A SINGLE-CHANNEL NOISE REDUCTION FILTERING/SMOOTHING TECHNIQUE IN THE TIME DOMAIN

Ningning Pan[1], Jacob Benesty[2], and Jingdong Chen[1]

[1] Center of Intelligent Acoustics and Immersive Communications, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China
{nnpan@mail.nwpu.edu.cn, jingdongchen@ieee.org}

[2] INRS-EMT, University of Quebec, 800 de la Gauchetiere Ouest, Suite 6900, Montreal, QC H5A 1K6, Canada
benesty@emt.inrs.ca

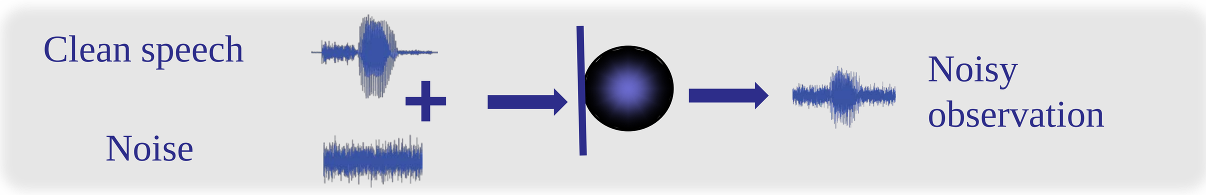
INTRODUCTION

Noise reduction: recover a speech of interest from noisy observation.

The proposed method:

- It first applies a time smoothing window to the noisy signal.
- Noise reduction filters are applied to the smoothed noisy signal.

SIGNAL MODEL AND PROBLEM FORMULATION



Microphone signal in the time domain:

$$y(k) = x(k) + v(k)$$

$x(k)$: desired signal; $v(k)$: additive noise

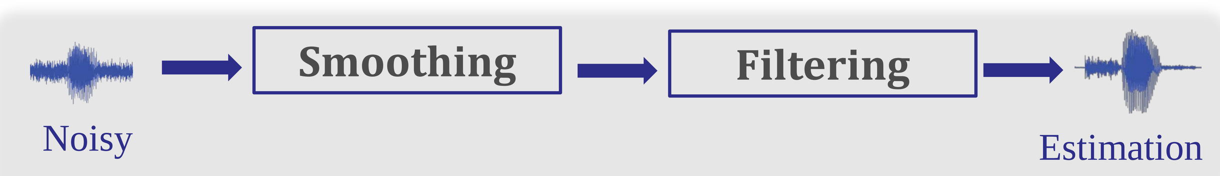
Considering past and future samples:

$$\mathbf{Y}(k) = \begin{bmatrix} y(k) & y(k+1) & \cdots & y(k+N-1) \\ y(k-1) & y(k) & \cdots & y(k+N-2) \\ \vdots & \vdots & \ddots & \vdots \\ y(k-L+1) & y(k-L+2) & \cdots & y(k+N-L) \end{bmatrix}$$

$$\mathbf{Y}(k) = \mathbf{X}(k) + \mathbf{V}(k)$$

Problem: estimate $x(k)$ from $\mathbf{Y}(k)$

LINEAR FILTERING/SMOOTHING FOR NOISE REDUCTION



Estimator:

$$z(k) = \mathbf{h}^T\mathbf{Y}(k)\mathbf{w}$$
$$= \mathbf{h}^T\mathbf{X}(k)\mathbf{w} + \mathbf{h}^T\mathbf{V}(k)\mathbf{w}$$
$$= x_{\text{fd}}(k) + v_{\text{rn}}(k)$$

$\mathbf{w}$: real-valued smoothing vector with length N

$\mathbf{h}$: noise reduction filter with length L

The variance:

$$\sigma_z^2 = E\left[z^2(k)\right] = \mathbf{h}^T\mathbf{R}_{\mathbf{Yw}}\mathbf{h} = \sigma_{x_{\text{fd}}}^2 + \sigma_{v_{\text{rn}}}^2$$

$\mathbf{R}_{\mathbf{Yw}}$: correlation matrix of $\mathbf{Y}(k)\mathbf{w}$, $\mathbf{R}_{\mathbf{Yw}} = \mathbf{R}_{\mathbf{Xw}} + \mathbf{R}_{\mathbf{Vw}}$

$\sigma_{x_{\text{fd}}}^2 + \sigma_{v_{\text{rn}}}^2$: variance of $x_{\text{fd}}(k)$ and $v_{\text{rn}}(k)$

PERFORMANCE MEASURES

input SNR

$$\text{iSNR} = \frac{\sigma_x^2}{\sigma_v^2}$$

output SNR

$$\text{oSNR}(\mathbf{h}) = \frac{\sigma_{x_{\text{fd}}}^2}{\sigma_{v_{\text{rn}}}^2} = \frac{\mathbf{h}^T\mathbf{R}_{\mathbf{Xw}}\mathbf{h}}{\mathbf{h}^T\mathbf{R}_{\mathbf{Vw}}\mathbf{h}}.$$

$\mathbf{h}$ should be found in such a way that $\text{oSNR}(\mathbf{h}) > \text{iSNR}$.

Signal distortion index

$$\upsilon_{\text{d}}(\mathbf{h}) = \frac{E\left\{\left[x(k) - \mathbf{h}^T\mathbf{X}(k)\mathbf{w}\right]^2\right\}}{\sigma_x^2}$$

MSE CRITERION

Error signal:

$$e(k) = z(k) - x(k) = \mathbf{h}^T\mathbf{Y}(k)\mathbf{w} - x(k) = e_{\text{d}}(k) + e_{\text{n}}(k)$$

$$e_{\text{d}}(k) = \mathbf{h}^T\mathbf{X}(k)\mathbf{w} - x(k)$$

$$e_{\text{n}}(k) = \mathbf{h}^T\mathbf{V}(k)\mathbf{w}$$

MSE criterion:

$$J(\mathbf{h}) = E\left[e^2(k)\right] = J_{\text{d}}(\mathbf{h}) + J_{\text{n}}(\mathbf{h})$$

$$J_{\text{d}}(\mathbf{h}) = E\left[e_{\text{d}}^2(k)\right]$$

$$J_{\text{n}}(\mathbf{h}) = E\left[e_{\text{n}}^2(k)\right]$$

OPTIMAL FILTERS

Maximum SNR filter

$$\mathbf{h}_{\max} = \varsigma\mathbf{t}_{\max}$$

$\mathbf{t}_{\max}$: Eigenvector corresponding to the maximum eigenvalue of the matrix product $\mathbf{R}_{\mathbf{Vw}}^{-1}\mathbf{R}_{\mathbf{Xw}}$

Substituting $\mathbf{h}_{\max}$ into the distortion-based MSE, we get

$$J_{\text{d}}(\mathbf{h}_{\max}) = \sigma_x^2 - 2\varsigma\mathbf{t}_{\max}^T\mathbf{r}_{\mathbf{Yw}x} + \varsigma^2\mathbf{t}_{\max}^T\mathbf{R}_{\mathbf{Xw}}\mathbf{t}_{\max}$$

The optimal value of ς:

$$\varsigma = \frac{\mathbf{t}_{\max}^T\mathbf{r}_{\mathbf{Yw}x}}{\mathbf{t}_{\max}^T\mathbf{R}_{\mathbf{Xw}}\mathbf{t}_{\max}}$$

$$\mathbf{h}_{\max} = \frac{\mathbf{t}_{\max}\mathbf{t}_{\max}^T\mathbf{r}_{\mathbf{Yw}x}}{\mathbf{t}_{\max}^T\mathbf{R}_{\mathbf{Xw}}\mathbf{t}_{\max}}$$

Minimum distortion filter

Minimize $J_{\text{d}}(\mathbf{h})$ → $\mathbf{h}_{\text{MD}} = \mathbf{R}_{\mathbf{Xw}}^{-1}\mathbf{r}_{\mathbf{Yw}x}$

Wiener filter

Minimize $J(\mathbf{h})$ → $\mathbf{h}_{\text{W}} = \mathbf{R}_{\mathbf{Yw}}^{-1}\mathbf{r}_{\mathbf{Yw}x}$

Tradeoff filter

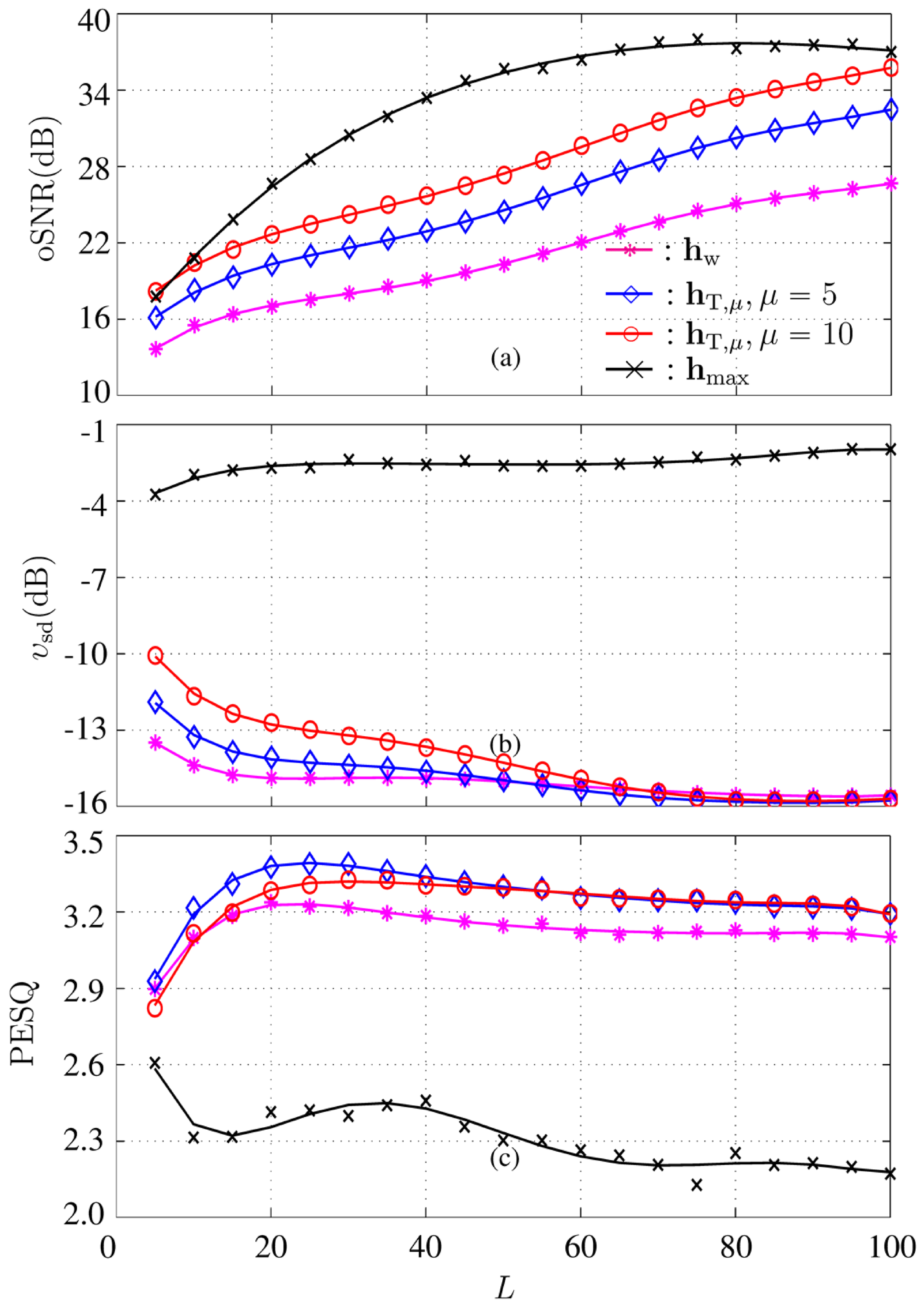
$$\min_{\mathbf{h}} J_{\text{d}}(\mathbf{h}) \quad \text{subject to} \quad J_{\text{n}}(\mathbf{h}) = \aleph\sigma_v^2$$

$$\mathbf{h}_{\text{T},\mu} = \left(\mathbf{R}_{\mathbf{Xw}} + \mu\mathbf{R}_{\mathbf{Vw}}\right)^{-1}\mathbf{r}_{\mathbf{Yw}x}$$

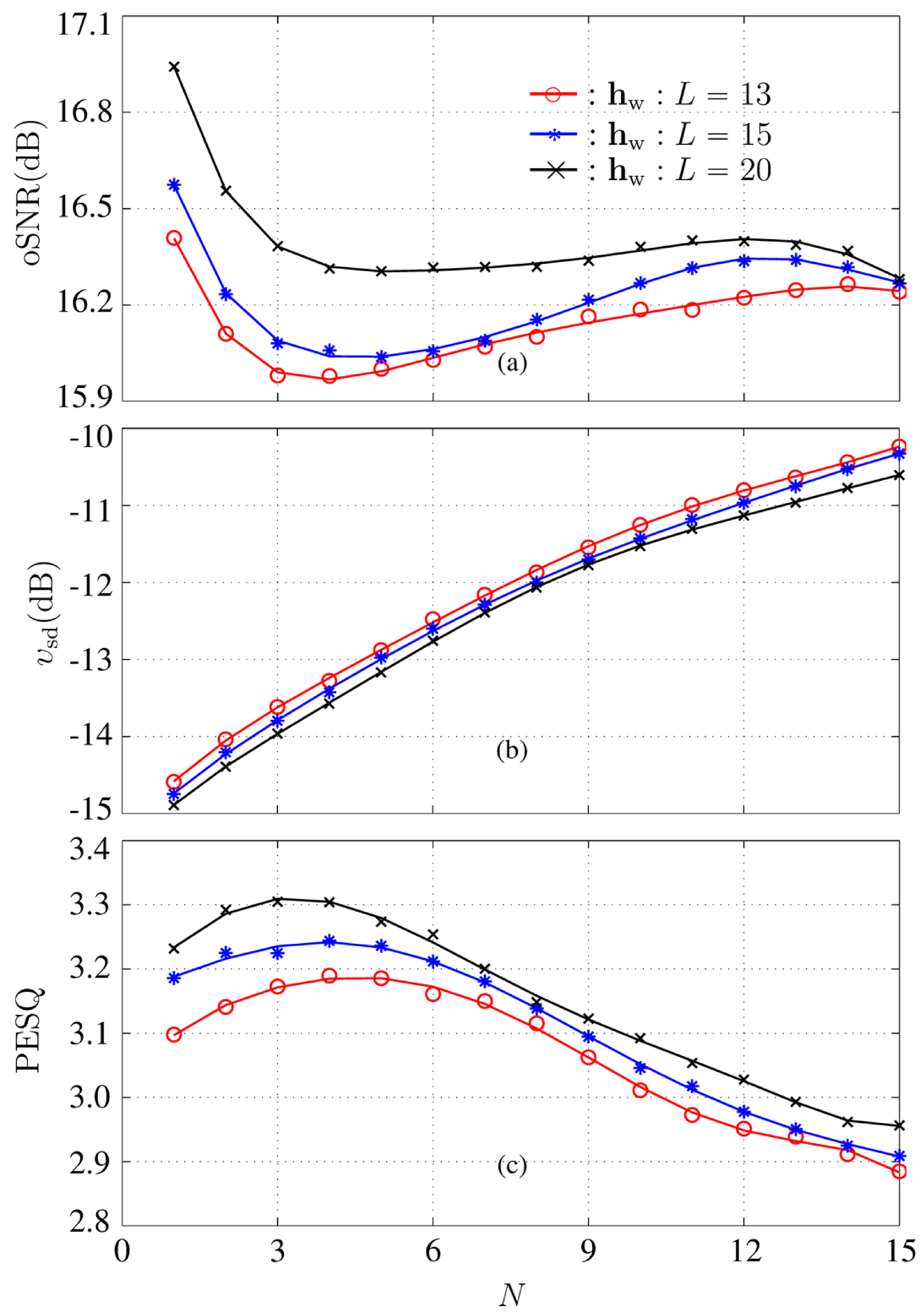
SIMULATIONS

Implementation:

clean speech	20 sentences from TIMIT database
noise	recorded in a Sedan car
smoothing window	Hann window
Sampling rate	8 kHz
iSNR	10 dB



The performance of the Wiener, tradeoff, and maximum SNR filters (without the time smoothing technique, i.e., $N = 1$) as a function of the filter length, L. The PESQ score of the noisy signal is 2.375.



Performance of the Wiener filter as a function of time smoothing length, N: (a) output SNR, (b) speech distortion index, and (c) PESQ. Simulation conditions: iSNR = 10 dB, $L = 13, 15, 20$.

CONCLUSIONS

- A linear estimation approach to single-channel noise reduction in the time domain with the ability to smooth and filter the observation signal at the same time is presented.
- Three different filters, maximum SNR, Wiener, and tradeoff filters, are derived.
- Results show advantages of smoothing.