SPEECH AND CHANNEL CODING FOR THE HALF-RATE GSM CHANNEL

Ira A. Gerson, Mark A. Jasiuk, Joseph M. Nowack, and Eric H. Winter

Corporate Systems Research Laboratories
Motorola Inc.

ABSTRACT - A VSELP speech coder and accompanying channel coder designed for the half-rate GSM channel are described. This system is being adopted by ETSI for the half-rate GSM channel. The speech coder employs a novel strategy for vector quantization of the reflection coefficients (r_i), which combines high coding efficiency, low codebook search complexity, and low storage requirements. A computationally streamlined version of a zero-pole spectral noise weighting function is implemented. An adaptive pitch prefilter and an adaptive spectral postfilter are used to improve the speech coder's performance for both the tandemed and non-tandemed cases. Error protection is provided by convolutional codes. Error detection is achieved through the use of a 3 bit CRC and Window Error Detection (WED).

1. INTRODUCTION

The full-rate speech codec for the GSM system utilizes a 13 kb/s RPE-LTP speech coder (GSM, 1990), (Vary, et al, 1988). The 13 kb/s speech data is then encoded to a data rate of 22.8 kb/s using forward error correcting/detecting (FEC) codes for transmission over the radio channel. One requirement for the half-rate codec is that the combined half-rate speech codec with channel coding utilize 11.4 kb/s. In addition the speech and channel codec must maintain the full-rate codec's quality for both unimpaired radio channels and radio channels with various degrees of impairment (channel error conditions).

In allocating 5600 bps to the VSELP speech coder and 5800 bps to FEC, several techniques, designed to maximize speech coder performance given the bit rate constraint, are introduced. In Section 2 a novel method for efficient vector quantization of the reflection coefficients is described. Section 3 outlines a computationally streamlined implementation of a zero-pole spectral noise weighting function, while Section 4 details the structure of the 5600 bps VSELP speech coder and its bit allocation. The FEC design is dealt with in Section 5.

2. QUANTIZATION OF REFLECTION COEFFICIENTS

To maintain low spectral distortion for an LPC quantizer utilizing fewer bits, a novel strategy for coding the LPC filter parameters, based upon vector quantization of the reflection coefficients (r_i), is employed. The error criterion used to select a reflection coefficient vector from the codebook, is the minimization of the Itakura distance. A computationally simplified version of the Fixed Point Lattice Technique (FLAT) [Cumani, 1982), (Gerson, 1985], based upon the autocorrelation function, is used to compute the residual error out of the last lattice stage corresponding to the reflection coefficient vector being evaluated.

The 28 bits allocated for coding the LP parameters are distributed among three codebooks, to reduce the search complexity and the storage required by the reflection coefficient vector quantizer. The segments of the vector quantizer span reflection coefficients r_1-r_3, r_4-r_6, and r_7-r_{10} respectively. The bit allocations for the vector quantizer segments are:

Q_1	11 bits
Q_2	9 bits
Q_3	8 bits

The codebooks are searched sequentially. The quantization-within-recursion property inherent to lattice techniques means that the reflection coefficient vector selected at the current segment can partially compensate for the distortion due to reflection coefficient quantization at the preceding segment(s).

The concept of a prequantizer is introduced to approach the performance of an exhaustive reflection coefficient vector codebook search while avoiding the complexity such a search would require. At each segment a prequantizer, defined by fewer bits than the vector quantizer, is employed. The prequantizer size at each segment is:

P_1	6 bits
P_2	5 bits
P_3	4 bits

Each vector in the prequantizer is associated with a set of vectors in the quantization codebook. The residual error due to each prequantizer vector at a given segment is calculated and four prequantizer vectors yielding the lowest error energies are identified. The index of each selected prequantizer vector is used to calculate an offset into the vector quantizer table at which the contiguous subset of quantizer vectors associated with that prequantizer vector begins. The size of each vector quantizer subset at the k-th segment is given by:

$$S_k = \frac{2^{Q_k}}{2^{P_k}} \tag{1}$$

The four sets of quantizer vectors associated with the four best prequantizer vectors are searched for the quantizer vector which results in the lowest value of residual energy. This process is repeated sequentially at each of the segments, resulting in a significant reduction in computational complexity compared to the full search. Thus at the first segment 64 prequantizer vectors and 128 quantizer vectors are evaluated (vs. 2048 for the full search), 32 prequantizer vectors and 64 quantizer vectors are evaluated at the second segment (vs. 512), and 16 prequantizer vectors and 64 quantizer vectors are evaluated at the third segment (vs. 256). The degradation in performance relative to a full search is negligible.

The memory required for the reflection coefficient vector codebooks is reduced by storing 8 bit codes instead of the reflection coefficient values. These codes are used to extract vector element values from a high resolution, scalar lookup table with 256 entries, which contains reflection coefficient values obtained by uniform sampling of the inverse sine function.

An autocorrelation version of the FLAT algorithm, AFLAT, is used to compute the residual error energy for a reflection coefficient vector being evaluated. Like FLAT, it has the ability to partially compensate for the reflection coefficient quantization error from the previous lattice stages, when computing optimal reflection coefficients or selecting a reflection coefficient vector from a vector quantizer at the current segment. This improvement can be significant for frames with high reflection coefficient quantization error.

The optimal reflection coefficients, computed via the FLAT technique with bandwidth expansion, are converted to R(i), an autocorrelation vector, prior to quantization. Define the initial conditions for the AFLAT recursion:

$$\overline{P}_0(i) = R(i), \qquad 0 \le i \le N_p - 1 \tag{2}$$

$$\overline{V}_0(i) = R(|i+1|), \qquad 1 - N_p \le i \le N_p - 1 \tag{3}$$

Initialize k, the vector quantizer segment index:

$$k = 1 \tag{4}$$

Let $l_l(k)$ be the index of the first lattice stage in the k-th segment, and $l_h(k)$ be the index of the last lattice stage in the k-th segment. The recursion for evaluating the residual energy out of lattice stage $l_h(k)$ at the k-th segment given $\tilde{r}$, a reflection coefficient vector from the prequantizer or the reflection coefficient vector from the quantizer, is described next.

Initialize j, the index of the lattice stage, to point to the beginning of the k-th segment:

$$j = I_l(k) \tag{5}$$

Set the initial conditions P_{j-1} and V_{j-1} to:

$$P_{j-1}(i) = \overline{P}_{j-1}(i), \qquad 0 \le i \le I_h(k) - I_l(k) \tag{6}$$

$$V_{j-1}(i) = \overline{V}_{j-1}(i), \qquad -I_h(k) + I_l(k) \le i \le I_h(k) - I_l(k) \tag{7}$$

Compute the values of V_j and P_j arrays using:

$$P_j(i) = (1+\hat{r}_j^2)\, P_{j-1}(i) + \hat{r}_j \left[\, V_{j-1}(i) + V_{j-1}(-i)\right], \qquad 0 \le i \le I_h(k) - j - 1 \tag{8}$$

$$V_j(i) = V_{j-1}(i+1) + \hat{r}_j^2\, V_{j-1}(-i-1) + 2\hat{r}_j\, P_{j-1}(|i+1|), \qquad 1 + j - I_h(k) \le i \le I_h(k) - j - 1 \tag{9}$$

Increment j:

$$j = j + 1 \tag{10}$$

If $j < I_h(k)$ go to (8).

The residual error out of lattice stage $I_h(k)$, given the reflection coefficient vector $\hat{r}$, is computed using (8):

$$E_r = P_{I_h(k)}(0) \tag{11}$$

Given the AFLAT recursion outlined, the residual error due to each vector from the prequantizer at the k-th segment is evaluated, the four subsets of quantizer vectors to search are identified, and residual error due to each quantizer vector from the selected four subsets is computed. The index of $\tilde{r}$, the quantizer vector which minimized E_r over all the quantizer vectors in the four subsets, is encoded with Q_k bits.

If $k < 3$ then the initial conditions for the recursion at segment k+1 need to be computed. Set j, the lattice stage index, equal to:

$$j = I_l(k) \tag{12}$$

Compute:

$$\overline{P}_j(i) = (1+\tilde{r}_j^2)\, \overline{P}_{j-1}(i) + \tilde{r}_j \left[\, \overline{V}_{j-1}(i) + \overline{V}_{j-1}(-i)\right], \qquad 0 \le i \le N_p - j - 1 \tag{13}$$

$$\overline{V}_j(i) = \overline{V}_{j-1}(i+1) + \tilde{r}_j^2\, \overline{V}_{j-1}(-i-1) + 2\tilde{r}_j\, \overline{P}_{j-1}(|i+1|), \qquad 1 + j - N_p \le i \le N_p - j - 1 \tag{14}$$

Increment j,

$$j = j + 1 \tag{15}$$

If $j \le I_h(k)$ go to (13).

Increment k, the vector quantizer segment index:

$$k = k + 1 \tag{16}$$

If $k \le 3$ go to (5). Otherwise the indices of the reflection coefficient vectors for the three segments have been chosen, and the reflection coefficient vector quantization is complete.

3. SPECTRAL NOISE WEIGHTING FILTER

To more precisely control the shape of the quantization noise, a spectral noise weighting function, based upon a tenth order zero-pole transfer function, is employed [Chen et al, 1992). The parameters of this function were selected to achieve a good compromise between tandemed and non-tandemed speech coder performance. To realize the benefits of this flexible noise weighting function while reducing the computational complexity which would result from implementing it directly, a computationally streamlined version of the weighting function was implemented. A single tenth order combined spectrally noise weighted LP synthesis filter is used to model the cascade of the LP synthesis filter and the zero and pole spectral noise weighting filters.

Define $\hat{H}(z)$, the interim spectrally noise weighted synthesis filter, as:

$$\hat{H}(z) = \left[1-\sum_{i=1}^{N_p} \alpha_i \lambda_1^i z^{-i}\right]^{-1} \left[1-\sum_{i=1}^{N_p} \alpha_i \lambda_2^i z^{-i}\right]\left[1-\sum_{i=1}^{N_p} \alpha_i \lambda_3^i z^{-i}\right]^{-1} \qquad (17)$$

where α_i's are the direct form LP coefficients. λ_1, λ_2, and λ_3 are the spectral noise weighting filter parameters which determine the amount of spectral noise weighting to be applied. The value of each λ_j is bounded by:

$$0 \leq \lambda_j \leq 1 \qquad (18)$$

$\hat{h}(n)$, the impulse response of $\hat{H}(z)$, is computed for N samples (where N is the subframe length), and the autocorrelation sequence of $\hat{h}(n)$ is calculated using:

$$R_{\hat{h}}(k) = \sum_{n=k}^{N-1} \hat{h}(n)\,\hat{h}(n-k),\ 0 \leq k \leq N_p \qquad (19)$$

From $R_{\hat{h}}(k)$, the coefficients of the combined spectrally noise weighted synthesis filter, $\tilde{H}(z)$, are computed via the Levinson recursion once per frame:

$$\tilde{H}(z) = \left[1-\sum_{i=1}^{N_p} \tilde{\alpha}_i z^{-i}\right]^{-1} \qquad (20)$$

The method for the spectral noise weighting filter coefficient update mimics how the direct form LPC filter coefficients are updated at subframes of a frame. That is, either interpolated or uninterpolated spectral noise weighting filter coefficients are used at each subframe, dependent on the soft interpolation bit (see Section 4).

4. 5600 BPS VSELP CODER

The 5.6 kb/s VSELP coder employs sub-sample resolution lags, spectral and harmonic noise weighting, frame lag trajectory based lag encoding, and multimode frame classification [Gerson & Jasiuk, 1992). The coder uses a 20 msec frame with four 5 msec subframes per frame given input speech sampled at 8 kHz. Table 1 shows the bit allocations for the 5.6 kb/s coder.

The ten reflection coefficients are vector quantized using a total of 28 bits per frame (see Section 3). A soft interpolation bit indicates whether the direct form LPC coefficients are interpolated for that frame. This bit is determined by evaluating both the interpolated and uninterpolated sets of direct form filter coefficients for the frame and selecting the set which minimizes the residual energy over the whole frame.

PARAMETER	BITS PER SUBFRAME	BITS PER FRAME
LPC coefficients		28
Soft Interpolation		1
energy - $R_q(0)$		5
Mode		2
Excitation Code I	7	28
Excitation Code H	7	28
G_S-P_0 Code	5	20
TOTAL	19	112

Mode = 0

PARAMETER	BITS PER SUBFRAME	BITS PER FRAME
LPC coefficients		28
Soft Interpolation		1
energy - $R_q(0)$		5
Mode		2
Lag (Subframe 1)	8	8
Lag Delta Codes (Subframes 2, 3, 4)	4	12
Excitation Code J	9	36
G_S-P_0 Code	5	20
TOTAL	-	112

Mode = 1, 2 or 3

Table 1. Bit Allocations for the 5.6 kb/s coder

Two bits are used to specify the voicing mode, based on the open-loop long term predictor (LTP) gain. If voicing mode 0 is selected, corresponding to an unvoiced speech frame, the adaptive codebook and the trained 9 bit VSELP codebook are replaced by two trained 7 bit VSELP codebooks. If mode 1, 2, or 3 is chosen, the adaptive codebook is employed, using frame lag trajectory based lag encoding (Gerson & Jasiuk, 1992). The frame lag trajectory, defined to be a sequence of subframe lags within the frame, is globally optimized, open-loop, over all the subframes in the frame and allows for a closed-loop lag search at each subframe to refine the lag estimate. The first subframe's lag is coded independently using 8 bits, and the lags at subframes 2, 3, and 4 are delta coded relative to the preceding subframe's coded lag with 4 bits. A lag deviation of -8 to +7 allowable lag levels from the current subframe to the next is allowed by the delta encoder. Three allowable lag levels, centered about the lag defined by the selected frame lag trajectory, are evaluated closed-loop for each subframe. The allowable lags are given by Table 2.

LAG RANGE	RESOLUTION
21 to 22 2/3	1/3
23 to 34 5/6	1/6
35 to 49 2/3	1/3
50 to 89 1/2	1/2
90 to 142	1

Table 2. Lag Quantizer

The overall frame energy is coded with five bits in 2 dB steps, for 64 dB of dynamic range. The coder gains are vector quantized relative to the frame energy with 5 bits (Gerson & Jasiuk, 1990), (Gerson & Jasiuk, 1992). The G_S-P_0 code represents the vector quantized values of two parameters, G_S and P_0, which are transformations of the excitation gains. G_S represents the subframe energy relative to the frame energy, $R_q(0)$. P_0 represents the relative contribution of the first excitation source (adaptive codebook for modes 1, 2, 3, first VSELP codebook for mode 0) to the total excitation and is bounded in value by 0 and 1.0. Separate G_S-P_0 codebooks are derived for each of the four voicing modes.

The speech coder's performance was further enhanced by the inclusion of an adaptive spectral postfilter and an adaptive pitch prefilter (Gerson & Jasiuk, 1990). By applying a moderate amount of spectral postfiltering and pitch prefiltering, the quality of the processed speech is noticeably improved for both the non-tandemed and the tandemed cases.

5. FORWARD ERROR CONTROL

The forward error control is achieved using (1) decoding error prevention and (2) bad frame mitigation. To prevent decoding errors, forward error control coding and interleaving are used. To avoid the severe degradation in speech quality due to a decoding error in the most important speech bits, two methods of bad frame detection are used in conjunction with a bad frame mitigation strategy.

Of the 112 speech coder bits per frame, 95 are encoded along with 3 CRC bits and 6 tail bits. The optimal free distance R=1/3, K=7 convolutional code encodes the 104 bits. The CRC bits are encoded at rate 1/3 while the other 101 bits are coded at rate 1/2 via puncturing. The same code is applied regardless of the speech coder mode, although the order of the speech bits into the convolutional encoder depends on whether the frame is voiced (modes 1, 2 and 3) or unvoiced (mode 0). The convolutional encoder outputs a bit stream of 211 bits which is then appended with the 17 uncoded bits. The resulting 228 bits are then dispersed over 4 TCH-HS time slots so that each slot contains bits from two speech coder frames. The interleaver mapping seeks to prevent a burst of channel errors from affecting adjacent bits in the decoder. In the receiver, soft-decision Viterbi decoding is performed.

After decoding, the 3 bit CRC is used to check the validity of the 22 most important speech bits assigned to the end of the trellis. In addition, the Window Error Detection (WED) technique is performed over the same 22 decoded bits as well as the CRC bits. A reliability estimate of the decoded bits is compared with an adaptive threshold, where the estimate is the difference in the path metrics between pairs of candidate paths along the surviving path in the Viterbi decoder. The minimum path metric difference along the surviving path within the 25 bit window is called the WED output. If the WED output is less than the threshold, a bad frame is declared. The threshold adapts based on an estimate of the channel conditions, obtained by averaging the WED outputs over the last 30 frames. If either the WED or the CRC detects an unreliable window, then a bad frame masking strategy is invoked. The bad frame masking strategy consists of an 8-state machine which may repeat speech parameters from previously decoded good speech frames or may attenuate or mute the speech depending on the decoding history.

6. CONCLUSIONS

An 11400 bps speech coder and FEC system have been described. This speech coder, when combined with forward error control, meets all of the requirements for the half-rate GSM speech codec. This combined speech/channel codec is currently being adopted for the GSM half-rate standard.

7. REFERENCES

Chen, J.-H. ; Cox, R.V.; Lin, Y.-C. ; Jayant, N.; Melchner, M. J. (1992) "A Low-Delay CELP Coder for CCITT 16 kb/s Speech Coding Standard," IEEE Journal on Selected Areas in Communications, vol. 10, no. 5, pp. 830-849, June 1992.

Cumani, A. (1982) "On a Covariance-Lattice Algorithm for Linear Prediction," Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 651-654, May 1982.

Gerson, I. A. (1985) "Method and Means of Determining Coefficients for Linear Predictive Coding," U. S. Patent #4,544,919, Oct. 1985.

Gerson, I. A. & Jasiuk, M. A. (1990) "Vector Sum Excited Linear Prediction (VSELP) Speech Coding at 8 kb/s," Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 461-464, April 1990.

Gerson, I. A. & Jasiuk, M.A. (1992) "Techniques for Improving the Performance of CELP-Type Speech Coders," IEEE Journal on Selected Areas in Communications, vol. 10, no. 5, pp. 858-865, June 1992.

GSM 6.10 "GSM Full Rate Speech Transcoding," ETSI, Jan. 1990.

Vary,P.; Hellwig,K.; Hofmann, R.; Sluyter, R.; Galland, C. & Rosso, M. (1988) "Speech Codec for the European Mobile Radio System," Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 227-230, April 1988.