Signal representations: Cepstrum

- **Source-filter separation for sound production**
  - For speech, source corresponds to excitation by a pulse train for voiced phonemes and to turbulence (noise) for unvoiced phonemes, and filter corresponds to vocal tract (resonators)
  - For music, source corresponds to vibrations (e.g. vibrating strings in plucked or bowed string instrument) and filter corresponds to the body of the instrument
  - Overall signal reaching the ear is the convolution of source with the impulse response of filter

$$y(t) = x(t) * h(t) = h(t) * x(t) = \int_{-\infty}^{+\infty} x(t - t')h(t')dt'$$

- **Cepstral analysis attempts to separate source from filter, hence it can be viewed as deconvolution**

Speech production illustration

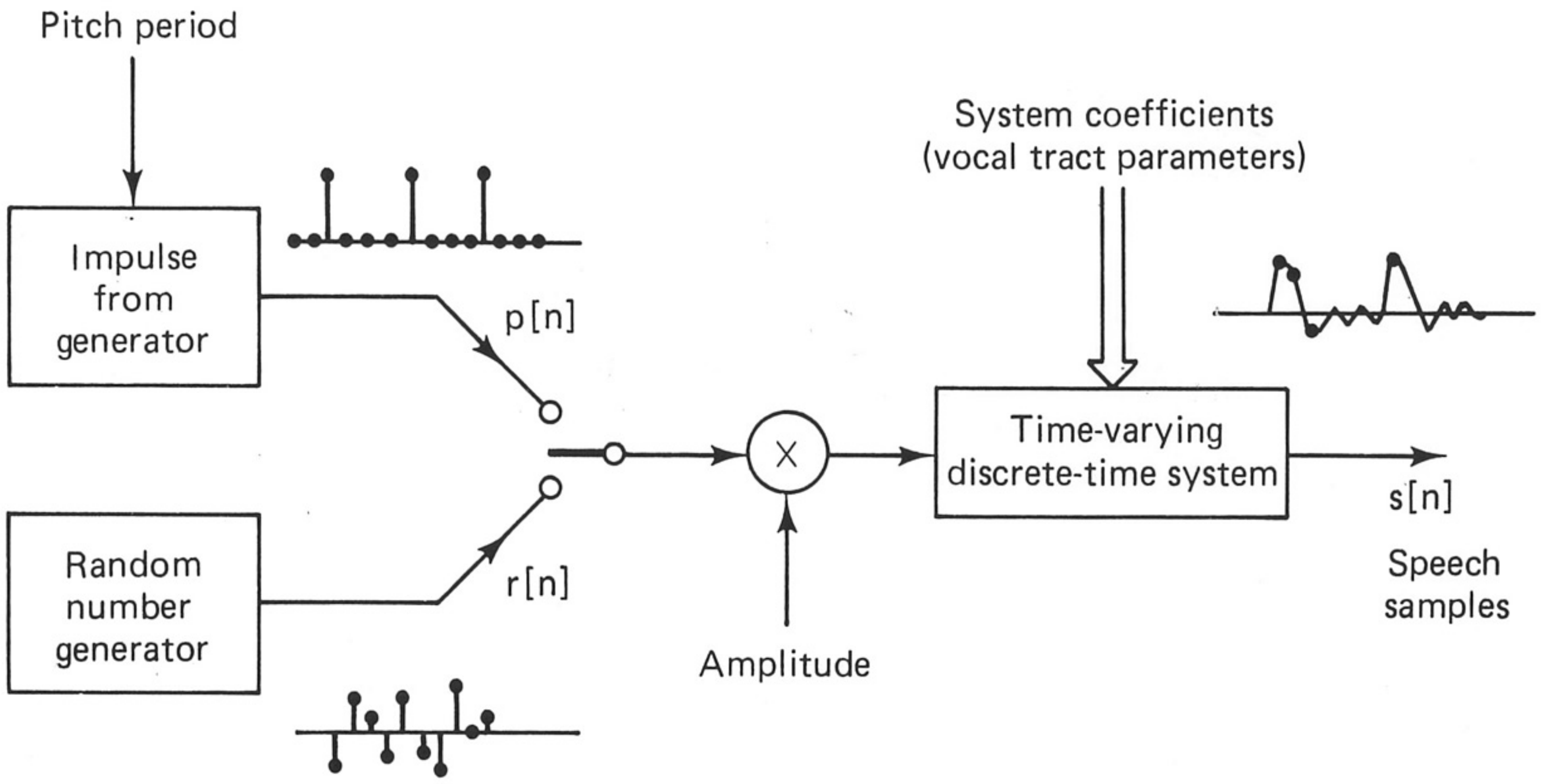


Figure 12.24 Discrete-time model of speech production.

Real cepstrum

- **For speech, the spectral magnitude can be written as**

$$|X(\omega)| = |V(\omega)||E(\omega)|$$

  - Taking the logarithm yields

$$\log|X(\omega)| = \log|V(\omega)| + \log|E(\omega)|$$

- **Observation for speech production**
  - The *E* term corresponds to an event (e.g. a pulse train with a frequency of 100 Hz) more extended in time than the impulse response of the vocal tract. Analogously, *E* corresponds to “carrier” and *V* corresponds to “envelope” in the frequency domain. In other words, *E* varies more quickly with respect to ω than *V*
  - Hence, one can apply some kind of “filter” to separate “high-frequency” components from “low-frequency” components, thus *E* term and *V* term

Real cepstrum (cont.)

- **Change of notations because the variable is frequency rather than time**
  - Filtering -> liftering
  - Frequency response -> quefrency response
  - Spectrum -> cepstrum
  - High (low) frequency components -> high (low) time components or high (low) quefrency components

Real cepstrum (cont.)

- **The log-operation converts a multiplicative term into an additive term, which can be operated upon by a linear operation such as filtering. The cepstrum is defined as the inverse Fourier transform**

$$c(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega n} \log\left|X(\omega)\right| d\omega$$

  - $c(n)$ is called the *n*th cepstral coefficient
  - Given separated cepstra for excitation and vocal tract, they can be inverted to give original spectral magnitudes
  - Only a moderate number of cepstral coefficients (e.g. 10-14) is needed for many applications, including speech recognition
- **Complex cepstrum exists as well**

Cepstral analysis illustration

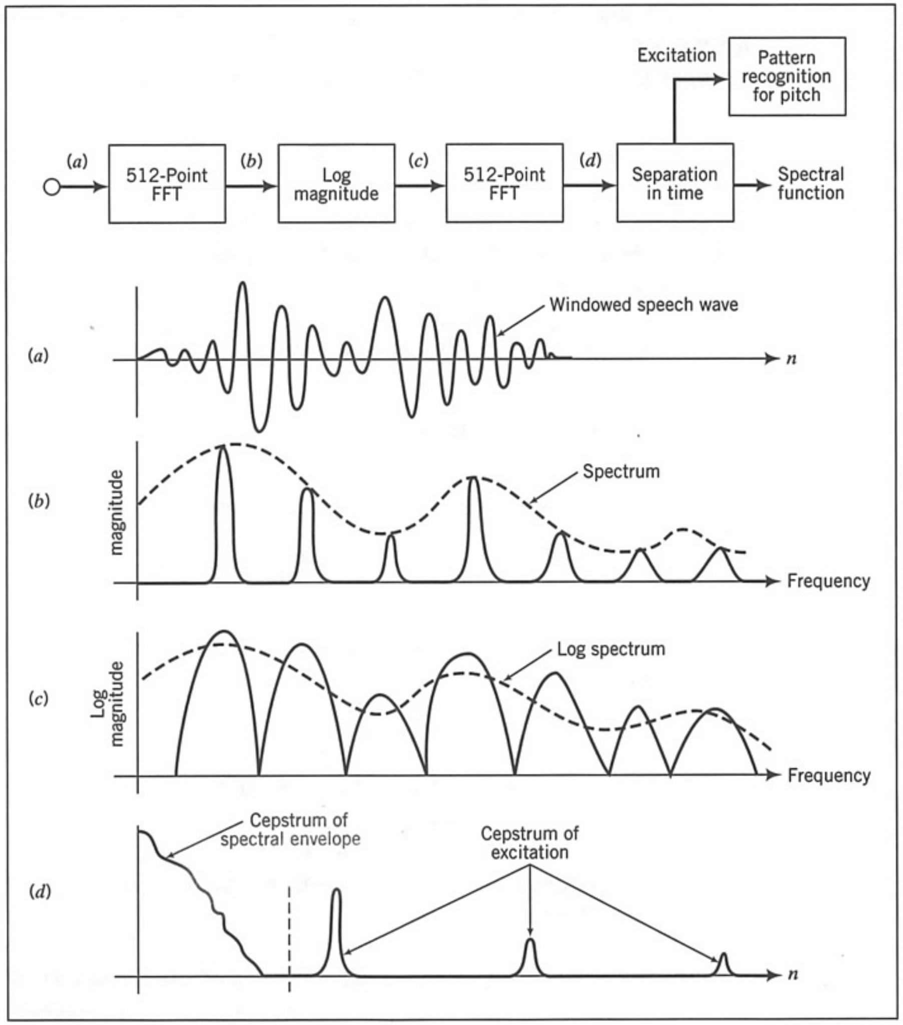


FIGURE 20.1 Cepstral analysis.

Linear predictive coding (LPC) for speech modeling

- **The vocal tract can be modeled as a cascaded set of acoustic tubes, each corresponding to a resonator**
- **Furthermore, each resonator corresponds to a formant**
  - Complete vowel spectrum can be reasonably represented by six resonators
- **A direct implementation of the spectral model is written as an all-pole filter in the complex *z* domain (*z*-transform is the discrete-time counterpart of the Laplace transform - generalized form of the Fourier transform):**

$$H(z) = \frac{1}{1 - \sum_{j=1}^{P} a_j z^{-j}}$$

  - P is twice the number of resonators, a_j's are coefficients

LPC illustration

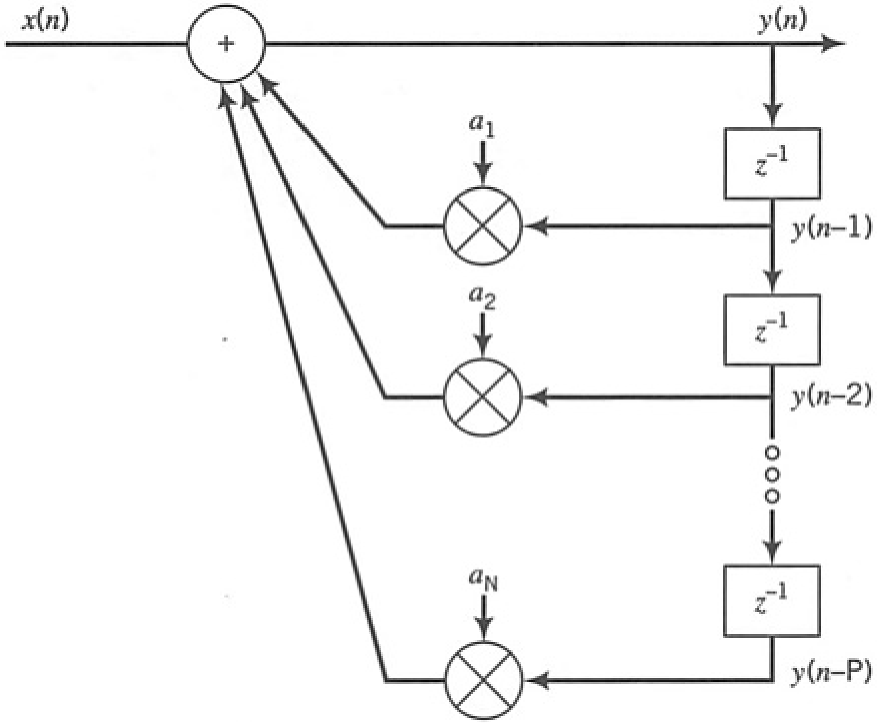


FIGURE 21.2 All-pole model for the generation of a discrete-time sequence.

LPC (cont.)

- **In the above system, the discrete-time response *y*(*n*) to the excitation *x*(*n*) can be written as**

$$y(n) = x(n) + \sum_{j=1}^{P} a_j y(n-j)$$

- **In LPC, the coefficients are computed to give an approximation to the original signal. That is, one attempts to *predict* the speech signal by a linear, weighted sum of its previous values:**

$$\hat{y}(n) = \sum_{j=1}^{P} a_j y(n-j)$$

  - $\hat{y}(n)$ is the linear predictor of $y(n)$
  - The coefficients that produce the best approximation are called the linear prediction coefficients

LPC (cont.)

- **The difference between the predictor and the original signal is called the error signal, residual error, LPC residual, or prediction error**
  - $e(n) = y(n) - y?(n)$ can be viewed as an approximation to the excitation signal

Residual error illustration

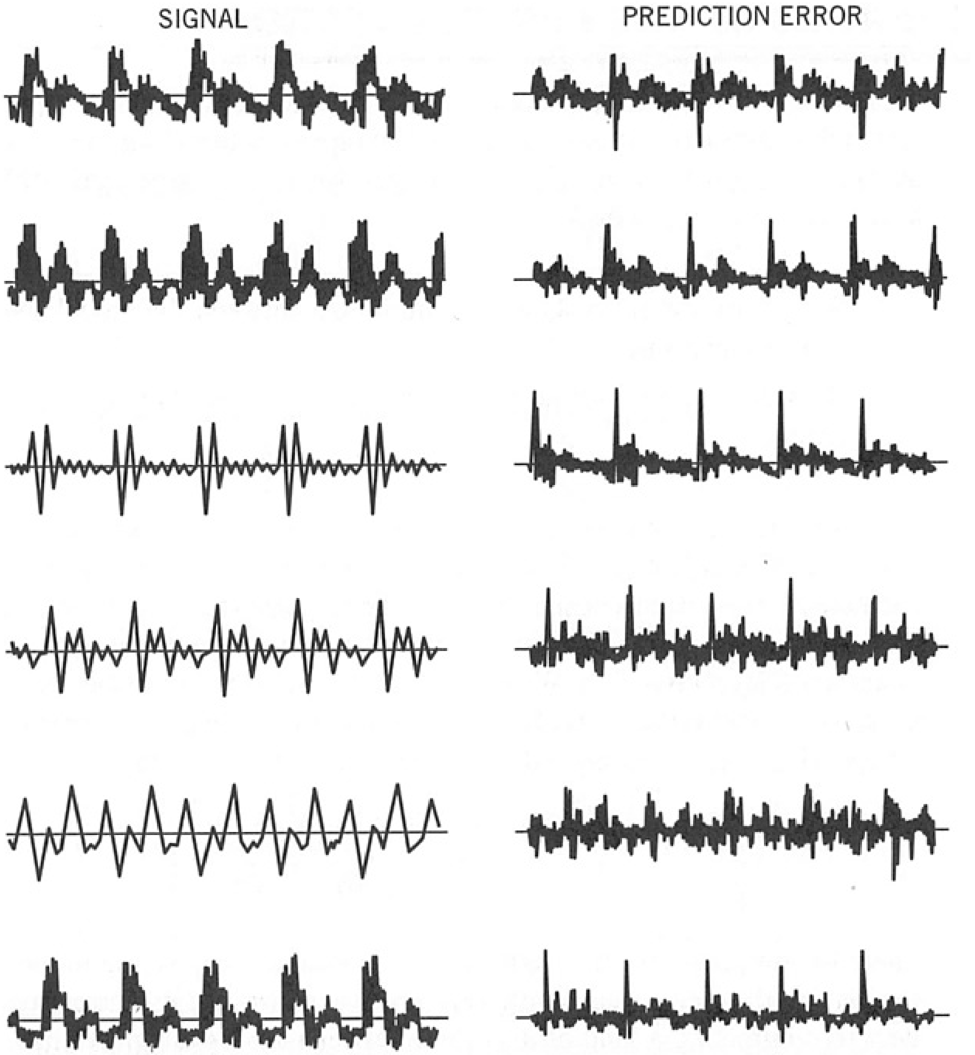


FIGURE 21.3 Residual error waveforms for several vowels. From [4].

LPC (cont.)

- **Computing the coefficients can be viewed as an optimization problem, where square error is generally used**

$$D = \sum_{n=0}^{N-1} e^2(n) = \sum_{n=0}^{N-1} [y(n) - \sum_{j=1}^{P} a_j y(n-j)]^2$$

- **Various methods can be employed to find coefficients, including gradient descent**

LPC (cont.)

- **Properties of LPC representation**
  - For a harmonic signal, the (spectral) model spectrum tends to follow (hug) harmonic peaks, but not harmonic valleys, hence yielding an estimate of the envelope of the signal spectrum
  - Too many coefficients will yield a good fit to signal spectrum, but miss spectral envelope. On the other hand, too few coefficients will miss formants. A reasonable number is between 10 and 20.
  - Prediction error is significantly higher for unvoiced speech
- **Compared to Fourier and cepstral analysis, LPC is more directly related to vocal tract characteristics**

More LPC illustrations

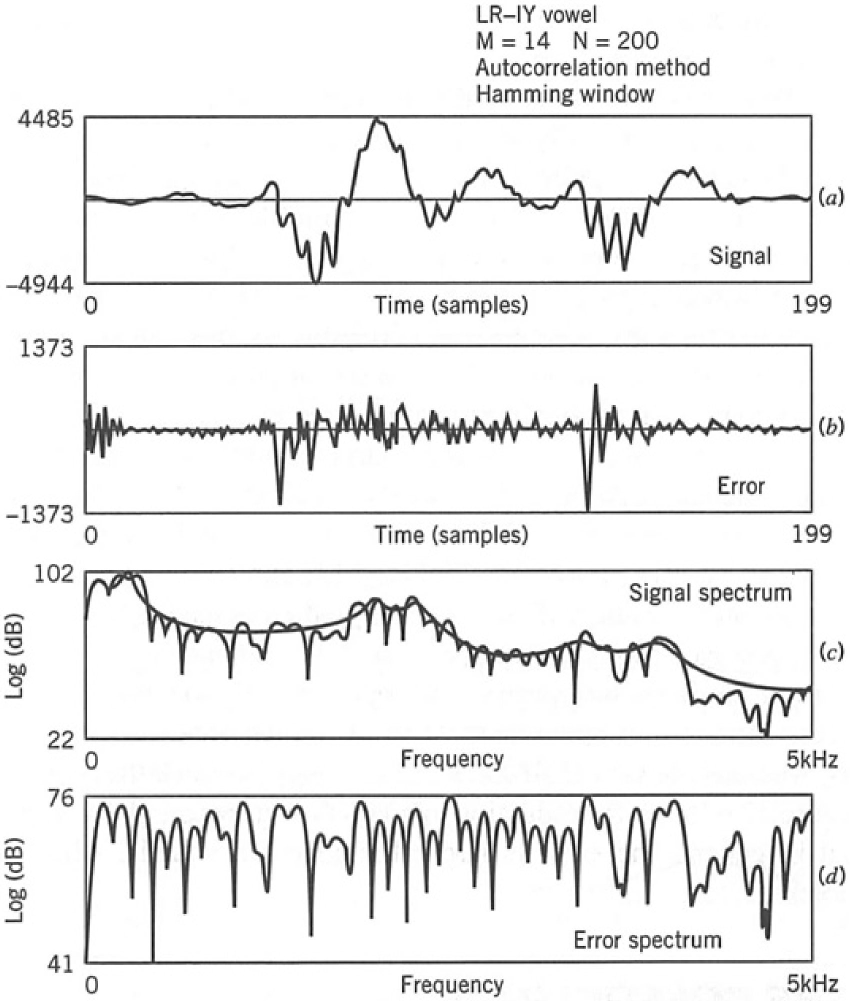


FIGURE 21.4 Example of (a) a windowed speech signal, (b) the LPC error signal, (c) the signal spectrum with the LPC spectral envelope superimposed, and (d) the LPC error spectrum. From [4].

More LPC illustrations

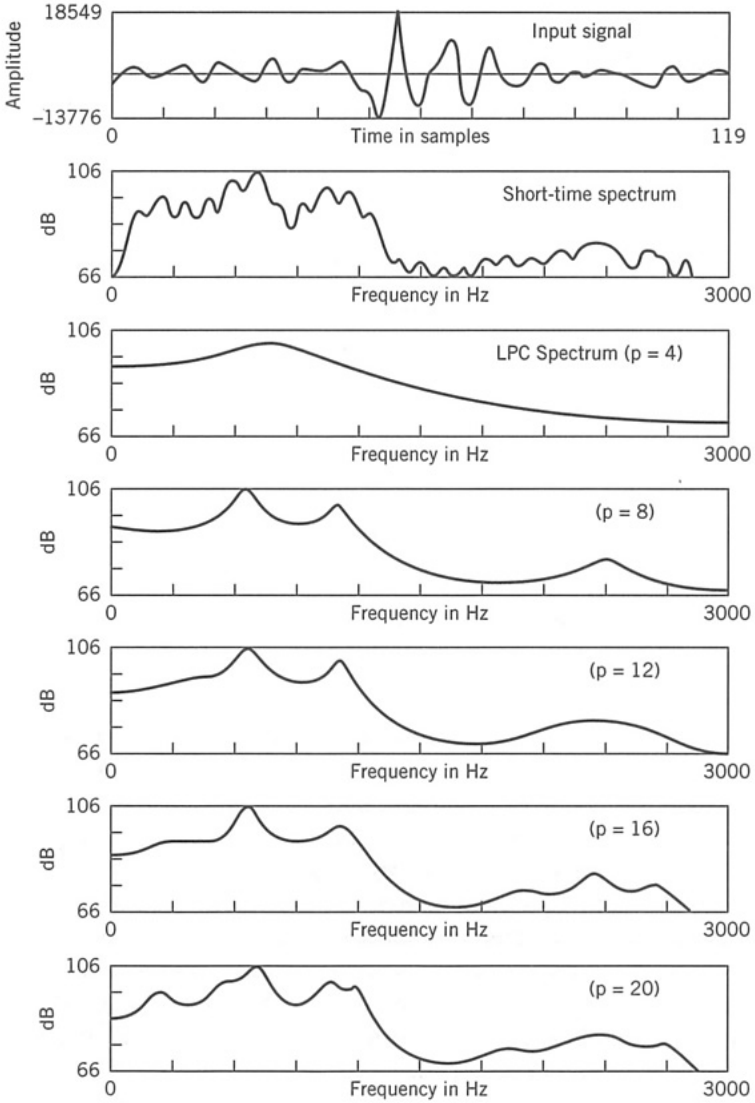


FIGURE 21.5 LPC speech spectra for different model orders. From [4].

More LPC illustrations

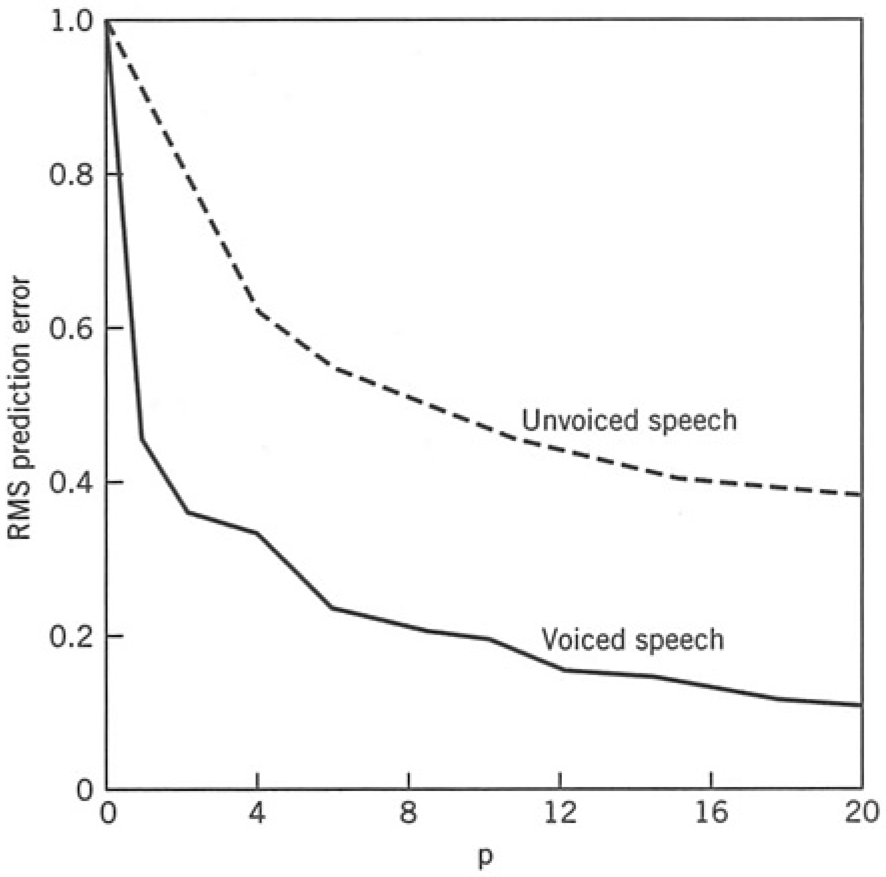
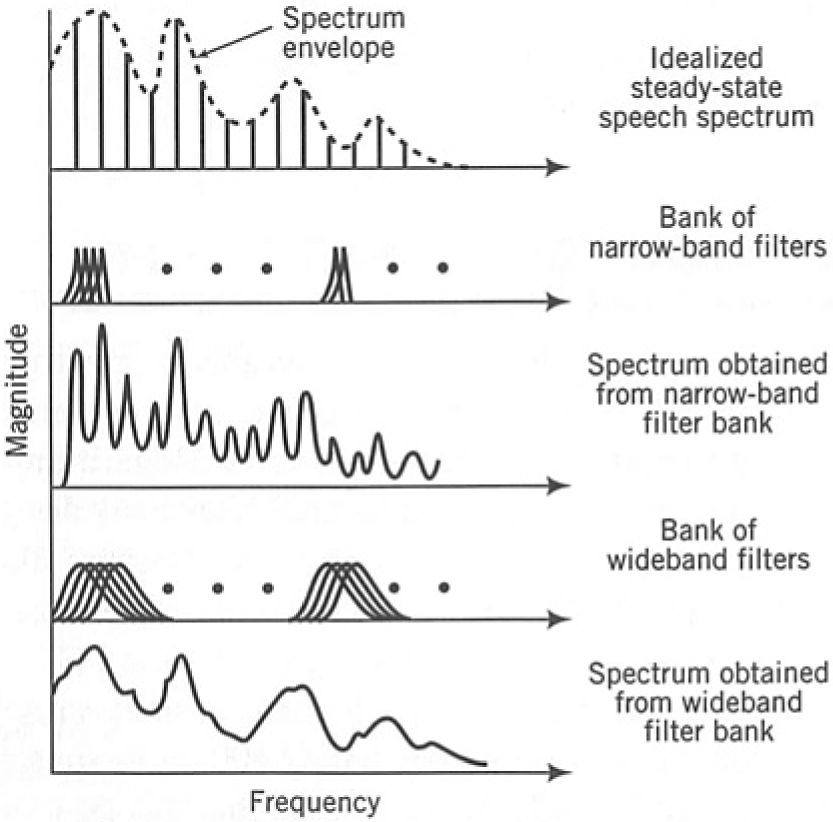


FIGURE 21.6 Root-mean-square prediction error for different model orders. From [4].

Spectral analysis via filterbanks


Spectrum envelope
Idealized steady-state speech spectrum
Bank of narrow-band filters
Spectrum obtained from narrow-band filter bank
Bank of wideband filters
Spectrum obtained from wideband filter bank
Magnitude
Frequency

Comparison of (Idealized) measured spectra for wide and narrow filter-bank analyzers

FIGURE 19.10 Narrow-band and wideband spectral analyses for an idealized speech sound.

Summary table

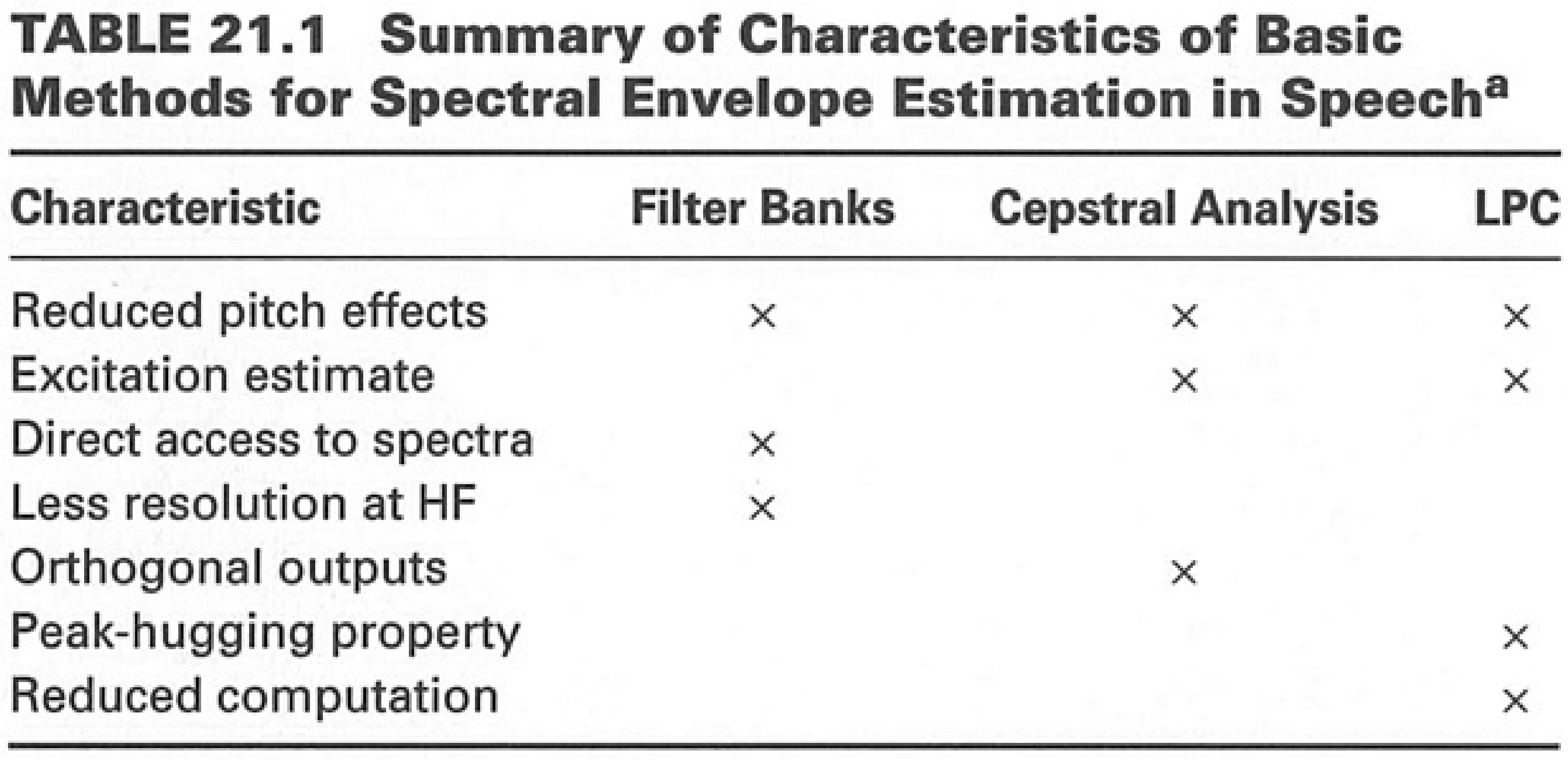


TABLE 21.1 Summary of Characteristics of Basic Methods for Spectral Envelope Estimation in Speech[a]

Characteristic	Filter Banks	Cepstral Analysis	LPC
Reduced pitch effects	×	×	×
Excitation estimate		×	×
Direct access to spectra	×		
Less resolution at HF	×		
Orthogonal outputs		×	
Peak-hugging property			×
Reduced computation			×