ECE 797:

Speech and Audio Processing

Hand-out for Lecture #11

Thursday, April 1, 2004

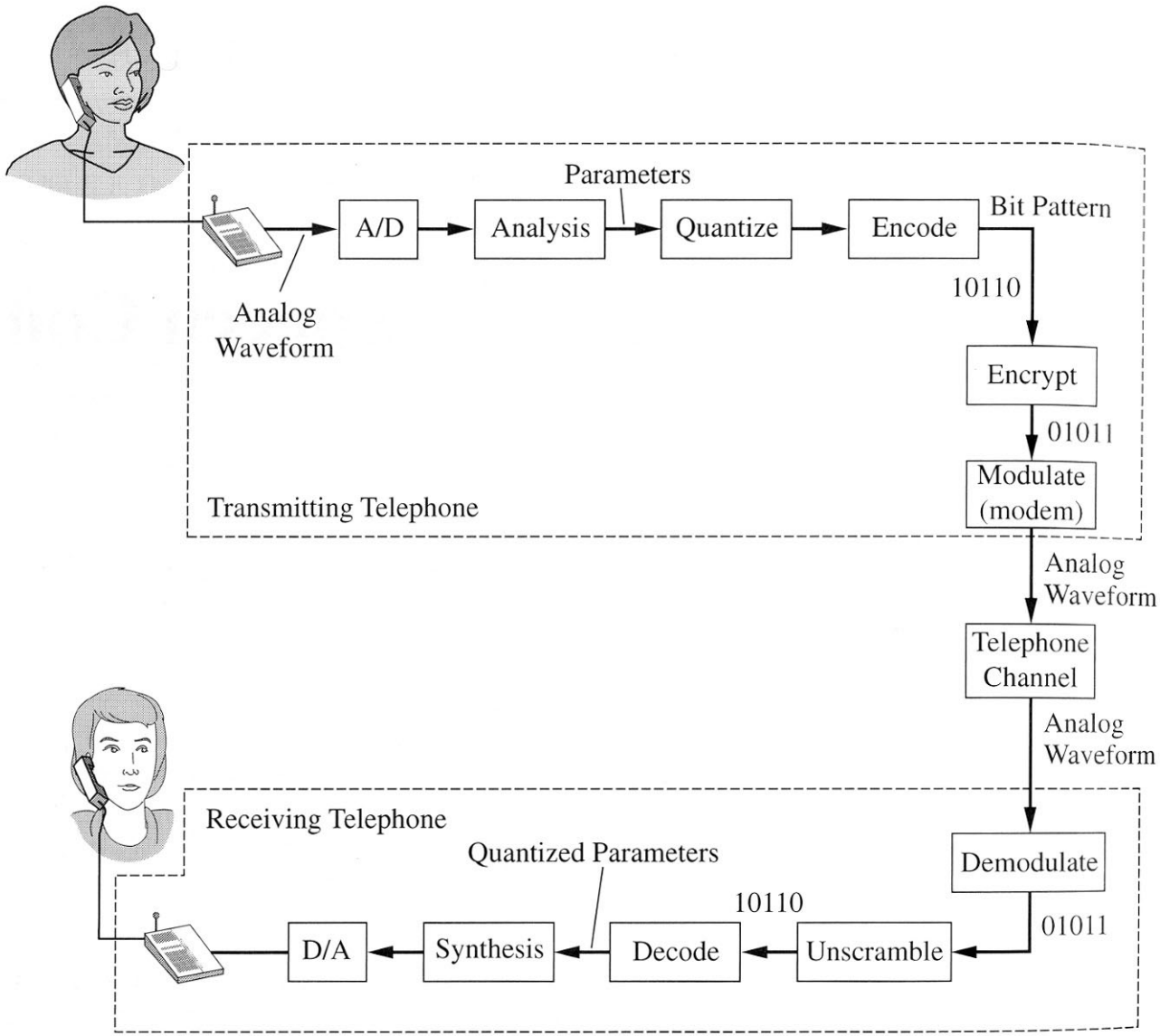

Figure 12.1 An example digital telephone communication system.

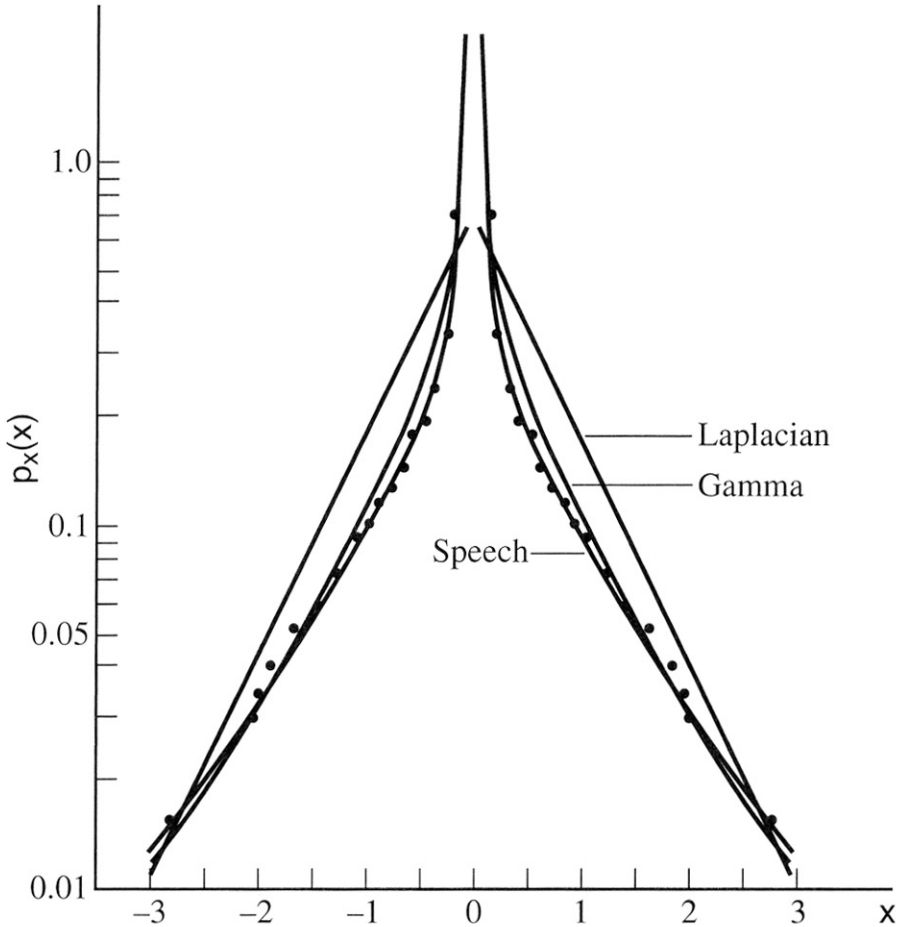

Figure 12.2 Comparison of histograms from real speech and gamma and Laplacian probability density fits to real speech. The densities are normalized to have mean $m_x = 0$ and variance $\sigma_x^2 = 1$. Dots (and the corresponding fitted curve) denote the histogram of the speech.

SOURCE: M.D. Paez and T.H. Glisson, "Minimum Mean-Squared Error Quantization in Speech" [66]. ©1972, IEEE. Used by permission.

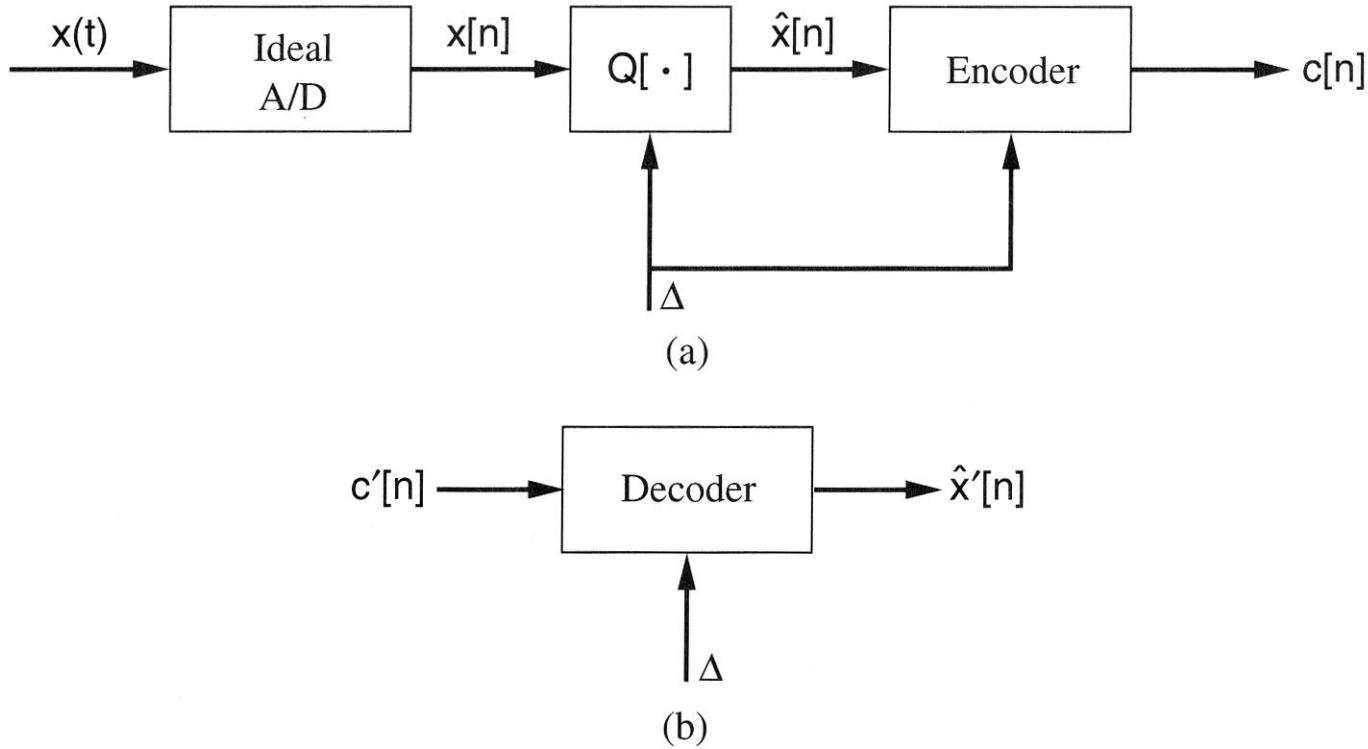

Figure 12.3 Waveform (a) coding and (b) decoding. Coding involves quantizing $x[n]$ to obtain a sequence of samples $\hat{x}[n]$ and encoding them to codewords $c[n]$. Decoding takes a sequence of codewords $c'[n]$ back to a sequence of quantized samples $\hat{x}'[n]$. $c'[n]$ denotes the codeword $c[n]$ that may be distorted by a channel and Δ is the quantization step size.

SOURCE: L.R. Rabiner and R.W. Schafer, *Digital Processing of Speech Signals* [71].

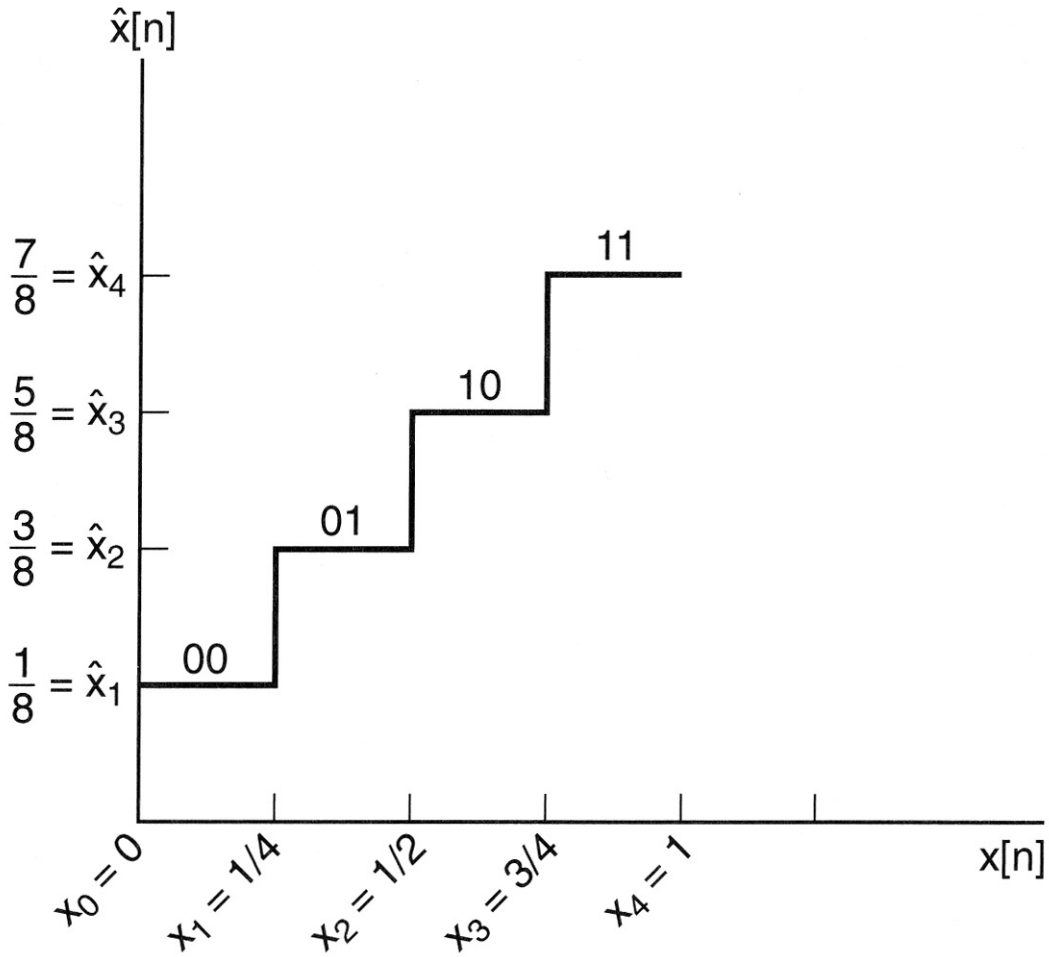

Figure 12.4 An example of uniform 2-bit quantization where the reconstruction and decision levels are uniformly spaced. The number of reconstruction levels $M = 4$ and the input falls in the range [0, 1].

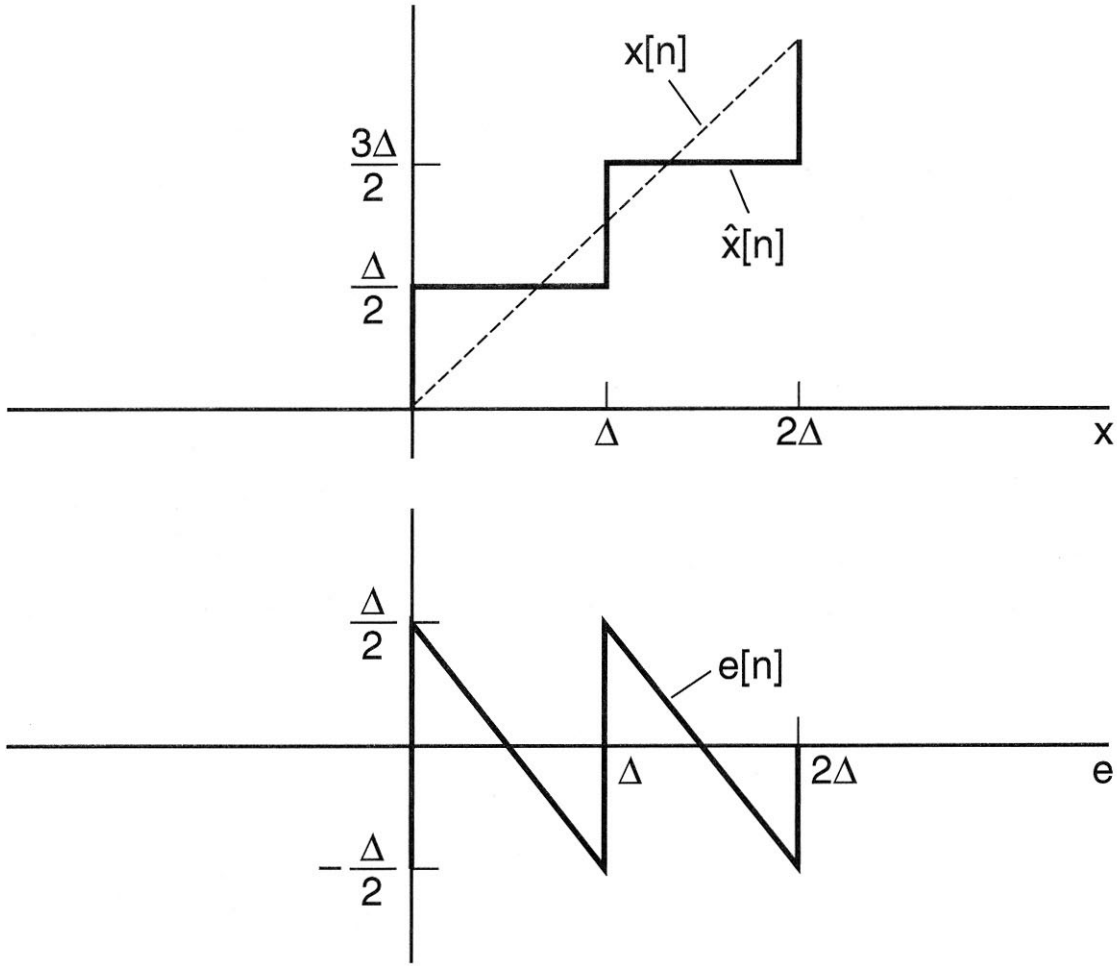

Figure 12.5 Quantization noise for a linearly changing input $x[n] = n$. The quantization noise is given by $e[n] = x[n] - \hat{x}[n]$, where $\hat{x}[n] = Q(x[n])$, with Q the quantization operator. For this case of a linearly-changing input, the quantization noise is seen to be signal-dependent.

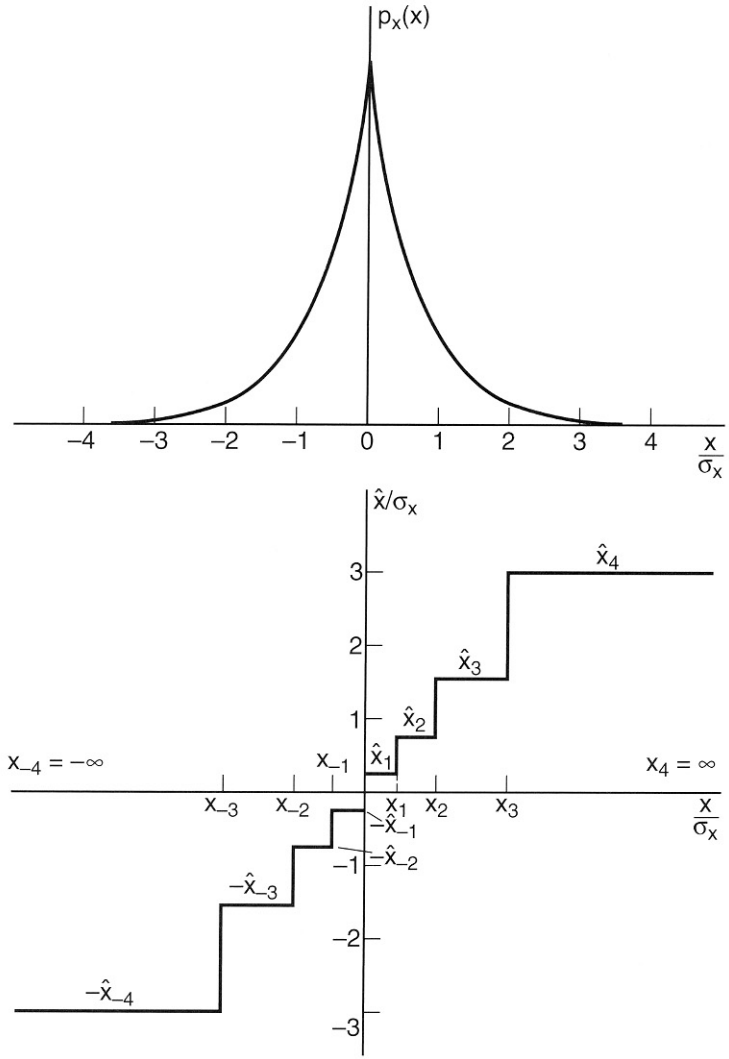

Figure 12.6 3-bit nonuniform quantizer: (a) Laplacian pdf; (b) decision and reconstruction levels.

SOURCE: L.R. Rabiner and R.W. Schafer, *Digital Processing of Speech Signals* [71]. ©1978, Pearson Education, Inc. Used by permission.

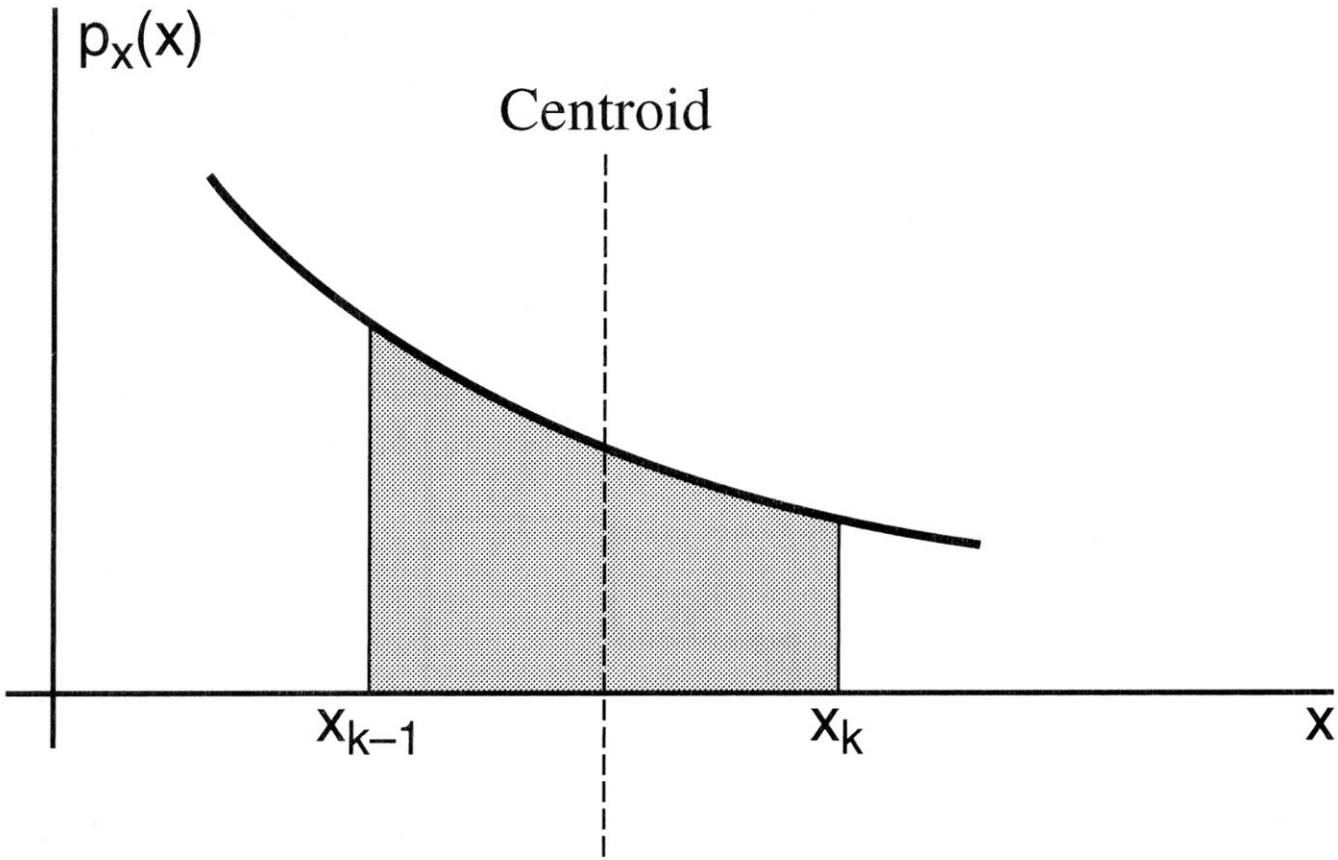

Figure 12.9 Centroid calculation in determining the optimal MSE reconstruction level.

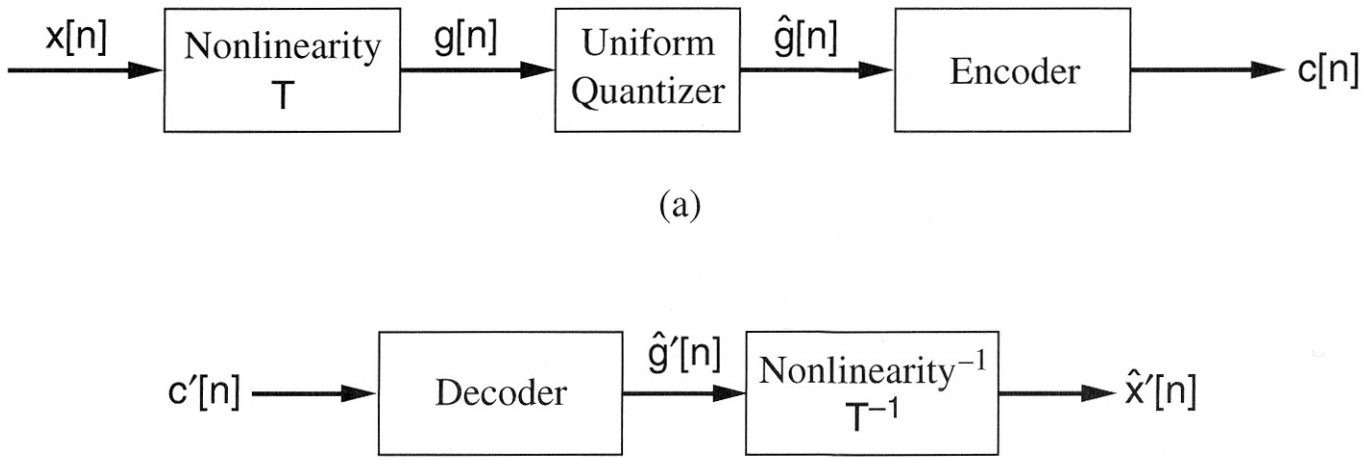

Figure 12.10 The method of companding in coding and decoding: (a) coding stage consisting of a nonlinearity followed by uniform quantization and encoding; (b) an inverse nonlinearity occurring after decoding.

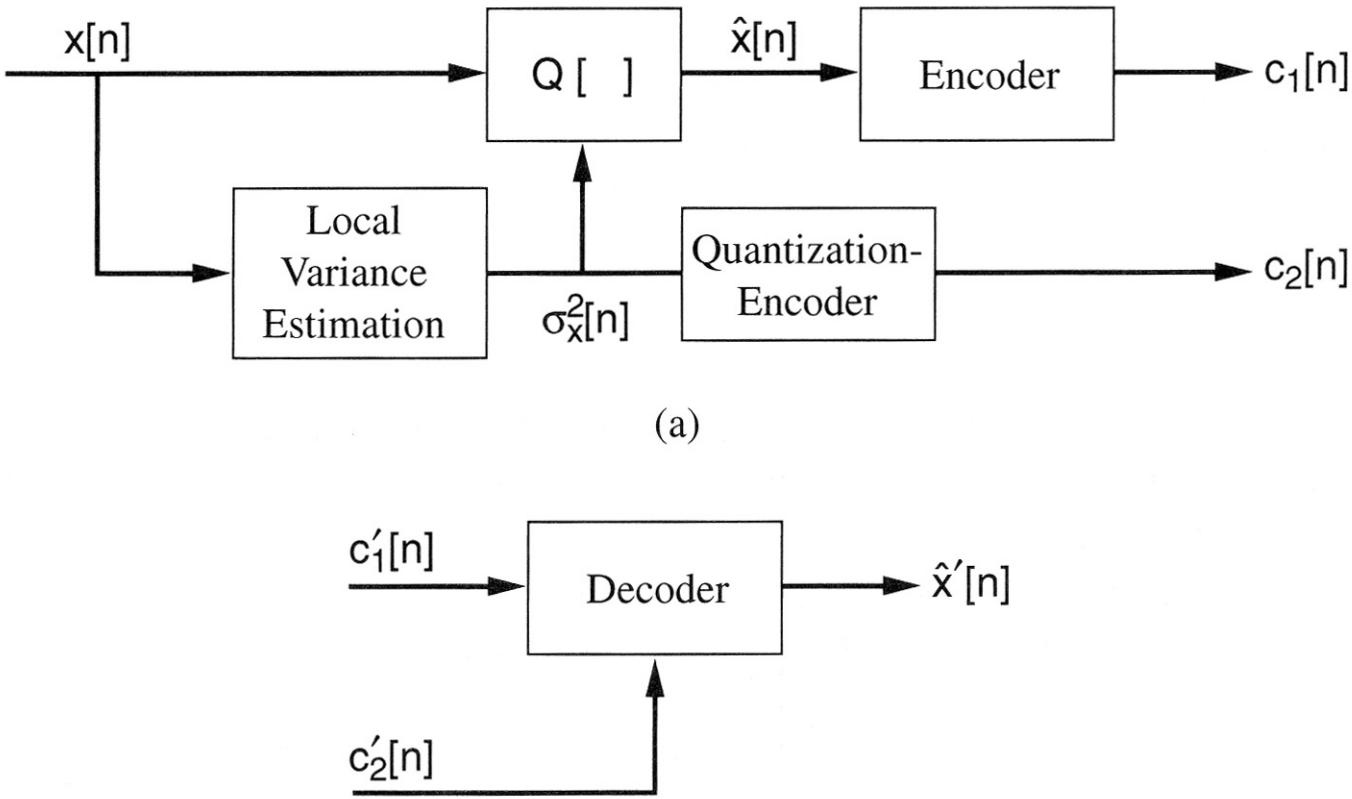

Figure 12.11 Adapting a nonuniform quantizer to a local pdf. By measuring the local variance $\sigma_x^2[n]$, we characterize the assumed Gaussian pdf. $c_1[n]$ and $c_2[n]$ are codewords for the quantized signal $\hat{x}[n]$ and time-varying variance $\hat{\sigma}^2[n]$, respectively. This feed-forward structure is one of a number of adaptive quantizers that exploit local variance.

Table 12.1 Comparison of 3-bit adaptive and nonadaptive quantization schemes [60]. Adaptive schemes use feed-forward adaptation.

SOURCE: Table from L.R. Rabiner and R.W. Schafer, *Digital Processing of Speech Signals* [71]. ©1978, Pearson Education, Inc. Used by permission. Data by Noll [60].

Nonuniform Quantizers	*Nonadaptive SNR (dB)*	*Adaptive* ($M = 128$) *SNR (dB)*	*Adaptive* ($M = 1024$) *SNR (dB)*
μ-law ($\mu = 100$, $x_{\max} = 8\sigma_x$)	9.5	–	–
Gaussian	7.3	15.0	12.1
Laplacian	9.9	13.3	12.8
Uniform Quantizers			
Gaussian	6.7	14.7	11.3
Laplacian	7.4	13.4	11.5

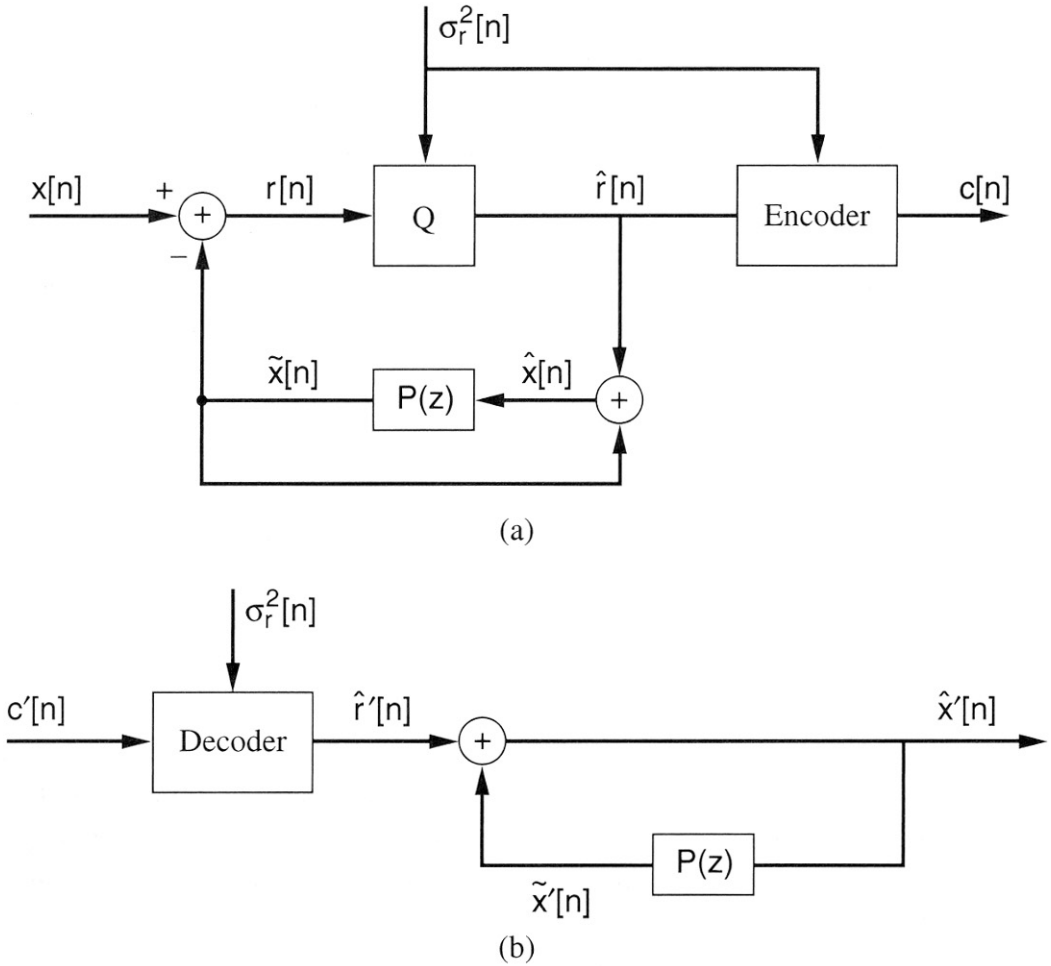

Figure 12.12 Differential coding (a) and decoding (b) schemes. The coefficients of the predictor $P(z)$, as well as the variance of the residual $r[n]$, must be quantized, coded, and decoded (not shown in the figure).

SOURCE: L.R. Rabiner and R.W. Schafer, *Digital Processing of Speech Signals* [71]. ©1978, Pearson Education, Inc. Used by permission.

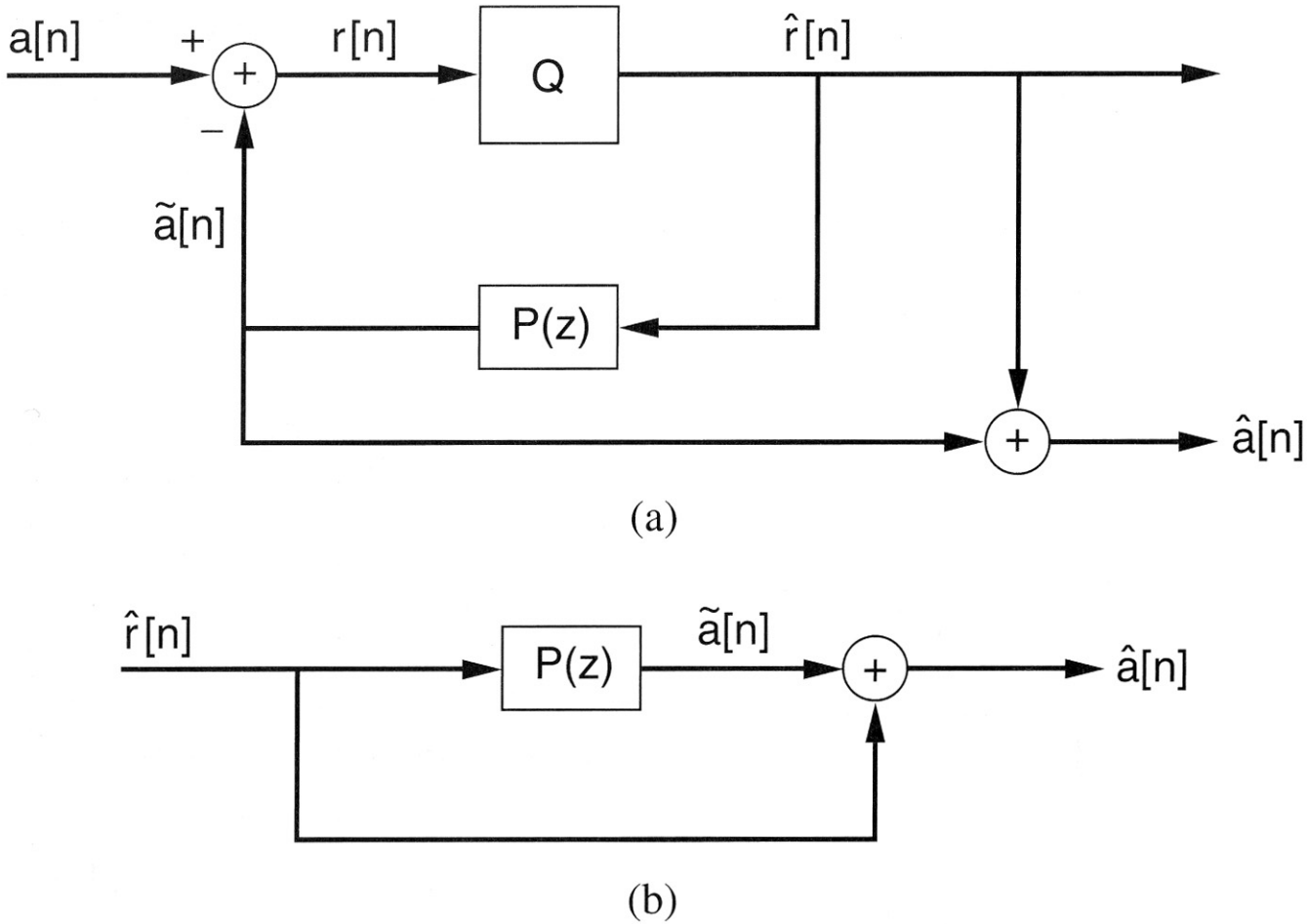

Figure 12.13 Differential coding (a) and decoding (b) schemes with moving average predictor.

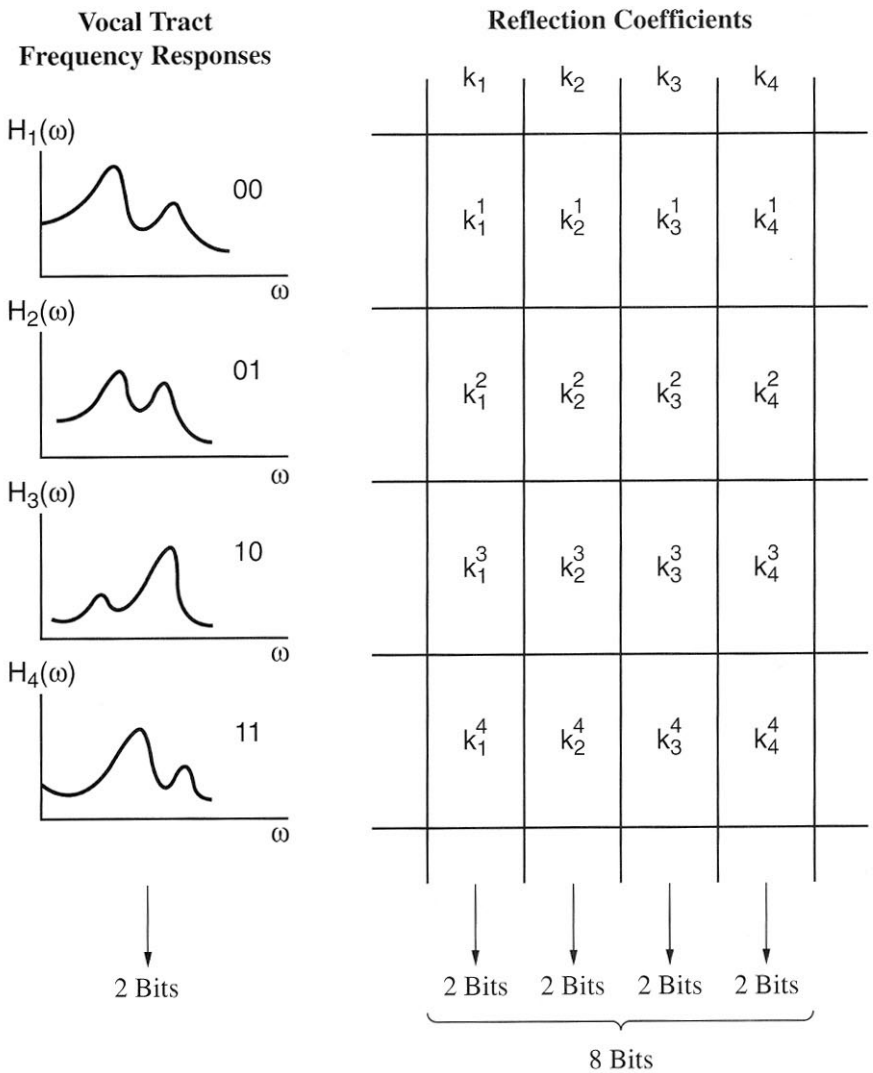

Figure 12.14 Comparison of bit rates required by scalar and vector representations of reflection coefficients for four two-pole vocal tract frequency responses. The four required reflection coefficients are highly correlated, the vocal tract being limited to only four configurations.

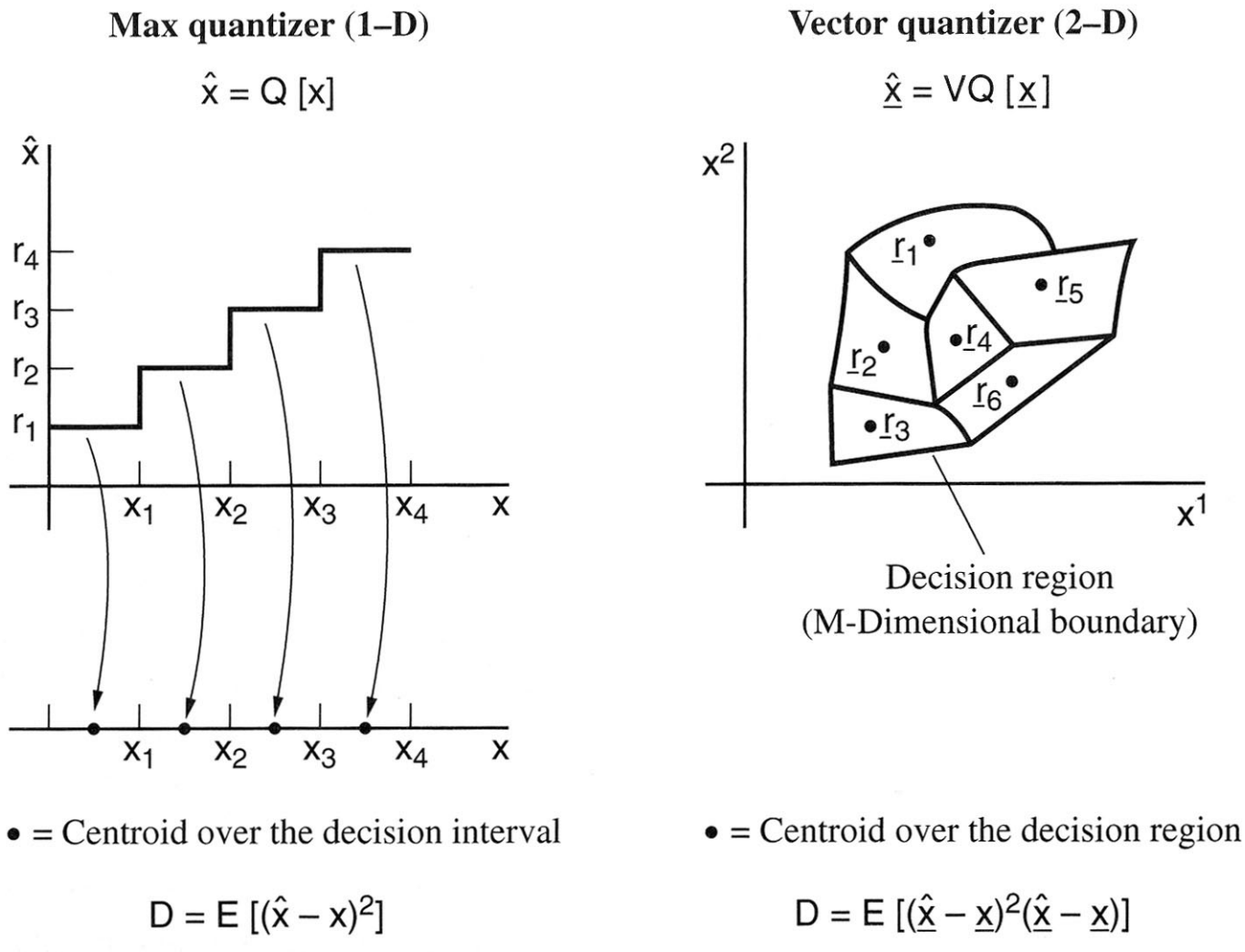

Figure 12.15 Comparison of scalar and vector quantization.

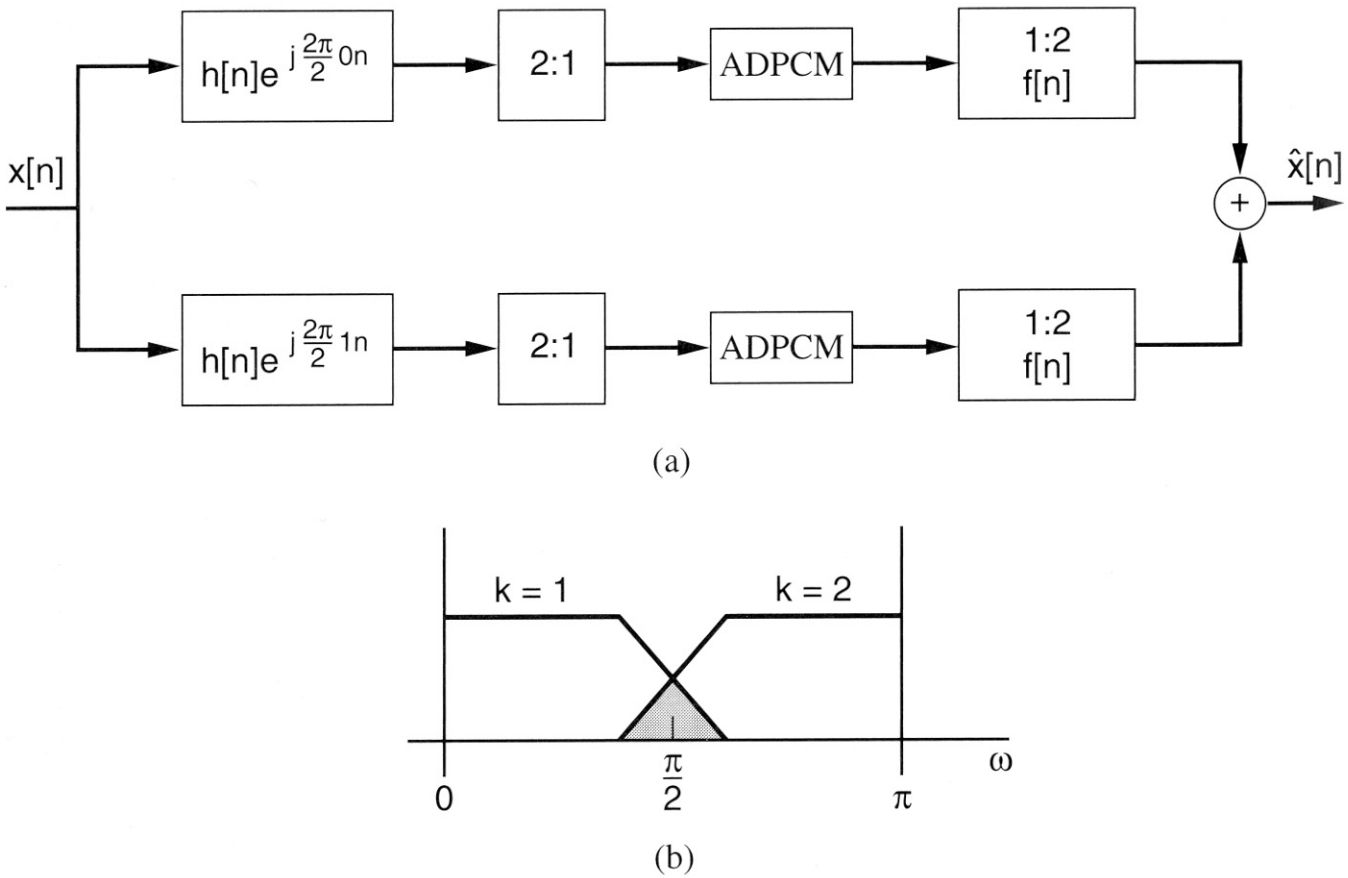

Figure 12.17 Example of a two-band subband coder using two overlapping quadrature mirror filters and ADPCM coding of the filter outputs: (a) complete analysis, coding/decoding, and synthesis configuration; (b) overlapping quadrature mirror filters.

SOURCE: J.M. Tribolet and R.E. Crochiere, "Frequency Domain Coding of Speech" [81]. ©1979, IEEE. Used by permission.

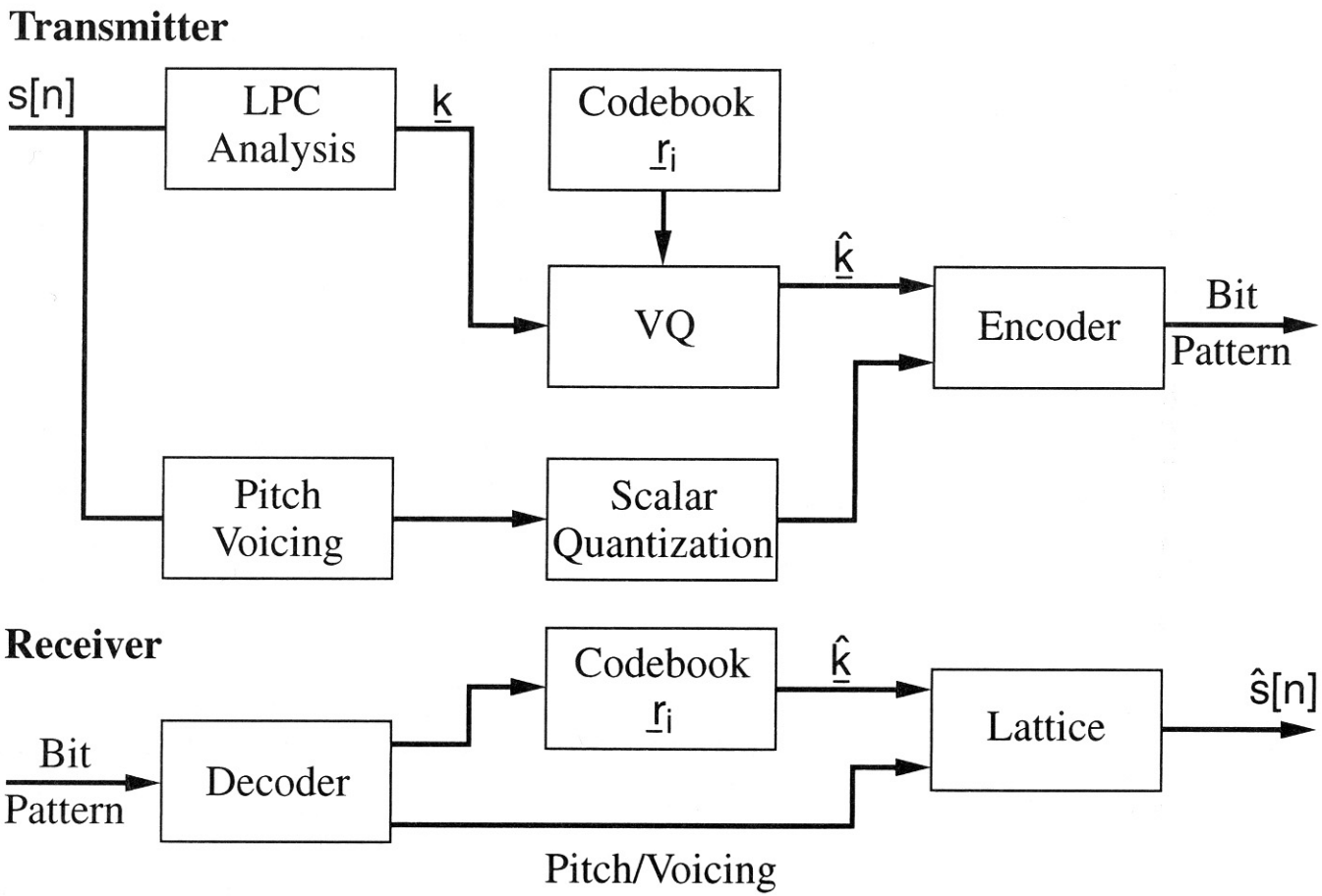

Figure 12.22 A VQ LPC vocoder. In this example, the synthesizer uses an all-pole lattice structure.

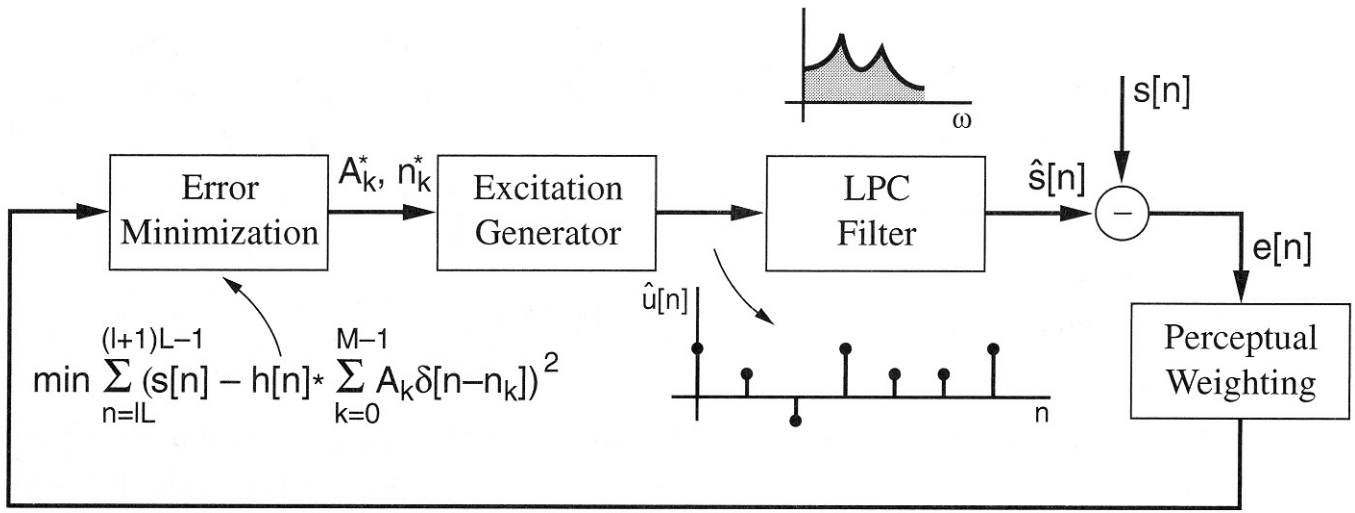

Figure 12.23 Multi-pulse linear prediction as an analysis-by-synthesis approach.

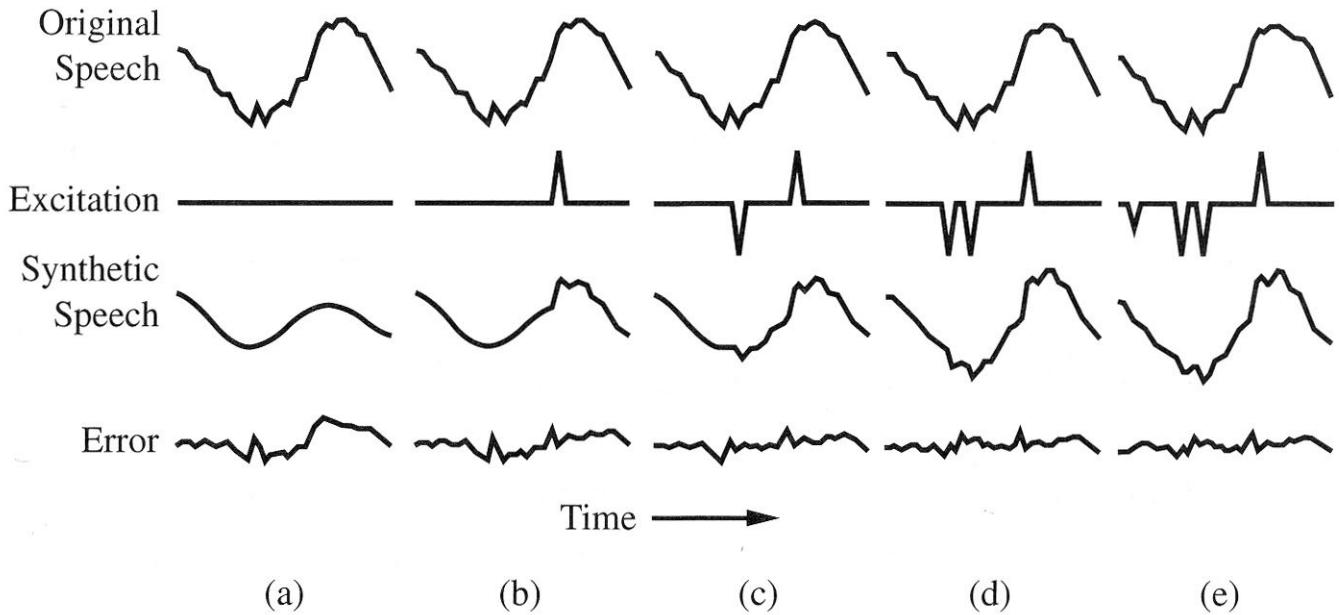

Figure 12.24 Illustration of the suboptimal (iterative) impulse-by-impulse multi-pulse solution. Each panel shows the waveform of a 5-ms frame interval, the excitation, the memory hangover from the previous frame, and the error resulting from estimating and subtracting successive impulses, and shows the error decreasing with each additional impulse.

SOURCE: B.S. Atal and J.R. Remde, "A New Model of LPC Excitation of Producing Natural-Sounding Speech at Low Bit Rates," [6]. ©1982, IEEE. Used by permission.

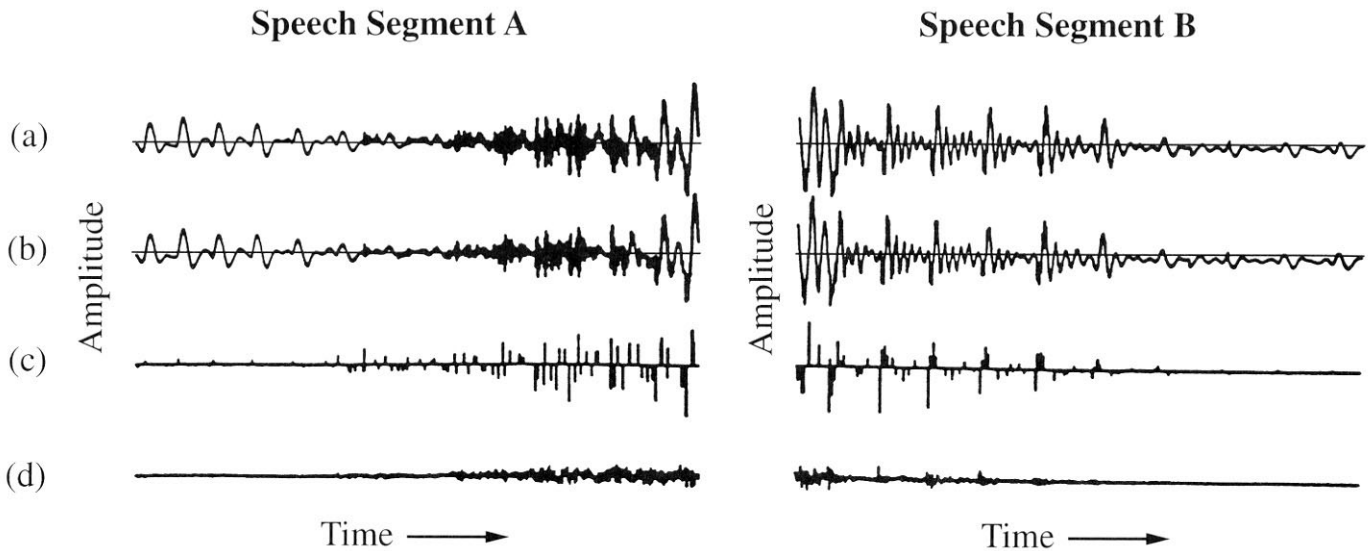

Figure 12.25 Two illustrations of multi-pulse synthesis. Each case shows the (a) original speech, (b) synthetic speech, (c) excitation signal, and (d) error.

SOURCE: B.S. Atal and J.R. Remde, "A New Model of LPC Excitation of Producing Natural-Sounding Speech at Low Bit Rates," [6]. ©1982, IEEE. Used by permission.

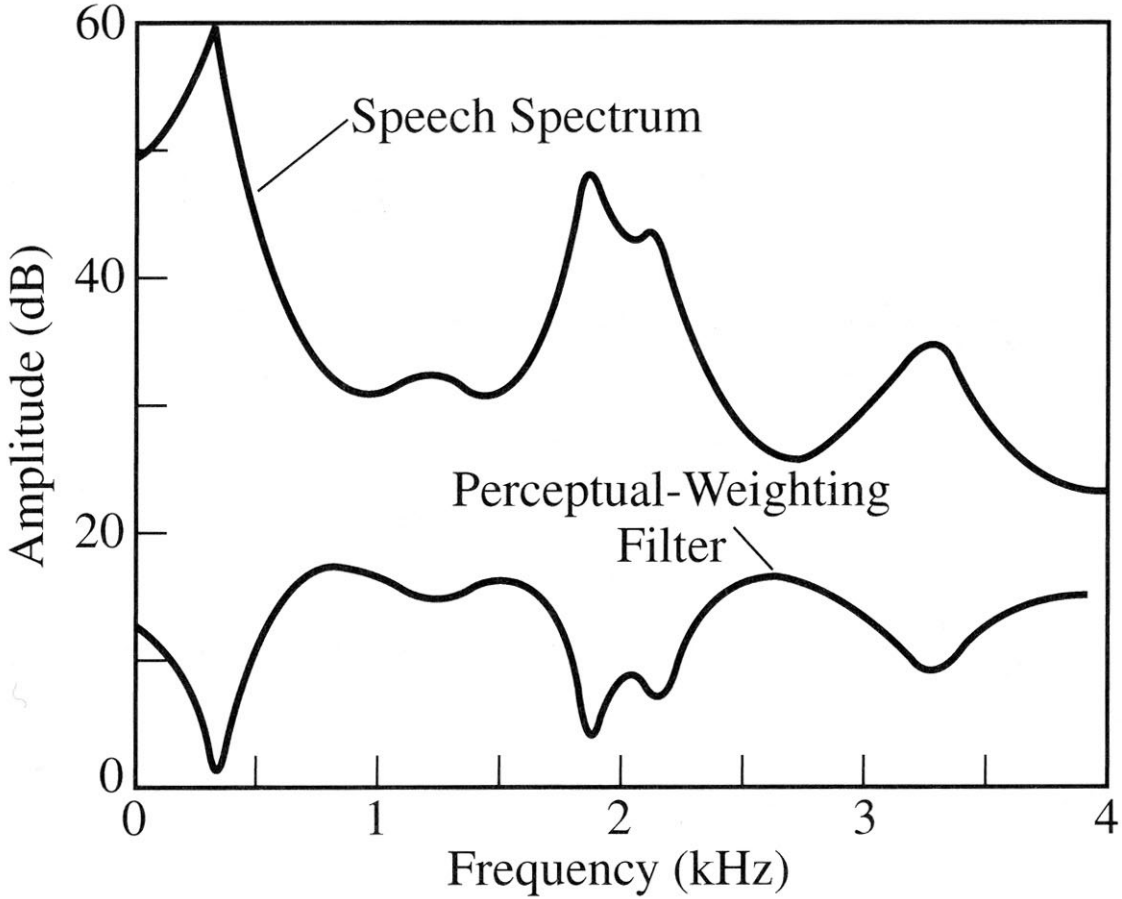

Figure 12.26 Example of perceptual weighting filter.

SOURCE: B.S. Atal and J.R. Remde, "A New Model of LPC Excitation of Producing Natural-Sounding Speech at Low Bit Rates," [6]. ©1982, IEEE. Used by permission.

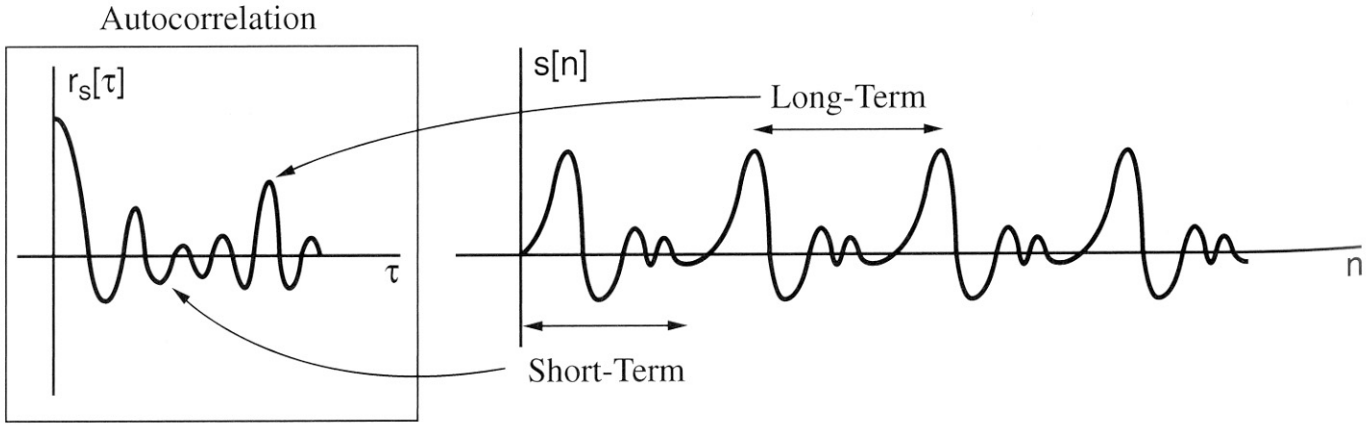

Figure 12.27 Illustration of short- and long-term correlation in the speech waveform.

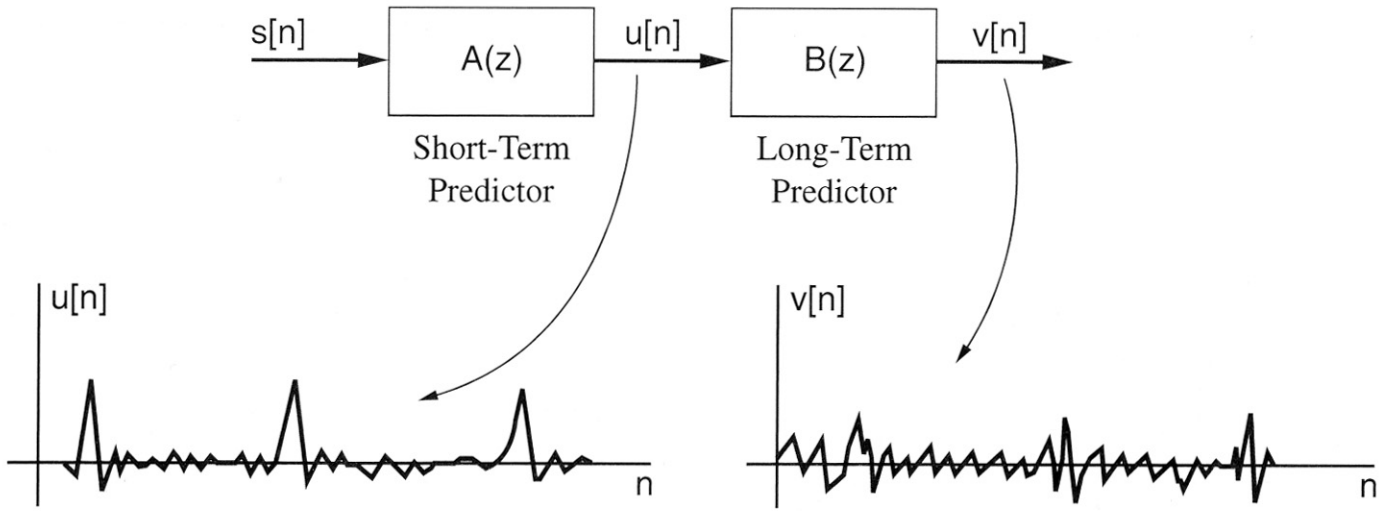

Figure 12.28 Illustration of short- and long-term prediction. The residual sequence $u[n]$ is the short-term prediction error and the residual $v[n]$ is the long-term prediction error.

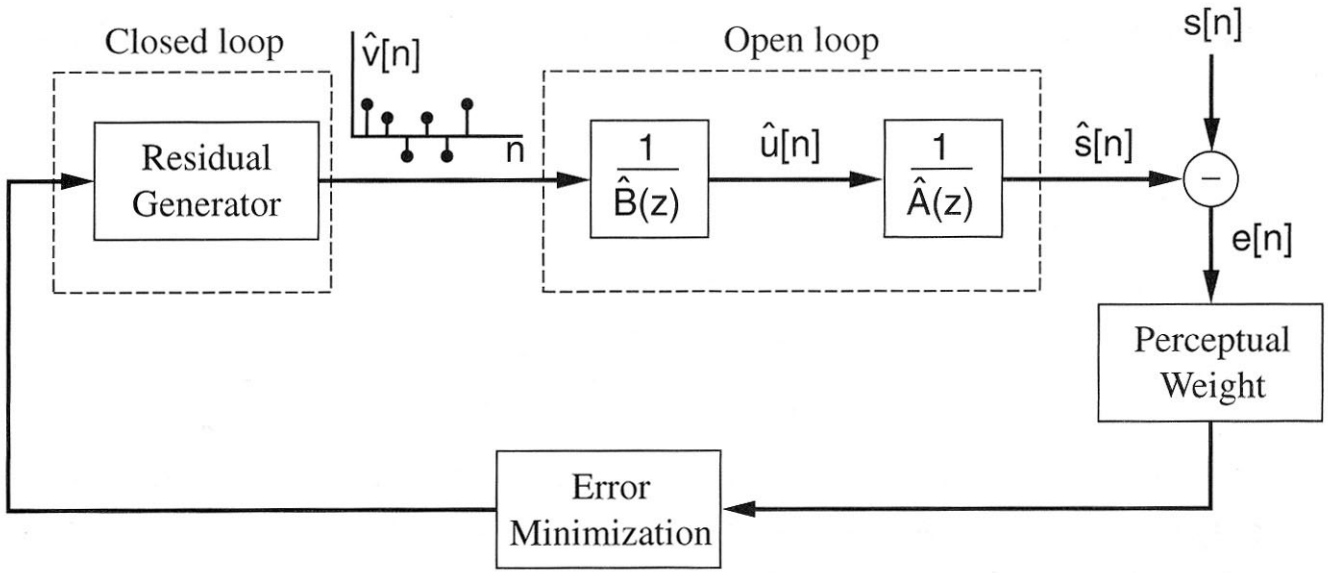

Figure 12.30 Closed-loop/open-loop multi-pulse analysis/synthesis with short- and long-term prediction. The polynomials $\hat{A}(z)$ and $\hat{B}(z)$ represent the short- and long-term predictor estimates with possibly quantized coefficients.