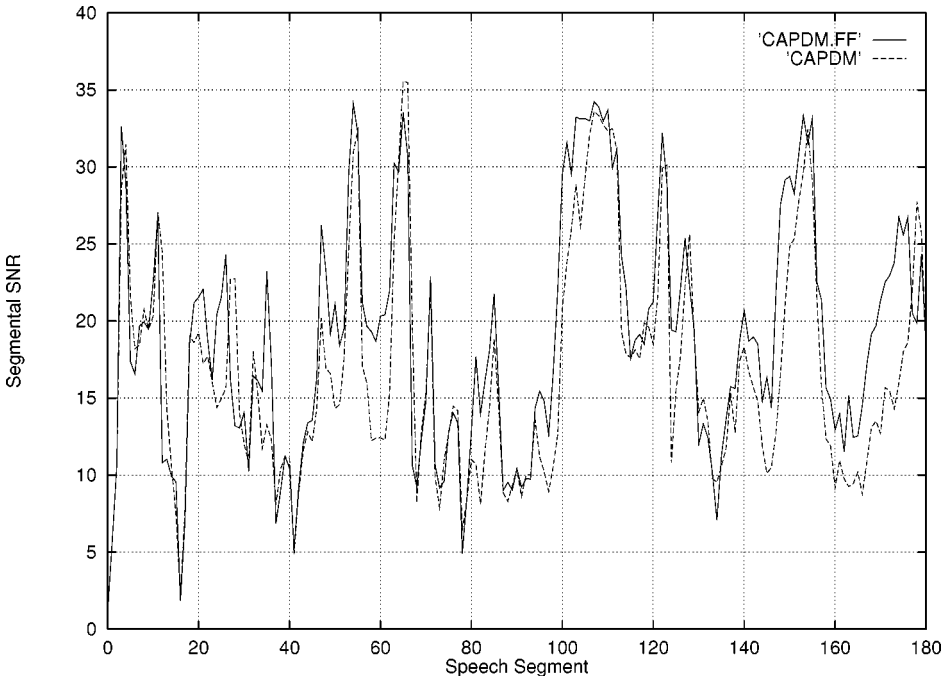


Fig. 12. SEGSNR of CAPDM and CAPDM.FF.

detected, it is used by a long-term adaptive prediction filter in addition to a short-term adaptive prediction filter to predict the current speech sample, using (11) and (12)

$$xe0(n) = \sum_{i=1}^{N} a0_i(n) * xe(n-i) + \sum_{j=1}^{L} b0_j(n) * xe(n-P-j) + \Delta_0(n) \quad (11)$$

$$xe1(n) = \sum_{i=1}^{N} a1i(n) * xe(n-i) + \sum_{j=1}^{L} b1_j(n) * xe(n-P-j) + \Delta_1(n) \quad (12)$$

where P is the estimated pitch period and L is the order of long-term prediction filter. The adaptation of long-term adaptive prediction filter coefficients is done in a similar way as in the short-term case. The pitch periods must be encoded and multiplexed with the encoded speech bit stream and transmitted to the receiver. The number of bits to encode a pitch period is eight, which is a small overhead compared with the packet size, say, 250. At a receiver, the pitch period encoded bits are simply demultiplexed from the received packet.

B. Feedback Pitch Detection

The block diagrams of CAPDM.FB encoder and decoder are shown in Figs. 9 and 10. Pitch periods are detected from previously predicted speech samples instead of the original speech samples. Here, a feedback pitch detection method is used instead of the AMDF method. The block diagram of this method is shown in Fig. 11. The predicted speech samples are sent to a delay line, and these delayed samples are multiplied with the current sample. The output of each multiplier is then filtered with a first-order IIR LPF. Finally, the outputs of the IIR's are compared and the maximum determines the location of the pitch period. This detector is in fact a simplified correlation detector. Since pitch periods are detected in a feedback manner, no side information is transmitted. In contrast to the feedforward approach, in which the pitch periods are updated in every packet, pitch periods are updated sample-by-sample in the feedback approach.

After a pitch period is detected, it is used by the long-term adaptive prediction filter as in the feedforward case. At a receiver, pitch periods are derived from the reconstructed speech samples just as in the encoder case.

IV. CAPDM Performance Evaluation in an Ideal Channel

We evaluated the CAPDM codec performance for both error-free and packet loss situations. Two male and two female speech waveform samples at 16 kb/s, each for about 3–5 s long, are used in our simulation. The codec performance is evaluated by both an objective measure called the segmental signal-to-noise ratio (SEGSNR) and a subjective measure [22] called the MOS. Different versions of CAPDM are compared with the continuous variable slope delta (CVSD) codec at 16 kb/s [20] and the G.721 ADPCM codec at 32 kb/s.

A. Error-Free Case

First, we compare the SEGSNR performance of CAPDM and CAPDM.FF, as shown in Fig. 12. As pitch-related redundancy is exploited in CAPDM.FF, its performance is improved by about 2–3 dB over the original CAPDM. On the other hand, we observe from Fig. 13 that the SEGSNR performance of CAPDM.FF and CAPDM.FB is nearly the same. The pitch

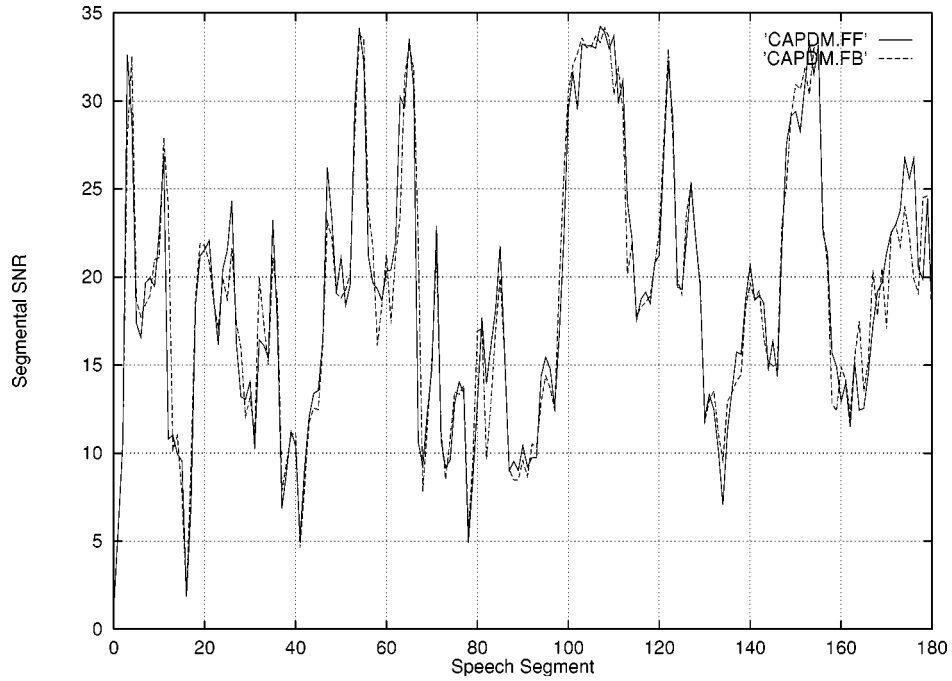


Fig. 13. SEGSNR of CAPDM.FF and CAPDM.FB.

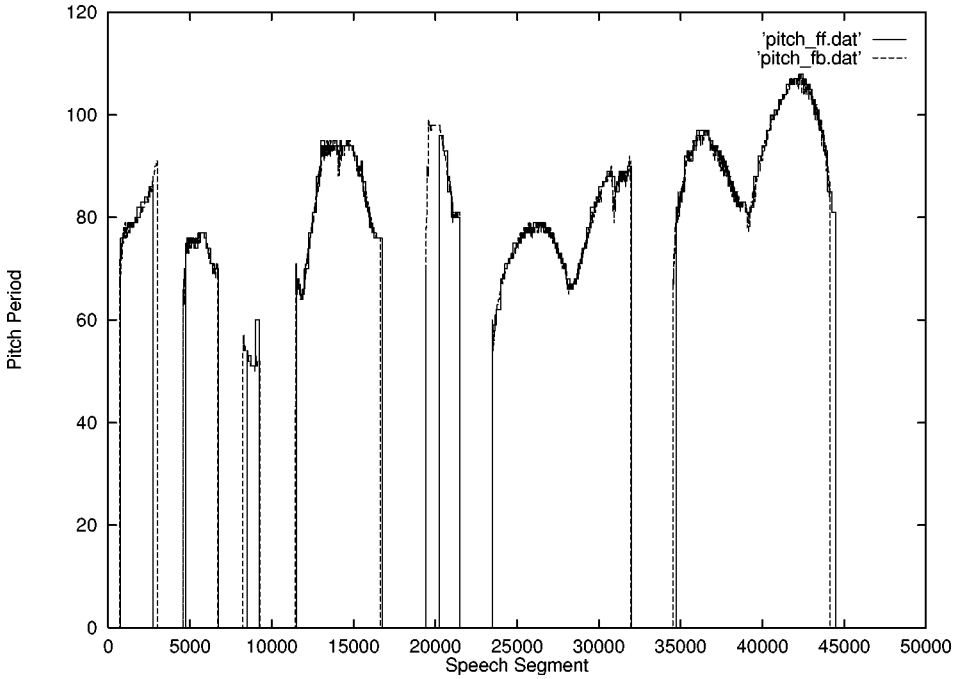


Fig. 14. The detected pitch periods using feedforward and feedback pitch detection methods.

periods detected by the two methods are shown in Fig. 14. Clearly, the detected pitch periods are almost the same. The feedback pitch detection method detects more pitch periods because it is operated on a sample-by-sample basis.

We make the following comments after comparing the performance of CAPDM.FF and CAPDM.FB.

1) Performance: These two codecs have nearly the same performance.
2) Delay: CAPDM.FF induces more coding delay than CAPDM.FB. While CAPDM.FF needs to perform AMDF over a 250-sample window, the pitch detection of CAPDM.FB is done on sample-by-sample basis.
3) Overhead: While CAPDM.FF needs eight extra bits as side information to encode a pitch period, CAPDM.FB does not. However, this is a small overhead compared with the packet size.
4) Complexity: The main difference in complexity between CAPDM.FB and CAPDM.FF is in their derivation of pitch periods at a decoder. While CAPDM.FB needs to do the pitch detection again at a decoder, CAPDM.FF does not.

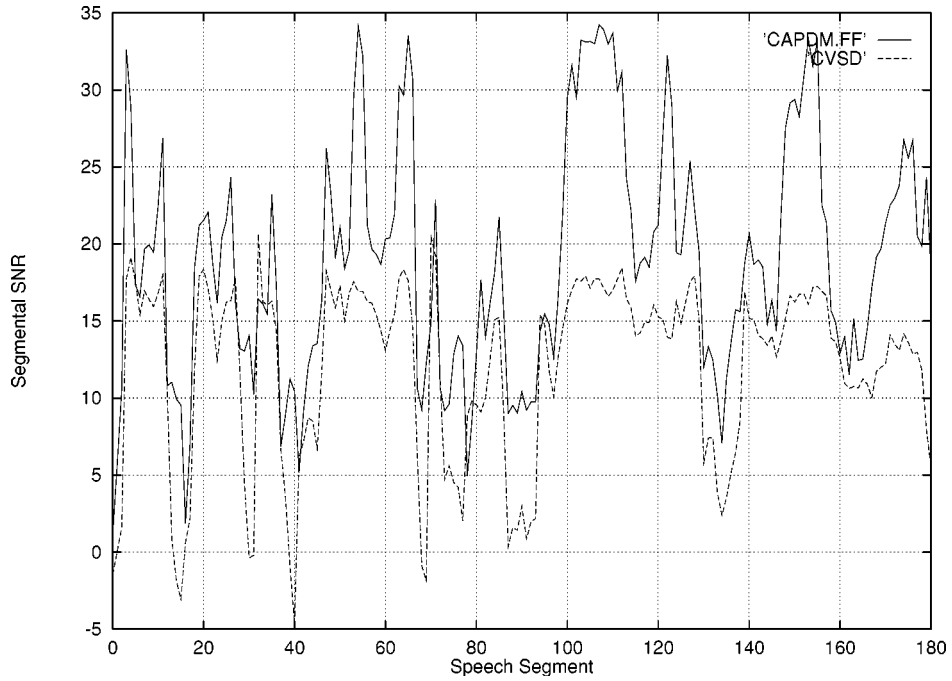


Fig. 15. Comparison of SEGSNR for CAPDM.FF and CVSD.

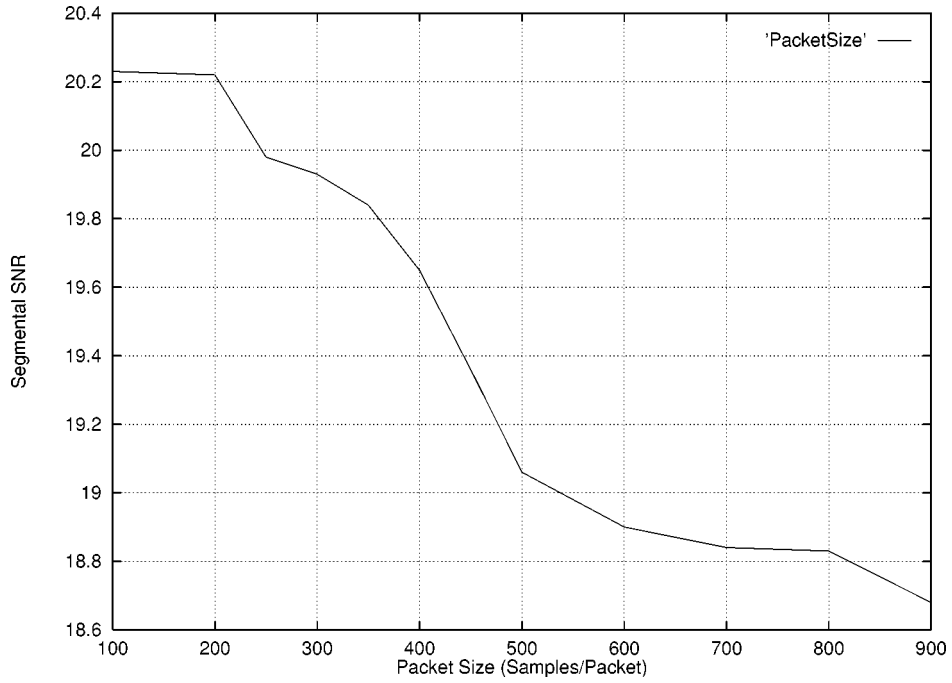


Fig. 16. SEGSNR for different packet sizes.

5) Packet loss considerations: Because a CAPDM.FB decoder needs to detect pitch periods again from received voice packets, it encounters more severe degradation in performance in packet loss situations. For CAPDM.FF, its pitch periods are carried as side information and they usually vary slowly over consecutive voice packets. We can use the pitch-based replication method to recover a lost speech packet. Thus, from both performance and stability point of view, CAPDM.FF is much better CAPDM.FB.

Due to its less complexity and better performance in packet loss, CAPDM.FF was adopted as the new CAPDM codec in the following simulations.

We now compare the SEGSNR performance between CAPDM.FF and CVSD at 16 kb/s in Fig. 15. It is observed that CAPDM.FF achieves a large SEGSNR gain over CVSD.

We also evaluated the SEGSNR's of CAPDM.FF at different packet sizes, from 100 to 900 samples (from 6 to 58 ms) per packet. From Fig. 16, we observe that as the packet size of speech waveform increased, the SEGSNR decreased. This is

TABLE II
SEGSNR PERFORMANCES WITH DIFFERENT LPF CUTOFF FREQUENCIES

Cutoff Frequency	4000	3500	3000	2500	2000
SEGSNR	20.2	20.7	21.2	21.7	22.5

TABLE III
SEGSNR PERFORMANCES FOR FOUR DIFFERENT SPEECH WAVEFORMS

Sample	F1	F2	M1	M2	Avg.
CAPDM.FF	21.2	23.0	21.3	20.8	21.6
CAPDM	17.97	19.60	19.84	18.64	19.01
CVSD	13.03	14.30	14.60	14.53	14.12
ADPCM(32 Kbit/s)	27.50	29.20	27.00	30.10	28.50

TABLE IV
MOS TESTS FOR FOUR DIFFERENT SPEECH WAVEFORMS

Sample	F1	F2	M1	M2	Avg.
CAPDM.FF	4.10	3.98	4.10	3.52	3.98
CAPDM.FF with APF	4.25	4.0	4.5	4.0	4.19
CAPDM	3.94	3.82	3.82	3.30	3.72
CAPDM with APF	4.15	3.9	4.2	3.75	4.0
CVSD	3.30	3.34	3.20	2.76	3.15
ADPCM(32 Kbit/s)	4.57	4.25	4.20	4.15	4.29

because human speech is nonstationary and a fixed pitch period is not valid for a large packet size. We suggest that packet sizes under 250 are better choices. In Table II, we list SEGSNR's of CAPDM.FF for different low-pass-filtered (using a third-order Butterworth LPF) speech signals. At a cutoff frequency of 3 KHz, the SEGSNR is about 21 dB, and this is the cutoff frequency used hereafter.

Simulation results with four different speech samples for CAPDM, CAPDM.FF, CVSD, and ADPCM are listed in Table III. On average, CAPDM.FF is 2.5 dB better than CAPDM and 6 dB better than CVSD. In Table IV, we list informal mean opinion score (MOS) listening test results for four simulated speech samples. It is observed that CAPDM.FF achieves an MOS at about 4.0 points. The improvement over CAPDM through pitch prediction is 0.26 points in MOS.

B. Adaptive Postfiltering

Although the SEGSNR of CAPDM.FF at 16 kb/s exceeds 21 dB on average, there is still small but perceivable quantization noise in the decoded speech samples. In order to further improve its subjective quality, noise reduction techniques are considered. The perceptual weighting filter proposed by Atal [23], [24] was first used to reshape the quantization noise spectrum according to the masking effect of human ear perception. Unfortunately, this method is not applicable to CAPDM.FF, since its quantization noise is not white. As a result, we use the adaptive postfiltering method [28] instead.

Fig. 17 shows the block diagram of a CAPDM codec with adaptive postfiltering (APF). The transfer function of the APF is given below

$$H(z) = \frac{1}{\sum_{k=1}^{N} \alpha^k a_k z^{-k}}, \qquad 0 < \alpha < 1 \tag{13}$$

where $a_k's$ are the coefficients of adaptive prediction filter. The APF attenuates noise in the spectral valleys of a speech signal and allows more noise in the spectral formants. This method has been used successfully in ADPCM and APC speech coders [28]. As APF reduces perceivable noise, it also attenuates the high-frequency components of output speech and causes a muffling effect. A high-frequency booster is utilized to brighten the postfiltered speech. Finally, an LPF with 3-KHz cutoff frequency is used to suppress out-of-band noise. We observe from Table IV that the subjective quality of CAPDM after postfiltering is enhanced by about 0.2 in MOS score.

V. CAPDM PERFORMANCE EVALUATION IN A NOISY CHANNEL

Due to the difficulties in maintaining the continuity of codec coefficients at the beginning of each packet, we isolate transmission packets by reinitializing the coefficients of both the stepsize estimation unit and the adaptive prediction units for each packet. These coefficients include the last reference stepsize, ($T = 20$), the basic reference stepsize, ($S = 30$), and the last predictor coefficients [$a_i(1) = 0.6$ and $a_i(i) = 0, 2 \le i \le N$]. These reinitialized coefficients are determined experimentally. Through packet isolation, the codec coefficients in different packets are independent of each other and the only relationship between two consecutive packets is in the use of the first packet for waveform reconstruction in the second packet. Our simulation shows that packet isolation greatly enhances the stability of CAPDM.FF, and the performance degradation caused by it is only 1.5 dB.

For a lost packet, a pitch-based waveform substitution technique is used to generate a replacement packet in order to enhance the codec performance. Several waveform substitution techniques have been proposed to alleviate the packet loss problem for PCM [3]. It was shown that pitch-based waveform substitution methods outperform other candidates [4]. With CAPDM.FF, pitch periods are available from correctly received packets. Therefore, we can recover missing packets by replacing the lost packets according to known pitch periods.

A. Packet Lost Case

We simulated the codec performance at packet loss rates up to 10%. The lost packets are selected randomly according to an uniform distribution. First, we evaluate the effect of packet isolation on CAPDM. The simulated SEGSNR's versus packet loss rate are plotted in Fig. 18. It is observed that as codec coefficients are reinitialized for each packet, the codec is more resistant to packet loss.

Next, we compare the performances of CAPDM, CAPDM.FF, and CVSD. In Fig. 19, we plot their SEGSNR's at different packet loss rates. With packet isolation, the SEGSNR performance of CAPDM.FF is above 15 dB when packet loss rate is below 3%. Compared with CAPDM, CAPDM.FF is about 3 dB better than CAPDM at these packet loss rates. As for CVSD, it is very robust in channel errors and its performance degradation is much slower. However, at packet loss rates below 3%, CAPDM.FF performs much better than CVSD.

Next, we compare the SEGSNR performances of CAPDM.FF at 16 kb/s and G.721 ADPCM at 32 kb/s in

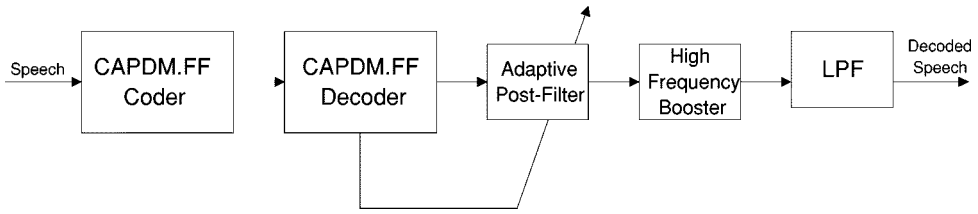


Fig. 17. CAPDM codec with adaptive postfiltering.

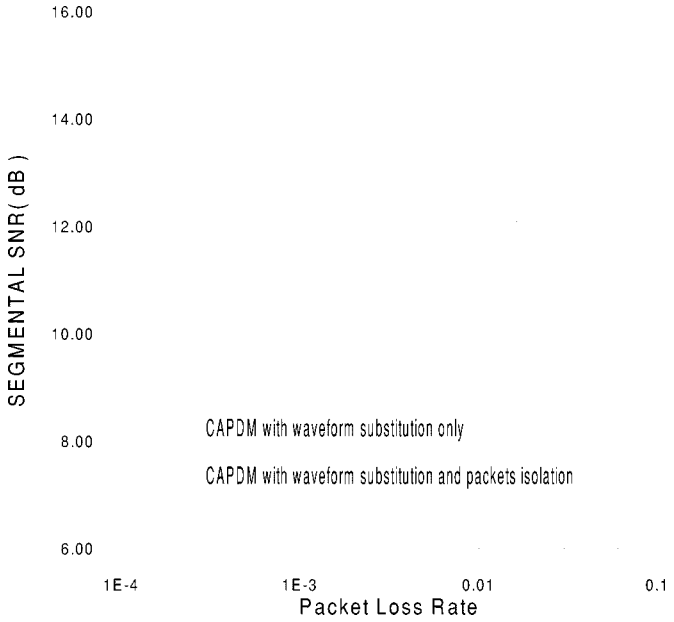


Fig. 18. The improvement by isolating packets.

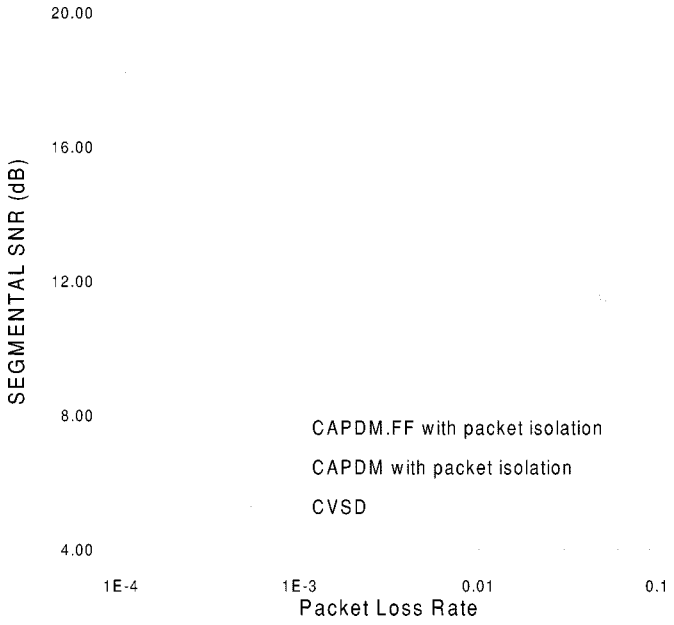


Fig. 19. SEGSNR's for CAPDM.FF, CAPDM, and CVSD under packet loss.

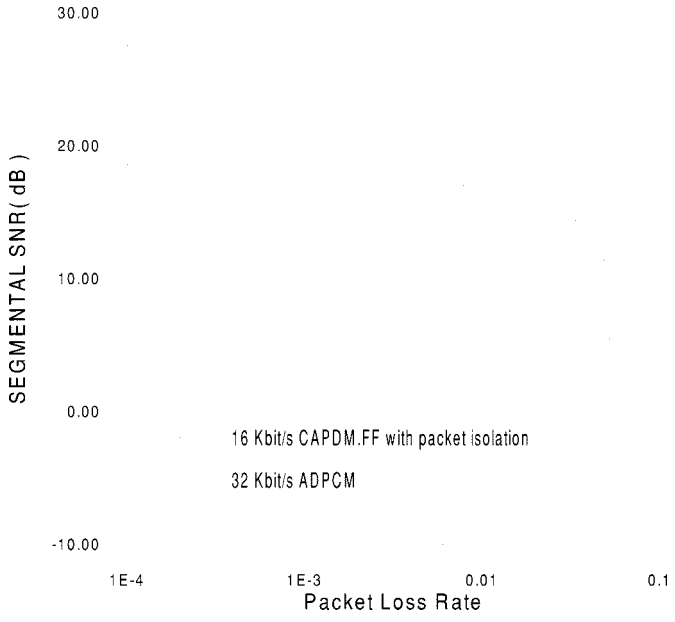


Fig. 20. SEGSNR's for 16-kb/s CAPDM.FF and 32-kb/s ADPCM under packet loss.

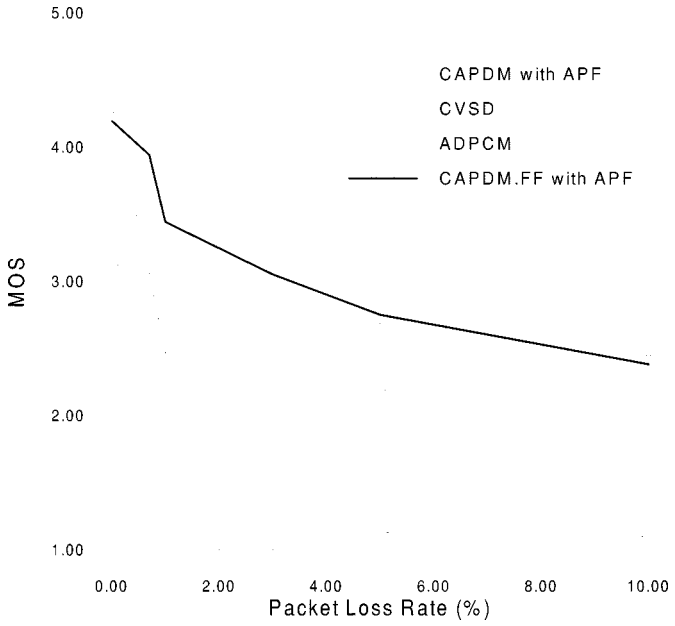


Fig. 21. MOS evaluation for CAPDM.FF with APF and CAPDM with APF, CVSD, and ADPCM under packet loss.

Fig. 20. The SEGSNR performance of ADPCM is much worse than CAPDM.FF when packet loss rate is increased above 3%. In Fig. 21, we compare the MOS performances of CAPDM.FF (with APF), CAPDM (with APF), CVSD, and ADPCM. The MOS performance degradation of ADPCM is much more severe than the other three codecs. There is a kind of impulse noise heard when ADPCM packets are lost. For CAPDM.FF with APF, its MOS score degrades smoothly and remains above 3.0 points at 3.0% packet loss rate.

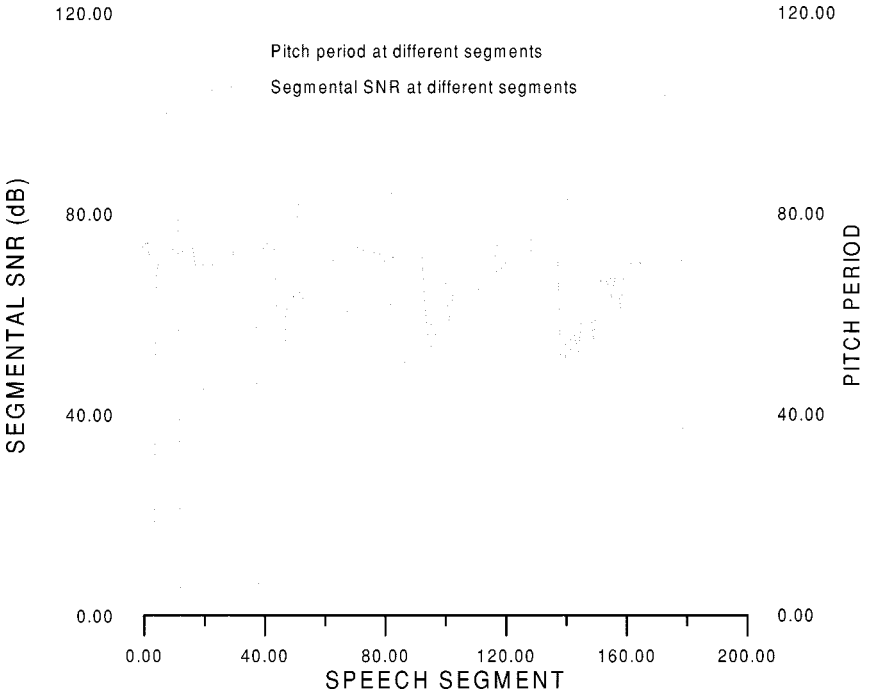


Fig. 22. The effect of one packet loss at different segments.

In Fig. 22, we demonstrate the effect of packet loss at different speech segments on CAPDM.FF codec performance for one speech sentence example. Each speech segment is selected in term for the lost packet. It is observed that CAPDM.FF codec performance becomes worse when the lost packet occurs in the transition region of unvoiced-to-voiced segments. This is because the lost packet results in a loss of the first few pitch periods and the immediately following packet could not predict the correct speech waveform accordingly.

From this observation, we feel that some error protection mechanisms might be used in these unvoiced-to-voiced transition regions, e.g., ARQ can be used when packets are lost during these transitions. On the other hand, ARQ can be used for all the lost packets without increasing too much traffic. In Fig. 23, we show the CAPDM.FF performance improvements through selected ARQ and nonselected ARQ, compared with CAPDM.FF without ARQ. The performance improvements through using nonselected ARQ are significant even at 10% packet loss rate. In this case, we only need to reserve about additional 10% channel capacity to achieve this improved performance.

In Table V, we summarize the SEGSNR performances of CAPDM.FF with different LPF bandwidth. In the packet lost case, the codec coefficients are adjusted for packet loss situations and are different from the coefficients used in the error-free case. There is about 2-dB degradation when CAPDM.FF codec is specifically designed for a packet loss channel.

B. Complexity

The complexity of CAPDM.FF is evaluated according to the algorithms presented in Sections II and III with an eleventh-order short-term filter, a fifth-order long-term filter, and the AMDF pitch detection algorithm. The estimated computation complexity of different codecs in MIPS are summarized in Table VI [28].

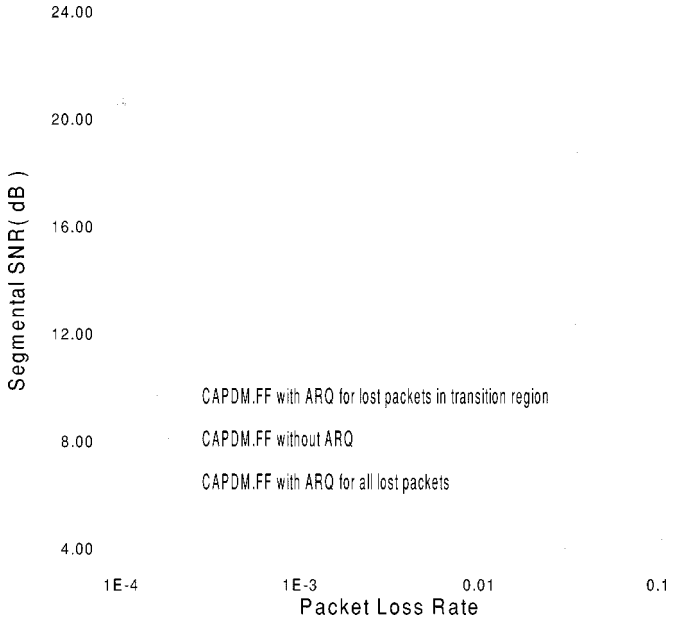


Fig. 23. The CAPDM.FF SEGSNR performance improvements by using ARQ under packet loss.

TABLE V
SUMMARY OF SEGSNR OF CAPDM.FF WITH DIFFERENT COEFFICIENTS

Coeff.	LPF BW	f1	f2	m1	m2	Avg.
Error Free	4000	20.2	22.0	20.2	19.6	20.5
Case	3500	20.7	22.5	20.8	20.3	21.1
	3000	21.2	23.0	21.3	20.8	21.6
	2500	21.7	23.7	21.8	21.3	22.1
	2000	22.5	24.6	22.6	22.0	22.9
Packet Lost	4000	17.4	19.6	18.5	18.9	18.6
Case	3500	17.8	20.1	18.8	19.2	19.0
	3000	18.2	20.7	19.2	19.6	19.4
	2500	18.8	21.3	19.8	20.1	20.0
	2000	19.6	22.1	20.4	20.7	20.7

TABLE VI
COMPLEXITY OF CAPDM.FF AND THE SELECTED ALGORITHMS

Algorithm	Bit Rate	Pith	Prediction	Stepsize	MIPS
CAPDM.FF	16 K	6.3	1.4	0.1	7.8
CAPDM	16 K	0.0	1.4	0.1	1.5
ADPCM	32 K				2.0
LD-CELP	16 K				19.0
RPE-LPC (GSM)	13 K				6.0

For CAPDM.FF, most computation efforts (about 80%) are spent in pitch detection. A CAPDM codec without pitch detection requires only 1.5 MIPS. The complexity of CAPDM.FF could be reduced by using a smaller observation window size for pitch detection. The complexity of CAPDM.FF is less than LD-CELP by about 11 MIPS and is roughly equal to a GSM speech coder. As the window size for pitch detection is halved, the CAPDM.FF complexity can be reduced to below that of a GSM codec.

VI. CONCLUSIONS

In this paper, we presented a new CAPDM codec structure at 16 kb/s. This codec, a hybrid of delta modulation (DM) and APC, consists of a stepsize estimation unit, a pitch detection unit, and an adaptive prediction unit. It determines the encoded bit by comparing the two estimates for bits 0 and 1 with the incoming speech sample and picking the closer one. This encoding procedure combines the features of one-step look ahead, syllabic companding, instantaneous companding, and adaptive prediction. From the discussions we know that CAPDM.FF has a flexible structure and the codec coefficients can be adjusted to achieve near toll quality speech encoding. Our simulation results demonstrated that the adoption of pitch prediction in CAPDM improves its performance by about 2 dB.

For PCN applications, the encoded speech samples are packetized for digital transmissions. Both the influences and the recovery mechanisms of lost packets are considered in the paper. We found that both waveform substitution and packet isolation are critical to the codec performance under packet loss. These techniques might be similarly useful for other APC coding schemes. The perceptual effect [27] of CAPDM.FF is also studied in the paper. Through the use of APF, the MOS performance of CAPDM.FF is improved by about 0.2 points. Our simulations show that the performance of CAPDM.FF with APF degrades smoothly even when packet loss rate approaches 10%. We also learned that the codec performance is more sensitive to certain locations of packet loss, especially when the packet loss happens in the regions of unvoiced-to-voiced transition. An ARQ scheme is suggested to protect this transition region and improve the codec performance. Another related topic is voice activity detection [13] which can be exploited to increase PCN channel capacity. This topic will be treated in the future.

REFERENCES

[1] R. Potter, "Personal communications for the mass market," *Telecommun. Int. Ed.*, vol. 24, no. 9, 1990.

[2] *32 kbit/s Adaptive Differential Pulse Code Modulation Recommendation G.721*, Nov. 1988.

[3] D. J. Goodman, G. B. Lockhart, O. J. Wasem, and W.-C. Wong, "Waveform substitution techniques for recovering missing speech segments in packet voice communications," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, Dec. 1986.

[4] O. J. Wasem, D. J. Goodman, C. A. Dvorak, and H. G. Page, "The effect of waveform substitution on the quality of PCM packet communications," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, Mar. 1988.

[5] N. Erdol, C. Castelluccia, and A. Zilouchian, "Recovery of missing speech packets using the short-time energy and zero-crossing measurements," *IEEE Trans. Speech Audio Processing*, vol. 1, July 1993.

[6] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, Jan. 1992.

[7] A. M. Kondoz, *Digital Speech-Coding for Low Bit Rate Communication Systems*, U.K.: Wiley, 1994.

[8] R. Ramachandran and P. Kabal, "Pitch prediction filters in speech coding," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 467–478, 1989.

[9] J. Gruber and L. Strawczynski, "Subjective effects of variable delay and speech clipping in dynamically managed voice systems," *IEEE Trans. Commun.*, vol. COM-33, pp. 801–808, Aug. 1985.

[10] J. Gruber and N. Le, "Performance requirements for integrated voice/data networks," *IEEE J. Select. Areas Commun.*, vol. SAC-1, pp. 981–1005, Dec. 1983.

[11] M. Decina and D. Vlack, "Voice by the packet?," *IEEE J. Select. Areas Commun.*, vol. SAC-1, pp. 961–962, Dec. 1983.

[12] O. G. Jaffe, "Reconstruction of missing packets of PCM and ADPCM encoded speech," M.Sc. thesis, MIT, Cambridge, MA, June 1986.

[13] S. Hatamian, "Enhanced speech activity detection for mobile telephony," in *IEEE 42nd VTS Conf.*, vol. 1, 1992, pp. 159–162.

[14] R. P. Ramachandran and P. Kabal, "Stability and performance analysis of pitch filter in speech coders," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1937–945, July 1987.

[15] C. H. Liu, "A new controlled adaptive prediction delta speech coder for wireless PCN applications," Master's thesis, National Chiao Tung Univ., Taiwan, R.O.C., June 1996.

[16] C.-C. Huang, "Controlled adaptive prediction delta modulation in mobile radio voice communications," in *IEEE VTC*, 1988, pp. 158–162.

[17] V. R. Viswanathan and A. L. Higgins, "Design of a robust LPC coder for speech transmission over 9.6 Kbit/s noisy channels," *IEICE Trans. Commun.*, pp. 663–673, April 1982.

[18] N. Scheinberg *et al.*, "A one-stage look-ahead algorithm for delta modulation," *IEEE Trans. Commun.*, pp. 861–863, July 1984.

[19] D. L. Cohn and J. L. Melsa, "The residual encoder-an improved ADPCM system for speech digitization," *IEEE Trans. Commun.*, pp. 935–941, Sep. 1975.

[20] M. Melnick, "Intelligibility performance of a variable slope delta modulation," in *Proc. IEEE Int. Conf. Comm.*, June 1973, pp. 46.5–46.7.

[21] C.-C. Huang, "Computer simulation and evaluation of mobile radio voice communication systems," Ph.D. dissertation, Univ. Calif., Berkeley, CA, 1984.

[22] W. Daumer, "Subjective evaluation of several efficient speech coders," *IEEE Trans. Commun.*, vol. COM-30, p. 655, Apr. 1982.

[23] B. S. Atal and M. R. Schroeder, "Predictive coding of speech signals and subjective error criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, June 1979.

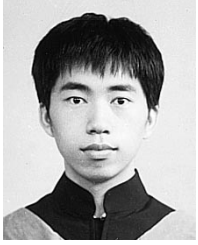
[24] B. S. Atal, "Predictive coding of speech at low bit rates," *IEEE Trans. Commun.*, vol. COM-30, April 1982.

[25] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

[26] D. J. Goodman, "Cellular packet communications," *IEEE Trans. Commun.*, vol. 38, Aug. 1990.

[27] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[28] A. S. Spanias, "Speech coding: A tutorial review," *Proc. IEEE*, vol. 82, Oct. 1994.



Chia-Horng Liu was born in Taiwan, R.O.C. He received the B.S. and M.S. degrees from National Chiao Tung University, Hsinchu, Taiwan, both in communication engineering, in 1994 and 1996, respectively. He is currently working toward the Ph.D. degree at National Chiao Tung University.

Chia-Chi Huang was born in Taiwan, R.O.C. He received the B.S. degree from National Taiwan University, Taiwan, in 1977 and the M.S. and Ph.D. degrees from the University of California, Berkeley, both in electrical engineering, in 1980 and 1984, respectively.

From 1984 to 1988, he was an RF and Communication System Engineer with the Corporate Research and Development Center, General Electric Company, Schenectady, NY, where he worked on mobile radio communications. From 1989 to 1992, he was with the IBM T. J. Watson Research Center, Yorktown Heights, NY, as a Research Staff Member working on indoor radio communications. Since September 1992, he has been with the Department of Communications, National Chiao Tung University, Hsinchu, Taiwan, as an Associate Professor.