Long-term prediction is based on the observation that primary pitch pulses due to glottal pulses are correlated and *predictable* over consecutive pitch periods, so that

$$s[n] \approx bs[n - P]$$

where P is the pitch period and b is a scale factor. In fact, we can consider the speech waveform as having a *short-term and long-term correlation*. As illustrated in Figure 12.27, the short-term correlation (with which we are already familiar from our linear prediction analysis of Chapter 5) occurs over the duration of the vocal tract response *within* a pitch period, while the long-term correlation occurs *across* consecutive pitch periods. The approach that we take, therefore, is to first remove short-term correlation by short-term prediction followed by removing long-term correlation by long-term prediction.

The short-term prediction-error filter is the pth-order polynomial $A(z) = 1 - P(z) = \sum_{k=1}^{p} a_k z^{-1}$, where p is typically in the range 10-16. The result of the short-term prediction error is a residual function $u[n]$ that includes primary pitch pulses (long-term correlation). The long-term prediction-error filter is of the form

$$B(z) = 1 - bz^{-P}$$

where bz^{-P} is the long-term predictor in the z-domain. The output of the long-term prediction-error filter is a residual

$$v[n] = u[n] - bu[n - P]$$

with fewer large (long-term correlated) pulses to code than in $u[n]$ (Figure 12.28). After removing the long-term prediction contribution, the residual $v[n]$ forms the basis for an efficient coding of a multi-pulse excitation. Having the short-term predictor and long-term predictor, we can then invert the process and recover the original speech waveform as shown in Figure 12.29 where we assume knowledge of the residual $v[n]$ as well as inverse filters $A(z)$ and $B(z)$. In synthesis, with a frame-by-frame implementation, the memory hangover from a previous frame is added into the result of filtering with $\frac{1}{A(z)}$ and $\frac{1}{B(z)}$ on the current frame.

In estimating the long-term predictor, we must estimate both the pitch period P and the scale factor b. The pitch period can be estimated independently with any pitch estimator. However, it is preferred to tie the estimation of P to the prediction problem because our goal is to remove pulses correlated over consecutive periods, reducing the prediction error. In the time domain, the long-term prediction error filter $B(z) = 1 - bz^{-P}$ is expressed by

$$b[n] = \delta[n] - b\delta[n - P].$$

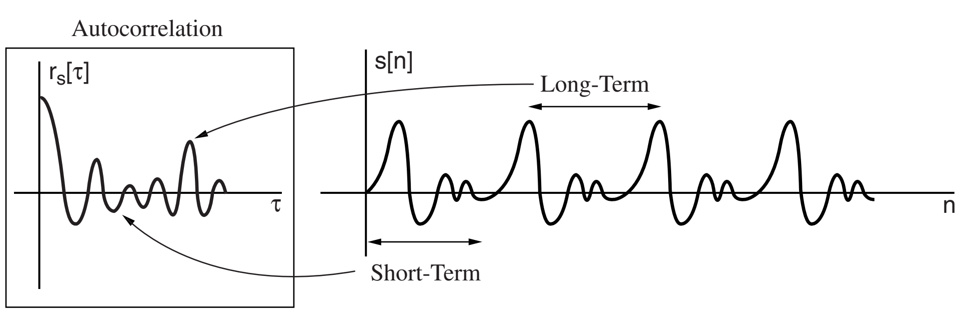

Figure 12.27 Illustration of short- and long-term correlation in the speech waveform.