Short-Term Autocorrelation and Covariance

Recall our definition of the autocorrelation function:

$$r(k) = \frac{1}{N}\sum_{n=k}^{N-1} s(n)s(n-k) = \frac{1}{N}\sum_{n=0}^{N-1-k} s(n)s(n+k) \qquad 0 < k < p$$

Note:

- this can be regarded as a dot product of $s(n)$ and $s(n+i)$.
- let's not forget preemphasis, windowing, and centering the window w.r.t. the frame, and that scaling is optional.

What would C++ code look like:

array-style:

```
for(k=0; k<M; k++) {
  r[k] = 0.0;
  for(n=0; n<N-k; n++) {
    r[k] += s[n] * s[n+k]
  }
}
```

pointer-style:

```
for(k=0; k<M; k++) {
  *r = 0.0; s = sig; sk = sig + k;
  for(n=0; n<N-i; n++) {
    *r += (s++) * (sk++);
  }
}
```

We note that we can save some multiplications by reusing products:

$$r[3] = s[3](s[0]+s[6]) + s[4](s[1]+s[7]) + \ldots + s[N]s[N-3]$$

This is known as the *factored autocorrelation* computation.

It saves about 25% CPU, replacing multiplications with additions and more complicated indexing.

Similarly, recall our definition of the covariance function:

$$c(k,l) = \frac{1}{N}\sum_{n=p}^{N-1} s(n-k)s(n-l) \qquad 0 < l < k < p$$

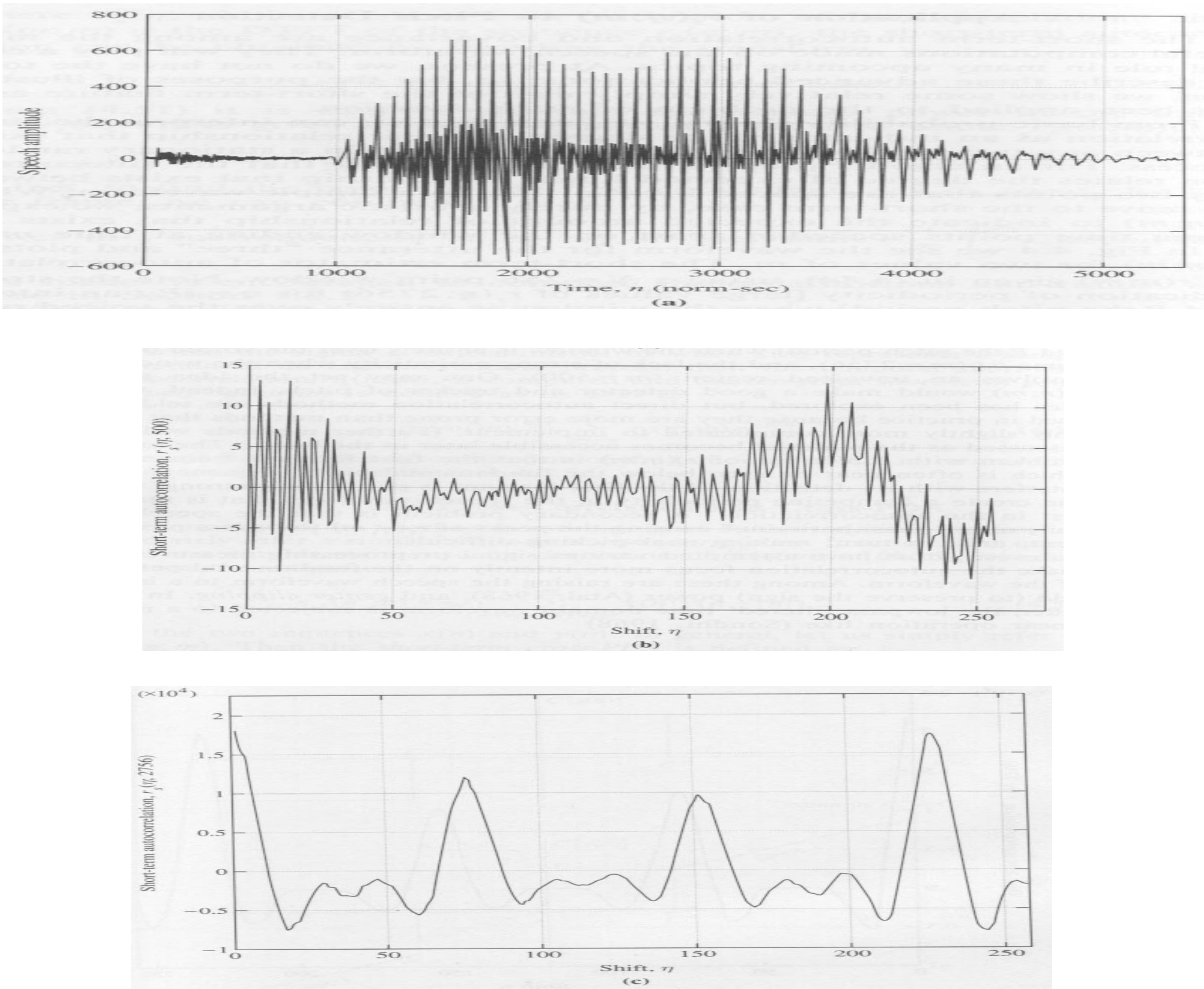
Note:

- we use N-p points
- symmetric so that only the $k \geq l$ terms need to be computed

This can be simplified using the recursion:

$$c(k,l) = c(k-1,l-1) + s(p-k)s(p-l) - s(N-k)s(N-l)$$

Example of the Autocorrelation Function



Autocorrelation functions for the word “three” comparing the consonant portion of the waveform to the vowel (256-point Hamming window).

Note:

- shape for the low order lags - what does this correspond to?
- regularity of peaks for the vowel — why?
- exponentially-decaying shape — which harmonic?
- what does a negative correlation value mean?

Other Short-Term Spectral Measures

Power Density Spectrum:

$$P_s(f) = \sum_{k=-\infty}^{\infty} r(k) e^{-j2\pi fk}$$

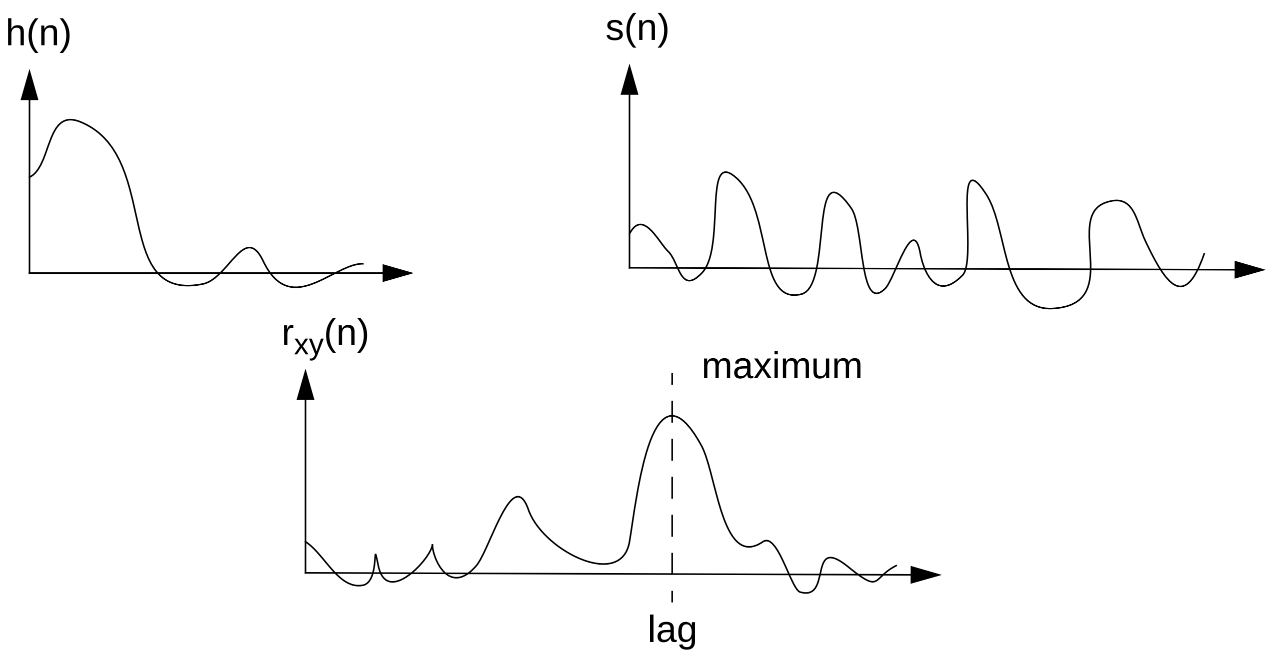
Cross-Correlation:

$$r_{xy}(k) = \sum_{n=0}^{N-1} x(n) y(n-k)$$

Cross-Power Density Spectrum:

$$P_{xy}(f) = \sum_{k=-\infty}^{\infty} r_{xy}(k) e^{-j2\pi fk}$$

Example: Lag Estimation



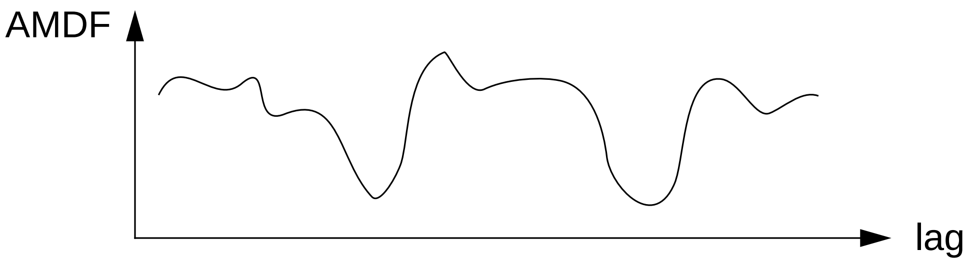
The cross-correlation can be used to find the location of the best match, in a least-square error sense, between $h(n)$ and $s(n)$ (for example, in multipulse linear predictive coding).

Other Interesting Short-Term Time-Domain Measures

Average Magnitude Difference Function (AMDF):

$$AMDF(k) = \frac{1}{N}\sum_{n=k}^{N-1}|s(n)-s(n-k)|$$

Why? What does this look like:



Zero-Crossing Measure:

$$Z_s(n) = \frac{1}{N}\sum_{m=1}^{N-1}\frac{|\mathrm{sgn}\{s(n-m)\}-\mathrm{sgn}\{s(n-m-1)\}|}{2}$$

Gives a crude estimate of the fundamental frequency.

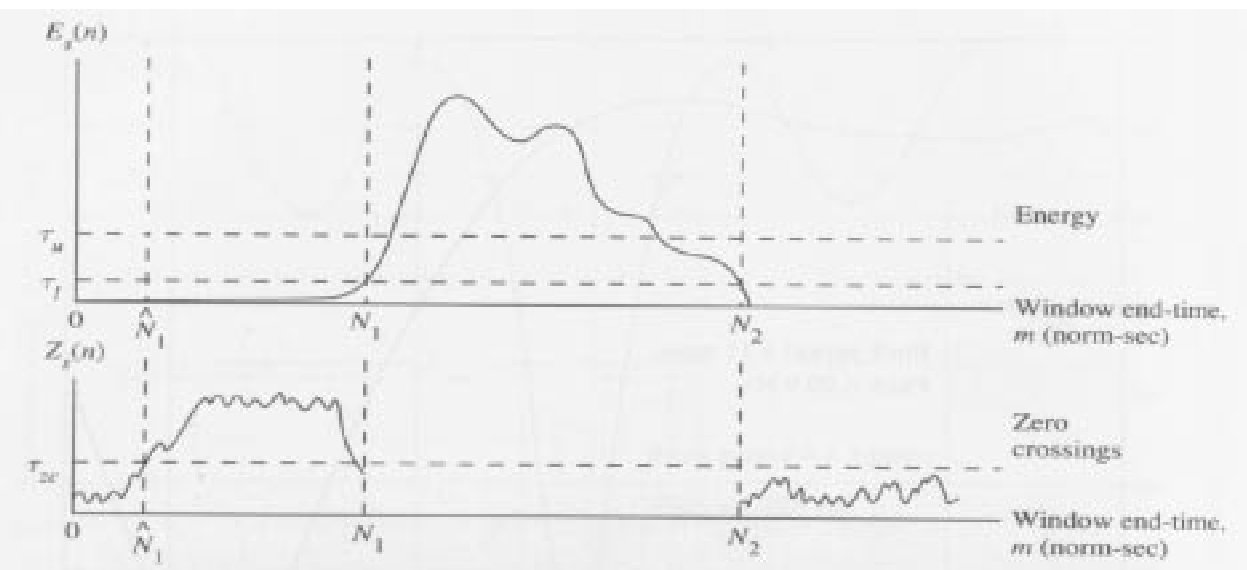
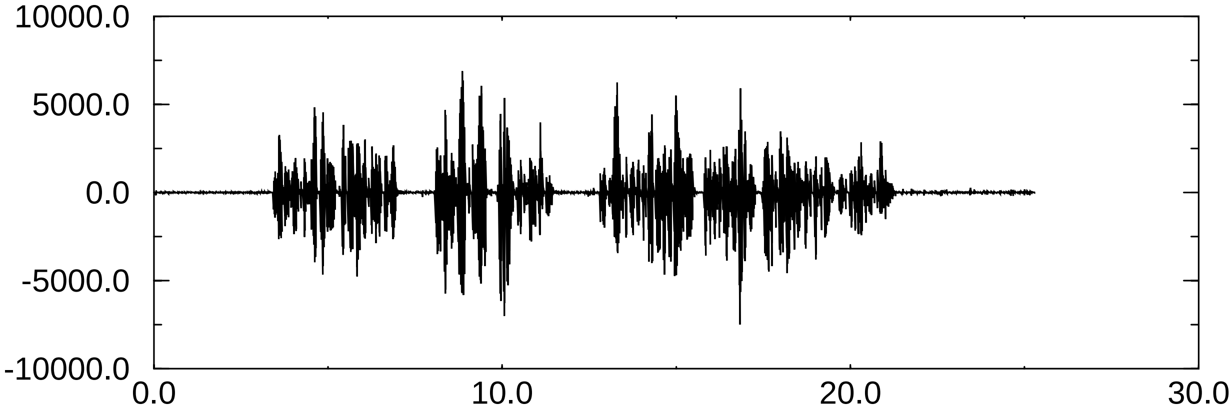


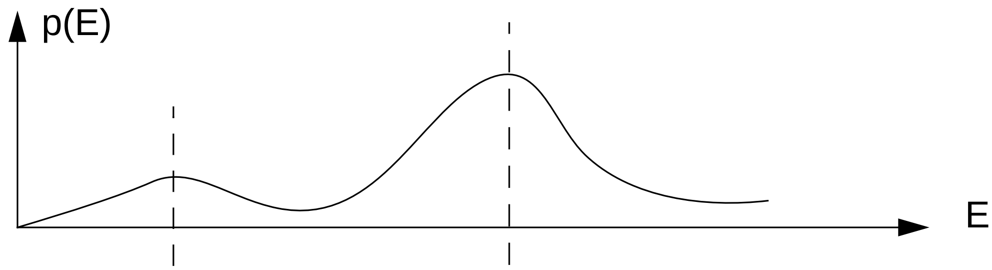
FIGURE 4.9. Short-term energy and zero crossing measures plotted for the word "four." The utterance has a strong fricative at the beginning; hence, the zero crossing level is initially high, and the energy low. The opposite is true as the signal enters the voiced portion of the utterance. After Rabiner and Sambur (1975).

An Application of Short-Term Power: Speech SNR Measurement

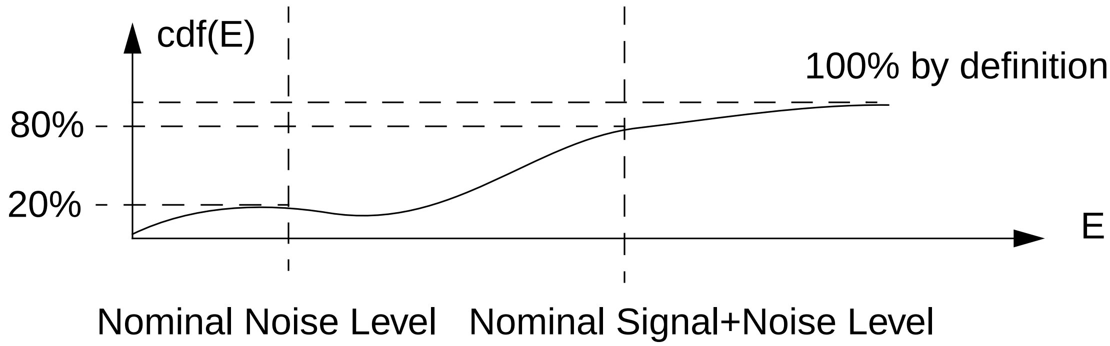
Problem: Can we estimate the SNR from a speech file?



Energy Histogram:



Cumulative Distribution:



The SNR can defined as:

$$SNR = 10\log_{10}\frac{E_s}{E_n} = 10\log_{10}\frac{(E_s + E_n) - E_n}{E_n}$$

What percentiles to use?

Typically, 80%/20%, 85%/15%, or 95%/15% are used.