AMSC 663-4 Project Proposal Fall 2008

Speech Enhancement: Reduction of Additive Noise in the Digital Processing of Speech

Avner Halevy
Professor Radu Balan

Background

- Speech: let’s talk about it!
- The enemy: noise (it’s bad for you!)
- Noise: where does it come from?
  - Noisy environment
  - Noise added in processing
    - Recording
    - Transmission
    - Reproduction

The Challenge

- Reduce noise without distorting the clean signal
- Improve
  - Quality
  - Intelligibility
- Our focus: additive white Gaussian noise

$$y(n) = x(n) + d(n)$$

The Method: Overview

- Short time analysis
  - Speech is highly non stationary
  - Break down the signal into small pieces (frames)
  - Transform into frequency domain (Fourier)
  - Alternative schemes: wavelets, or nonlinear
- Process using one of several algorithms (the heart)
- Synthesis: reconstructing an enhanced signal

Spectral Subtraction - Words

- Estimate the magnitude of the noise spectrum when speech is absent
- Subtract the estimate from the magnitude of the spectrum of the noisy signal
- Keep the noisy phase
- Inverse Fourier to obtain enhanced signal in time domain

Spectral Subtraction - Symbols

$$y(n) = x(n) + d(n)$$

$$Y(\omega) = X(\omega) + D(\omega)$$

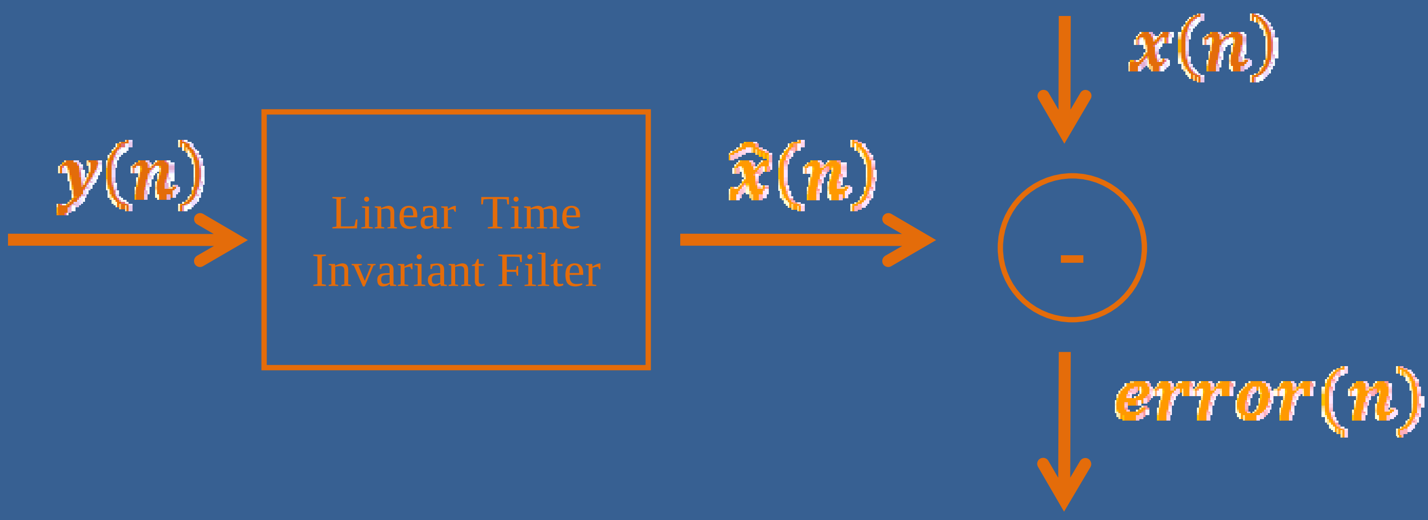
$$Y(\omega) = |Y(\omega)|e^{i\varphi_y(\omega)}$$

$$\hat{X}(\omega) = [|Y(\omega)| - |\hat{D}(\omega)|]e^{i\varphi_y(\omega)}$$

$$\hat{x}(n) = inverse\ Fourier\ \{\ \hat{X}(\omega)\ \}$$

Wiener Filtering - Basic

- Construct linear time invariant filter
- Optimal: minimizes mean square error



Iterative Wiener Filtering

- Assume a speech production model determined by a finite number of parameters
- Given an estimate of the clean signal, compute the parameters using linear prediction
- Using the parameters, compute optimal filter
- Estimate clean signal using the updated filter
- Repeat, using new estimate of clean signal

Additional Details

- Implementation on PC using MATLAB
- Validation using artificially constructed signals
- Testing (objective) using NOIZEUS
  - Intelligibility
  - Quality

$$SNR_{seg} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \frac{\sum_{n=Nm}^{Nm+N-1} x^2(n)}{\sum_{n=Nm}^{Nm+N-1} (x(n) - \hat{x}(n))^2}$$

$$fwSNR_{seg} = \frac{10}{M} \sum_{m=0}^{M-1} \frac{\sum_{j=1}^{K} B_j \log_{10}[\frac{E_c(m,j)}{E_e(m,j)}]}{\sum_{j=1}^{K} B_j}$$

Bibliography

- [1] Deller, J., Hansen, J., and Proakis, J. (2000) *Discrete Time Processing of Speech Signals*, New York, NY: Institute of Electrical and Electronics Engineers
- [2] Quatieri, T. (2002) *Discrete Time Speech Signal Processing*, Upper Saddle River, NJ: Prentice Hall
- [3] Loizou, P. (2007) *Speech Enhancement: Theory and Practice*, Boca Raton, FL: Taylor & Francis Group

Questions?

Speech now, or forever hold your peace

(hey, it’ll make noise reduction much easier!)