Enhance Speech with Temporal Sparse Noise by Robust Kalman Filter

Zhu Liang YU *, Fei WU * , Ling CEN †, Zhenghui GU * and Yuanqing LI *
* College of Automation Science and Engineering, South China University of Technology, Guangzhou
E-mail: zlyu@scut.edu.cn
† Institute of InfoComn Research, A*STAR, Singapore
E-mail: lcen@i2r.a-star.edu.sg

***Abstract*—Single channel speech enhancement is an important problem in practice. One of the well used single channel speech enhancement method, spectral subtraction, can only work for stationary noise. Another method based on Kalman filtering is able to work with non-stationary signals. However, it can only produce optimal estimation of speech signal which is corrupted by Gaussian noise. In practice, speech acquisition also suffers from non-stationary temporal sparse noise. In this paper, we propose a method based on robust Kalman filter to remove not only the stationary noise but also such kind of temporal sparse noise. Formulated as a convex optimization problem, the robust Kalman filtering based method can be solved efficiently by Interior Point Method (IPM). Numerical results show that the proposed method is robust against temporal sparse noises.**

I. Introduction

In practice, speech signals are always corrupted by various noises. The presence of noise affects the intelligibility and quality of speech signals. Hence, enhancing speech degraded by noises is an important problem in many signal processing applications, like hearing aids, mobile communications and speech recognition, etc. The goal of speech enhancement is to improve both the intelligibility and the quality of speech, by attenuating the noises without substantially degrading the speech. A large number of methods on speech enhancement have been reported in the literatures [1], [2], including single channel speech enhancement method [1] and multi-channel speech enhancement method like microphone array [2] etc.

Single channel speech enhancement methods, like spectral subtraction [3] and stationary Wiener filtering method [1] are widely applied in practice because their simple hardware structures and efficient algorithms. Spectral subtraction method has drawbacks on "music noise" effects in the enhanced speech. Also, since speech is nonstationary in nature, stationary Wiener filter does not perform very well. To overcome this problem, a Kalman filtering based method was proposed in [4]. The Kalman filtering based method exploits the speech production model in processing and allows for nonstationarity of speech. With the existence of efficient Kalman filtering algorithm, this method as well as its derivations [5], [6] have been well applied to speech enhancement.

We know that the Kalman filter [7], [8] is derived based on two assumptions, linearity and Gaussian noise. The assumption on Gaussian noise holds for many applications and the Kalman filter based method produces high quality enhanced speech signals. However, after carefully checking the recorded speech signal, we can find that the observed signal is always corrupted by zero-mean Gaussian and sparse (Laplacian) noises. The sparse noise often has Laplacian distribution. The conventional Kalman filter treats the Laplacian noise as Gaussian noise so that it cannot achieves optimal denoising performance. In this paper, we propose an speech enhancement method based on robust Kalman filter. The robust Kalman filter [9] is designed to deal with Gaussian and Laplacian noises and formulated as a convex optimization problem. Efficient numerical methods, like Interior Point Method (IPM), are available to solve such convex optimization problem [10]. Some numerical results are shown to demonstrate the superior performance of the proposed method.

II. Brief Review on Speech Enhancement Using Kalman Filter

Using Kalman filter for speech enhancement was first introduced in [4]. The derivation of Kalman filter normally assumes the process noise and observation noise are both uncorrelated and have normal distributions. This implies that the Kalman based method is best suitable for reduction of white Gaussian noise. To deal with color noises, extended method can be found in [5].

In frame based processing, it is assumed that speech signal is stationary during each frame. For each frame, the speech signal can be represented by an autoregressive (AR) process and all the AR parameters remain constant across the segment. Since an AR process is the output of an all-pole linear system driven by white noise, the speech signal at kth time instant, $s(k)$, is given as

$$s(k) = \sum_{i=1}^{p} a_i s(k-i) + u(k) \tag{1}$$

where $a_i, i = 1, \cdots, p$ are the AR model parameters, p is the order of AR process, $u(k)$ is the white process noise. In Kalman filter based speech enhancement approach, we also assume that it is Gaussian white noise.

The AR process in (1) can be easily transformed into the following state-space model as

$$\mathbf{x}(k) = \mathbf{A}\mathbf{x}(k-1) + \mathbf{B}u(k) \tag{2}$$

where the state vector $\mathbf{x}(k)$, state transition matrix $\mathbf{A}$ and input matrix $\mathbf{B}$ are defined as

$$\begin{aligned}\mathbf{x}(k) &= \begin{bmatrix} s(k) & s(k-1) & \cdots & s(k-p+1) \end{bmatrix}^T \\ \mathbf{A} &= \begin{bmatrix} -a_1 & -a_2 & \cdots & -a_p \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 1 & \cdots & 0 \\ 0 & \cdots & 1 & 0 \end{bmatrix} \\ \mathbf{B} &= \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T \end{aligned} \tag{3}$$

In practice, we only have noise corrupted speech $y(k)$ available for processing

$$y(k) = s(k) + n(k) \tag{4}$$

where $n(k)$ is the observation noise. In conventional Kalman filtering problem, it is assume that $n(k)$ has normal distribution $\mathcal{N}(0, \sigma_v^2)$. By expressing (4) in state-space model, we have

$$y(k) = \mathbf{C}\mathbf{x}(k) + n(k) \tag{5}$$

where the observation matrix $\mathbf{C}$ is defined as

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \tag{6}$$

In speech enhancement problem, we only have observations $\{y(1), y(2), \cdots, y(k)\}$ available, what we want is to estimate the state vector $\mathbf{x}(k)$ from the observations $\{y(1), y(2), \cdots, y(k)\}$. With the state and observation equations (2) and (5), if the parameters $a_i, i = 1, \cdots, p$ are known, it is the clear that Kalman filter can readily be applied for an estimate of the state-vector $\mathbf{x}(k)$ [7], [8]. Hence, Kalman filtering for speech enhancement can be implemented in two steps:

1) Estimate the AR parameters, process and observation noise covariances based on the noisy observations [11]. Since the objective of this paper is to show the effectiveness of using robust Kalman filter in speech enhancement, we assume that the AR model parameters are estimated from clean speech [4], [12], [13].
2) Apply the Kalman filtering algorithm using the estimated parameters values.

If we denote $\hat{\mathbf{x}}(k|k)$ and $\hat{\mathbf{x}}(k|k-1)$ as the estimates of state-vector $\mathbf{x}(k)$ given the measurements up to $y(k)$ and $y(k-1)$, and Σ denotes the steady-state error covariance associated with predicting the next state. The time update of Kalman filtering is given as

$$\hat{\mathbf{x}}(k|k-1) = \mathbf{A}\hat{\mathbf{x}}(k-1|k-1) \tag{7}$$

which propagates forward the state estimate at time $k-1$, after the measurement $y(k-1)$, to the state estimate at time k, but before the measurement $y(k)$ is known. With the measurement $y(k)$ is available, the measurement update is given as follows

$$\hat{\mathbf{x}}(k|k) = \hat{\mathbf{x}}(k|k) + \Sigma\mathbf{C}^T(\mathbf{C}\Sigma\mathbf{C}^T + \sigma_v^2)^{-1}\tilde{y}(k) \tag{8}$$

where $\tilde{y}(k) = y(k) - \mathbf{C}\hat{\mathbf{x}}(k|k-1)$ is the innovation.

III. Robust Kalman Filtering Based Speech Enhancement Method

In practice, we find that the observation noise contained in the acquired speech signal can be modeled as Gaussian noise mixed with Laplacian noise. Hence, the observation equation in (5) is modified as

$$y(k) = \mathbf{C}\mathbf{x}(k) + n_g(k) + n_s(k) \tag{9}$$

where $n_g(k)$ denotes the Gaussian measurement noise $\mathcal{N}(0, \sigma_v^2)$, $n_s(k)$ is the zero-mean Laplacian noise (sparse noise). With the above two noises, the standard Kalman filter is unable to find the optimal estimate of state vector.

In this paper, we use the robust Kalman filter [9] in speech enhancement problem. For easy understanding of the robust Kalman filter, we first reformulate the measurement update (8) into a convex optimization problem as follows

$$\begin{aligned} &\min\ n^2(k)/\sigma_v^2 + (\mathbf{x} - \hat{\mathbf{x}}(k|k-1))^T\Sigma^{-1}(\mathbf{x} - \hat{\mathbf{x}}(k|k-1)) \\ &s.t.\ \ \mathbf{y}(k) = \mathbf{C}\mathbf{x} + n(k) \end{aligned} \tag{10}$$

The optimization problem (10) is derived based on the maximization of a posterior $p(x(k)|y(k))$. The first term in the cost function is a loss term corresponding to the noise $n(k)$, and the second is a loss term associated with the state estimate deviated from the prior. The problem in (10) produces the same estimate as (8).

With the observation equation given in (9), the robust Kalman filter is given as

$$\begin{aligned} &\min\ n_g^2(k)/\sigma_v^2 + (\mathbf{x} - \hat{\mathbf{x}}(k|k-1))^T\Sigma^{-1}(\mathbf{x} - \hat{\mathbf{x}}(k|k-1)) \\ &\qquad + \lambda||n_s(k)||_1 \\ &s.t.\ \ \mathbf{y}(k) = \mathbf{C}\mathbf{x} + n_g(k) + n_s(k) \end{aligned} \tag{11}$$

where $||\mathbf{x}||_1$ denotes the $\mathcal{L}_1$ norm

$$||\mathbf{x}||_1 = \sum_{i=1}^{N} |x_i|,$$

and the parameter $\lambda \geq 0$ is used to adjust the sparsity of $n_s(k)$. If the variable λ is large enough, the value of $n_s(k)$ in the above optimization problem will become zero. In such case, the robust Kalman filter is the same as the standard Kalman filter.

Since (11) is a convex optimization problem [10], it is not difficult to find the optimal values of the variables $n_g(k), n_s(k)$ and $\mathbf{x}$. However, we know that if more variables are involved in optimization problem, the numerical method for solving (11), such interior point method [10], [14], is less efficient. In order to solve this problem, we can manually eliminates the variables in (11). After applying some mathematical manipulations, the problem in (11) can be formulated as

$$\min\ (\tilde{y}(k) - n_s(k))^T\mathbf{Q}(\tilde{y}(k) - n_s(k)) + \lambda||n_s(k)||_1 \tag{12}$$

where

$$\begin{aligned}\tilde{y}(k) &= y(k) - \mathbf{C}\hat{\mathbf{x}}(k|k-1) \\ \mathbf{L} &= \Sigma\mathbf{C}^T(\mathbf{C}\Sigma\mathbf{C}^T + \sigma_v^2)^{-1} \\ \mathbf{Q} &= ||\mathbf{I} - \mathbf{C}\mathbf{L}||_2^2/\sigma_v^2 + \mathbf{L}^T\Sigma^{-1}\mathbf{L}\end{aligned} \quad (13)$$

After solving the optimization problem in (12), we can recover the speech signal from the state-vector

$$\hat{\mathbf{x}}(k|k) = \hat{\mathbf{x}}(k|k-1) + \mathbf{L}(\tilde{y}(k) - n_s(k)) \quad (14)$$

IV. Numerical Results

In this section, we demonstrate some numerical examples to show the performance of the proposed method. The speech used in simulations are sampled with rate $10k$Hz. The order of AR model used in simulation is 13. The output signal-to-noise ratio (SNR) of the enhanced speech at different input SNR are studied using the standard Kalman filtering based method and the proposed method. In the simulation of measurement noise, the sparse noise are simulated to randomly distributed in time axis with probability 0.05 and has amplitude in Gaussian distribution.

In the first experiment, a clean speech signal shown in Fig. 1 are corrupted by Gaussian noise of variance 1.0 and sparse noise with amplitude variance 2.25. The noise corrupted signal is shown in Fig. 2. It is clear that this speech signal is seriously corrupted by noise and its input SNR is $-15.5193dB$. The speech enhanced by the standard Kalman filtering method and the proposed robust Kalman filtering method are shown in Fig. 3 and Fig. 4, respectively. Comparing the original speech signal with the enhanced speech signals in Fig. 3 and Fig. 4, it is clear that the speech enhanced by the standard Kalman filtering method has shrunk amplitude and some residual sparse noises. Moreover, from the difference between the enhanced speech and the original one shown in Fig. 5, we can find that the speech signal enhanced by the standard Kalman filtering method has significant difference to the origianl one. The signal error of the proposed method is significantly lower than that of the standard Kalman filtering method. After listening test, we find that the signal enhanced by the robust Kalman filtering method also has high speech quality.

To further compare the performance of the methods under different input SNR, in the second simulation experiment, we compare the output SNR of these method with different input SNR. The SNR is changed by setting different amplitude of the sparse noise. The results shown in Table I indicate that when the sparse noise is weak, the proposed method has similar performance as the standard Kalman filtering based method. This is because in such case, the Gaussian noise is the dominant noise, the robust Kalman filter performs similarly as the standard one. However, when the sparse noise increases, for example, when the input SNR is $-12.2303dB$, the output SNR of the proposed method is significantly higher than that of the standard Kalman filtering method.

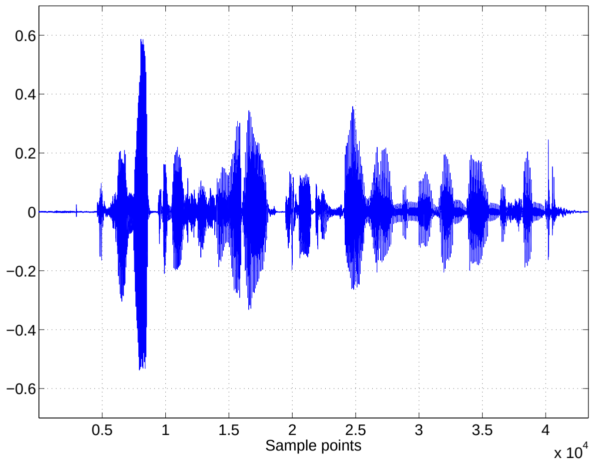

Fig. 1. The clean speech signal

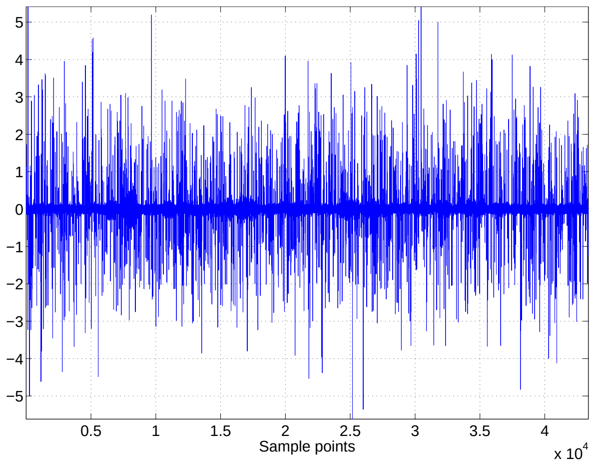

Fig. 2. The speech signal with Guassian and Sparse noises

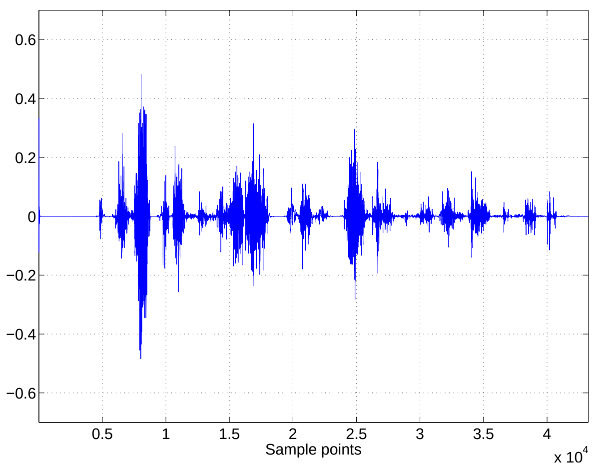

Fig. 3. The estimated speech signal by standard Kalman filtering method

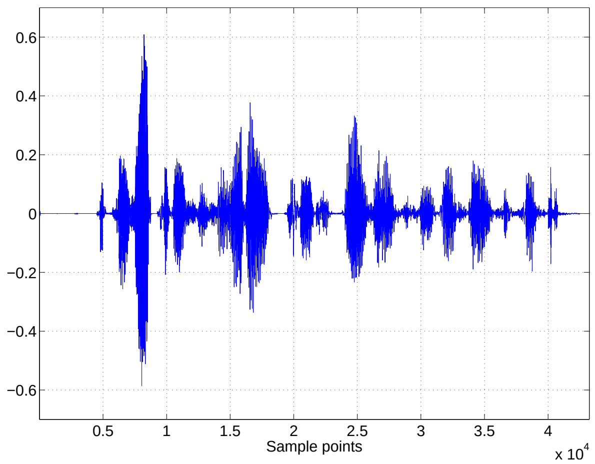

Fig. 4. The estimated speech signal by the proposed robust Kalman filtering method

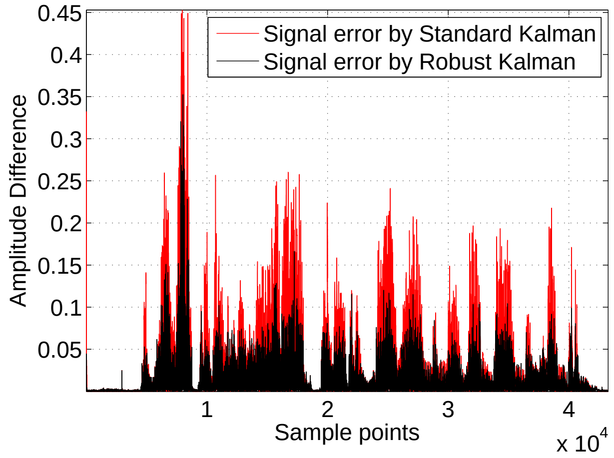

Fig. 5. Comparison of the signal error produced by the proposed robust Kalman filtering method and the standard Kalman filtering method

V. Conclusions and Discussions

A robust Kalman filtering based method is proposed for speech enhancement. This method is especially useful when the speech signal is corrupted by not only conventional Gaussian noise but also sparse noise. In this method, we firstly use conventional method to estimate the AR parameters of the speech segments, then use these AR parameters in Kalman filtering. Since the sparse noise only affects the observation equation, in the robust Kalman filtering method, we still use the conventional time updating approach, but modify the measurement updating approach. The simulation results show the effectiveness of the proposed method. It has significantly higher output SNR than the standard Kalman filtering based method. Although the measurement update approach is formulated in convex optimization problem, its computational complexity is higher than the conventional measurement updating approach in Kalman filter. How to derive an efficient method to get the solution of the robust Kalman filter is one of our future research topic. Also in the the estimation of AR parameters, we do not consider effects of the sparse noise contained in the observed speech. How to improve the performance of AR parameters with the existence of sparse noise is also an important problem in our future research.

Input SNR	Standard Kalman	Robust Kalman
-0.1586	9.5742	9.4606
-0.6402	9.2184	9.2544
-2.1082	8.1398	8.8342
-6.7830	5.0359	8.3517
-12.2303	1.5013	8.2773

TABLE I
Comparison of output SNR at different input SNR by the standard Kalman filtering method and the proposed robust Kalman filtering method. (The unit of this table is dB)

Acknowledgment

This work was supported in part by NSFC (60802068 and 60825306), Program for New Century Excellent Talents in University (NCET-10-0370), the Fundamental Research Funds for the Central Universities, SCUT (2009ZZ0055 and 2011ZM0124), the Specialized Research Fund for the Doctoral Program of Higher Education (20090172120039) and SRF for ROCS, SEM.

References

[1] J. S. Lim, Ed., *Speech Enhancement*. Englewood Cliffs, N.J.: Prentice-Hall,, 1983.

[2] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*. Springer, 2001.

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

[4] K. Paliwal and A. Basu, “A speech enhancement method based on kalman filtering,” in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP’87.*, ser. Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP’87., vol. 12, 1987, pp. 177–180.

[5] S. Gannot, D. Burshtein, and E. Weinstein, “Iterative and sequential kalman filter-based speech enhancement algorithms,” *IEEE Trans. Speech Audio Process.*, vol. 6, no. 4, pp. 373–385, 1998.

[6] B. Kybic, “Kalman filtering and speech enhancement,” *Master’s thesis, éecole Polytechnique Féedéerale De Lausanne*, 1998.

[7] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Serial D, pp. 35–45, 1960.

[8] M. S. Grewal and A. P. Andrews, *Kalman Filtering: Theory and Practice Using Matlab*. John Wiley & Sons, Inc., 2001.

[9] J. Mattingley and S. Boyd, “Real-time convex optimization in signal processing,” *IEEE Signal Proc. Mag.*, vol. 27, no. 3, pp. 50–61, 2010.

[10] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[11] J. Lim and A. Oppenheim, “All-pole modeling of degraded speech,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 26, no. 3, pp. 197 – 210, 1978.

[12] Y. Ephraim, D. Malah, and B. H. Juang, “On the application of hidden markov models for enhancing noisy speech,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 1846–1856, 1989.

[13] Y. Ephraim, “A bayesian estimation approach for speech enhancement using hidden markov models,” *IEEE Trans. Signal Process.*, vol. 40, no. 4, pp. 725–735, 1992.

[14] Y. Nesterov and A. Nemirovskii, *Interior-point polynomial algorithms in convex programming*. Society for Industrial Mathematics, 1987.