Time-Domain Parameters for Processing Speech

Time-Domain Parameters for Speech Processing

- Short-time Average Energy
- Short-time Average Zero Crossing Rate (ZCR)
- Short-time Autocorrelation function
- Short-time Average Magnitude Difference Function (AMDF)

Short-Time Average Energy (STAE)

- Segment speech to phonemes & syllables (finer variations STAE contour)
- Voiced/Unvoiced segment detection (larger variations of STAE contour)
- Word & Phrase boundaries or Silence detection
- Begin-End detection
- Multiplex several conversations by exploiting the pause regions

Short-Time Average Energy (STAE)

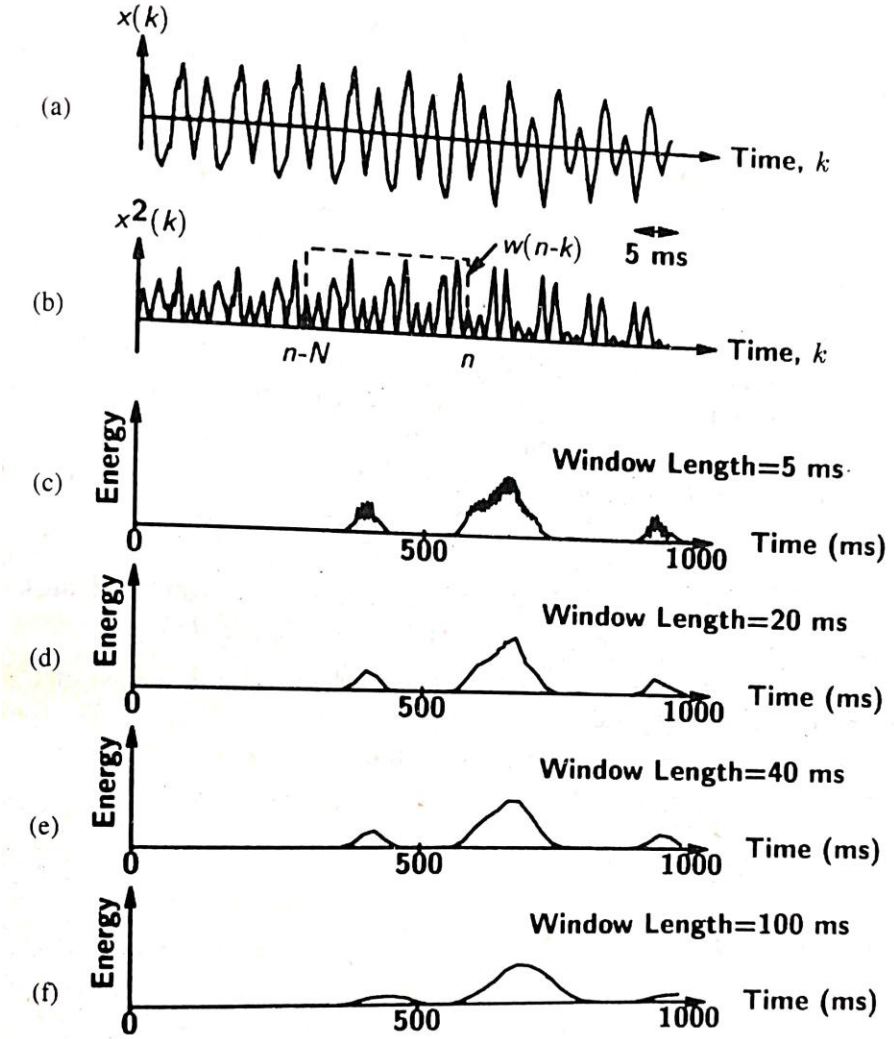
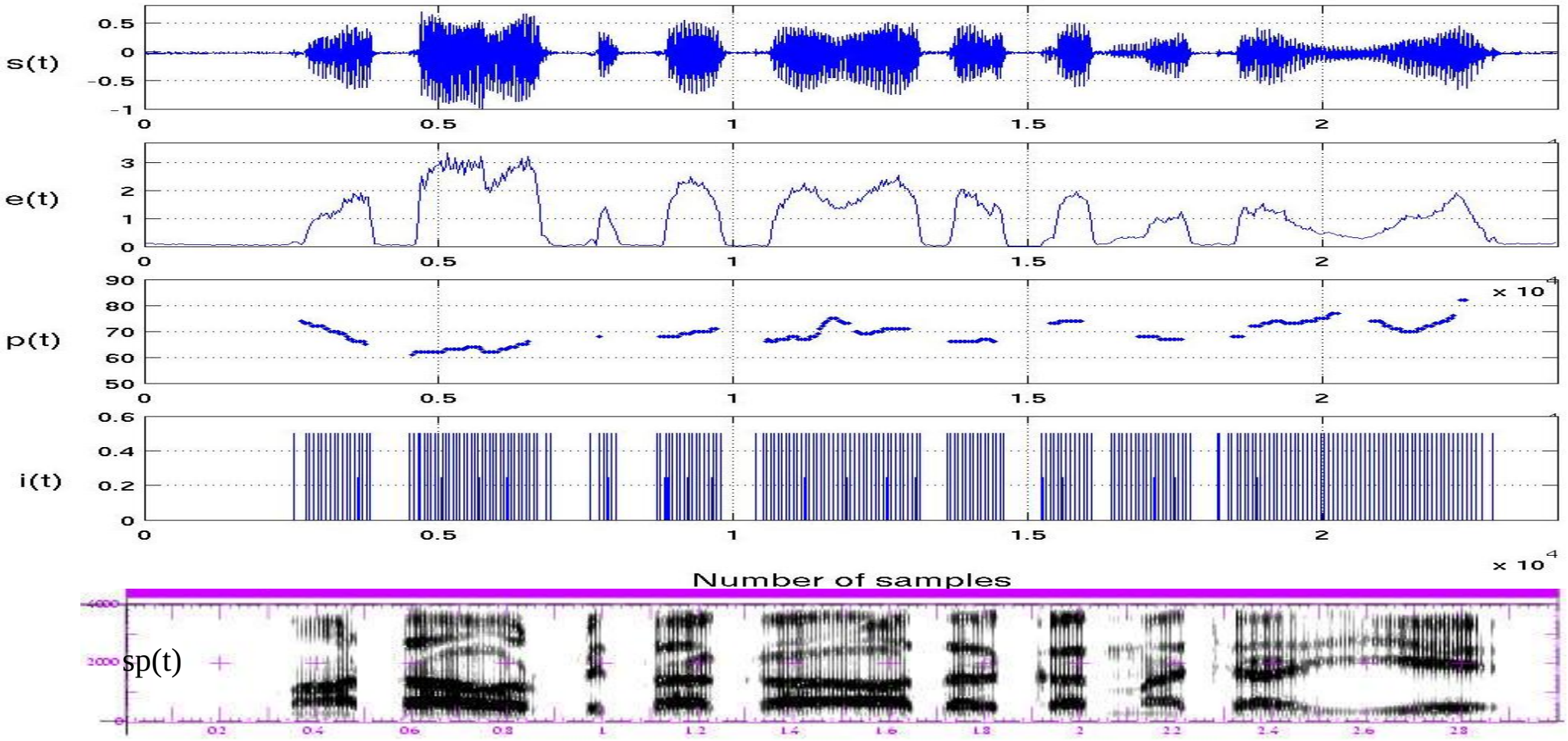


Figure 6.5 Illustration of the computation of short-time energy: (a) 50 ms of a vowel, (b) the squared version of (a), with a superimposed window of length N samples delayed n samples, (c–f) energy function for a 1 s utterance, using rectangular windows of different lengths.

Illustration of knowledge sources in speech (Contd..)



s(t): speech waveform, e(t): energy contour, p(t): pitch contour, i(t): instants of significant excitation and sp(t) : spectrogram

Short-Time Average Zero Crossing Rate

- Indirect spectral estimation of narrow-band signal
  - ✓ Computationally efficient
  - ✓ $F_0 = \frac{\text{ZCR}}{2}$ x $(\#\frac{\text{cycles}}{\text{sec}})$
  - ✓ $F_0 = \frac{\text{ZCR}}{2}$ x F_s
- ZCR correlates with the average frequency of major energy concentration
- High ZCR → Unvoiced (4900)
- Low ZCR → Voiced (1400)
- Difficulty for detection of voiced fricatives
  - ✓ Periodic energy in the voiced bar vs. Unvoiced energy due to HF noise
  - ✓ \Z\ → low zcr & \v\ → high zcr
- Application to speech recognition

Short-Time Average Zero Crossing Rate

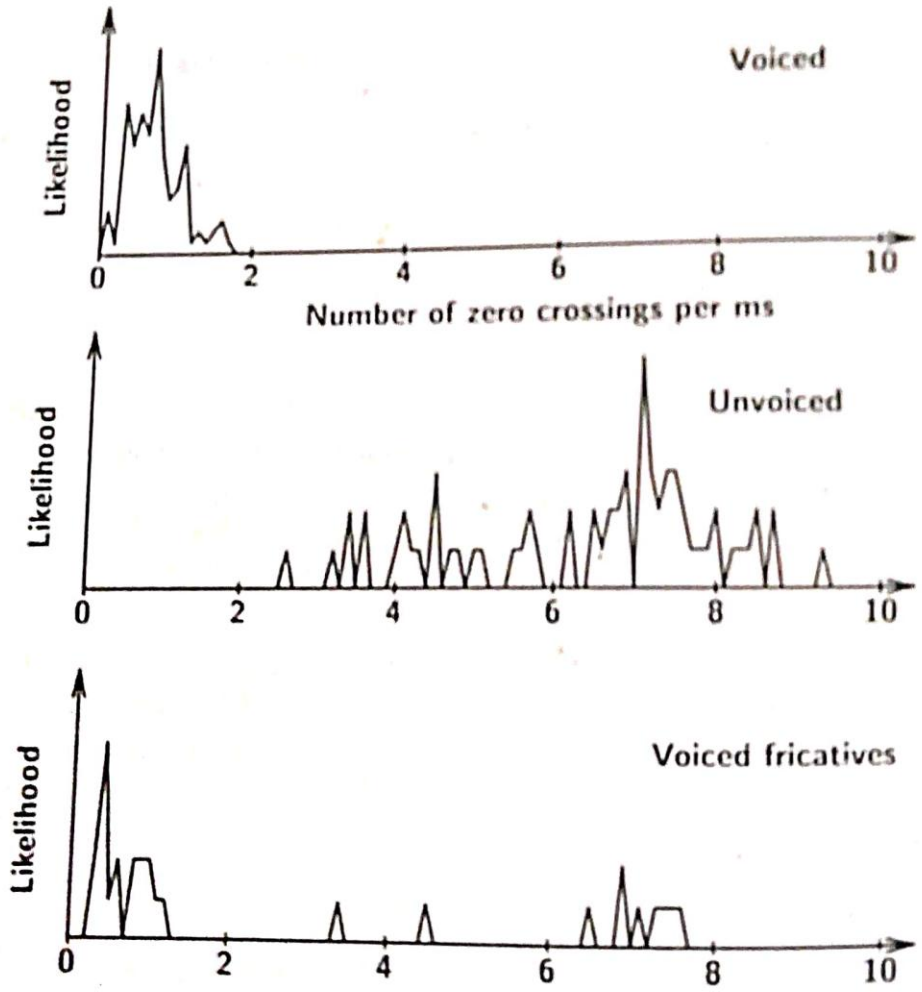


Figure 6.6 Typical distribution of zero-crossings for voiced sonorants, for unvoiced frication, and for voiced frication.

Short-Time Autocorrelation Function (STACF)

$$R(k) = \sum s(n)s(n-k)$$

$$R(k) = IFT\left\{\left|S\left(e^{jw}\right)\right|^2\right\}$$

- F0 Estimation
- LP analysis (Autocorrelation coefficients)
- Voiced/Unvoiced region detection
- Computational Complexity
  - ✓ F0 Estimation : Larger K
  - ✓ LP-Spectrum : Low values of K

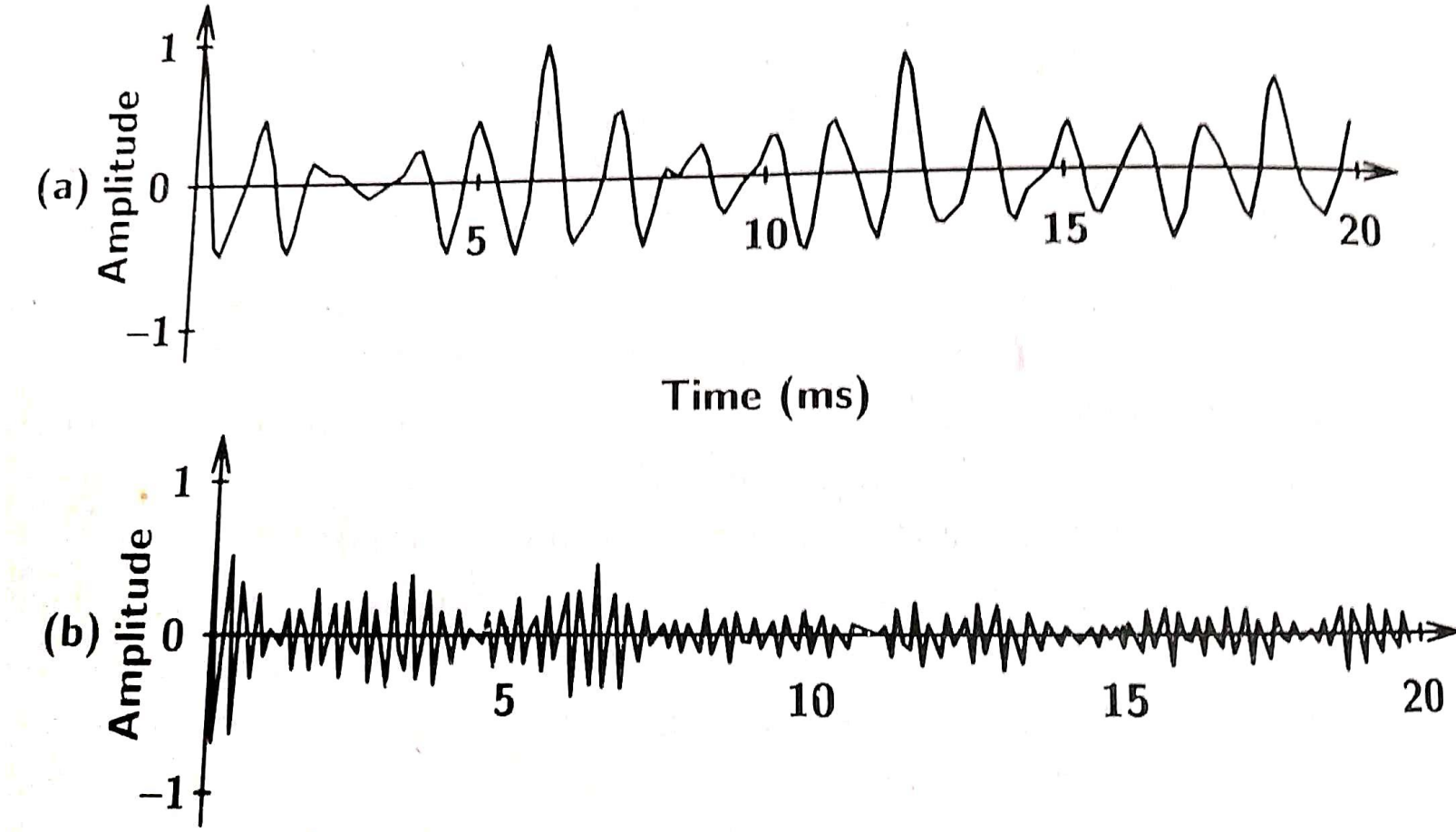


Figure 6.7 Typical autocorrelation function for (a) voiced speech and (b) unvoiced speech, using a 20 ms rectangular window ($N = 201$).

Short-Time Average Magnitude Difference Function

$$AMDF(k) = \sum |s(n) - s(n-k)|$$

- Performs similar objective as ACF
- Computationally efficient

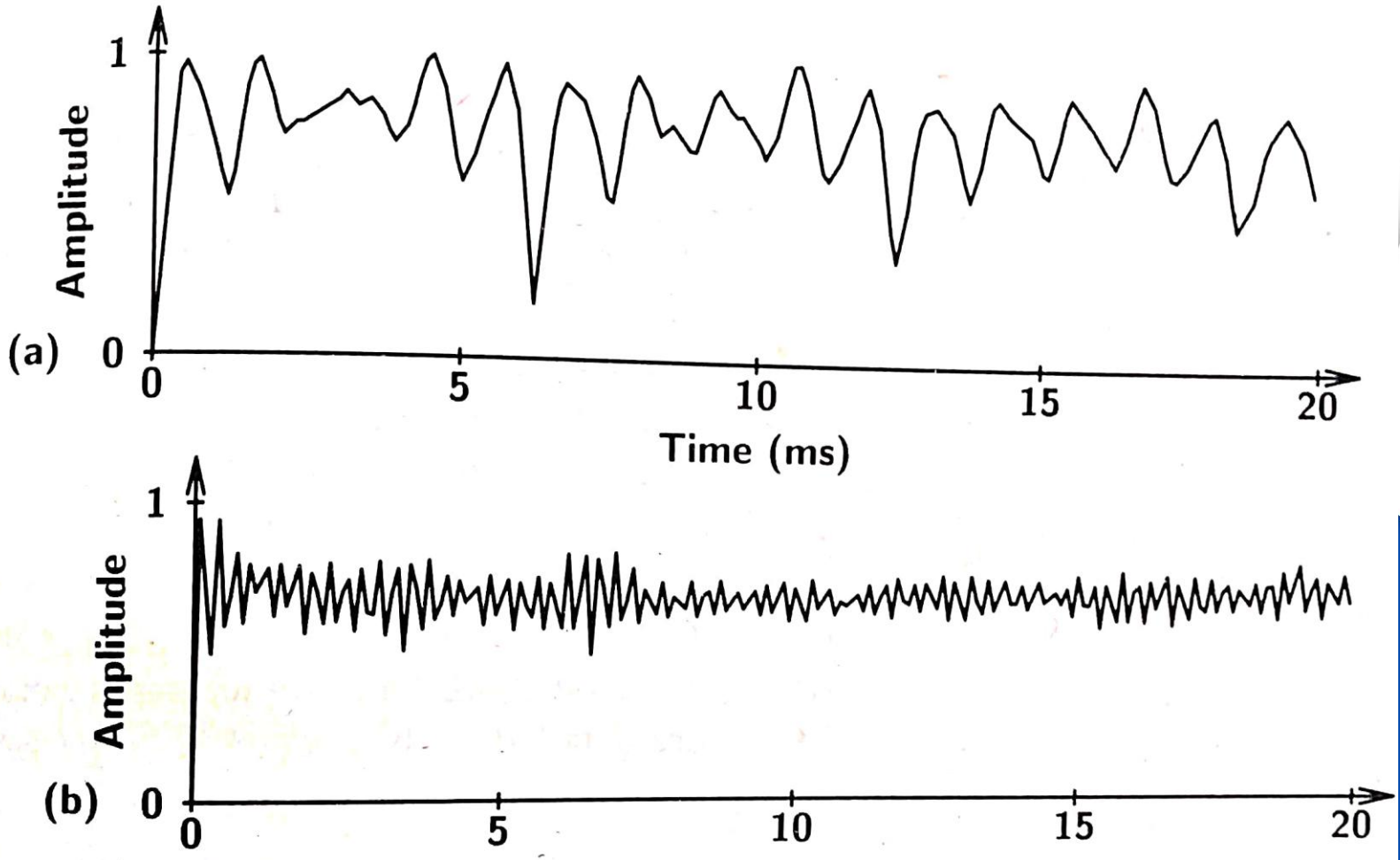


Figure 6.8 AMDF function (normalized to 1.0) for the same speech segments as in Figure 6.7.