

SPEECH ENHANCEMENT: CONCEPT AND METHODOLOGY

Presented by Dominic K. C. Ho

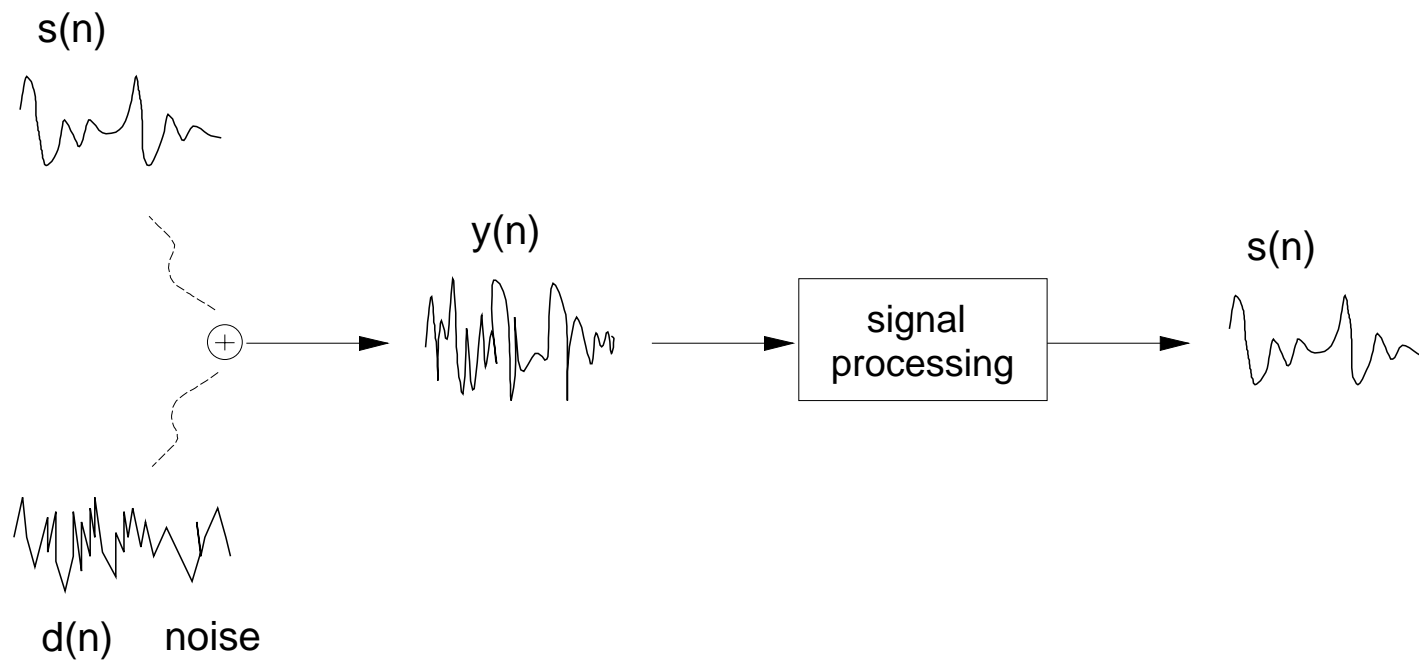
Demo prepared by Tong Wang

University of Missouri-Columbia

Background

Problem:

- recover $s(n)$ from $y(n) = s(n) + d(n)$



References:

- J. S. Lim and A. V. Oppenheim, "Enhancement and Bandwidth Compression of Noisy Speech," *Proc. IEEE*, vol. 67, No. 2, pp. 1586-1604, Dec. 1979.
- J. H. L. Hansen and M. A. Clements, "Constrained Iterative Speech Enhancement with Application to Speech Recognition," *IEEE Trans. Signal Processing*, vol. 39, No. 4, pp. 795-805, Apr. 1991.

Applications:

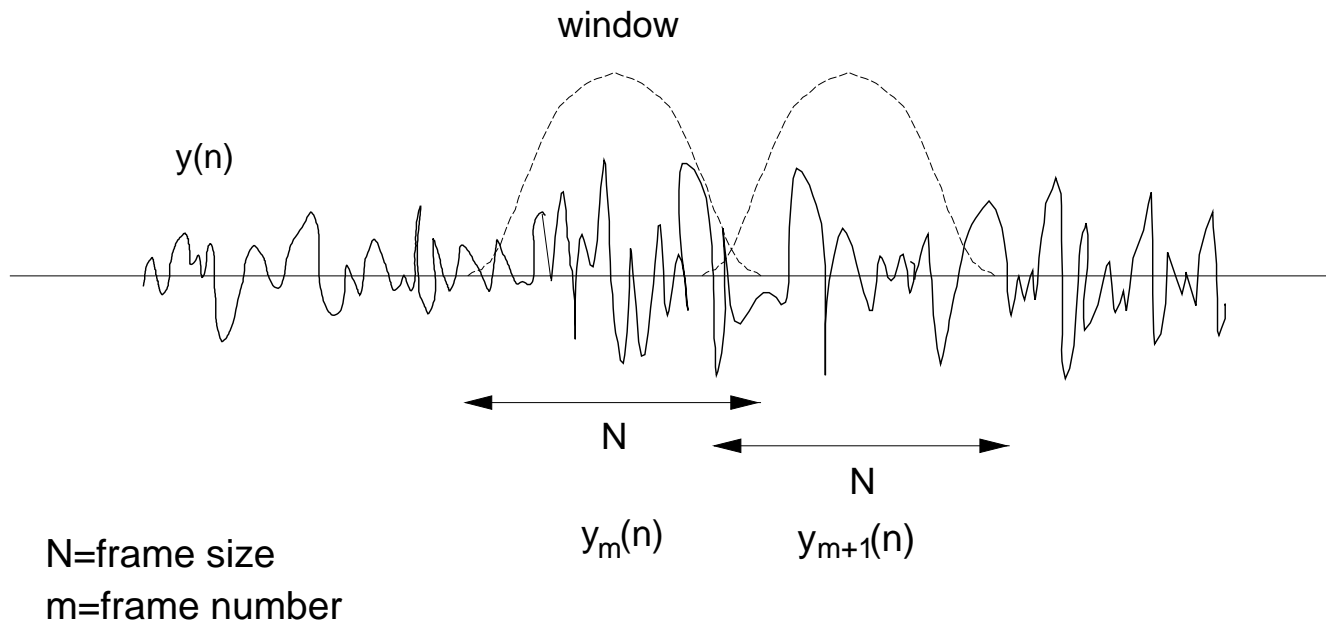
- improving speech quality or intelligibility in multimedia and wireless communications
- communications between pilot and air traffic control tower
- speech recognition

Speech Enhancement Depends on:

- good signal processing technique
- human perceptual factor,
speech quality and intelligibility are dependent on short term spectral amplitude and insensitive to spectral phase

Processing of Speech Signal:

- speech is stationary over a short period of time (10ms to 20ms)
- frame by frame processing



- $y_m(n) = s_m(n) + d_m(n)$, $0 \leq n \leq N - 1$

Methods:

- Spectral Subtraction
- Wiener Filtering
- Iterative Wiener Filtering
- Improved Iterative Wiener Filtering
- Constrained Iterative Wiener Filtering

Spectral Subtraction

- Subtracting noise power spectrum from noisy signal power spectrum
- Assumption: noise power spectral density (PSD) $P_d(\omega)$ is known ($P_d(\omega) = E[|D(\omega)|^2]$)
- Concept:

$$y_m(n) = s_m(n) + d_m(n)$$

$$Y_m(\omega) = S_m(\omega) + D_m(\omega)$$

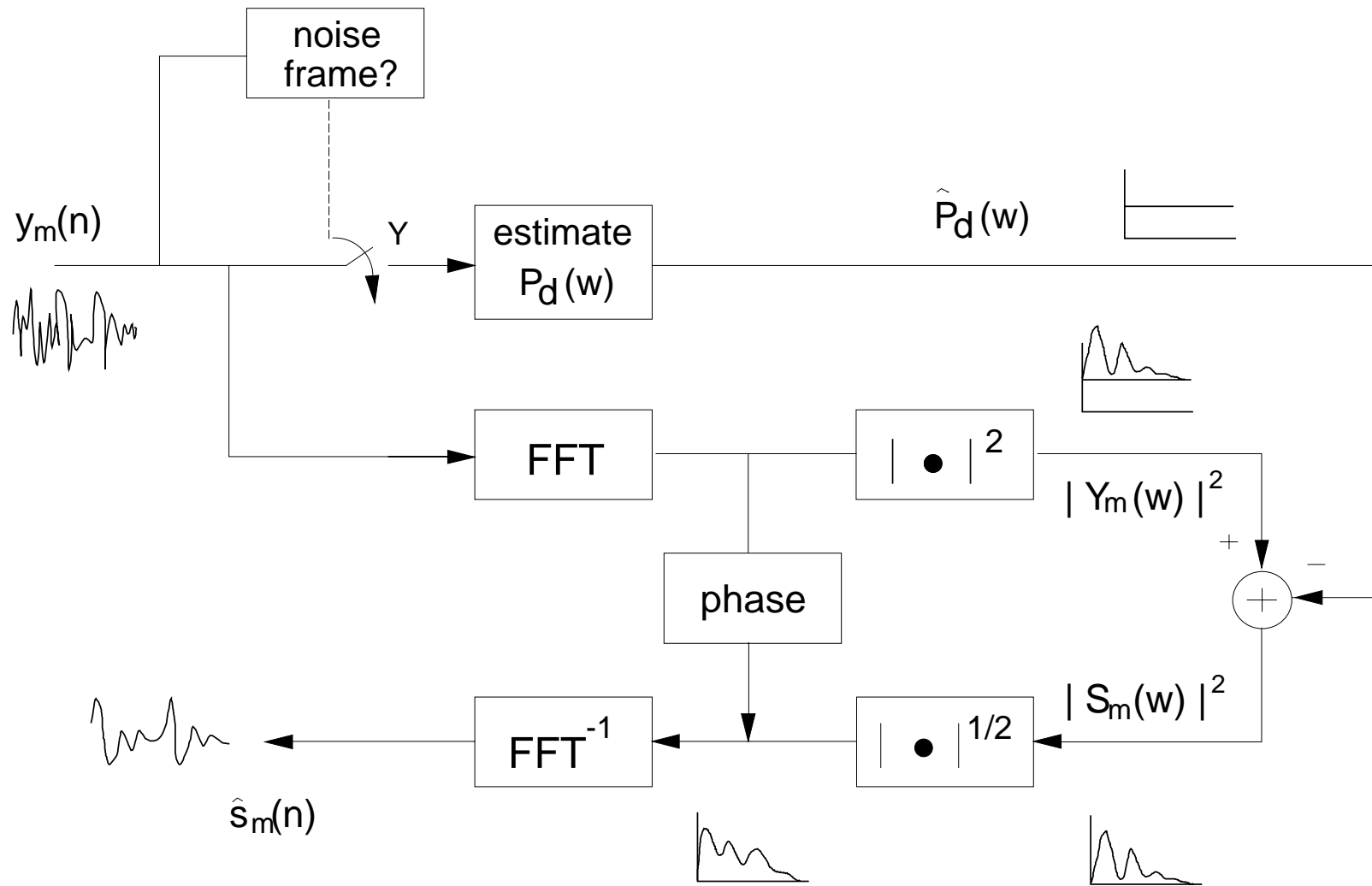
$$|Y_m(\omega)|^2 = |S_m(\omega)|^2 + |D_m(\omega)|^2 + S_m(\omega)D_m(\omega)^* + S_m(\omega)^*D_m(\omega)$$

$$\approx |S_m(\omega)|^2 + |D_m(\omega)|^2$$

$$|\hat{S}_m(\omega)|^2 \approx |Y_m(\omega)|^2 - P_d(\omega)$$

$$\hat{S}_m(\omega) = |\hat{S}_m(\omega)| \angle Y_m(\omega) \Rightarrow \hat{s}_m(n) = F^{-1}\{\hat{S}_m(\omega)\}$$

Block Diagram:

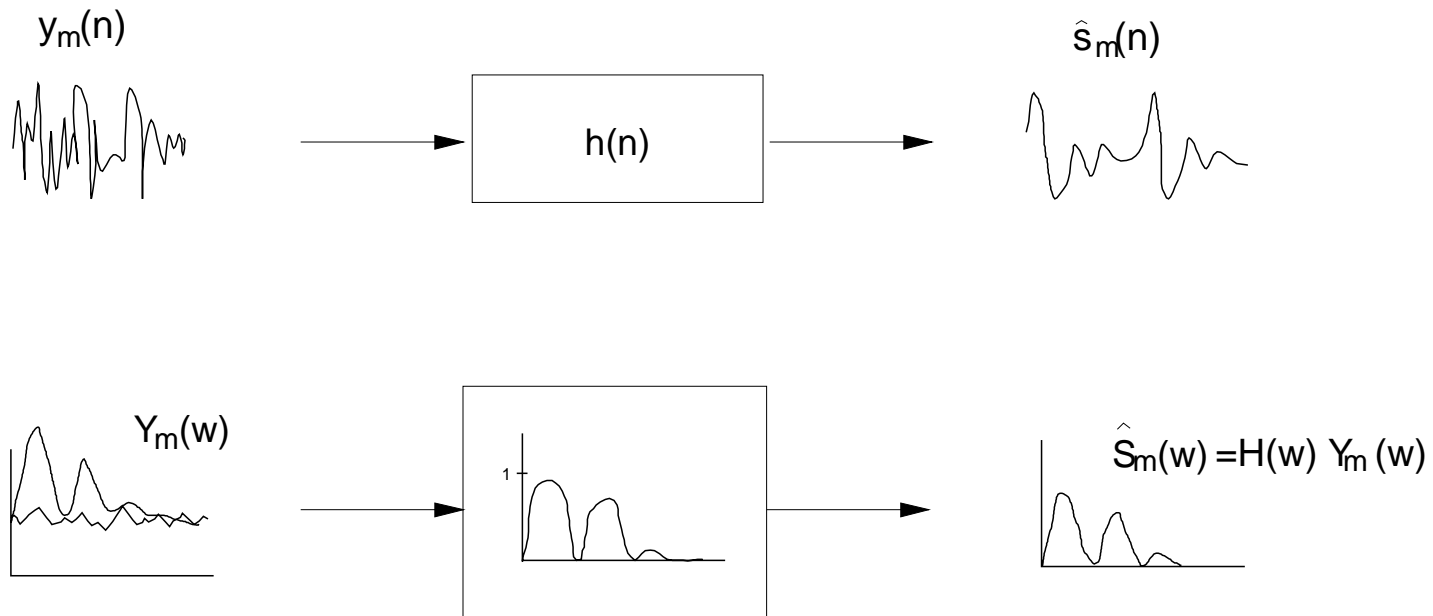


Demo:

- Clean Speech
- Speech + Noise
- Processed by Spectral Subtraction

Wiener Filtering

- Concept:



$$H(\omega) = \frac{P_s(\omega)}{P_y(\omega)}$$

- $H(\omega)$ weights spectrum according to SNR at different frequencies

- $H(\omega)$ weights spectrum according to SNR at different frequencies
- $$H(\omega) = \frac{P_s(\omega)}{P_y(\omega)} = \frac{P_s(\omega)}{P_s(\omega) + P_d(\omega)} \approx \begin{cases} 1, & P_s(\omega) \gg P_d(\omega) \\ 0, & P_s(\omega) \ll P_d(\omega) \end{cases}$$
- Wiener filter minimizes $E[\{s_m(n) - \hat{s}_m(n)\}^2]$, optimum in the mean-square error sense

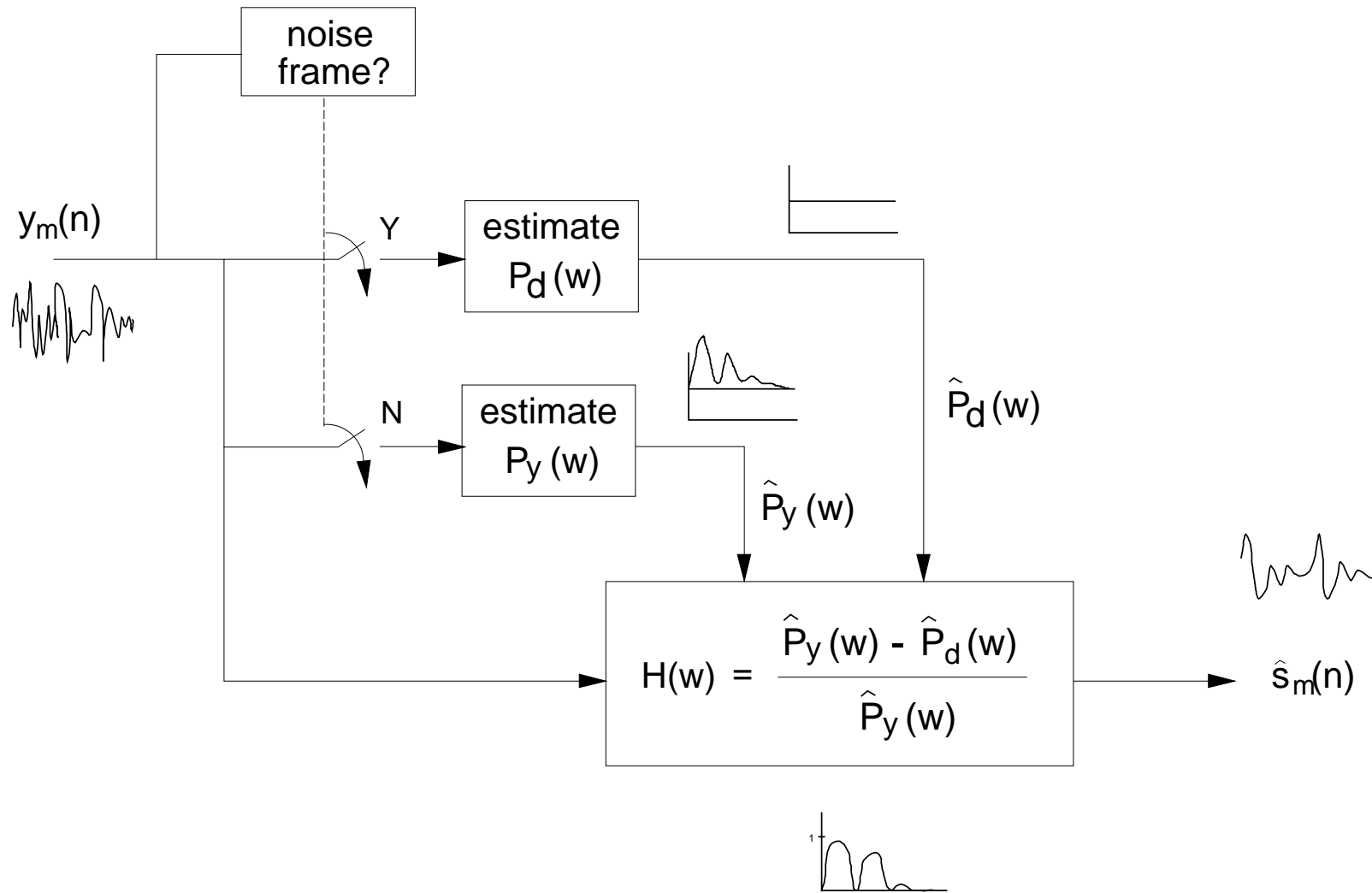
Approximation:

- $\hat{P}_y(\omega) = \sum |Y_m(\omega)|^2$ over noisy speech segments
- $\hat{P}_d(\omega) = \sum |Y_m(\omega)|^2$ over noise segments
- $\hat{P}_s(\omega) = \hat{P}_y(\omega) - \hat{P}_d(\omega)$

Demo:

- Clean Speech
- Speech + Noise
- Processed by Ideal Wiener Filtering

Block Diagram:



Demo:

- Clean Speech
- Speech + Noise
- Processed by Wiener Filtering

Iterative Wiener Filtering

- Estimating $\hat{P}_s(\omega)$ by $\hat{P}_y(\omega) - \hat{P}_d(\omega)$ may not be good
- Can do better by computing $\hat{P}_s(\omega)$ from the Wiener filter output
- Algorithm:

$$\hat{P}_s(\omega)_0 = \hat{P}_y(\omega) - \hat{P}_d(\omega)$$

$$i = 0$$

repeat

$$H(\omega)_i = \frac{\hat{P}_s(\omega)_i}{\hat{P}_s(\omega)_i + \hat{P}_d(\omega)}$$

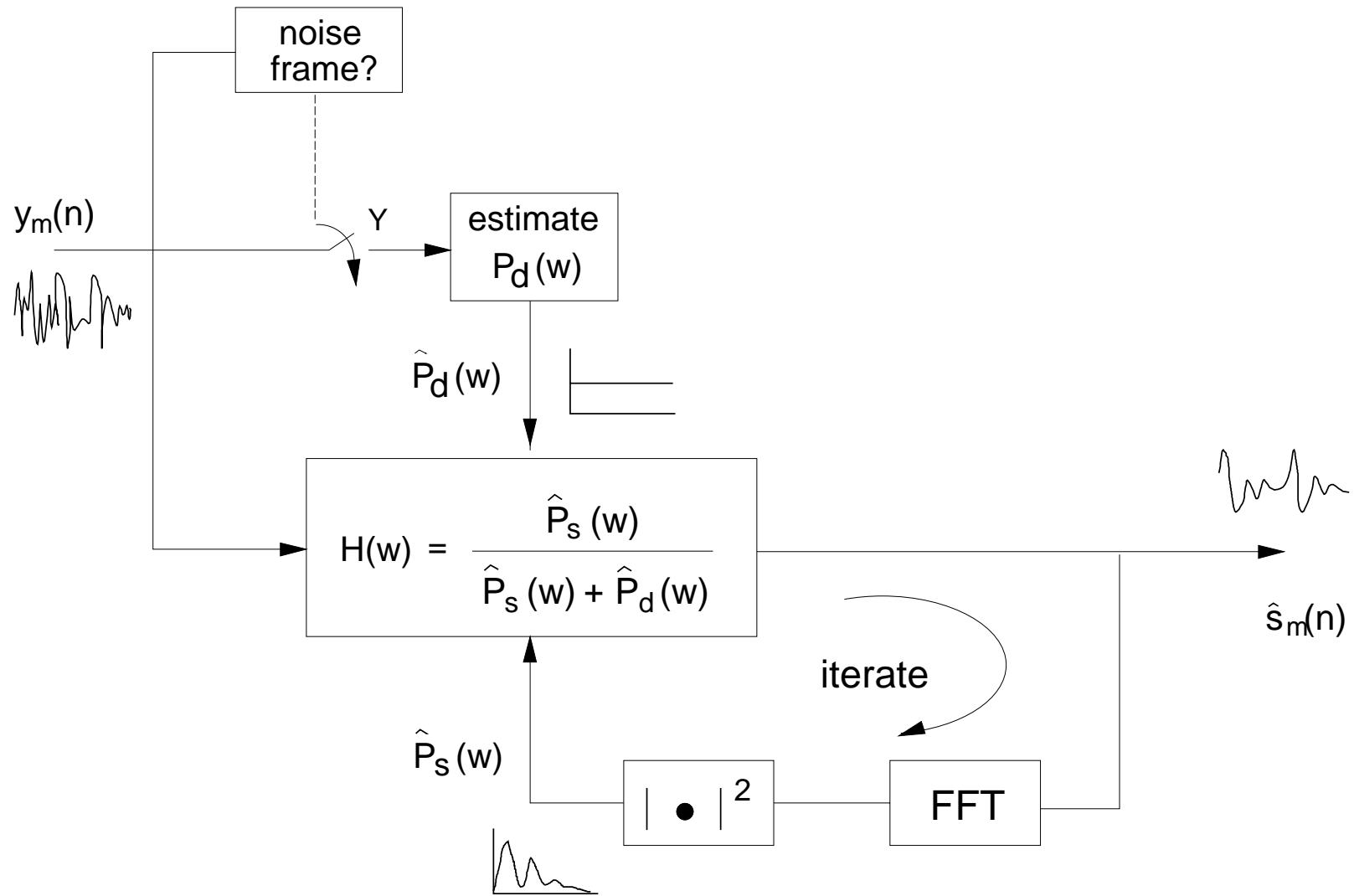
$$S_m(\omega)_{i+1} = H(\omega)_i Y_m(\omega)$$

$$\hat{P}_s(\omega)_{i+1} = |S_m(\omega)_{i+1}|^2$$

$$i=i+1$$

until convergence

Block Diagram:



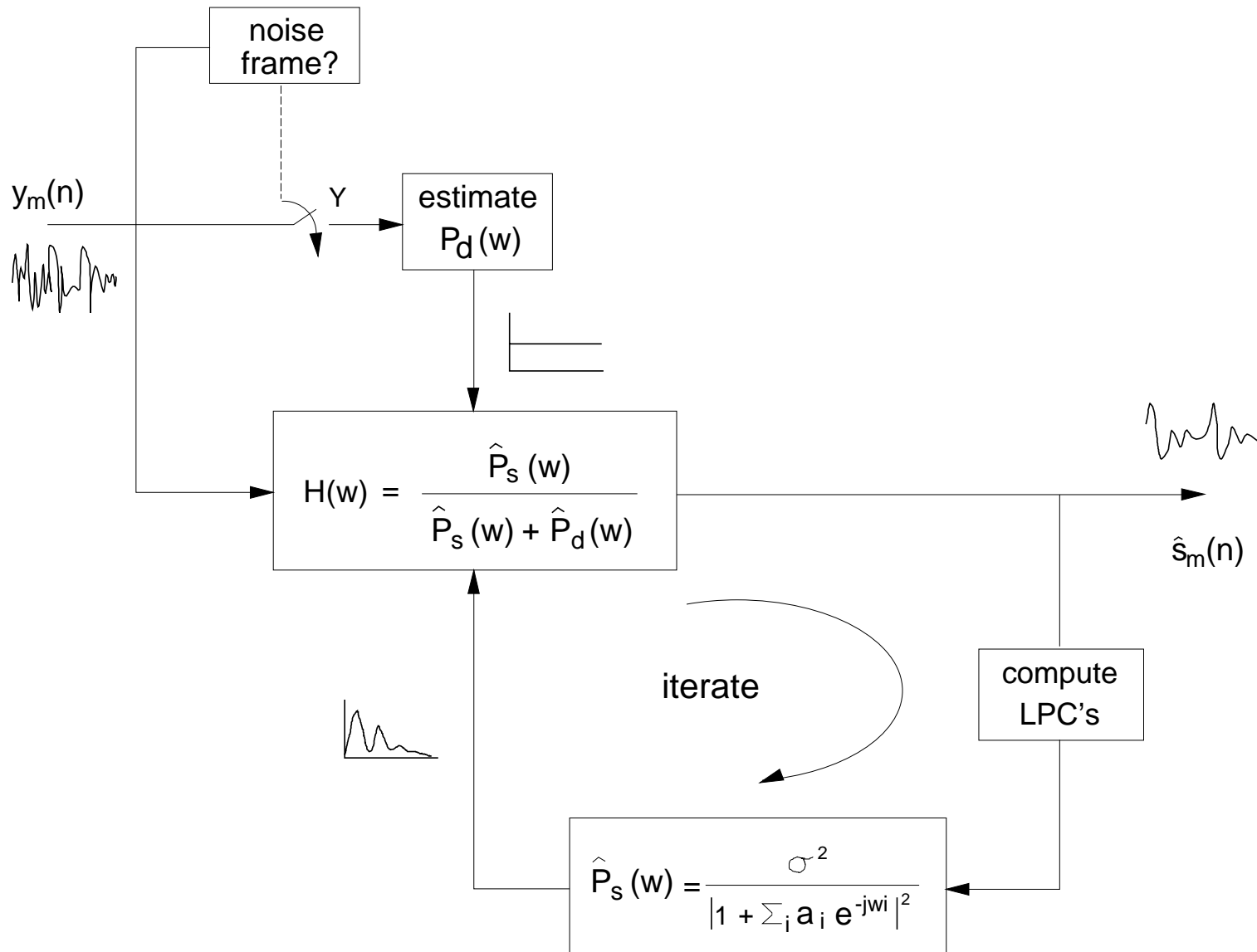
Improved Iterative Wiener Filtering

- Improve estimation of $\hat{P}_s(\omega)$ using constraint from speech production model
- Linear prediction model of speech

$$s(n) = -a_1s(n-1) - a_2s(n-2) - \dots - a_p s(n-P) + \varepsilon(n)$$

$$\Rightarrow P_s(\omega) = \frac{\sigma_\varepsilon^2}{|1 + \sum_{i=1}^P a_i e^{-j\omega i}|^2}$$

Block Diagram:



Demo:

- Clean Speech
- Speech + Noise
- Processed by Improved Iterative Wiener Filtering

Constrained Iterative Wiener Filtering

- apply a priori speech characteristics to impose interframe and interframe constraints on the speech spectrum

Summary

- Speech enhancement is important in human to human or human to machine communications
- Two classes of speech enhancement methods: spectral subtraction and Wiener filtering
- Wiener filtering is an optimum filter in the mean-square error sense
- Wiener filtering, assuming known signal and noise spectra, gives an upper bound in performance
- Imposing constraints from speech production model and speech characteristics produce better signal spectrum estimation and hence improve performance