Error Protection to IS-96 Variable Rate CELP Speech Coding[1]

Sue-Jean Li*, Min-Chin Yang*, Pao-Chi Chang*, and Hong Shen Wang**

**Department of Electrical Engineering, National Central University*
Chung-Li, Taiwan, ROC
E-mail: pcchang@roger.ee.ncu.edu.tw

***Media Technology Corporation, Taipei, Taiwan, ROC*
E-mail: hswang@roger.ee.ncu.edu.tw

Abstract

IS-96 describes a variable rate code-excited linear prediction (CELP) speech coder in code division multiple access (CDMA) digital cellular environment. In this work we focus on the study of the error protection methods in CDMA cellular environment. We find that although the error protection technique described in IS-96 can improve the speech quality in noisy channels considerably, the error concealment effect in IS-96 can be further improved by a better bit protection pattern without increasing the processing complexity.

I. Introduction

Telecommunications Industry Association (TIA) and Electronic Industry Association (EIA) adopted IS-96 [1] as an interim standard of speech coding for the CDMA digital cellular system, defined by IS-95 [2], [3]. IS-96 speech coder, originally proposed by Qualcomm Inc. as QCELP [4], [5], is a variable rate coder in which any one of four rates, 1, 1/2, 1/4, 1/8, corresponding to 8, 4, 2, 0.8 kbps, respectively, is chosen for each 20 ms frame based on the speech activities. Variable rate coding in CDMA environment is a very efficient technique which reduces not only the required bandwidth but also the average transmitting power which generates interferences to other users [6], [7].

In the cellular environment, the bit error rate can not be guaranteed to be low enough so that the error effect can be totally ignored due to the uncontrolled characteristics of fading channels. In general, there are two error recovery methods for the digital transmission, the error detection and retransmission, as well as the forward error correction (FEC) [8], [9]. For realtime applications, such as voice conversations, the delay resulting from retransmission is usually not tolerable. Thus only the forward error correction method is considered in this work. Further more, in a practical speech coding system, it may not be possible to use FEC to protect all data except most important bits. Therefore, FEC alone only protects a small part of total data from limited number of bit errors.

At higher bit error rates, these uncorrected errors can still deteriorate the speech quality significantly. In addition to FEC, IS-96 proposes an error concealment scheme, referred as the detection-erasure method, which detects errors by a cyclic redundancy code (CRC) on the whole frames or from the front stages of the receiver and marks frames with uncorrected errors to be erased frames. All data in an erased frame are treated as unreliable and discarded. The speech coder uses the information from the previous frame to construct the speech with less energy and flatter spectrum. This can reduce the annoying pop and ping sounds in the synthesized speech.

In this work, we study the error protection scheme of IS-96 and propose a better bit protection pattern which yields better quality in most cases. We measure the bit error sensitivity of each CELP parameter, and find that the bit protection pattern used in IS-96 is not necessarily the best. We adjust the protection pattern based on the bit error sensitivities and perform corresponding modifications in error concealment. As a result, both segmental SNR (SSNR) and subjective listening tests confirm the improvement of the modification. In addition, this modification does not increase any hardware and software complexity.

In section II, we describe general aspects of IS-96 variable rate speech coder. In particular, its error protection and concealment method is described in section III. We propose the improved bit protection scheme in section IV. The simulation results are discussed in section V. Finally, conclusions are given in section VI.

II. IS-96 Variable Rate CELP

A. General Description

The IS-96 variable rate CELP uses the analysis-by-synthesis approach to extract and quantize the speech parameters and the residual signals. The input speech is at first sampled at a rate of 8000 samples per second and is quantized to a uniform PCM format with 14 bits. The speech codec encoding procedure determines the input parameters for the decoders with the criterion of minimizing the

[1] This work was supported by the National Science Council, under Grant NSC-84-2221-E-008-006.

perceptual distortion between the synthesis speech and the original speech. The encoding procedure also includes quantizing the parameters and packing them into data packets for transmission.

B. Variable Rate Coding

The speech codec dynamically selects one of four data rates every 20 ms, depending on the speech activities. It makes an initial rate selection based on the frame energy $R(0)$, defined as $R(0)=\sum_{n=0}^{L-1} S^2(n)$, and a set of three thresholds, $T_1(B_i)$, $T_2(B_i)$, and $T_3(B_i)$, where the three thresholds are determined based upon a second order polynomial of the background noise level B_i every frame. The background noise level B_i is computed for the i-th frame from the information in the previous frame as the following.

$$B_i = \min\left\{R(0)_{prev},\ 5059644,\ \max\left(1.00547B_{i-1},\ B_{i-1}+1\right)\right\}$$

Note that the background noise level can only increase slightly but drop abruptly in time. To determine the rate in a frame $R(0)$ is compared with the three thresholds. If $R(0)$ is greater than $T_3(B_i)$, Rate 1 is selected. If $R(0)$ is greater than $T_2(B_i)$ but less than $T_3(B_i)$, Rate 1/2 is selected. If $R(0)$ is greater than $T_1(B_i)$ but less than $T_2(B_i)$, Rate 1/4 is selected. Otherwise, Rate 1/8 is selected.

C. Encoder

The speech is segmented into 20 ms frames, consisting of 160 samples. The linear prediction coding (LPC) parameters are updated once per frame. The number of bits used to encode the LPC parameter is a function of the selected data rate. The pitch parameters are updated at a rate of four, two, one, and zero times per frame for Rate 1, 1/2, 1/4 and 1/8 respectively. Similarly, the codebook parameters are updated eight, four, two, and one times, respectively.

The LPC formant synthesis filter removes the short term redundancies in the speech. The LPC coefficients are obtained from the autocorrelation function. Then the LPC coefficients are transformed into line spectrum pair (LSP) frequencies for the reasons of the good quantization, interpolation, and stability properties of LSPs. The pitch filter removes the long term redundancies in the speech. The pitch lag, L, is quantized from 17 to 143 samples using 7 bits for each pitch update. The pitch gain, b, is represented by 3 bits and ranges from 0 to 2. A codebook (CB) is used to vector quantize the residual signal. The excitation codebook consists of 128 code vectors. The 8 appropriate codebook gains in dB are predicted by the last codebook subframe predictor memories.

D. Decoder

When the receiver receives a packet, which consists of a frame of speech, it will determine the transmission rate and assess the quality of the received packet. If the transmission rate can not be satisfactorily determined, the multiplex sublayer informs the receiver speech codec of an erasure. In addition, the receiver speech codec may declare an erasure when it receives a Rate 1 packet with bits errors exceeding one detected by the BCH code.

The speech decoder structure is shown in Figure 1. If the receiver receives a no-error packet, it synthesizes the speech as follows. First a vector is taken from the codebook. For Rate 1/8, a pseudorandom vector is generated. For all other rates, a vector specified by an index I is taken from the codebook, which is a table of vectors. This vector is multiplied by a gain G, and then is filtered by the long-term pitch synthesis filter. The output is filtered by the LPC formant synthesis filter to reconstruct the speech signal.

III. IS-96 Error Protection and Error Concealment Method

IS-96 provides both FEC and detection-erasure methods for the error protection. At Rate 1, 11 parity check bits are added to each frame to provide error correction and detection for the 18 most perceptually significant bits. A BCH(28,18) code, a shortened BCH(31,21) code, which has the minimum distance of 5, is used to protect these 18 bits.

In IS-96, the 18 most significant bits are defined as the MSBs of the 10 quantized LSPs and the second MSBs[2] of the 8 quantized log codebook gains. As we described in section I, both FEC and detection-erasure methods have improvements in noisy channel conditions. To make the error protection method robust to different error patterns, the BCH code is used in the way to correct one error and detect up to 3 errors.

To provide the error detection capability for these bits other than the 18 bits, a 12-bit outer CRC is used to protect the entire frame of Rate 1, which is already BCH coded. For Rate 1/2, an 8-bit CRC is used instead. If the received packet quality is not good enough, which is determined from the multiplex sublayer of the receiver or CRC errors, the speech codec can perform the error concealment by the parameters of the current or the previous frames .

The receiving speech codec checks Rate 1 packets for bit errors, shown as Fig. 2, before the synthesis procedure is performed. If the CRC code shows no error in a packet, the received frame is decoded as a normal Rate 1 frame. In certain cases, the decoder may receive a Rate 1 packet with bit errors. If the CRC detects errors and the BCH code indicates no error or only 1 bit in error, the bit in error is corrected and the LSP and codebook data, which are partially protected by the BCH code, are used as a normal Rate 1. The pitch lag is repeated from the last pitch subframe of the

[2] The first MSBs are the sign bits which show less impact to the quality deterioation.

previous frame. The pitch gain is first saturated at 0.75 and then decayed toward 0. The decay factor is 0.75, 0.5, 0.25, and 0 for the first, second, third, and fourth or more consecutive such packets.

If the BCH code indicates more than 1 bits in error, the speech decoder treats the frame as an erasure. In an erasure frame, the memories in the LSP predictors are multiplied by 0.90625 and LSP frequencies are regenerated from these memories. The previous pitch lag is used with the pitch gain first saturated at 0.9 and then decayed towards 0. The decay factor is 0.9, 0.6, 0.3, and 0 for the first, second, third, and fourth or more consecutive such packets. The codebook index is randomly chosen and the current codebook gain in dB is set to $\lfloor \hat{G}_l \times 0.7 \rfloor$, where $\hat{G}_l$ is the previous codebook gain in dB, and $\lfloor x \rfloor$ indicates the largest integer less than or equal to x.

IV. Improved Bit Protection Pattern

To verify the significance of the 18 bits protected by a BCH code in IS-96, we perform the fixed-bit position inversion analysis [10] to locate the most perceptually significant bits. In this experiment, we assume that a fixed bit position in a frame is always decoded with error and then compute the decoded speech SSNR reduction. The result for a male speaker is shown as in Figure 3. We observe that the MSBs of LSPs and pitch gains have the most significant impact to SSNR reduction. In particular, errors in MSBs of LSPs results in pop and ping sounds which even affect the recognition of speech contents seriously. In contrast, although errors in pitch lags have impact to SSNR reduction, no particular bits dominate the quality reduction. The subjective listening tests also show that its impact to the speech quality is not as serious as the SSNR reduction indicates.

Although the second MSBs of codebook gains have 6 to 8 dB SSNR reduction, subjectively it only results in lower volume sounds, and almost has no effect to the speech recognition. On the other hand, the MSBs of the pitch gain yield more serious SSNR reduction, and the decoded speech generates very annoying crack sounds.

To verify the above observation, we perform the following error propagation experiments which measure the error impacts from an erroneous frame to its consecutive frames. A male speech sequence of 18.25 sec in length is used in the tests. We add errors to a fixed bit in frame 1, 11, 21, ..., and measure the average SSNR reduction of the current and following frames. Fig.4 shows the error propagation property of MSB of LSPi, Fig.4(a) is the SSNR in the case of no error. Fig.5 shows the impact of the error in pitch gains in 4 pitch subframes. Fig.6 shows the impact of the error in CB gain second MSB in 8 codebook subframes.

We observe that LSP parameters have most significant error impact to the current and following frames. We also note that the pitch gain has more significant impact than the CB gain. In all cases the error propagation effect reduce substantially after 6 frames.

IS-96 chooses 10 MSBs of LSP and 8 second MSBs of CB gains from 8 subframes as the 18 most significant bits. The possible reasons are that the LSP and the CB gain are coded with prediction in the coding process. In the decoder the error of these parameters may affect the internal states of the predictors, and hence the errors will propagate to succeeding frames. Based on the above fixed position bit inversion analysis and the error propagation experiments, however, we find that pitch gains have more serious impact than CB gains in terms of the subjective quality. Thus we propose a new protection bit pattern which include 10 MSBs of LSP as well as 4 MSBs and 4 second MSBs of the pitch gains of the 4 subframes in a frame. In other words, the BCH code is now used to protect the LSP and the pitch parameters instead of the LSP and the codebook parameters as in the IS-96.

Consequently, the error concealment procedure needs minor modifications to reflect the new protection pattern. In the decoder, an erasure packet is still treated the same as in IS-96 for error concealment. For a Rate 1 packet with bit errors that the error is corrected by the BCH code, the LSP and pitch parameters are used as a normal Rate 1 packet. However, the unprotected codebook gain is attenuated whose value in dB is set to $\lfloor \hat{G}_l \times 0.7 \rfloor$, where $\hat{G}_l$ is the previous codebook gain in dB, and the codebook index is chosen randomly.

V. Experiment Results

We perform simulations to compare the speech quality processed by IS-96 and the new bit protection scheme. The test sequences include both male and female, English and Mandarin sentences.

In a fading channel, errors are likely generated in a burst mode. However, data interleaving and randomization in the encoder can break the burst errors into random errors in the interleaving range. Thus, in these simulations errors are generated randomly in a range. We perform experiments at 1%, 3%, 5%, and 7% random bit error rate in the error interval of 2, 4 and 6 frames. This corresponds to different burst lengths with the two-frame interleaver used in IS-95.

To perform meaningful results, the error bursts are generated around voice active frames. SSNR are computed as the average of SSNRs of the error frames and succeeding six frames to include the error propagation property. Three error protection schemes are tested. They are (1) LSP and CB gain protection without error concealment (LCB), (2) LSP and CB gain protection with error concealment as in IS-96 (IS-96), and (3) LSP and pitch gain protection with error concealment (LP-con), as proposed in this paper.

The results are shown in Table 1. In most cases LP-con have higher SSNR than IS-96. Informal subjective listening tests also confirm that LP-con yields better speech quality. At relatively high error rates, e.g., 3% or higher, the error concealment technique improves the quality significantly, compared to the scheme without error concealment. At low error rate, e.g., 1%, however, error correction without error concealment has even higher SSNR than that with error concealment. The is due to the fact that once a frame is classified as an erasure, whole data in a frame are discarded even though most of them are correct.

VI. Conclusion

The speech codec needs a special design to be used in fading channels which generate errors in a time-varying fashion. We performed a study on the error protection based on IS-96, and propose a new protection pattern which can further improve the coded speech quality. The new scheme only needs to change the bit protection pattern and minor modifications in the error concealment procedure. No overhead in the computational complexity is needed.

VII. REFERENCES

[1] "Speech Service Option Standard for Wideband Spread Spectrum Digital Cellular System", TIA/EIA Interim Standard IS-96, Telecommunications Industry Association, April 1994.

[2] "Mobile Station Compatibility Standard for Dual Mode Wideband Spread Spectrum Cellular System", TIA/EIA Interim Standard IS-95, Telecommunications Industry Association, 1993.

[3] J. S. Lee, "Overview of the Technical Basis of Qualcomm's CDMA Cellular Telephone System Design, — A View of North American TIA/EIA IS-95", IEEE Singapore ICCS '94. Conference Proceedings. vol. 2, pp. 353-358, 1994.

[4] W. Gardner, P. Jacobs and C. Lee, "QCELP: A Variable Rate Speech Coder for CDMA Digital Cellular", Speech and Audio Coding for Wireless and Network Applications, Kluwer Academic Publishers, pp. 85-92, 1993.

[5] J. H. Park and Y. M. Kim, "Analysis and Optimization of Speech Coder Algorithm for CDMA Digital Cellular", IEEE Singapore ICCS '94. Conference Proceedings. vol. 3, pp. 870-874, 1994.

[6] A. Gersho and E. Paksoy, "Variable Rate Speech Coding for Cellular Networks", Speech and Audio Coding for Wireless and Network Applications, Kluwer Academic Publishers, pp. 77-84, 1993.

[7] A. Gersho and E. Paksoy, "An Overview of Variable Rate Speech Coding for Cellular Networks", IEEE 1992 International Conference on Selected Topics in Wireless Communications, pp. 172-175, 1992.

[8] J. P. Campbell, Jr., V. C. Welch and T. E. Tremain, "An Expandable Error-Protected 4800 bps CELP Coder (U.S. Federal Standard 4800 bps Voice Coder)", IEEE International Conference on Acoustics, Speech and Signal Processing. vol. 2, pp. 735-738, 1989.

[9] F. F. Tzeng and P. G. Sherwood, "Error Protection for Low Rate Speech Trans-mission over a Mobile Satellite Channel", IEEE Global Telecommunications Conference and Exhibition, vol. 3, pp. 1810-1814, 1990.

[10] V. K. Varma and D. W. Lin, "Transmitting Digital Speech in a Portable Radio Environment", Advances in Speech Coding, Kluwer Academic Publishers, pp. 97-107, 1991.

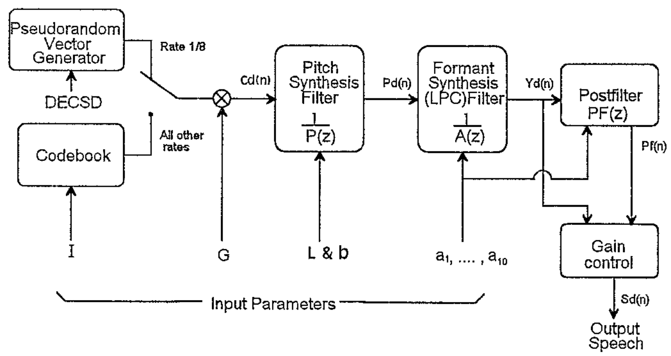

Fig.1: IS-96 CELP decoder block diagram.

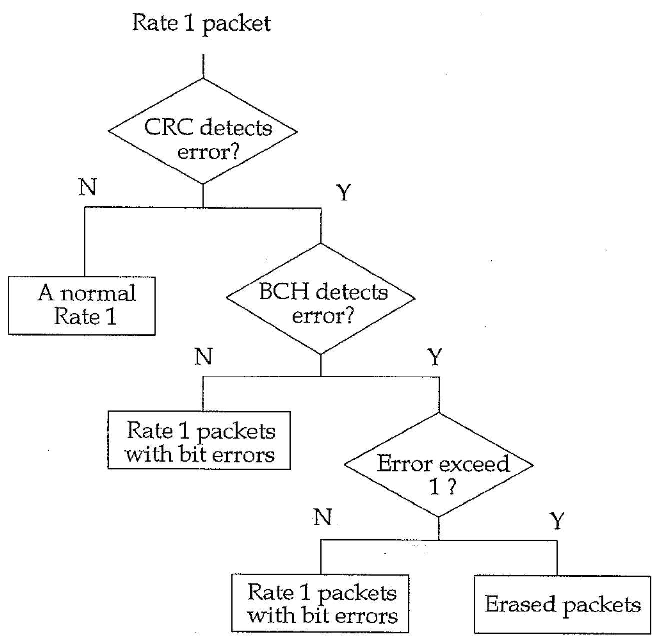

Fig.2: Quality evaluation for received packets.

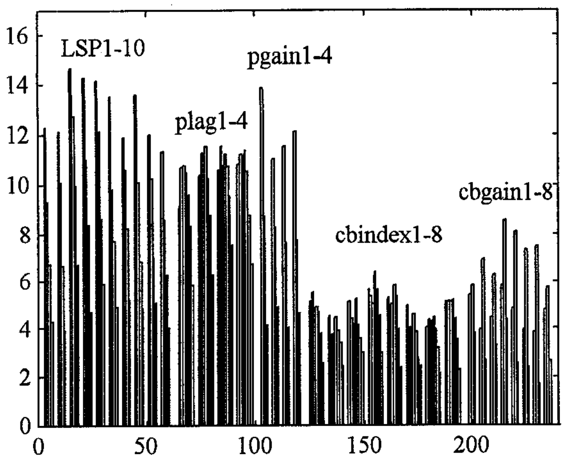

Fig.3: Bit error sensitivities for CELP parameters. Each bar in the figure represents the SSNR reduction of bit errors at the fixed position.

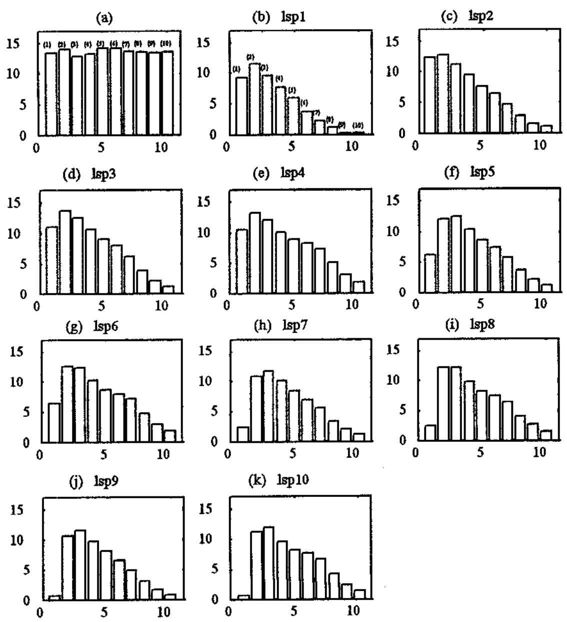

Fig.4 : (a) SSNR of frames with no bit error, (b) ~ (k) SSNR reduction of frames by LSP*i* MSB errors

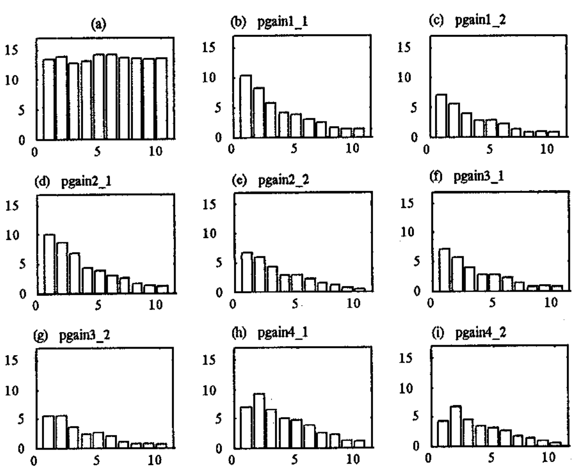

Fig.5 : (a) SSNR of frames with no bit error, (b) ~ (k) SSNR reduction of frames by pgain*i* MSB and second MSB errors

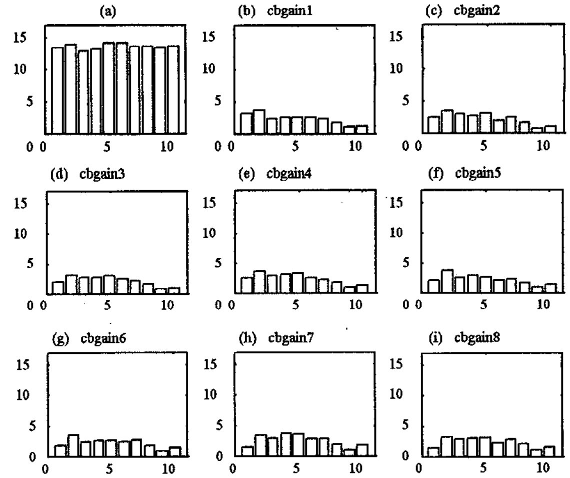

Fig.6 : (a) SSNR of frames with no bit error, (b) ~ (k) SSNR reduction of frames by cbgain*i* second MSB errors

	2 frames	4 frames	6 frames
clean	12.50	11.98	12.05
LCB	6.54	4.56	4.26
IS-96	4.55	2.78	3.06
LP-con	5.38	4.14	2.87

(a). data error rate = 1%

	2 frames	4 frames	6 frames
clean	12.50	11.98	12.05
LCB	2.71	1.84	1.20
IS-96	2.60	1.77	2.05
LP-con	2.65	1.92	2.13

(b). data error rate = 3%

	2 frames	4 frames	6 frames
clean	12.50	11.98	12.05
LCB	1.81	1.10	0.66
IS-96	2.14	1.34	1.78
LP-con	2.87	1.25	1.78

(c). data error rate = 5%

	2 frames	4 frames	6 frames
clean	12.50	11.98	12.05
LCB	1.64	0.58	0.10
IS-96	2.02	1.27	1.50
LP-con	2.40	1.42	1.91

(d). data error rate = 7%

Table 1: SSNR in burst intervals with different error concealment schemes at various BERs and burst lengths.

LCB: LSP and CB gain protection without error concealment, IS-96: LSP and CB gain protection with error concealment as in IS-96, LP-con: LSP and pitch gain protection with error concealment.