A Single-Channel Noise Reduction Filtering/Smoothing Technique in the Time Domain

Conference Paper · April 2018

DOI: 10.1109/ICASSP.2018.8462371

3 authors:

Ningning Pan
Northwestern Polytechnical University
2 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Jacob Benesty
Institut National de la Recherche Scientifique
613 PUBLICATIONS **12,800** CITATIONS
SEE PROFILE

Jingdong Chen
Institute of Electrical and Electronics Engineers
307 PUBLICATIONS **4,721** CITATIONS
SEE PROFILE

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
[2]: INRS-EMT, University of Quebec, Montreal, QC H5A 1K6, Canada

ABSTRACT

In this paper, we present a single-channel smoothing-and-filtering technique for noise reduction in the time domain. Unlike traditional noise reduction methods, which directly apply a noise reduction filter to the noisy signal, the developed technique achieves noise reduction in two steps. It first applies a time smoothing window to the noisy signal, which, on the one hand, can help reduce high frequency noise and, on the other hand, can help leverage the correlation between successive signal samples. A noise reduction filter is then applied to the smoothed noisy signal to estimate the speech signal of interest. Three optimal and suboptimal noise reduction filters are derived, including the Wiener, maximum signal-to-noise-ratio (SNR), and tradeoff filters. Simulation results reveal that the developed method can produce better noise reduction performance, i.e., higher gains in the perceptual-evaluation-of-speech-quality (PESQ) score, than the traditional methods without smoothing.

Index Terms— Noise reduction, speech enhancement, single channel, time-domain smoothing, optimal filtering.

1. SIGNAL MODEL AND PROBLEM FORMULATION

In the noise reduction problem considered in this paper, the noisy observation or microphone signal is given by [1], [2]

$$y(k) = x(k) + v(k), \tag{1}$$

where k is the discrete-time index, $x(k)$ is the clean speech signal (also called the desired signal), and $v(k)$ is the unwanted additive noise, which is assumed to be uncorrelated with $x(k)$. All signals are considered to be zero mean, real, and broadband.

By considering past and future time samples of the observations, we can define an observation matrix of size $L \times N$:

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

This matrix will be used in the rest of this paper. We define the matrices $\mathbf{X}(k)$ and $\mathbf{V}(k)$ in a similar way but with the clean and noise signals, respectively.

Then, the objective of single-channel noise reduction in the time domain is the estimation of the desired signal, $x(k)$, from some of the data contained in $\mathbf{Y}(k)$, in the best possible way. In the following, we show how to combine smoothing and filtering to achieve this goal.

2. LINEAR FILTERING/SMOOTHING FOR NOISE REDUCTION

In this section, we explain the linear estimation technique for the derivation of single-channel noise reduction filters in the time domain that have the ability to smooth the observed signals at the same time.

Multiplying the matrix $\mathbf{Y}(k)$ by a real-valued window, $\mathbf{w}$, of length N, we observe that each component of the vector $\mathbf{Y}(k)\mathbf{w}$ is smoothed along the time axis with future time samples. Then, the estimator that we propose is

$$\begin{aligned} z(k) &= \mathbf{h}^T\mathbf{Y}(k)\mathbf{w} \\ &= \mathbf{h}^T\mathbf{X}(k)\mathbf{w} + \mathbf{h}^T\mathbf{V}(k)\mathbf{w} \\ &= x_{\mathrm{fd}}(k) + v_{\mathrm{rn}}(k), \end{aligned} \tag{3}$$

where $z(k)$ is the estimate of $x(k)$, $\mathbf{h}$ is a real-valued linear filter of length L, the superscript ^{T} is the transpose operator, $x_{\mathrm{fd}}(k) = \mathbf{h}^T\mathbf{X}(k)\mathbf{w}$ is the filtered desired signal, and $v_{\mathrm{rn}}(k) = \mathbf{h}^T\mathbf{V}(k)\mathbf{w}$ is the residual noise. With this approach, the filtering is performed with past time samples while smoothing is performed with future time samples.

We deduce that the variance of $z(k)$ is

$$\begin{aligned} \sigma_z^2 &= E\left[z^2(k)\right] \\ &= \mathbf{h}^T\mathbf{R}_{\mathbf{Y}\mathbf{w}}\mathbf{h} \\ &= \sigma_{x_{\mathrm{fd}}}^2 + \sigma_{v_{\mathrm{rn}}}^2, \end{aligned} \tag{4}$$

where $E[\cdot]$ denotes mathematical expectation, $\mathbf{R}_{\mathbf{Y}\mathbf{w}} = E\left[\mathbf{Y}(k)\mathbf{w}\mathbf{w}^T\mathbf{Y}^T(k)\right]$ is the correlation matrix of $\mathbf{Y}(k)\mathbf{w}$, $\sigma_{x_{\mathrm{fd}}}^2 = \mathbf{h}^T\mathbf{R}_{\mathbf{X}\mathbf{w}}\mathbf{h}$ is the variance of the filtered desired signal, with $\mathbf{R}_{\mathbf{X}\mathbf{w}}$ being the correlation matrix of $\mathbf{X}(k)\mathbf{w}$, and $\sigma_{v_{\mathrm{rn}}}^2 = \mathbf{h}^T\mathbf{R}_{\mathbf{V}\mathbf{w}}\mathbf{h}$ is the variance of the residual noise, with $\mathbf{R}_{\mathbf{V}\mathbf{w}}$ being the correlation matrix of $\mathbf{V}(k)\mathbf{w}$.

3. PERFORMANCE MEASURES

According to the signal model given in (1), we define the input SNR as

$$\mathrm{iSNR} = \frac{\sigma_x^2}{\sigma_v^2}, \tag{5}$$

where $\sigma_x^2 = E\left[x^2(k)\right]$ and $\sigma_v^2 = E\left[v^2(k)\right]$ are the variances of $x(k)$ and $v(k)$, respectively.

The output SNR quantifies the SNR after the filtering/smoothing process. It is defined as

$$\mathrm{oSNR}\left(\mathbf{h}\right) = \frac{\sigma_{x_{\mathrm{fd}}}^2}{\sigma_{v_{\mathrm{rn}}}^2} = \frac{\mathbf{h}^T\mathbf{R}_{\mathbf{X}\mathbf{w}}\mathbf{h}}{\mathbf{h}^T\mathbf{R}_{\mathbf{V}\mathbf{w}}\mathbf{h}}. \tag{6}$$

This work was supported in part by the NSFC "Distinguished Young Scientists Fund" under grant No. 61425005 and the NSFC-ISF joint research program under Grant No. 61761146001.

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

To quantify the amount of noise being rejected by the filter, we define the noise reduction factor as [3, 4]

$$\xi_{\text{n}}(\mathbf{h}) = \frac{\sigma_v^2}{\mathbf{h}^T \mathbf{R}_{\mathbf{V}\mathbf{w}} \mathbf{h}}. \tag{7}$$

For optimal filters, we should have $\xi_{\text{n}}(\mathbf{h}) \geq 1$.

In practice, the filter adds distortion to the desired signal. In order to evaluate the level of this distortion, we define the desired signal reduction factor as [3, 4]

$$\xi_{\text{d}}(\mathbf{h}) = \frac{\sigma_x^2}{\mathbf{h}^T \mathbf{R}_{\mathbf{X}\mathbf{w}} \mathbf{h}}. \tag{8}$$

For optimal filters, we should have $\xi_{\text{d}}(\mathbf{h}) \geq 1$. The larger the value of $\xi_{\text{d}}(\mathbf{h})$, the more the desired signal is distorted.

By making the appropriate substitutions, one can derive the relationship:

$$\frac{\text{oSNR}(\mathbf{h})}{\text{iSNR}} = \frac{\xi_{\text{n}}(\mathbf{h})}{\xi_{\text{d}}(\mathbf{h})}. \tag{9}$$

This expression indicates the equivalence between gain/loss in SNR and distortion.

Another way to measure the distortion of the desired signal due to the filter is the desired signal distortion index, which is defined as the mean-squared error (MSE) between the desired signal and the filtered desired signal, normalized by the variance of the desired signal, i.e.,

$$\upsilon_{\text{d}}(\mathbf{h}) = \frac{E\left\{\left[x(k) - \mathbf{h}^T \mathbf{X}(k)\mathbf{w}\right]^2\right\}}{\sigma_x^2}. \tag{10}$$

The desired signal distortion index is usually upper bounded by 1 for optimal filters.

4. MSE CRITERION

In the time domain, the error signal between the estimated and desired signals is

$$e(k) = z(k) - x(k) = \mathbf{h}^T \mathbf{Y}(k)\mathbf{w} - x(k), \tag{11}$$

which can also be written as the sum of two uncorrelated error signals:

$$e(k) = e_{\text{d}}(k) + e_{\text{n}}(k), \tag{12}$$

where

$$e_{\text{d}}(k) = \mathbf{h}^T \mathbf{X}(k)\mathbf{w} - x(k) \tag{13}$$

is the distortion of the desired signal due to the filter and

$$e_{\text{n}}(k) = \mathbf{h}^T \mathbf{V}(k)\mathbf{w} \tag{14}$$

represents the residual noise. The MSE criterion is then

$$\begin{aligned} J(\mathbf{h}) &= E\left[e^2(k)\right] \\ &= \sigma_x^2 - 2\mathbf{h}^T \mathbf{r}_{\mathbf{Y}\mathbf{w}x} + \mathbf{h}^T \mathbf{R}_{\mathbf{Y}\mathbf{w}} \mathbf{h} \\ &= J_{\text{d}}(\mathbf{h}) + J_{\text{n}}(\mathbf{h}), \end{aligned} \tag{15}$$

where $\mathbf{r}_{\mathbf{Y}\mathbf{w}x} = E[\mathbf{Y}(k)\mathbf{w}x(k)]$,

$$J_{\text{d}}(\mathbf{h}) = E\left[e_{\text{d}}^2(k)\right] = \upsilon_{\text{d}}(\mathbf{h})\sigma_x^2, \tag{16}$$

and

$$J_{\text{n}}(\mathbf{h}) = E\left[e_{\text{n}}^2(k)\right] = \frac{\sigma_v^2}{\xi_{\text{n}}(\mathbf{h})}. \tag{17}$$

We deduce that

$$\begin{aligned} \frac{J_{\text{d}}(\mathbf{h})}{J_{\text{n}}(\mathbf{h})} &= \text{iSNR} \times \xi_{\text{n}}(\mathbf{h}) \times \upsilon_{\text{d}}(\mathbf{h}) \\ &= \text{oSNR}(\mathbf{h}) \times \xi_{\text{d}}(\mathbf{h}) \times \upsilon_{\text{d}}(\mathbf{h}). \end{aligned} \tag{18}$$

This shows how the different performance measures are related to the MSEs.

5. OPTIMAL FILTERS

In this section, we derive a class of single-channel noise reduction filters from the maximization of the output SNR and the minimization of the MSEs.

The maximum SNR filter, $\mathbf{h}_{\max}$, is obtained by maximizing the output SNR as defined in (6), from which we recognize the generalized Rayleigh quotient [5]. It is well known that this quotient is maximized with the eigenvector corresponding to the maximum eigenvalue of the matrix product $\mathbf{R}_{\mathbf{V}\mathbf{w}}^{-1}\mathbf{R}_{\mathbf{X}\mathbf{w}}$. Let us denote $\lambda_{\max}$ this maximum eigenvalue and $\mathbf{t}_{\max}$ the corresponding eigenvector. Then, it is clear that the maximum SNR filter is

$$\mathbf{h}_{\max} = \varsigma \mathbf{t}_{\max}, \tag{19}$$

where $\varsigma \neq 0$ is an arbitrary real-valued number. We also have

$$\text{oSNR}(\mathbf{h}_{\max}) = \lambda_{\max} \geq \text{iSNR} \tag{20}$$

and

$$\text{oSNR}(\mathbf{h}_{\max}) \geq \text{oSNR}(\mathbf{h}), \ \forall \mathbf{h}. \tag{21}$$

One of the best ways to find the parameter ς is by minimizing distortion. Substituting $\mathbf{h}_{\max}$ into the distortion-based MSE, we get

$$J_{\text{d}}(\mathbf{h}_{\max}) = \sigma_x^2 - 2\varsigma \mathbf{t}_{\max}^T \mathbf{r}_{\mathbf{Y}\mathbf{w}x} + \varsigma^2 \mathbf{t}_{\max}^T \mathbf{R}_{\mathbf{X}\mathbf{w}} \mathbf{t}_{\max}, \tag{22}$$

from which we find the optimal value of ς:

$$\varsigma = \frac{\mathbf{t}_{\max}^T \mathbf{r}_{\mathbf{Y}\mathbf{w}x}}{\mathbf{t}_{\max}^T \mathbf{R}_{\mathbf{X}\mathbf{w}} \mathbf{t}_{\max}}. \tag{23}$$

As a result, the maximum SNR filter with minimum distortion is

$$\mathbf{h}_{\max} = \frac{\mathbf{t}_{\max}\mathbf{t}_{\max}^T \mathbf{r}_{\mathbf{Y}\mathbf{w}x}}{\mathbf{t}_{\max}^T \mathbf{R}_{\mathbf{X}\mathbf{w}} \mathbf{t}_{\max}}. \tag{24}$$

The minimum distortion (MD) filter is obtained by minimizing $J_{\text{d}}(\mathbf{h})$. We get

$$\mathbf{h}_{\text{MD}} = \mathbf{R}_{\mathbf{X}\mathbf{w}}^{\dagger} \mathbf{r}_{\mathbf{Y}\mathbf{w}x}, \tag{25}$$

where $\mathbf{R}_{\mathbf{X}\mathbf{w}}^{\dagger}$ is the pseudo-inverse of $\mathbf{R}_{\mathbf{X}\mathbf{w}}$. If $\mathbf{R}_{\mathbf{X}\mathbf{w}}$ is of full rank, $\mathbf{R}_{\mathbf{X}\mathbf{w}}^{\dagger} = \mathbf{R}_{\mathbf{X}\mathbf{w}}^{-1}$ and $\mathbf{h}_{\text{MD}}$ becomes

$$\mathbf{h}_{\text{MD}} = \mathbf{R}_{\mathbf{X}\mathbf{w}}^{-1} \mathbf{r}_{\mathbf{Y}\mathbf{w}x}. \tag{26}$$

The Wiener filter is obtained from the optimization of the MSE criterion, $J(\mathbf{h})$. The minimization of $J(\mathbf{h})$ with respect to $\mathbf{h}$ leads to

$$\mathbf{h}_{\text{W}} = \mathbf{R}_{\mathbf{Y}\mathbf{w}}^{-1} \mathbf{r}_{\mathbf{Y}\mathbf{w}x}. \tag{27}$$

We should have

$$\mathrm{oSNR}\left(\mathbf{h}_{\mathrm{W}}\right) \geq \mathrm{oSNR}\left(\mathbf{h}_{\mathrm{MD}}\right) \tag{28}$$

and

$$\xi_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{W}}\right) \geq \xi_{\mathrm{d}}\left(\mathbf{h}_{\mathrm{MD}}\right). \tag{29}$$

Another interesting approach that can compromise between noise reduction and desired signal distortion is the tradeoff filter obtained by

$$\min_{\mathbf{h}} J_{\mathrm{d}}\left(\mathbf{h}\right) \quad \text{subject to} \quad J_{\mathrm{n}}\left(\mathbf{h}\right) = \aleph\sigma_v^2, \tag{30}$$

where $0 \leq \aleph \leq 1$, to ensure that filtering achieves some degree of noise reduction. Assuming that the matrix $\mathbf{R}_{\mathbf{V}\mathbf{w}}$ is of full rank, which is generally true in practice, we find that the optimal filter is

$$\mathbf{h}_{\mathrm{T},\mu} = \left(\mathbf{R}_{\mathbf{X}\mathbf{w}} + \mu\mathbf{R}_{\mathbf{V}\mathbf{w}}\right)^{-1}\mathbf{r}_{\mathbf{Y}\mathbf{w}x}, \tag{31}$$

where $\mu \geq 0$ is a Lagrange multiplier. Clearly, for $\mu = 0$ and $\mu = 1$, we get the MD (if $\mathbf{R}_{\mathbf{X}\mathbf{w}}$ is invertible) and Wiener filters, respectively. For $\mu > 1$ (resp. $\mu < 1$), the tradeoff filter reduces more (resp. less) noise and introduces more (resp. less) distortion to the desired speech signal than the Wiener filter.

6. SIMULATIONS

In this section, simulations are carried out to evaluate the effectiveness of the developed filtering/smoothing technique for noise reduction in the time domain. The output SNR defined in (6), the speech distortion index defined in (10), and the perceptual evaluation of speech quality (PESQ) score [6] are adopted as the performance measures.

The clean speech signals (consisting of 20 sentences with 10 from a male speaker and the other 10 from a female speaker) are taken from the TIMIT database [7]. In our simulations, we consider noise reduction of narrow band signals, so all the signals are downsampled from the original sampling rate of 16 kHz to 8 kHz. To obtain the noisy speech, noise recorded in a Sedan car running at 50 MPH on a highway is added to the clean speech, which is properly scaled to control the input SNR level.

In order to implement the filters derived in the previous section, we need to know the correlation matrices of the noisy and noise signals. In this simulation, we compute these matrices directly from the respective signals using a recursive method [8], i.e.,

$$\widehat{\mathbf{R}}_{\mathbf{Y}\mathbf{w}}(k) = \alpha_y\widehat{\mathbf{R}}_{\mathbf{Y}\mathbf{w}}(k-1) + (1-\alpha_y)\mathbf{Y}(k)\mathbf{w}\mathbf{w}^T\mathbf{Y}^T(k), \tag{32}$$

and $\widehat{\mathbf{R}}_{\mathbf{V}\mathbf{w}}$ is computed the same way. Then, the correlation matrix of the speech signal is computed as $\widehat{\mathbf{R}}_{\mathbf{X}\mathbf{w}} = \widehat{\mathbf{R}}_{\mathbf{Y}\mathbf{w}} - \widehat{\mathbf{R}}_{\mathbf{V}\mathbf{w}}$. For the estimation of $\mathbf{r}_{\mathbf{Y}\mathbf{w}x}$, the two quantities $\mathbf{r}_{\mathbf{Y}\mathbf{w}y}$ and $\mathbf{r}_{\mathbf{V}\mathbf{w}v}$ are estimated first:

$$\widehat{\mathbf{r}}_{\mathbf{Y}\mathbf{w}y}(k) = \alpha_y\widehat{\mathbf{r}}_{\mathbf{Y}\mathbf{w}y}(k-1) + (1-\alpha_y)\mathbf{Y}(k)\mathbf{w}y(k), \tag{33}$$

$$\widehat{\mathbf{r}}_{\mathbf{V}\mathbf{w}v}(k) = \alpha_v\widehat{\mathbf{r}}_{\mathbf{V}\mathbf{w}v}(k-1) + (1-\alpha_v)\mathbf{V}(k)\mathbf{w}v(k), \tag{34}$$

where $\alpha_y \in (0,1)$ and $\alpha_v \in (0,1)$ are two forgetting factors, which control the influence of the previous data samples on the current estimate (the initial estimate is obtained from the first 1600 signal samples with a long-time average). Then, we have $\widehat{\mathbf{r}}_{\mathbf{Y}\mathbf{w}x}(k) = \widehat{\mathbf{r}}_{\mathbf{Y}\mathbf{w}y}(k) - \widehat{\mathbf{r}}_{\mathbf{V}\mathbf{w}v}(k)$. These estimated correlation matrices are then substituted into the deduced filters. In this simulation, the smoothing window $\mathbf{w}$ is set to be the Hann window.

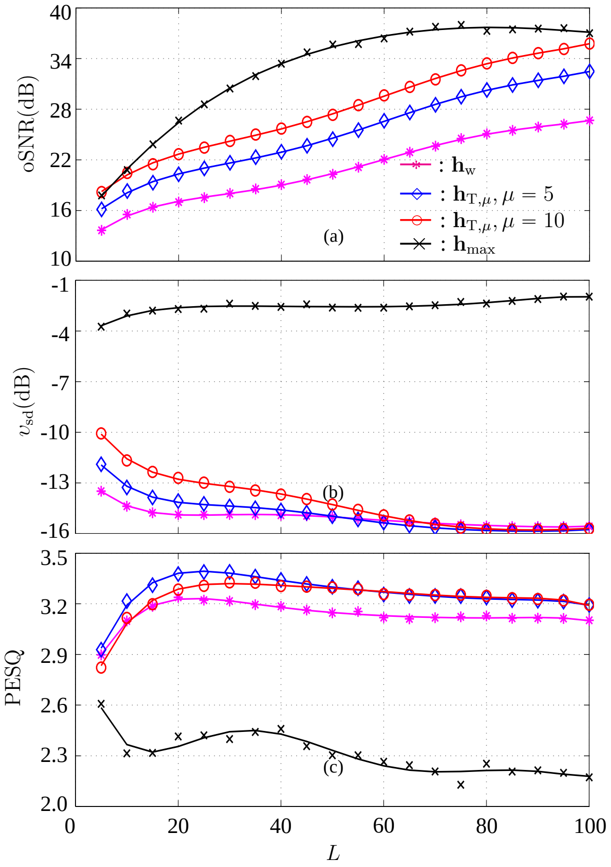

Fig. 1. Performance of the Wiener, tradeoff, and maximum SNR filters as a function of the filter length L in a car noise condition: (a) output SNR, (b) speech distortion index, and (c) PESQ. Simulation conditions: $\alpha_y = \alpha_v = 0.95$, iSNR $= 10$ dB, $N = 1$, and the PESQ score of the noisy signal is 2.375.

First, we study 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 input SNR is 10 dB and the value of all the forgetting factors is set to 0.95. The PESQ score of the noisy signal is 2.375. The results are plotted in Fig. 1. One can see from this figure that for the Wiener and tradeoff filters, both the output SNR and PESQ first increase with L, then decrease gradually, while the speech distortion index decreases monotonously. The results of the different tradeoff filters agree well with the theoretical analysis. For $\mu > 1$, the tradeoff filter reduces more noise but it introduces more distortion to the desired speech signal as compared to the Wiener filter. As for the maximum SNR filter, it achieves the maximum SNR gain in comparison with the Wiener and tradeoff filters, but it also generates the most speech distortion.

In the second simulation, we investigate the effect of smoothing on the noise reduction performance. Due to space limitation, we only present the results of the Wiener filter as a function of N. The input SNR is, again, 10 dB, the value of all the forgetting factors is set to be 0.95, and the filter length, L, is set to be $13, 15$, and 20. Please note that $N = 1$ corresponds to the case without time smoothing. The results are plotted in Fig. 2. One can see from this figure that the output SNR first decreases then increases with N while the speech distortion index increases monotonously with N.

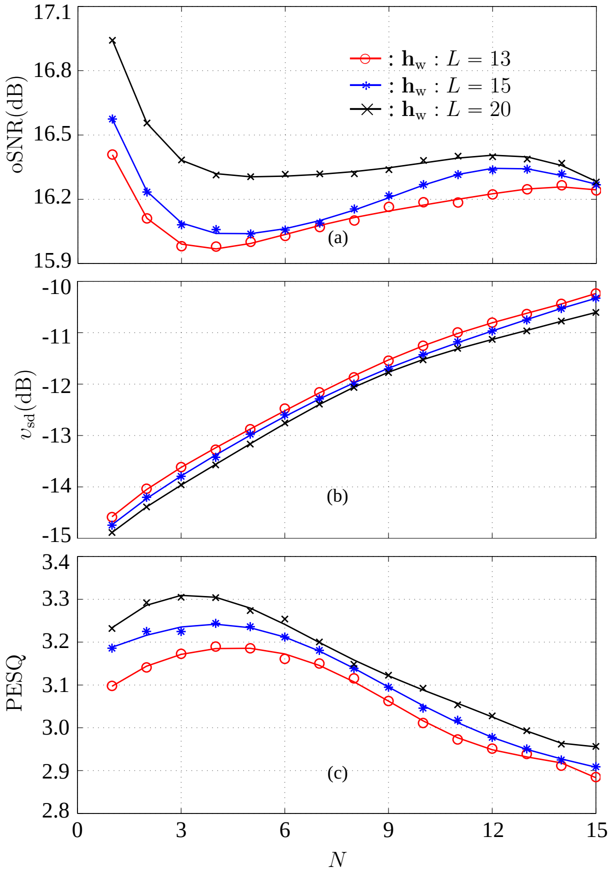

Fig. 2. Performance of the Wiener filter as a function of time smoothing length, N, in a car noise condition: (a) output SNR, (b) speech distortion index, and (c) PESQ. Simulation conditions: $\alpha_y = \alpha_v = 0.95$, iSNR $= 10$ dB, $L = 13, 15, 20$, and the PESQ score of the noisy signal is 2.375.

Moreover, the PESQ first increases and then decreases. With proper choice of the values of N and L, significant PESQ improvement are achieved, e.g., $L = 20, N = 3$, the gain in PESQ is close to 1. It is clearly seen that time smoothing technique can help improve performance. Besides improving noise reduction performance, time smoothing can also help reduce the computational complexity as a shorter filter length is needed to achieve similar performance, for instance, $L = 15$ and $N = 2$ gives almost the same PESQ score as $L = 20$ and $N = 1$.

7. CONCLUSIONS

In this paper, we presented a linear estimation technique for single-channel noise reduction in the time domain, which combines the optimal filtering and smoothing techniques together. Specifically, a smoothing window is firstly applied to the time-domain noisy signal so as to leverage the correlation between successive samples. Then, three different noise reduction filters are derived based on various criteria, including Wiener, maximum SNR, and tradeoff filters. Simulations were carried out to assess the proposed method. Significant PESQ improvement was observed with a proper choice of the time smoothing window and the filter length, which justified the effectiveness of the proposed smoothing-and-filtering approach. While it can be used to improve performance, the presented technique can also be used to reduce the computational complexity of the optimal filter technique as a shorter filter length is needed with smoothing to achieve a similar performance with a longer filter length without smoothing.

8. RELATION TO PRIOR WORK

Acoustic noise is omnipresent in our environments, which may bring detrimental effects to speech communication and human-machine interface systems such as cellular phones, automatic speech recognition (ASR), hearing aids, audio bridging, teleconferencing, and robotics. Noise reduction is the process of recovering a clean speech signal of interest from microphone observations (either a single microphone or multiple microphones) corrupted by additive noise [1, 2]. A great deal of efforts have been devoted to addressing this problem in the literature [9, 10, 11] and various methods have been proposed, including subspace methods [12, 13, 14], optimal filtering [3, 4, 15], statistical approach [16], spectral subtraction type of techniques [17, 18], and data driven based machine learning methods [19, 20], etc.

Some of the aforementioned methods conduct noise reduction in the time domain, while others operate in transform domains [21], among which the frequency domain or short-time-Fourier-transform domain is widely adopted [9, 10, 11, 22]. Generally, working in the frequency domain makes the implementation computationally efficient due to the fast Fourier transform (FFT), but “musical noise” [23] is a troublesome problem for noise reduction algorithms in this domain. In comparison, time-domain methods are more computationally expensive, but they do not suffer from the musical noise problem. Moreover, analysis of noise reduction performance in the time domain can be easier than in a transform domain from a statistical viewpoint. As a result, it is still important to study noise reduction in the time domain.

In this paper, we focused on the noise reduction problem in the time domain. One of the most straightforward ways to estimate a sample in this domain is to pass the noisy samples through a filtering vector. The core issue then becomes one of finding a good filter, which can improve the SNR without adding much distortion to the speech of interest. In the literature, many different good filters are derived, some are optimal from an optimization point of view while others are suboptimal, which can provide a better tradeoff between noise reduction and speech distortion. Examples of these filters include Wiener, maximum signal-to-noise-ratio (SNR) [15], linearly constraint minimum variance (LCMV) [24, 25], minimum variance distortionless (MVDR) [24], and tradeoff filters [4], which are all derived from the mean-squared-error (MSE) criterion but with different constraints [26].

In this paper, we proposed 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, thereby achieving noise reduction. Smoothing techniques are widely used in noise reduction in the frequency domain in order to leverage the correlation of adjacent frequency bins [27]. In this paper, the smoothing technique is introduced to the time-domain methods. The resulting approach consists of two steps. In the first step, a time-domain smoothing window, e.g., Hanning window, rectangular window, Hamming window etc., is applied to the noisy signal. Then, in the second step, a noise reduction filter is applied to the smoothed signal to further achieve noise reduction. We considered three different filters, i.e., maximum SNR, Wiener, and tradeoff filters, which are derived using different criteria.

9. REFERENCES

[1] P. C. Loizou, *Speech Enhancement: Theory and Practice*, Boca Raton Florida: CRC Press, 2007.

[2] J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Noise Reduction in Speech Processing*, Berlin, Germany: Springer-Verlag, 2009.

[3] J. Benesty, S. Makino, and J. Chen, *Speech Enhancement*. Berlin: Springer-Verlag, 2005.

[4] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1218–1234, July 2006.

[5] J. N. Franklin, *Matrix Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1968.

[6] *Mapping function for transforming raw result scores to MOS-LQO*, ITU-T Rec. P.862.1, 2003.

[7] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, "Darpa timit acoustic phonetic continuous speech corpus," `http://www.ldc.upenn.edu/Catalog/LDC93S1.html`.

[8] J. Benesty, J. Chen, and Y. Huang, "Time-delay estimation via linear interpolation and cross correlation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 12, no. 5, pp. 509–519, Sept. 2004.

[9] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, no. 12, pp. 1586–1604, Dec. 1979.

[10] M. Brandstein and Eds. D. B. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, Berlin, Germany: Springer-Verlag, 2001.

[11] J. Chen, J. Benesty, Y. Huang, and E. J. Diethorn, "Fundamentals of noise reduction," in *Springer Handbook on Speech Processing and Speech Communication*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds., pp. 843–871. Berlin, Germany: Springer-Verlag, 2008.

[12] Y. Ephraima and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 3, no. 4, pp. 256–266, July 2005.

[13] Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 11, no. 4, pp. 334–341, July 2003.

[14] J. Benesty, J. R. Jensen, M. G. Christensen, and J. Chen, *Speech Enhancement: A Signal Subspace Perspective*, Academic Press, 2014.

[15] G. Huang, J. Benesty, T. Long, and J. Chen, "A family of maximum SNR filters for noise reduction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 22, no. 12, pp. 2034–2047, Dec. 2014.

[16] S. Srinivasan, J. Samuelsson, and W. B. Kleijn, "Codebook-based Bayesian speech enhancement for nonstationary environments," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 15, no. 2, pp. 441–452, Feb. 2007.

[17] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 27, no. 2, pp. 113–120, Apr. 1979.

[18] R. Miyazaki, H. Saruwatari, T. Inoue, Y. Takahashi, K. Shikano, and K. Kondo, "Musical-noise-free speech enhancement based on optimized iterative spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 20, no. 7, pp. 2080–2094, Sept. 2012.

[19] Y. Xu, J. Du, L. Dai, and C. Lee, "A regression approach to speech enhancement based on deep neural networks," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 23, no. 1, pp. 7–19, Jan. 2015.

[20] X. Zhang and D.-L. Wang, "A deep ensemble learning method for nonaural speech separation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 5, pp. 967–977, May 2016.

[21] J. Benesty, J. Chen, and Y. Huang, "Noise reduction algorithms in a generalized transform domain," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 17, no. 6, pp. 1109–1123, Aug. 2009.

[22] K. K. Paliwal, B. Schwerin, and K. K. Wójcicki, "Speech enhancement using a minimum mean-square error short-time spectral modulation magnitude estimator," *Speech Commun.*, vol. 54, pp. 282–305, Jan. 2012.

[23] T. Inoue, H. Saruwatari, Y. Takahashi, K. Shikano, and K. Kondo, "Theoretical analysis of musical noise in generalized spectral subtraction based on higher order statistics," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 19, no. 6, pp. 1770–1779, Dec. 2010.

[24] J. Benesty, J. Chen, Y. Huang, and T. Gaensler, "Time-domain noise reduction based on an orthogonal decomposition for desired signal extraction," *J. Acoust. Soc. Am.*, vol. 132, no. 1, pp. 452–446, July 2012.

[25] S. M. Nørholm, J. R. Jensen, and M. G. Christensen, "Enhancement and noise statistics estimation for non-stationary voiced speech," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 645–657, Apr. 2016.

[26] M. K. Becker and T. Gerkmann, "On MMSE-based estimation of amplitude and complex speech spectral coefficients under phase-uncertainty," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 12, pp. 2251–2262, Dec. 2016.

[27] I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 11, no. 5, pp. 466–475, Sept. 2003.