Mermelstein, P. “Speech and Channel Coding for North American TDMA Cellular Systems”
Mobile Communications Handbook
Ed. Suthan S. Suthersan
Boca Raton: CRC Press LLC, 1999

Speech and Channel Coding for North American TDMA Cellular Systems

Paul Mermelstein
INRS-Télécommunications
University of Québec

28.1 Introduction

The goals of this chapter are to give the reader a tutorial introduction and high-level understanding of the techniques employed for speech transmission by the IS-54 digital cellular standard. It builds on the information provided in the standards document but is not meant to be a replacement for it. Separate standards cover the control channel used for the setup of calls and their handoff to neighboring cells, as well as the encoding of data signals for transmission. For detailed implementation information the reader should consult the most recent standards document [9].

IS-54 provides for encoding bidirectional speech signals digitally and transmitting them over cellular and microcellular mobile radio systems. It retains the 30-kHz channel spacing of the earlier advanced mobile telephone service (AMPS), which uses analog frequency modulation for speech transmission and frequency shift keying for signalling. The two directions of transmission use frequencies some 45 MHz apart in the band between 824 and 894 MHz. AMPS employs one channel per conversation in each direction, a technique known as frequency division multiple access (FDMA). IS-54 employs time division multiple access (TDMA) by allowing three, and in the future six, simultaneous transmissions to share each frequency band. Because the overall 30-kHz channelization of the allocated 25 MHz of spectrum in each direction is retained, it is also known as a FDMA-TDMA system. In contrast, the later IS-95 standard employs code division multiple access (CDMA) over bands of 1.23 MHz by combining several 30-kHz frequency channels.

Each frequency channel provides for transmission at a digital bit rate of 48.6 kb/s through use of differential quadrature-phase shift key (DQPSK) modulation at a 24.3-kBd channel rate. The channel is divided into six time slots every 40 ms. The full-rate voice coder employs every third time slot and utilizes 13 kb/s for combined speech and channel coding. The six slots provide for an eventual half-rate channel occupying one slot per 40 ms frame and utilizing only about 6.5 kb/s for each call. Thus, the simultaneous call carrying capacity with IS-54 is increased by a factor 3(factor 6 in the future) above that of AMPS. All digital transmission is expected to result in a reduction in transmitted power. The resulting reduction in intercell interference may allow more frequent reuse of the same frequency channels than the reuse pattern of seven cells for AMPS. Additional increases in erlang capacity (the total call-carrying capacity at a given blocking rate) may be available from the increased trunking efficiency achieved by the larger number of simultaneously available channels. The first systems employing dual-mode AMPS and TDMA service were put into operation in 1993.

In 1996 the TIA introduced the IS-641 enhanced full rate codec. This codec consists of 7.4 kb/s speech coding following the algebraic code-excited linear prediction (ACELP) technique [7], and 5.6 kb/s channel coding. The 13 kb/s coded information replaces the combined 13 kb/s for speech and channel coding introduced by the IS-54 standard. The new codec provides significant enhancements in terms of speech quality and robustness to transmission errors. The quality enhancement for clear channels results from the improved modeling of the stochastic excitation by means of an algebraic **codebook** instead of the two trained VSELP codebooks. Improved robustness to transmission errors is achieved by employing predictive quantization techniques for the linear-prediction filter and gain parameters, and increasing the number of bits protected by forward error correction.

28.2 Modulation of Digital Voice and Data Signals

The modulation method used in IS-54 is $\pi/4$ shifted differentially encoded quadrature phase-shift keying (DPSK). Symbols are transmitted as changes in phase rather than their absolute values. The binary data stream is converted to two binary streams X_k and Y_k formed from the odd- and even-numbered bits, respectively. The quadrature streams I_k and Q_k are formed according to

$$\begin{aligned} I_k &= I_{k-1}\cos\left[\Delta\phi\left(X_k, Y_k\right)\right] - Q_{k-1}\sin\left[\Delta\phi\left(X_k, Y_k\right)\right] \\ Q_k &= I_{k-1}\sin\left[\Delta\phi\left(X_k, Y_k\right)\right] + Q_{k-1}\cos\left[\Delta\phi\left(X_k, Y_k\right)\right] \end{aligned}$$

where I_{k-1} and Q_{k-1} are the amplitudes at the previous pulse time. The phase change $\Delta\phi$ takes the values $\pi/4$, $3\pi/4$, $-\pi/4$, and $-3\pi/4$ for the dibit (X_k, Y_k) symbols (0,0), (0,1), (1,0) and (1,1), respectively. This results in a rotation by $\pi/4$ between the constellations for odd and even symbols.

The differential encoding avoids the problem of 180° phase ambiguity that may otherwise result in estimation of the carrier phase.

The signals I_k and Q_k at the output of the differential phase encoder can take one of five values, $0, \pm 1, \pm 1/\sqrt{2}$ as indicated in the constellation of Fig. 28.1. The corresponding impulses are applied to the inputs of the I and Q baseband filters, which have linear phase and square root raised cosine frequency responses. The generic modulator circuit is shown in Fig. 28.2. The rolloff factor α determines the width of the transition band and its value is 0.35,

$$|H(f)| = \begin{cases} 1, & 0 \le f \le (1-\alpha)/2T \\ \sqrt{1/2\{1 - \sin[\pi(2fT-1)/2\alpha]\}}, & (1-\alpha)/2T \le f \le (1+\alpha)/2T \\ 0, & f > (1+\alpha)/2T \end{cases}$$

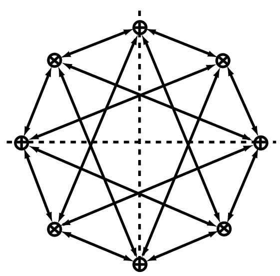

FIGURE 28.1: Constellation for $\pi/4$ shifted QPSK modulation. *Source*: TIA, 1992. Cellular System Dual-mode Mobile Station–Base Station Compatibility Standard TIA/EIA IS-54. With permission.

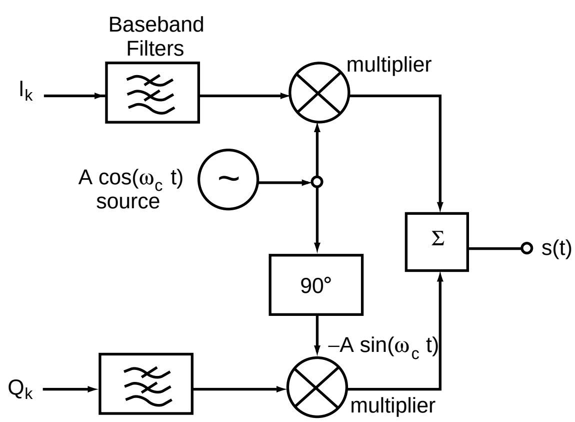

FIGURE 28.2: Generic modulation circuit for digital voice and data signals. *Source*: TIA, 1992. Cellular System Dual-mode Mobile Station–Base Station Compatibility Standard TIA/EIA IS-54.

28.3 Speech Coding Fundamentals

The IS-54 standard employs a vector-sum excited linear prediction (VSELP) coding technique. It represents a specific formulation of the much larger class of code-excited linear prediction (CELP) coders [2] that have proved effective in recent years for the coding of speech at moderate rates in the range 4–16 kb/s. VSELP provides reconstructed speech with a quality that is comparable to that available with frequency modulation and analog transmission over the AMPS system. The coding rate employed is 7.95 kb/s. Each of the six slots per frame carry 260 b of speech and channel coding information for a gross information rate of 13 kb/s. The 260 b correspond to 20 ms of real time speech, transmitted as a single burst.

For an excellent recent review of speech coding techniques for transmission, the reader is referred to Gersho, 1994 [3]. Most modern speech coders use a form of analysis by synthesis coding where the encoder determines the coded signal one segment at a time by feeding candidate excitation segments into a replica of a synthesis filter and selecting the segment that minimizes the distortion between the original and reproduced signals. Linear prediction coding (LPC) techniques [1] encode the speech signal by first finding an optimum linear filter to remove the short-time correlation, passing the signal through that LPC filter to obtain a residual signal, and encoding this residual using much fewer bits than would have been required to code the original signal with the same fidelity. In most cases the coding of the residual is divided into two steps. First, the long-time correlation due to the periodic pitch excitation is removed by means of an optimum one-tap filter with adjustable gain and lag. Next, the remaining residual signal, which now closely resembles a white-noise signal, is encoded. Code-excited linear predictors use one or more codebooks from which they select replicas of the residual of the input signal by means of a closed-loop error-minimization technique. The index of the codebook entry as well as the parameters of all the filters are transmitted to allow the speech signal to be reconstructed at the receiver. Most code-excited coders use trained codebooks. Starting with a codebook containing Guassian signal segments, entries that are found to be used rarely in coding a large body of speech data are iteratively eliminated to result in a smaller codebook that is considered more effective.

The speech signal can be considered quasistationary or stationary for the duration of the speech frame, of the order of 20 ms. The parameters of the short-term filter, the LPC coefficients, are determined by analysis of the autocorrelation function of a suitably windowed segment of the input signal. To allow accurate determination of the time-varying pitch lag as well as simplify the computations, each speech frame is divided into four 5-ms subframes. Independent pitch filter computations and residual coding operations are carried out for each subframe.

The speech decoder attempts to reconstruct the speech signal from the received information as best possible. It employs a codebook identical to that of the encoder for excitation generation and, in the absence of transmission errors, would produce an exact replica of the signal that produced the minimized error at the encoder. Transmission errors do occur, however, due, to signal fading and excessive interference. Since any attempt at retransmission would incur unacceptable signal delays, sufficient error protection is provided to allow correction of most transmission errors.

28.4 Channel Coding Considerations

The sharp limitations on available bandwidth for error protection argue for careful consideration of the sensitivity of the speech coding parameters to transmission errors. Pairwise interleaving of coded blocks and convolutional coding of a subset of the parameters permit correction of a limited number of transmission errors. In addition, a cyclic redundancy check (CRC) is used to determine whether the error correction was successful. The coded information is divided into three blocks of varying sensitivity to errors. Group 1 contains the most sensitive bits, mainly the parameters of the LPC filter and frame energy, and is protected by both error detection and correction bits. Group 2 is provided with error correction only. The third group, comprising mostly the fixed codebook indices, is not protected at all.

The speech signal contains significant temporal redundancy. Thus, speech frames within which errors have been detected may be reconstructed with the aid of previously correctly received information. A bad-frame masking procedure attempts to hide the effects of short fades by extrapolating the previously received parameters. Of course, if the errors persist, the decoded signal must be muted while an attempt is made to hand off the connection to a base station to/from which the mobile may experience better reception.

28.5 VSELP Encoder

A block diagram of the VSELP speech encoder [4] is shown in Fig. 28.3. The excitation signal is generated from three components, the output of a long term or pitch filter, as well as entries from two codebooks. A weighted synthesis filter generates a synthesized approximation to the frequency-weighted input signal. The weighted mean square error between these two signals is used to drive the error minimization process. This weighted error is considered to be a better approximation to the perceptually important noise components than the unweighted mean square error. The total weighted square error is minimized by adjusting the pitch lag and the codebook indices as well as their gains. The decoder follows the encoder closely and generates the excitation signal identically to the encoder but uses an unweighted linear-prediction synthesis filter to generate the decoded signal. A spectral postfilter is added after the synthesis filter to enhance the quality of the reconstructed speech.

The precise data rate of the speech coder is 7950 b/s or 159 b per time slot, each corresponding to 20 ms of signal in real time. These 159 b are allocated as follows: 1) short-term filter coefficients, 38 bits; 2) frame energy, 5 bits; 3) pitch lag, 28 bits; 4) codewords, 56 bits; and 5) gain values, 32 bits.

28.6 Linear Prediction Analysis and Quantization

The purpose of the LPC analysis filter is to whiten the spectrum of the input signal so that it can be better matched by the codebook outputs. The corresponding LPC synthesis filter $A(z)$ restores the short-time speech spectrum characteristics to the output signal. The transfer function of the tenth-order synthesis filter is given by

$$A(z) = \frac{1}{1 - \sum_{i=1}^{N_p} \alpha_i z^{-i}}$$

The filter predictor parameters $\alpha_1, \ldots, \alpha_{N_p}$ are not transmitted directly. Instead, a set of **reflection coefficients** $r_1, \ldots, r_{N_p}$ are computed and quantized. The predictor parameters are determined from the reflection coefficients using a well-known backward recursion algorithm [6].

A variety of algorithms are known that determine a set of reflection coefficients from a windowed input signal. One such algorithm is the fixed point **covariance lattice**, FLAT, which builds an optimum inverse lattice stage by stage. At each stage j, the sum of the mean-squared forward and backward residuals is minimized by selection of the best reflection coefficient r_j. The analysis window used is 170 samples long, centered with respect to the middle of the fourth 5-ms subframe of the 20-ms

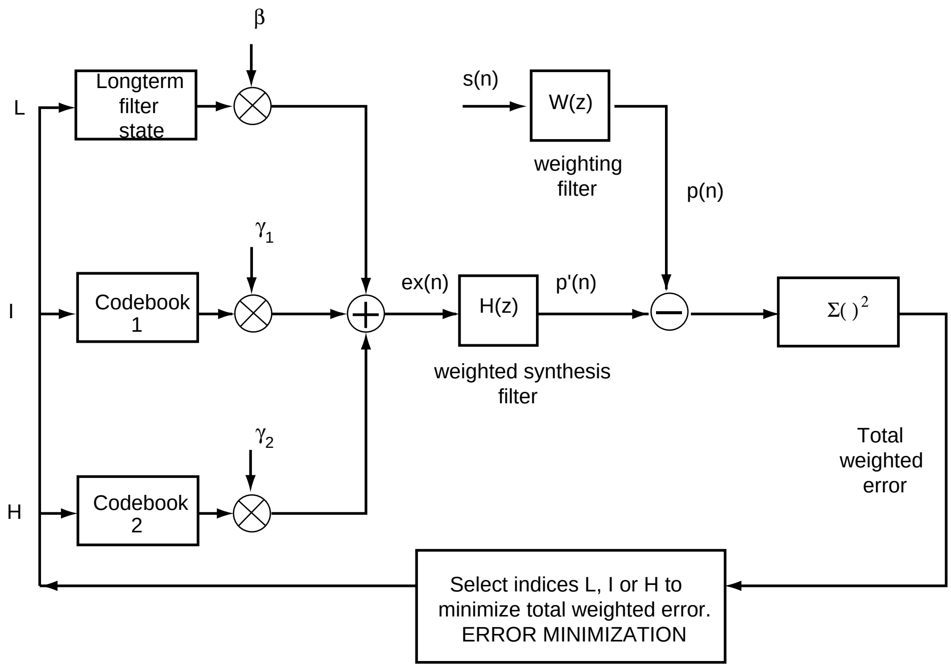

FIGURE 28.3: Black diagram of the speech encoder in VSELP. TIA. 1992. Cellular system Dual-mode Mobile Station–Base Station Compatibility Standard. TIA/EIA IS-54.

frame. Since this centerpoint is 20 samples from the end of the frame, 65 samples from the next frame to be coded are used in computing the reflection coefficient of the current frame. This introduces a lookahead delay of 8.125 ms.

The FLAT algorithm first computes the covariance matrix of the input speech for $N_A = 170$ and $N_p = 10$,

$$\phi(i,k) = \sum_{n=N_p}^{N_A-1} s(n-i)s(n-k), \qquad 0 \le i, \quad k \le N_p\ ,$$

Define the forward residual out of stage j as $f_j(n)$ and the backward residual as $b_j(n)$. Then the autocorrelation of the initial forward residual $F_0(i,k)$ is given by $\phi(i,k)$. The autocorrelation of the initial backward residual $B_0(i,k)$ is given by $\phi(i+1,k+1)$ and the initial cross correlation of the two residuals is given by $C_0(i,k) = \phi(i,k+1)$ for $0 \le i,k \le N_{p-1}$. Initially j is set to 1. The **reflection coefficient** at each stage is determined as the ratio of the cross correlation to the mean of the autocorrelations. A block diagram of the computations is shown in Fig. 28.4. By quantizing the reflection coefficients within the computation loops, reflection coefficients at subsequent stages are computed taking into account the quantization errors of the previous stages. Specifically,

$$\begin{aligned}
C'_{j-1} &= C_{j-1}(0,0) + C_{j-1}(N_p - j, N_p - j) \\
F'_{j-1} &= F_{j-1}(0,0) + F_{j-1}(N_p - j, N_p - j) \\
B'_{j-1} &= B_{j-1}(0,0) + B_{j-1}(N_p - j, N_p - j)
\end{aligned}$$

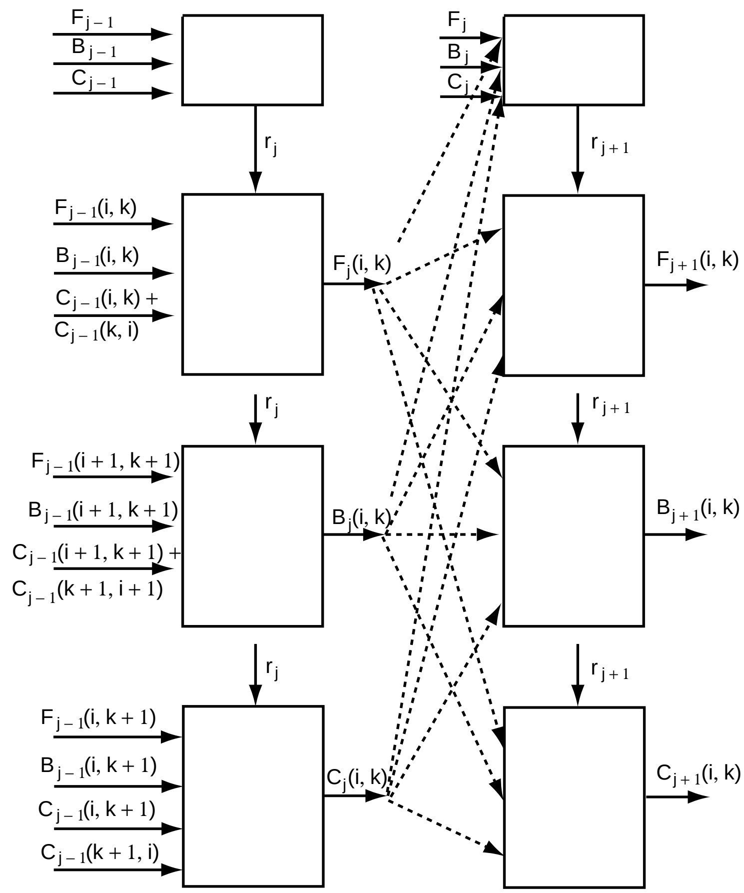

FIGURE 28.4: Block diagram for lattice covariance computations.

and

$$r_j = \frac{-2C'_{j-1}}{F'_{j-1} + B'_{j-1}}$$

Use of two sets of correlation values separated by $N_p - j$ samples provides additional stability to the computed reflection coefficients in case the input signal changes form rapidly.

Once a quantized reflection coefficient r_j has been determined, the resulting auto- and cross correlations can be determined iteratively as

$$\begin{aligned}
F_j(i,k) &= F_{j-1}(i,k) + r_j[C_{j-1}(i,k) + C_{j-1}(k,i)] + r_j^2 B_{j-1}(i,k) \\
B_j(i,k) &= B_{j-1}(i+1,k+1) + r_j[C_{j-1}(i+1,k+1) + C_{j-1}(k+1,i+1)] \\
&\quad + r_j^2 F_{j-1}(i+1,k+1)
\end{aligned}$$

and

$$\begin{aligned}
C_j(i,k) &= C_{j-1}(i,k+1) + r_j[B_{j-1}(i,k+1) + F_{j-1}(i,k+1)] \\
&\quad + r_j^2 C_{j-1}(k+1,i)
\end{aligned}$$

These computations are carried out iteratively for r_j, $j = 1, \ldots, N_p$.

28.7 Bandwidth Expansion

Poles with very narrow bandwidths may introduce undesirable distortions into the synthesized signal. Use of a binomial window with effective bandwidth of 80 Hz suffices to limit the ringing of the LPC filter and reduce the effect of the LPC filter selected for one frame on the signal reconstructed for subsequent frames. To achieve this, prior to searching for the reflection coefficients, the $\phi(i, k)$ is modified by use of a window function $w(j)$, $j = 1, \ldots, 10$, as follows:

$$\phi'(i, k) = \phi(i, k) w(|i - k|)$$

28.8 Quantizing and Encoding the Reflection Coefficients

The distortion introduced into the overall spectrum by quantizing the reflection coefficients diminishes as we move to higher orders in the reflection coefficients. Accordingly, more bits are assigned to the lower order coefficients. Specifically, 6, 5, 5, 4, 4, 3, 3, 3, 3, and 2 b are assigned to $r_1, \ldots, r_{10}$, respectively. Scalar quantization of the reflection coefficients is used in IS-54 because it is particularly simple. **Vector quantization** achieves additional quantizing efficiencies at the cost of significant added complexity.

It is important to preserve the smooth time evolution of the linear prediction filter. Both the encoder and decoder linearly interpolate the coefficients α_i for the first, second and third subframes of each frame using the coefficients determined for the previous and current frames. The fourth subframe uses the values computed for that frame.

28.9 VSELP Codebook Search

The codebook search operation selects indices for the long-term filter (pitch lag L) and the two codebooks I and H so as to minimize the total weighted error. This closed-loop search is the most computationally complex part of the encoding operation, and significant effort has been invested to minimize the complexity of these operations without degrading performance. To reduce complexity, simultaneous optimization of the codebook selections is replaced by a sequential optimization procedure, which considers the long-term filter search as the most significant and therefore executes it first. The two vector-sum codebooks are considered to contribute less and less to the minimization of the error, and their search follows in sequence. Subdivision of the total codebook into two vector sums simplifies the processing and makes the result less sensitive to errors in decoding the individual bits arising from transmission errors.

Entries from each of the two vector-sum codebooks can be expressed as the sum of basis vectors. By orthogonalizing these basis vectors to the previously selected codebook component(s), one ensures that the newly introduced components reduce the remaining errors. The subframes over which the codebook search is carried out are 5 ms or 40 samples long. An optimal search would need exploration of a 40-dimensional space. The vector-sum approximation limits the search to 14 dimensions after the optimal pitch lag has been selected. The search is further divided into two stages of 7 dimensions each. The two codebooks are specified in terms of the fourteen, 40-dimensional basis vectors stored at the encoder and decoder. The two 7-b indices indicate the required weights on the basic vectors to arrive at the two optimum codewords.

The codebook search can be viewed as selecting the three best directions in 40-dimensional space, which when summed result in the best approximation to the weighted input signal. The gains of the three components are determined through a separate error minimization process.

28.10 Long-Term Filter Search

The long-term filter is optimized by selection of a lag value that minimizes the error between the weighted input signal $p(n)$ and the past excitation signal filtered by the current weighted synthesis filter $H(z)$. There are 127 possible coded lag values provided corresponding to lags of 20–146 samples. One value is reserved for the case when all correlations between the input and the lagged residuals are negative and use of no long term filter output would be best. To simplify the convolution operation between the impulse response of the weighted synthesis filter and the past excitation, the impulse response is truncated to 21 samples or 2.5 ms. Once the lag is determined, the untruncated impulse response is used to compute the weighted long-term lag vector.

28.11 Orthogonalization of the Codebooks

Prior to the search of the first codebook, each filtered basis vector may be made orthogonal to the long-term filter output, the zero-state response of the weighted synthesis filter $H(z)$ to the long-term prediction vector. Each orthogonalized filtered basis vector is computed by subtracting its projection onto the long-term filter output from itself.

Similarly, the basis vectors of the second codebook can be orthogonalized with respect to both the long-term filter output and the first codebook output, the zero-state response of $H(z)$ to the previously selected summation of first-codebook basis vectors. In each case the codebook excitation can be reconstituted as

$$u_{k,i}(n) = \sum_{m=1}^{M} \theta_{im} v_{k,m}(n)$$

where $k = 1, 2$ for the two codebooks, $i = I$ or H the 7-b code vector received, $v_{k,m}$ are the two sets of basis vectors, and $\theta_{im} = +1$ if bit m of codeword $i = 1$ and -1 if bit m of codeword $i = 0$. Orthogonalization is not required at the decoder since the gains of the codebooks outputs are determined with respect to the weighted nonorthogonalized code vectors.

28.12 Quantizing the Excitation and Signal Gains

The three codebook gain values β, γ_1, and γ_2 are transformed to three new parameters GS, $P0$ and $P1$ for quantization purposes. GS is an energy offset parameter that equalizes the input and output signal energies. It adjusts the energy of the output of the LPC synthesis filter to equal the energy computed for the same subframe at the encoder input. $P0$ is the energy contribution of the long-term prediction vector as a fraction of the total excitation energy within the subframe. Similarly, $P1$ is the energy contribution of the code vector selected from the first codebook as a fraction of the total excitation energy of the subframe. The transformation reduces the dynamic range of the parameters to be encoded. An 8-b **vector quantizer** efficiently encodes the appropriate $(GS, P0, P1)$ vectors by selecting the vector which minimizes the weighted error. The received and decoded values β, γ_1, and γ_2 are computed from the received $(GS, P0, P1)$ vector and applied to reconstitute the decoded signal.

28.13 Channel Coding and Interleaving

The goals of channel coding are to reduce the impairments in the reconstructed speech due to transmission errors. The 159 b characterizing each 20-ms block of speech are divided into two classes, 77 in class 1 and 82 in class 2. Class 1 includes the bits in which errors result in a more significant impairment, whereas the speech quality is considered less sensitive to the class- 2 bits. Class 1 generally includes the gain, pitch lag, and more significant reflection coefficient bits. In addition, a 7-b cyclic redundancy check is applied to the 12 most perceptually significant bits of class 1 to indicate whether the error correction was successful. Failure of the CRC check at the receiver suggests that the received information is so erroneous that it would be better to discard it than use it. The error correction coding is illustrated in Fig. 28.5.

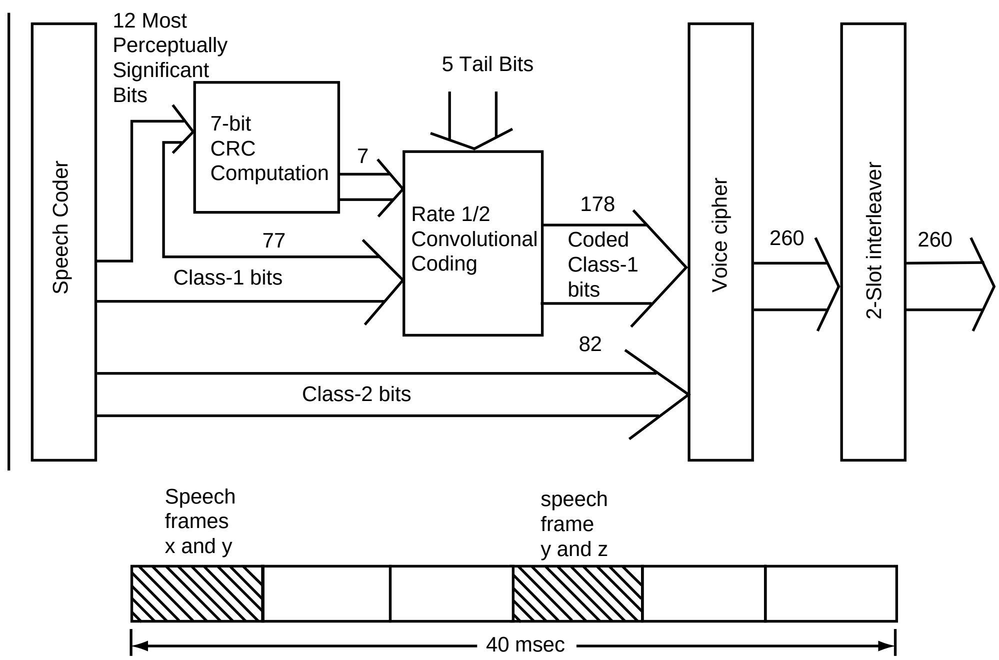

FIGURE 28.5: Error correction insertion for speech coder. Source TIA, 1992. Cellular Systems Dual-Mode Mobile Station–Base Station Compatibility Standards. TIA/EIA IS-54. With permission.

The error correction technique used is rate 1/2 convolutional coding with a constraint length of 5 [5]. A tail of 5 b is appended to the 84 b to be convolutionally encoded to result in a 178-b output. Inclusion of the tail bits ensures independent decoding of successive time slots and no propagation of errors between slots.

Interleaving the bits to be transmitted over two time slots is introduced to diminish the effects of short deep fades and to improve the error-correction capabilities of the channel coding technique. Two speech frames, the previous and the present, are interleaved so that the bits from each speech block span two transmission time slots separated by 20 ms. The interleaving attempts to separate the convolutionally coded class-1 bits from one frame as much as possible in time by inserting noncoded class-2 bits between them.

28.14 Bad Frame Masking

A CRC failure indicates that the received data is unusable, either due to transmission errors resulting from a fade, or from pre-emption of the time slot by a control message (fast associated control channel, FACCH). To mask the effects that may result from leaving a gap in the speech signal, a masking operation based on the temporal redundancy between adjacent speech blocks has been proposed. Such masking can at best bridge over short gaps but cannot recover loss of signal of longer duration. The bad frame masking operation may follow a finite state machine where each state indicates an operation appropriate to the elapsed duration of the fade to which it corresponds. The masking operation consists of copying the previous LPC information and attenuating the gain of the signal. State 6 corresponds to error sequences exceeding 100 ms, for which the output signal is muted. The result of such a masking operation is generation of an extrapolation in the gap to the previously received signal, significantly reducing the perceptual effects of short fades. No additional delay is introduced in the reconstructed signal. At the same time, the receiver will report a high frequency of bad frames leading the system to explore handoff possibilities immediately. A quick successful handoff will result in rapid signal recovery.

28.15 ACELP Encoder

The ACELP encoder employs linear prediction analysis and quantization techniques similar to those used in VSELP and discussed in Section 28.6. The frame structure of 20 ms frames and 5 ms subframes is preserved. Linear prediction analysis is carried out for every frame. The ACELP encoder uses a long-term filter similar to the one discussed in Section 28.10 and represented as an adaptive codebook. The nonpredictable part of the LPC residual is represented in terms of ACELP codebooks, which replace the two VSELP codebooks shown in Fig. 28.3.

Instead of encoding the reflection coefficients as in VSELP, the information is transformed into line-spectral frequency pairs (LSP) [8]. The LSPs can be derived from linear prediction coefficients, a 10th order analysis generating 10 line-spectral frequencies (LSF), 5 poles, and 5 zeroes. The LSFs can be vector quantized and the LPC coefficients recalculated from the quantized LSFs. As long as the interleaved order of the poles and zeroes is preserved, quantization of the LSPs preserves the stability of the LPC synthesis filters. The LSPs of any frame can be better predicted from the values calculated and transmitted corresponding to previous frames, resulting in additional advantages. The long-term means of the LSPs are calculated for a large body of speech data and stored at both the encoder and decoder. First-order moving-average prediction is then used for the mean-removed LSPs. The time-prediction technique also permits use of predicted values for the LSPs in case uncorrectable transmissions errors are encountered, resulting in reduced speech degradation. To simplify the vector quantization operations, each LSP vector is split into 3 subvectors of dimensions 3, 3, and 4. The three subvectors are quantized with 8, 9, and 9 bits respectively, corresponding to a total bit assignment of 26 bits per frame for LPC information.

28.16 Algebraic Codebook Structure and Search

Algebraic codebooks contain relatively few pulses having nonzero values leading to rapid search of the possible innovation vectors, the vectors which together with the ACB output form the excitation of the LPC filter for the current subframe. In this implementation the 40-position innovation vector contains only four nonzero pulses and each can take on only values $+1$ and -1. The 40 positions are divided into four tracks and one pulse is selected from each track. The tracks are generally equally spaced but differ in their starting value, thus the first pulse can take on positions 0, 5, 10, 15, 20, 25, 30, or 35 and the second has possible positions 1, 6, 11, 16, 21, 26, 31, or 36. The first three pulse positions are coded with 3 bits and the fourth pulse position (starting positions 3 or 4) with 4 bits, resulting in a 17-bit sequence for the algebraic code of each subframe.

The algebraic codebook is searched by minimizing the mean square error between the weighted input speech and the weighted synthesized speech over the time span of each subframe. In each case the weighting is that produced by a perceptual weighting filter that has the effect of shaping the spectrum of the synthesis error signal so that it is better masked by spectrum of the current speech signal.

28.17 Quantization of the Gains for ACELP Encoding

The adaptive codebook gain and the fixed (algebraic) codebook gains are vector quantized using a 7-bit codebook. The gain codebook search is performed by minimizing the mean-square of the weighted error between the original and the reconstructed speech, expressed as a function of the adaptive codebook gain and a fixed codebook correction factor. This correction factor represents the log energy difference between a predicted gain and an estimated gain. The predicted gain is computed using fourth-order moving-average prediction with fixed coefficients on the innovation energy of each subframe. The result is a smoothed energy profile even in the presence of modest quantization errors. As discussed above in case of the LSP quantization, the moving-average prediction serves to provide predicted values even when the current frame information is lost due to transmission errors. Degradations resulting from loss of one or two frames of information are thereby mitigated.

28.18 Channel Coding for ACELP Encoding

The channel coding and interleaving operations for ACELP speech coding are similar to those discussed in Section 28.13 for VSELP coding. The number of bits protected by both error-detection (parity) and error-correction convolutional coding is increased to 48 from 12. Rate 1/2 convolutional coding is used on the 108 more significant bits, 96 class-1 bits, 7 CRC bits and the 5 tail bits of the convolutional coder, resulting in 216 coded class-1 bits. Eight of the 216 bits are dropped by puncturing, yielding 208 coded class-1 bits which are then combined with 52 nonprotected class-2 bits. As compared to the channel coding of the VSELP encoder, the numbers of protected bits is increased and the number of unprotected bits is reduced while keeping the overall coding structure unchanged.

28.19 Conclusions

The IS-54 digital cellular standard specifies modulation and speech coding techniques for mobile cellular systems that allow the interoperation of terminals built by a variety of manufacturers and systems operated across the country by a number of different service providers. It permits speech communication with good quality in a transmission environment characterized by frequent multipath fading and significant intercell interference. Generally, the quality of the IS-54 decoded speech is better at the edges of a cell than the corresponding AMPS transmission due to the error mitigation resulting from channel coding. Near a base station or in the absence of significant fading and interference, the IS-54 speech quality is reported to be somewhat worse than AMPS due to the inherent limitations of the analysis–synthesis model in reconstructing arbitrary speech signals with limited bits. The

IS-641 standard coder achieves higher speech quality, particularly at the edges of heavily occupied cells where transmission errors may be more numerous. At this time no new systems following the IS-54 standard are being introduced. Most base-stations have been converted to transmit and receive on the IS-641 standard as well and use of IS-54 transmissions is dropping rapidly. At the time of its introduction in 1996 the IS-641 coder represented the state of the art in terms of toll quality speech coding near 8 kb/s, a significant improvement over the IS-54 coder introduced in 1990. These standards represent reasonable engineering compromises between high performance and complexity sufficiently low to permit single-chip implementations in mobile terminals.

Both IS-54 and IS-641 are considered second generation cellular standards. Third generation cellular systems promise higher call capacities through better exploitation of the time-varying transmission requirements of speech conversations, as well as improved modulation and coding in wider spectrum bandwidths that achieve similar bit-error ratios but reduce the required transmitted power. Until such systems are introduced, the second generation TDMA systems can be expected to provide many years of successful cellular and personal communications services.

Defining Terms

Codebook: A set of signal vectors available to both the encoder and decoder.

Covariance lattice algorithm: An algorithm for reduction of the covariance matrix of the signal consisting of several lattice stages, each stage implementing an optimal first-order filter with a single coefficient.

Reflection coefficient: A parameter of each stage of the lattice linear prediction filter that determines 1) a forward residual signal at the output of the filter-stage by subtracting from the forward residual at the input a linear function of the backward residual, also 2) a backward residual at the output of the filter stage by subtracting a linear function of the forward residual from the backward residual at the input.

Vector quantizer: A quantizer that assigns quantized vectors to a vector of parameters based on their current values by minimizing some error criterion.

References

[1] Atal, B.S. and Hanauer, S.L., Speech analysis and synthesis by linear prediction of the speech wave. *J. Acoust. Soc. Am.*, 50, 637–655, 1971.

[2] Atal, B.S. and Schroeder, M., Stochastic coding of speech signals at very low bit rates. *Proc. Int. Conf. Comm.*, 1610–1613, 1984.

[3] Gersho, A., Advances in speech and audio compression. *Proc. IEEE*, 82, 900–918, 1994.

[4] Gerson, I.A. and Jasiuk, M.A., Vector sum excited linear prediction (VSELP) speech coding at 8 kbps. *Int. Conf. Acoust. Speech and Sig. Proc.*, ICASSP90, 461–464, 1990.

[5] Lin S. and Costello, D., *Error Control Coding: Fundamentals and Application*, Prentice Hall, Englewood Cliffs, NJ, 1983.

[6] Makhoul, J., Linear prediction, a tutorial review. *Proc. IEEE*, 63, 561–580, 1975.

[7] Salami, R., Laflamme, C., Adoul, J.P., and Massaloux, D., A toll quality 8 kb/s speech codec for the personal communication system (PCS). *IEEE Trans. Vehic. Tech.*, 43, 808–816, 1994.

[8] Soong, F.K. and Juang, B.H., Line spectrum pair (LSP) and speech data compression. *Proc. ICASSP'84*, 1.10.1–1.10.4, 1984.

[9] Telecommunications Industry Association, EIA/TIA Interim Standard, Cellular System Dual-mode Mobile Station–Base Station Compatibility Standard IS-54B, TIA/EIA, Washington, D.C., 1992.

Further Information

For a general treatment of speech coding for telecommunications, see N.S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice Hall, Englewood, NJ, 1984. For a more detailed treatment of linear prediction techniques, see J. Markel and A. Gray, *Linear Prediction of Speech*, Springer–Verlag, NY, 1976.