ECE 797:

Speech and Audio Processing

Hand-out for Lecture #5

Thursday, February 12, 2004

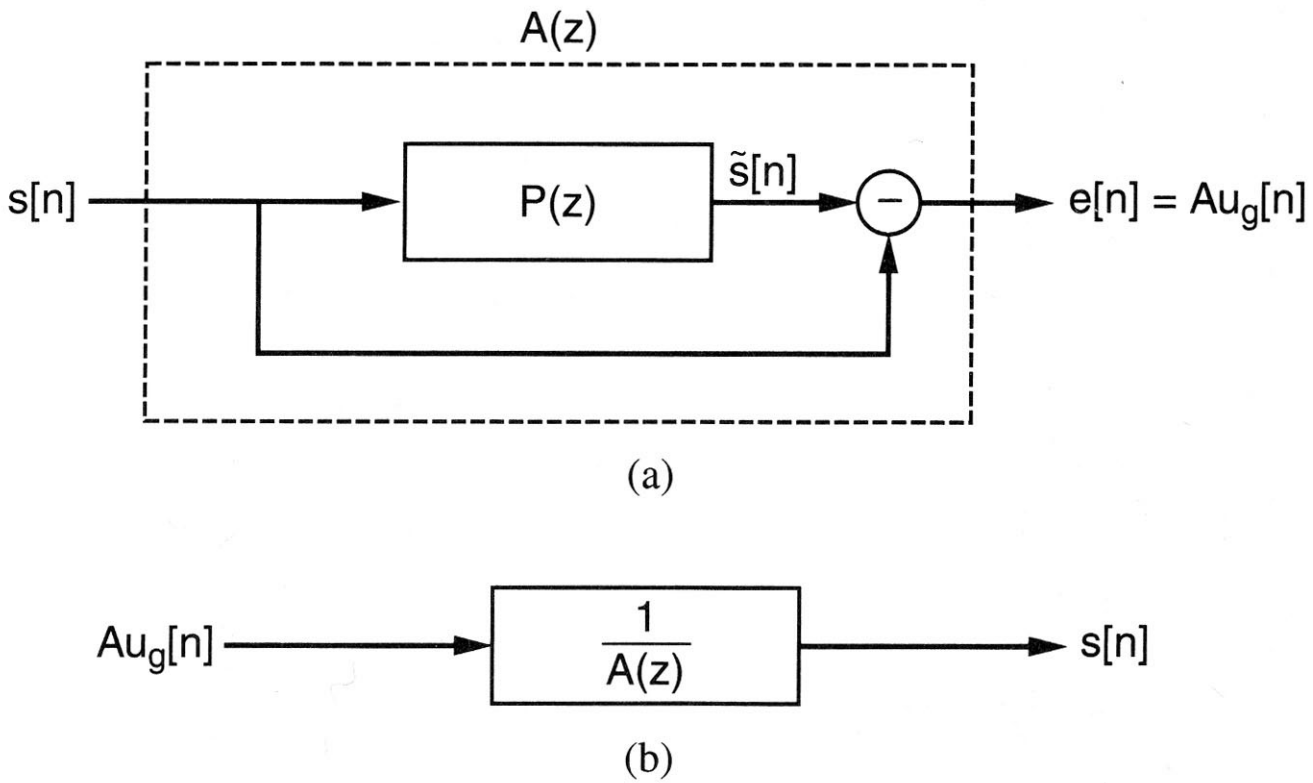

Figure 5.1 Filtering view of linear prediction: (a) prediction-error filter $A(z) = 1 - P(z)$; (b) recovery of $s[n]$ with $\frac{1}{A(z)}$ for $\alpha_k = a_k$. $A(z)$ is also considered the inverse filter because it can yield the input to an all-pole transfer function.

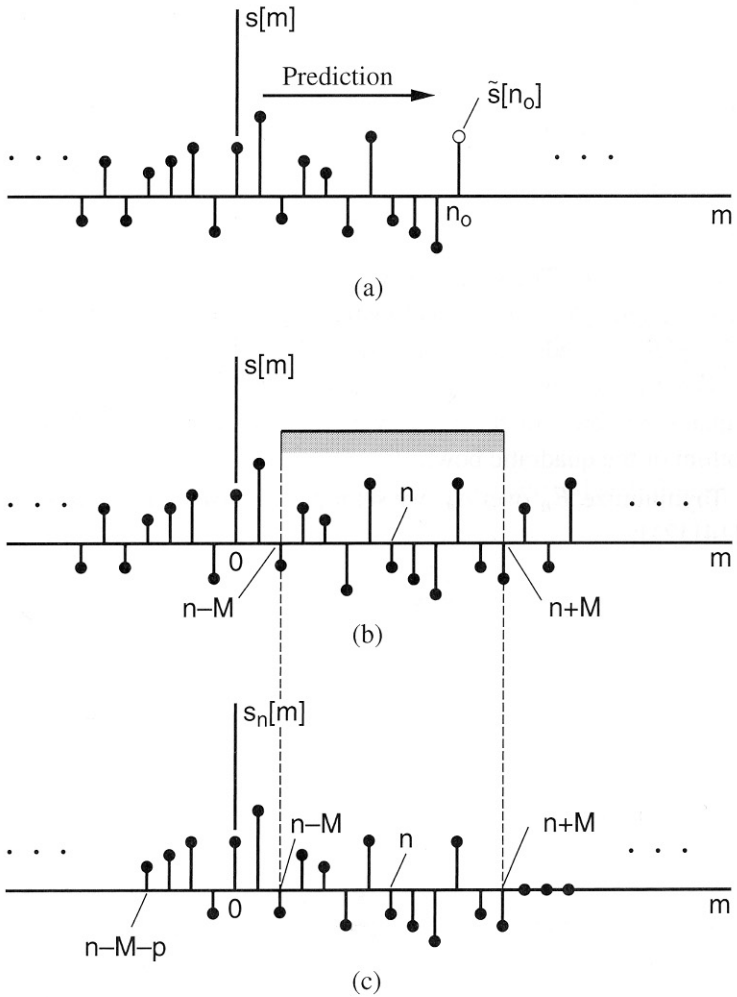

Figure 5.2 Short-time sequences used in linear prediction analysis: (a) prediction at time n_o; (b) samples in the vicinity of time n, i.e., samples over the interval $[n-M, n+M]$; (c) samples required for prediction of samples in the interval $[n-M, n+M]$. This set of samples, denoted by $s_n[m]$, includes samples inside and outside the interval $[n-M, n+M]$.

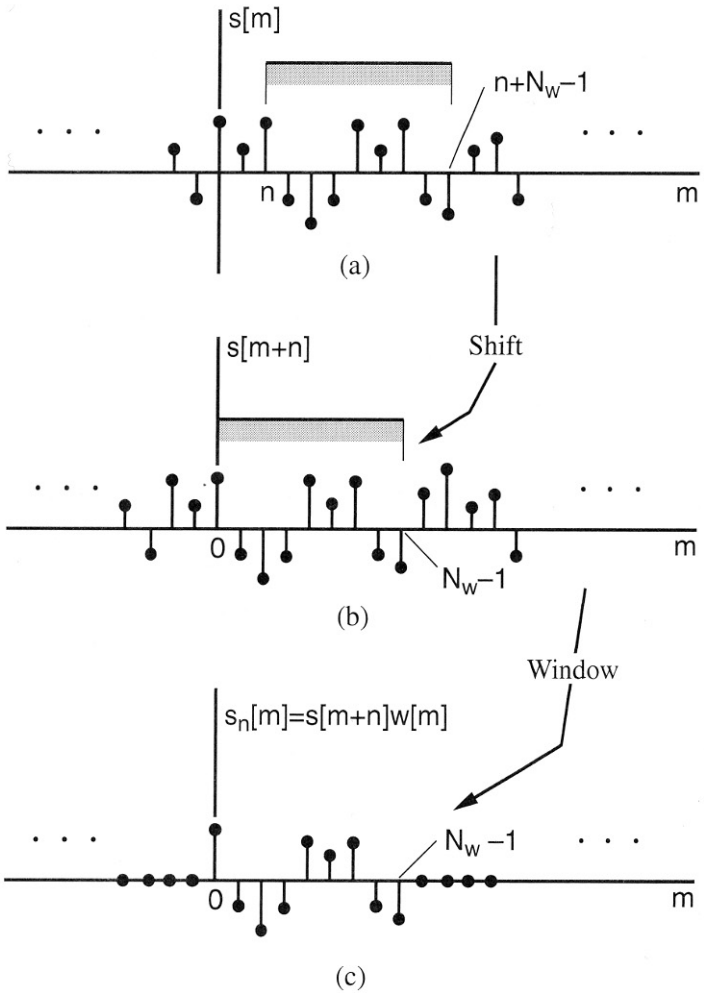

Figure 5.3 Formulation of the short-time sequence used in the autocorrelation method. In this interpretation, the waveform $s[m]$ [panel (a)] is shifted by n samples [panel (b)] and then windowed by an N_w–point rectangular sequence $w[m]$ [panel (c)]. The resulting sequence $s_n[m]$ is zero outside the interval $[0, N_w - 1]$.

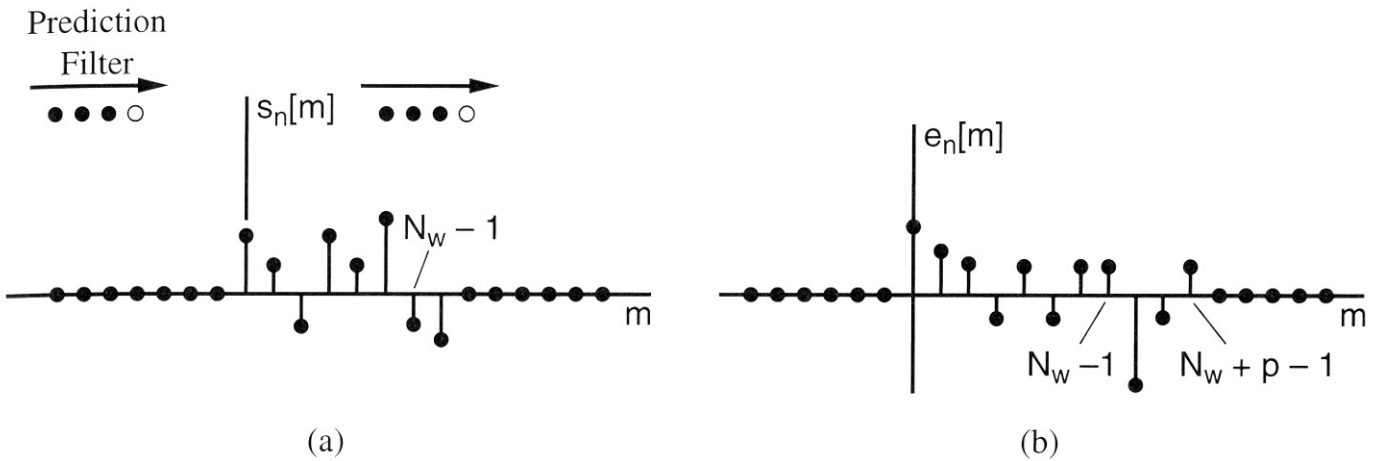

Figure 5.4 Example of a third-order predictor in the autocorrelation method of linear prediction: (a) sliding predictor filter; (b) prediction error. Prediction error is largest at the beginning and the end of the interval $[0, N_w + p - 1]$.

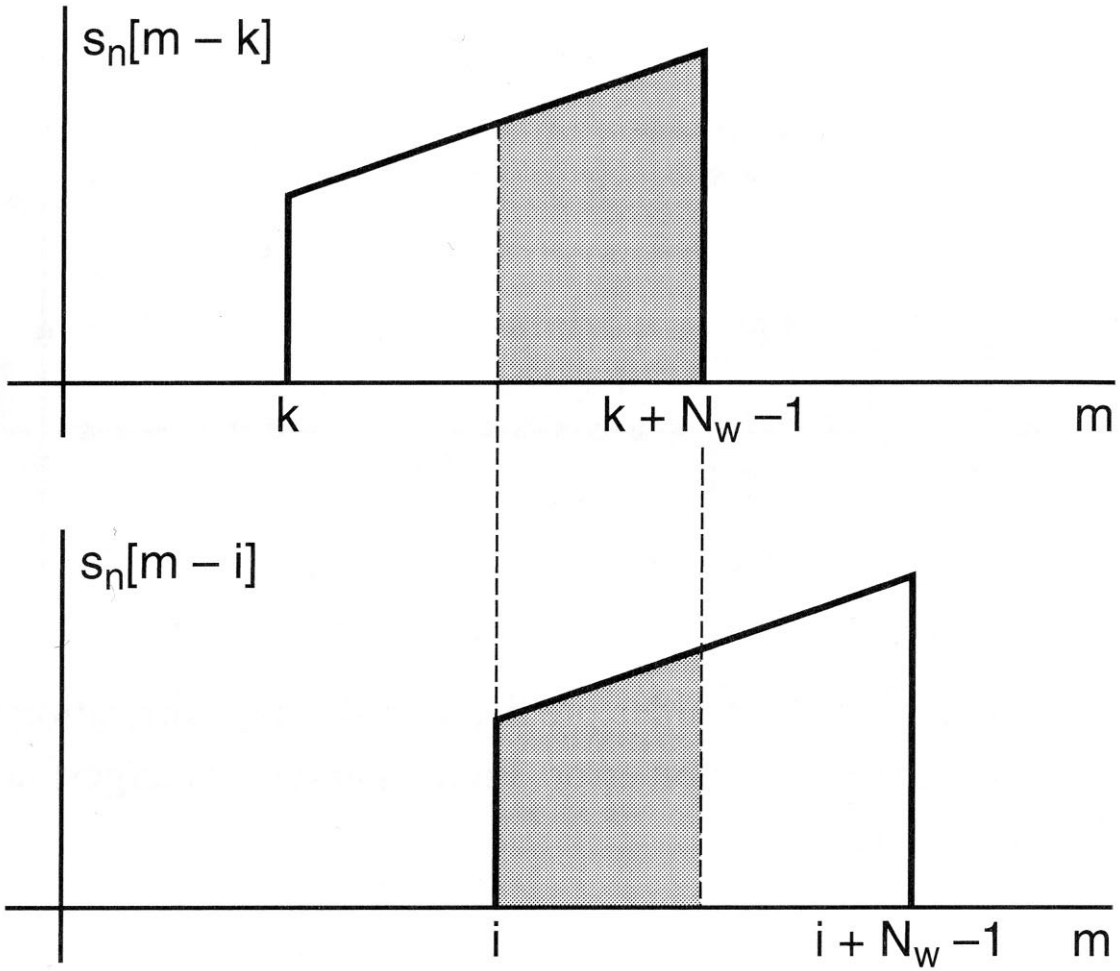

Figure 5.5 Construction of the autocorrelation function in the autocorrelation method. Overlapping regions of $s_n[m-k]$ and $s_n[m - i]$ are used in determining $\Phi_n[i, k]$.

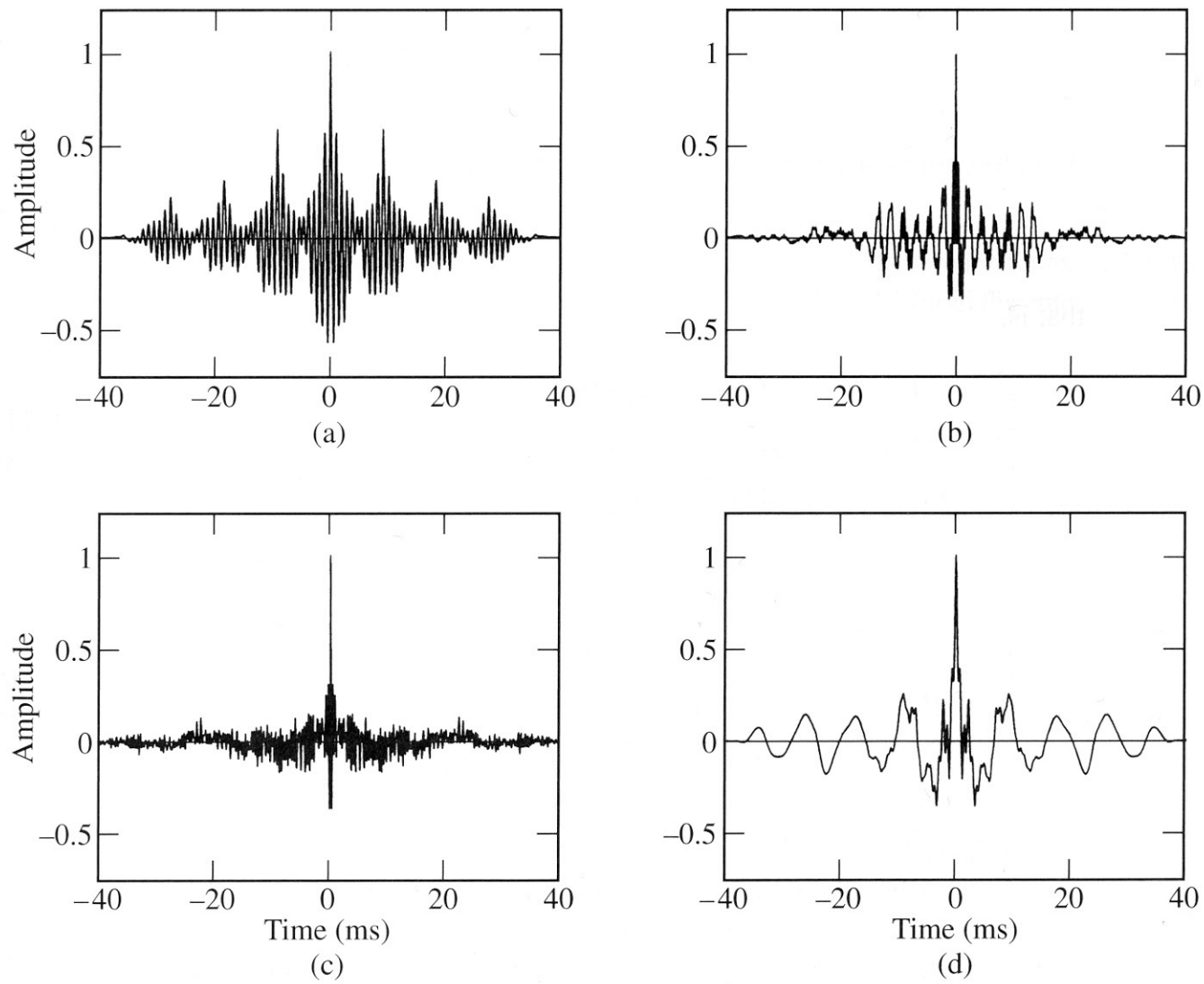

Figure 5.6 Illustration of autocorrelation functions of speech: (a) vowel /o/ in "p<u>o</u>p"; (b) unvoiced plosive /k/ in "ba<u>k</u>er"; (c) unvoiced fricative /f/ in "<u>f</u>ather"; (d) voiced plosive /g/ in "<u>g</u>o."

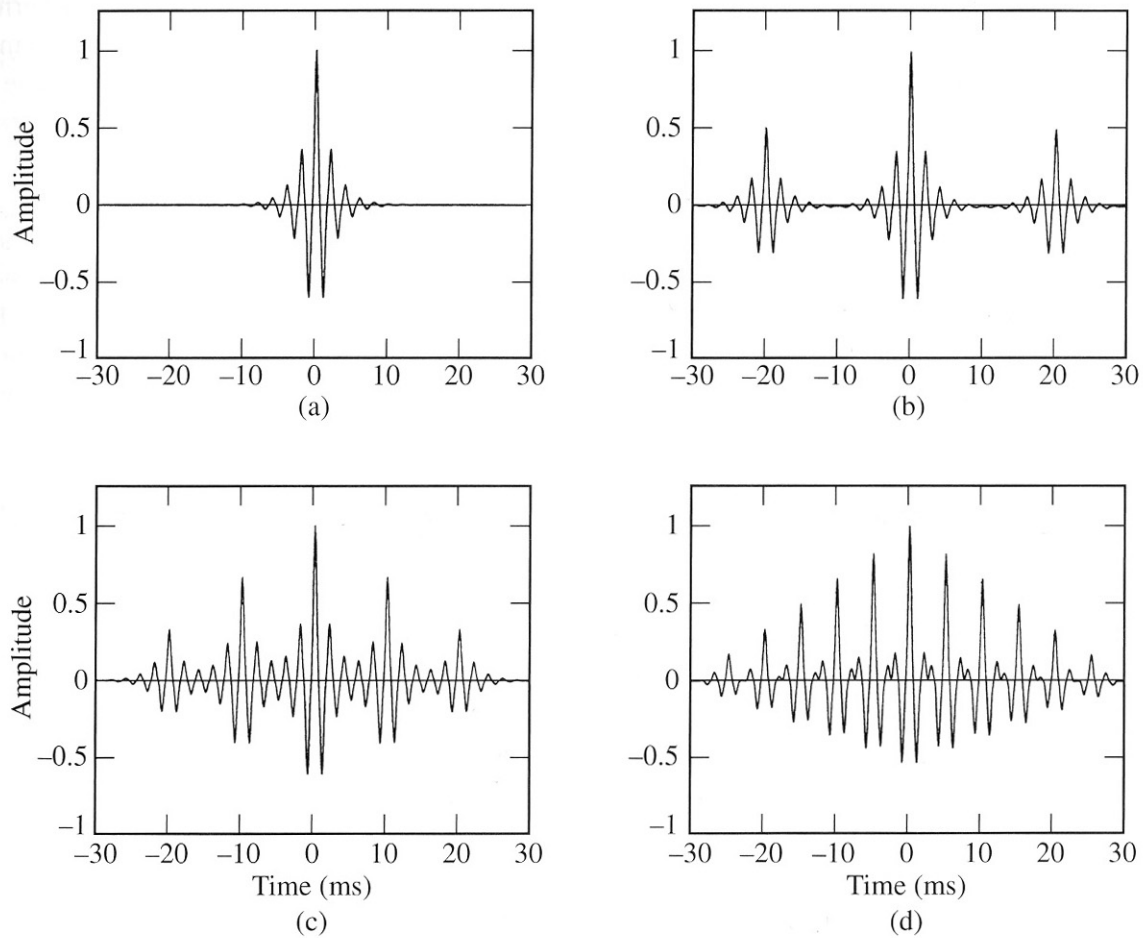

Figure 5.7 Illustration of autocorrelation "aliasing" in the autocorrelation method of linear prediction for a windowed periodic waveform: (a) $r_h[\tau]$; (b) $r_n[\tau]$ for 50-Hz pitch; (c) $r_n[\tau]$ for 100-Hz pitch; (d) $r_n[\tau]$ for 200-Hz pitch. Aliasing increases with rising pitch. For the 200-Hz pitch case, low-order autocorrelation coefficients are considerably different from those of $r_h[\tau]$, thus illustrating that accuracy of linear prediction analysis decreases with increasing pitch. In this example, $h[m]$ consists of two poles at 500 Hz and 1500 Hz and $s_n[m]$ is the convolution of $h[m]$ with a periodic impulse train, rectangularly windowed over a 40-ms duration.

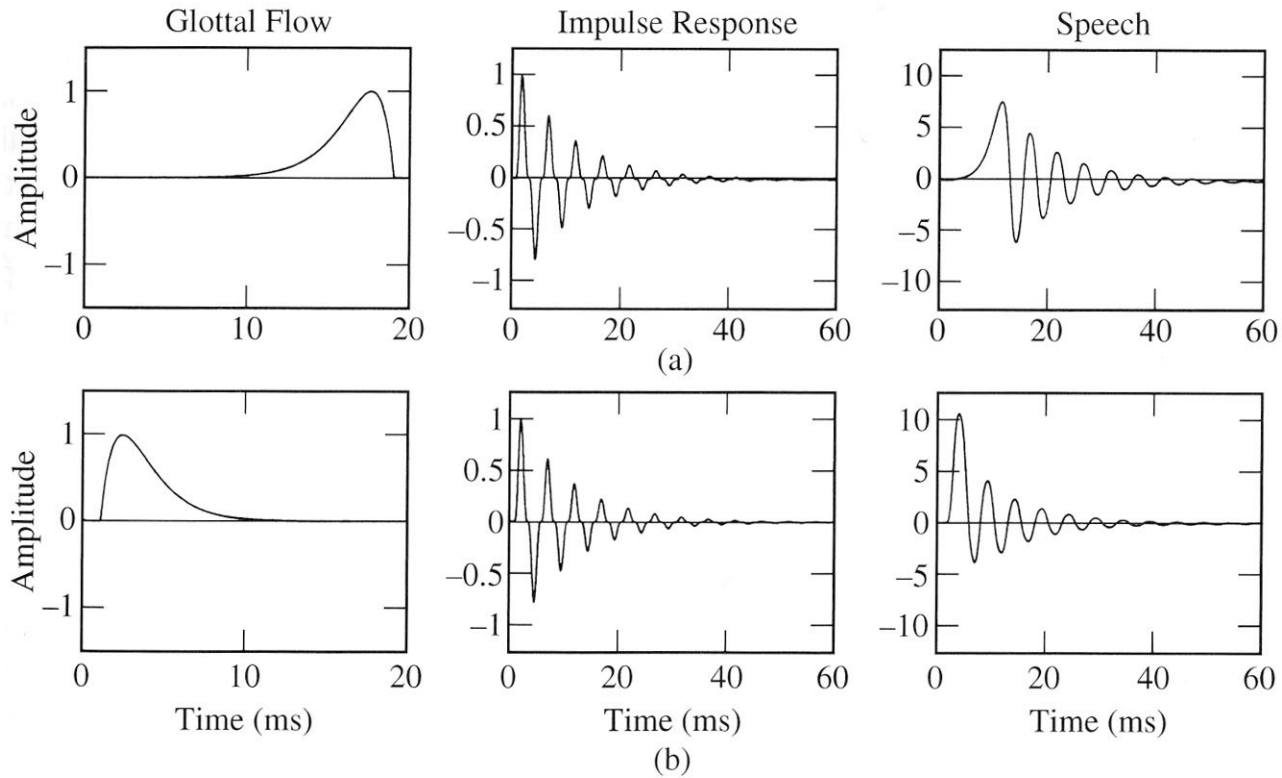

Figure 5.8 Transformation of glottal maximum-phase poles to minimum-phase poles by the autocorrelation method of linear prediction: (a) maximum-phase glottal flow (with displacement from the time origin), vocal tract impulse response, and resulting speech waveform for one excitation impulse; (b) minimum-phase glottal flow (with displacement from the time origin), vocal tract impulse response, and resulting speech waveform for one excitation impulse as obtained by the autocorrelation method of linear prediction. This example illustrates how the "attack" of an impulse response (and thus its perception) can be altered through linear prediction analysis. The simulated vocal tract consists of two resonances at 200 Hz and 600 Hz. The glottal flow waveform in (a) is the time-reversed decaying exponential $0.93^n u[n]$ (with $u[n]$ the unit step) convolved with itself.

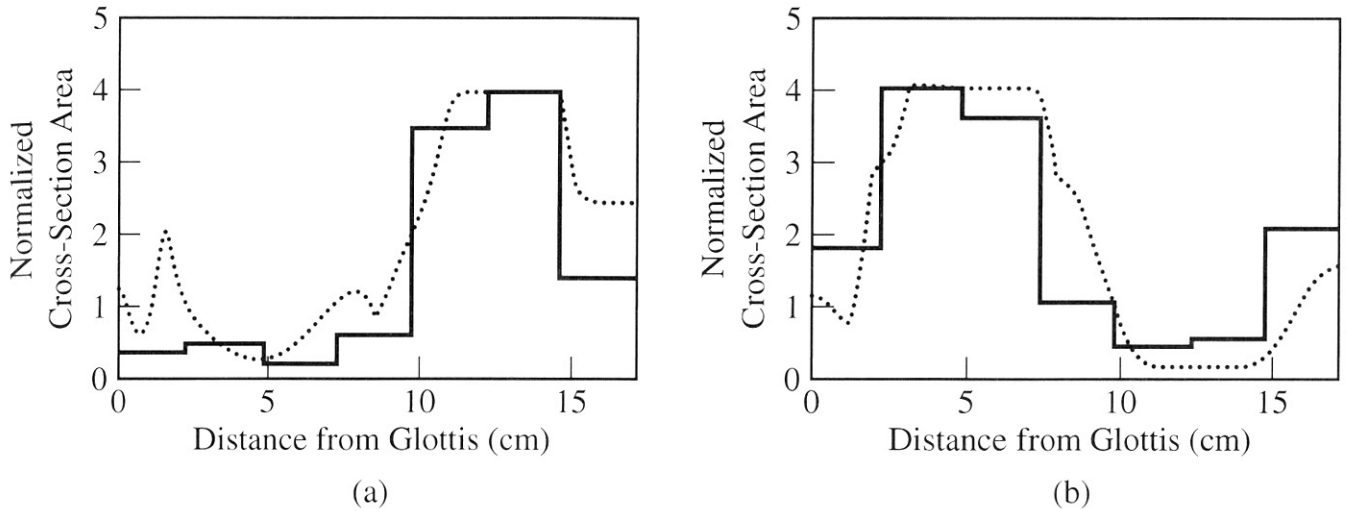

Figure 5.10 Estimation of vocal tract area functions using the autocorrelation method of linear prediction: (a) the vowel /a/; (b) the vowel /i/. The dotted curve is the cross-section measurement made for a Russian vowel (from Fant [27]), while the solid curve is the cross-section tube estimates based on linear prediction using the American vowel counterpart.

SOURCE: H. Wakita, "Estimation of the Vocal Tract Shape by Optimal Inverse Filtering and Acoustic/Articulatory Conversion Methods" [27]. ©1972, H. Wakita. Used by permission.

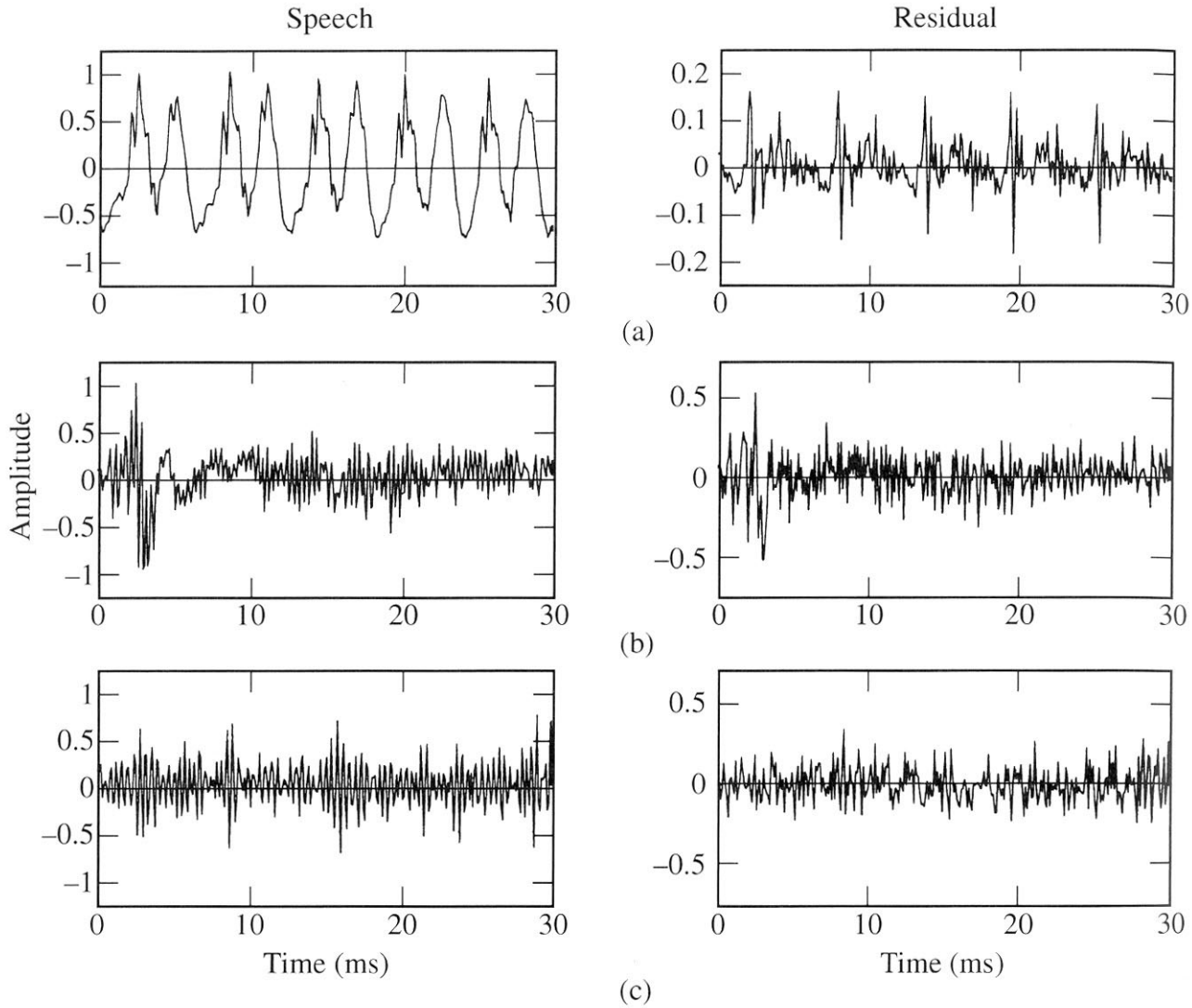

Figure 5.11 Prediction error residuals for (a) voiced; (b) unvoiced plosive; (c) unvoiced fricative sounds. A 14th-order predictor is used for 5000 Hz-bandwidth signals.

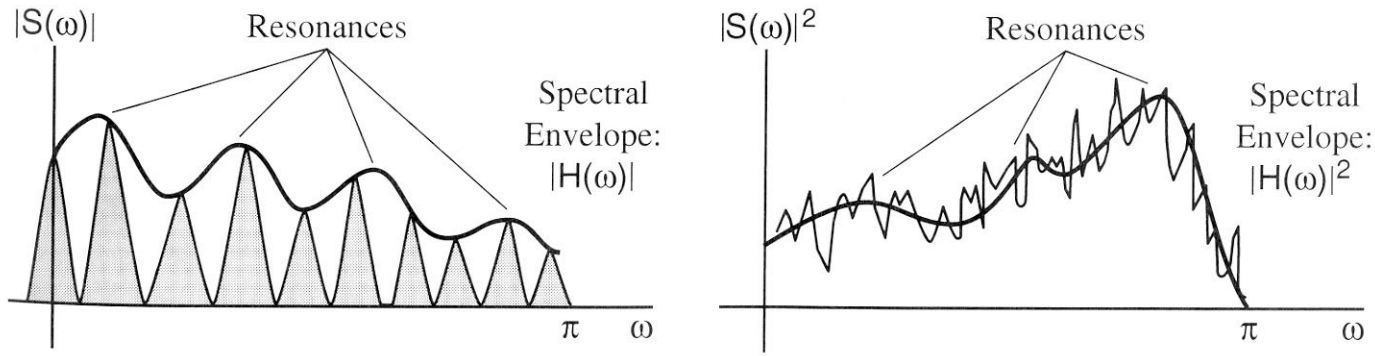

Figure 5.12 Schematics of spectra of periodic and stochastic speech sounds: (a) harmonic spectrum of a voiced sound. For simplicity, we assume $\frac{2\pi}{P} W(0) = 1$; (b) periodogram of a stochastic sound. For simplicity, we assume a unit variance white-noise source and that bias in the periodogram estimate is negligible.

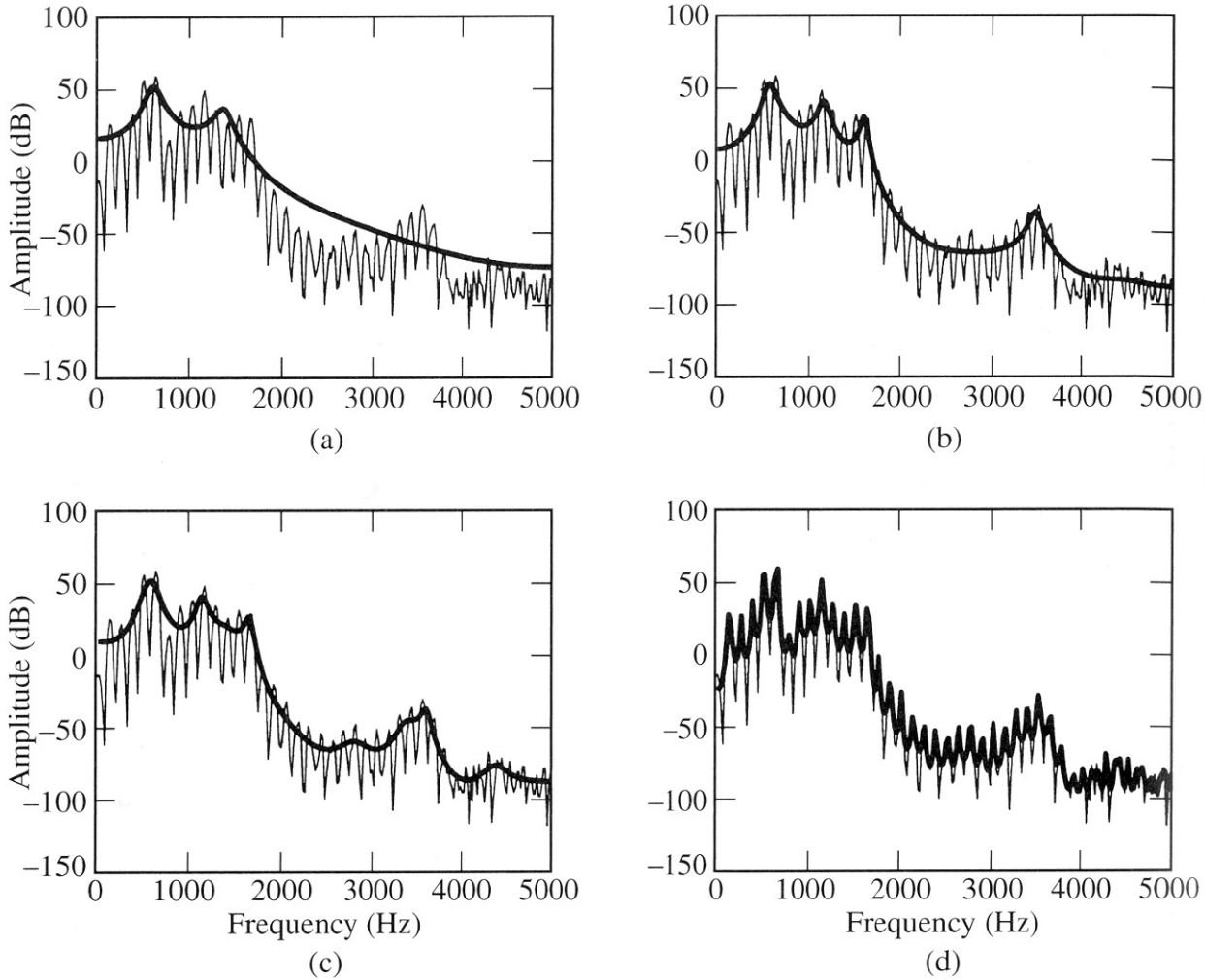

Figure 5.13 Linear prediction analysis of steady vowel sound with different model orders using the autocorrelation method: (a) order 6; (b) order 14; (c) order 24; (d) order 128. In each case, the all-pole spectral envelope (thick) is superimposed on the harmonic spectrum (thin), and the gain is computed according to Equation (5.30).

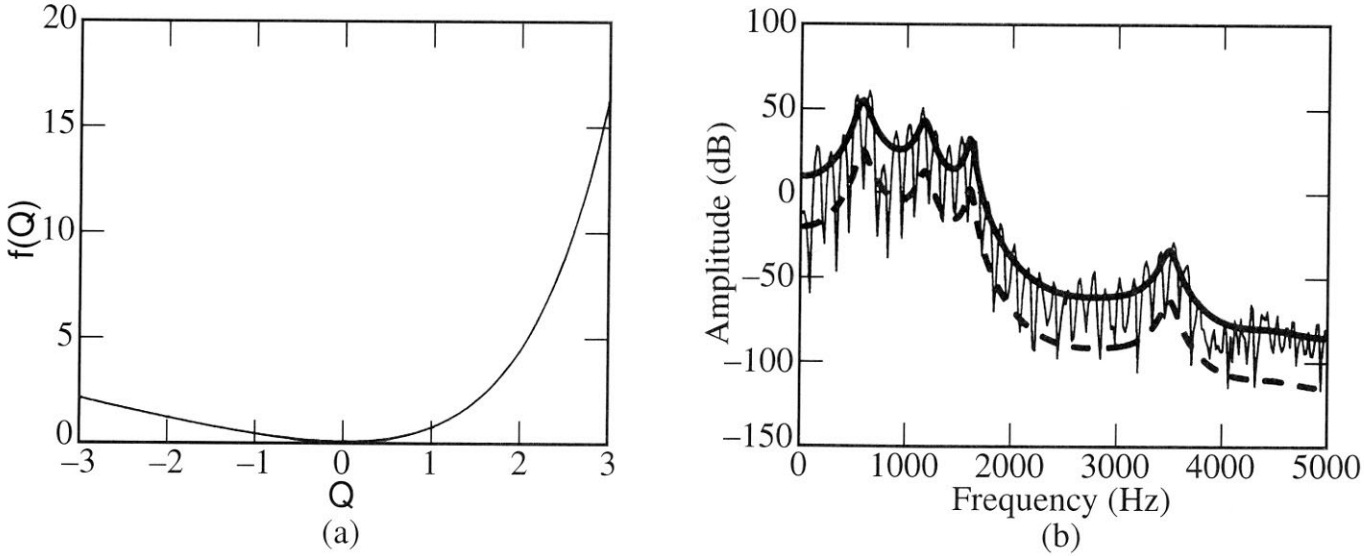

Figure 5.14 Illustration of favoring a match to spectral peaks in linear prediction analysis: (a) the asymmetric function $f(Q) = e^{Q} - Q - 1$; (b) schematic that shows favoring spectral peaks over spectral valleys. Upper (solid) envelope is favored over lower (dashed) envelope.

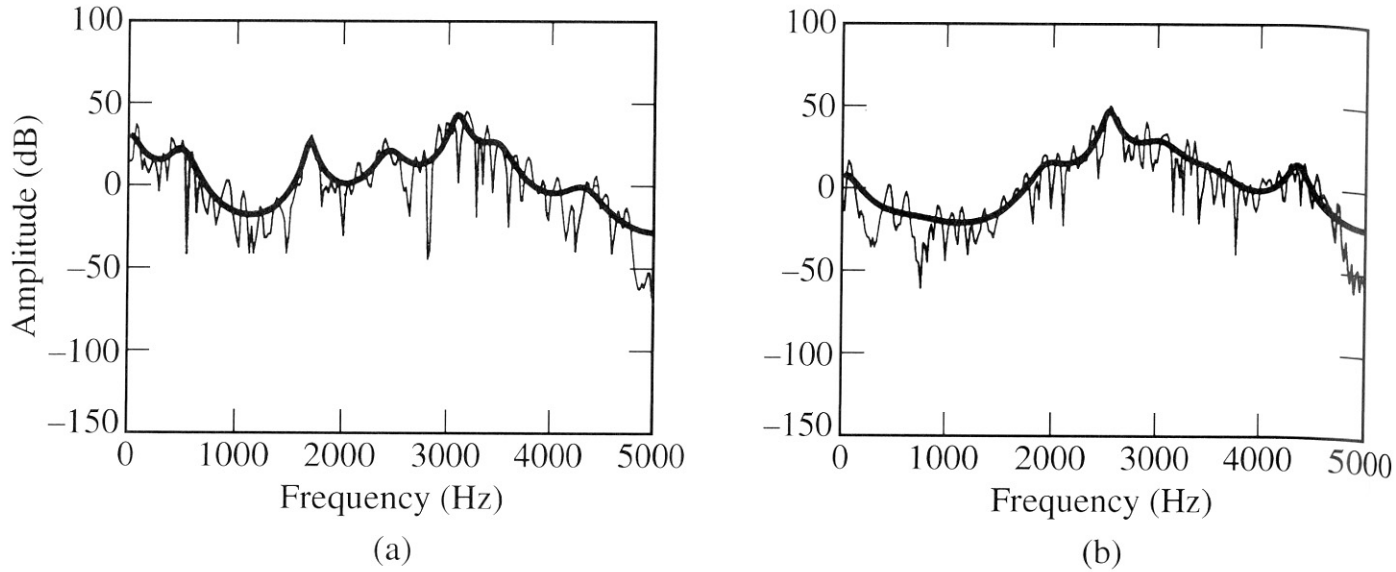

Figure 5.15 Spectral peak matching in linear prediction analysis for (a) an unvoiced plosive ("t" in "tea") and (b) an unvoiced fricative component of an affricate ("ch" in "which") of Figure 5.11. Order 14 was used in the autocorrelation method.

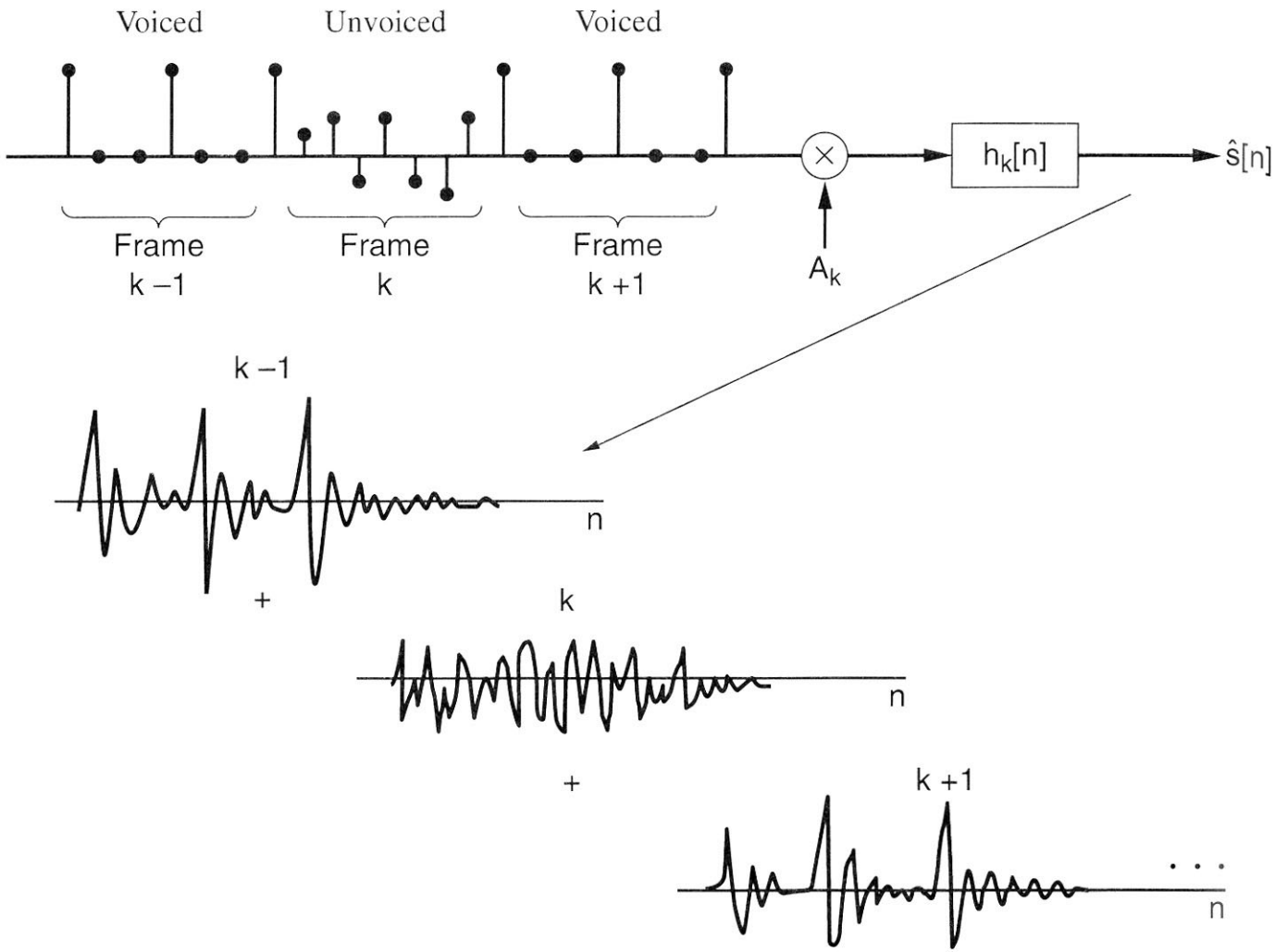

Figure 5.16 Overlap-add synthesis using an all-pole model. The waveform is generated frame by frame by convolutional synthesis and the filter output on each frame is overlapped and added with adjacent frame outputs.

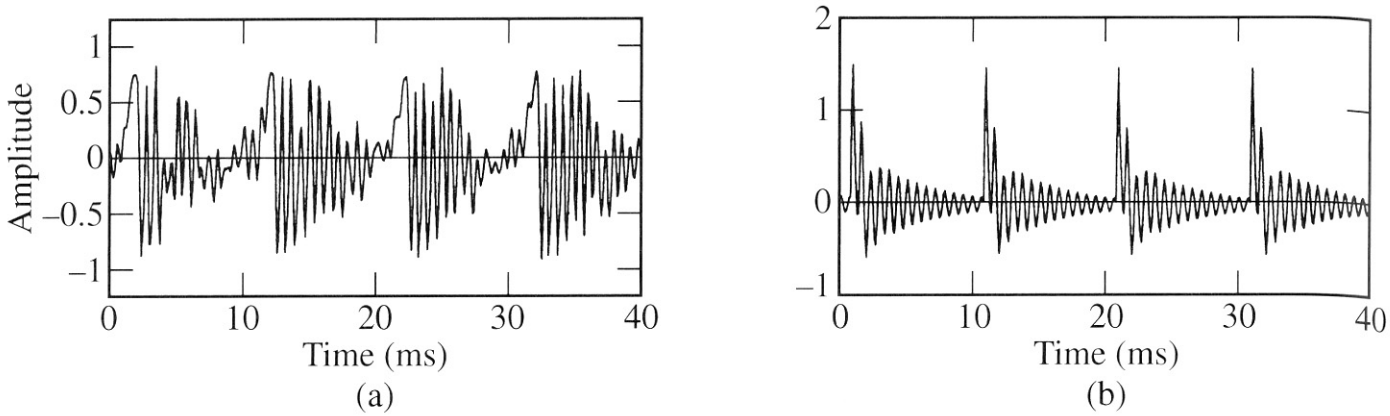

Figure 5.18 Speech reconstruction based on linear prediction analysis using the autocorrelation method: (a) original; (b) synthesized. The reconstruction was performed with overlap-add synthesis and using a 14th-order all-pole model.