# Speech Enhancement in Presence of Colored Noise Using An Improved Least Square Estimation

Youshen Xia and Pengyu Wang
College of Mathematics and Computer Science
Fuzhou University, China (ysxia@fzu.edu.cn)

**Abstract** **Speech enhancement in presence of colored noise has been of significant topics. This paper presents a novel parameter estimation-based method for speech enhancement under colored noise environments. Parameters of speech signal modeled as autoregressive process are estimated by using an improved least square estimation and then the speech signal is recovered from the Kalman filtering. Compared with the adaptive recursive estimation-based method for speech enhancement, the present method has a low computation complexity. Moreover, computed results shows that the presented parametric method has a better performance in having a significant gain in SNR than the adaptive recursive estimation-based method at different colored noise.**

**Keywords** Colored noise, noisy speech signal, speech enhancement, parametric method

## 1 Introduction

Speech enhancement has been studied because of its many applications, such as voice communication, voiced -control systems, and the transmitted speech signals, where received speech signals are corrupted by background noise which is either white or colored. The objective of speech enhancement is to restore the original signal based on a single sequence of noisy observations [1]. There are several types of methods for speech enhancement. The first type is the spectral subtraction method which employs nonparametric techniques [2-3]. The second type is the subspace method, which is based on well-known singular value decomposition techniques. Signal enhancement is to remove the noise subspace and to estimate the clean speech signal from the noisy speech subspace [4-5]. The third type is the parametric method. The speech signal is modeled as autoregressive (AR) process. After the AR parameters are estimated, the speech signal is then recovered from Kalman filtering [6-12]. Most of parametric methods use standard stationary Gaussian white noise assumption. In practices, noisy speech is contaminated by colored noise. So, speech enhancement in presence of colored noise has been of significant topics. A early parametric method for speech enhancement dealing with colored noise was proposed by Gibson et al [10]. Recently, Gabrea presented an adaptive parameter estimation method for speech enhancement in presence of colored noise. The adaptive parameter estimation method combined two techniques. The

© 2013. The authors - Published by Atlantis Press 

estimation of the driving processes variances is derived by an adapting approach proposed in control by Myers and Tapley [13]. The parameter estimation of the transition matrix, which contains the speech AR models parameters, was made using a adaptation of the robust recursive least square algorithm with variable forgetting factor proposed by Milosavljevic et al. [14]. All these parametric methods use an argument Kalman filtering and both noisy AR parameters and speech AR parameters need being estimated, which cause a high computation complexity.

This paper presents a novel parameter estimation-based method for speech enhancement in presence of colored noise. Parameters of speech signal modeled as autoregressive process are estimated by using an improved least square estimation [15,16] and then the speech signal is recovered from the non-argument Kalman filtering. Compared with the adaptive recursive estimation-based method, the present method has a low computation complexity. Moreover, computed results shows that the presented parametric method has a better performance in having a significant gain in SNR than the adaptive recursive estimation-based method at different colored noise.

## 2 Speech model and Kalman filtering

Consider clean speech signal $s(k)$, which is modeled as an autoregressive (AR) signal

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

where $\{a_i\}$ are the speech AR parameters, $s(k)$ is the kth sample of speech signal, $u(k)$ is the kth sample of the drive white noise with variance $\sigma_u^2$, and $p$ is the speech model order. The clean speech signal $s(k)$ is observed in the presence of the additive noise

$$y(k) = s(k) + v(k) \tag{2}$$

where $y(k)$ is the kth sample of the observation and $v(k)$ is colored noise with covariance matrix $R_v$, which is assumed to be uncorrelated with the drive noise sequence $u(k)$. In a special case that the observation noise is a Gaussian white noise, $R_v$ is a diagonal matrix and its diagonal elements represent the noise variances. The purpose of speech enhancement is to estimate the clean speech $s(k)$ from noisy speech observation $y(k)$.

Define a $p$-dimensional clean vector, state vector, measured noise vector, deriving noise vector as $\mathbf{x}(n) = [s(n-p+1),...,s(n-1),s(n)]^T$, $\mathbf{y}(n) = [y(n-p+1),...,y(n-1),y(n)]^T$, $\mathbf{v}(n) = [v(n-p+1),...,v(n-1),v(n)]^T$, $\mathbf{u}(n) = [u(n-p+1),...,u(n-1),u(n)]^T$, and the transition matrix as

$$F_a = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 \\ a_p & a_{p-1} & a_{p-2} & \dots & a_2 & a_1 \end{pmatrix}$$

respectively. Using a vector Kalman filter, the model of the measured speech signal is expressed as

$$\begin{cases} \mathbf{x}(n) = F_a \mathbf{x}(n-1) + G\mathbf{u}(n) \\ \mathbf{y}(n) = H_p \mathbf{x}(n) + \mathbf{v}(n) \end{cases} \tag{3}$$

where $H_p$ is a $p$-th order identity matrix and $G = [0,...,0,1]^T \in R^p$. Then the standard Kalman filter estimation and updating equations for speech enhancement are as follows:

$$\begin{cases} \mathbf{K}(n) = P(n|n-1)(R_v + P(n|n-1))^{-1} \\ P(n|n-1) = F_a P(n-1|n-1)F_a^T + \sigma_u^2 GG^T \\ \hat{\mathbf{x}}(n) = F_a \hat{\mathbf{x}}(n|n-1) + \mathbf{K}(n)\mathbf{e}(n) \\ P(n) = (I - \mathbf{K}(n))P(n|n-1) \end{cases} \tag{4}$$

where $\mathbf{e}(n) = \hat{\mathbf{x}}(n) - \hat{\mathbf{x}}(n|n-1)$, $\hat{\mathbf{x}}(n|n-1) = F_a\hat{\mathbf{x}}(n)$, $R_v$ is the covariance matrix of the measured colored noise $v$, $K(n)$ is the Kalman gain matrix, $\hat{\mathbf{x}}(n)$ represents the filtered estimate of state vector $\mathbf{x}(n)$, $P(n)$ is the filtered state error covariance matrix, and $P(n|n-1)$ is predicted state error correlation matrix.

## 3 Existing adaptive estimation-based parametric method

The adaptive estimation-based parametric method restores speech signals by an argument Kalman filtering, which was early developed by Gibsion et al. Recently, Gabrea presented an improvement on the parametric method by using both a adaptive technique and a recursive least square technique with variable forgetting for the estimation of the driving noise and the speech AR parameter. Let measured colored noise be modeled as $q$th-order AR process:

$$v(k) = \sum_{i=1}^{q} a_i v(k-i) + w(k) \tag{5}$$

where $w(k)$ is zero-mean Gaussian white noise with variance being $\sigma_w^2$. Then the colored noise speech model in both (1) and (2) can be represented by an argument state system:

$$\begin{cases} \mathbf{x}(n) = F\mathbf{x}(n-1) + \mathbf{d}(n) \\ y(n) = H^T\mathbf{x}(n) \end{cases} \tag{6}$$

where $\mathbf{d}(n) = [0,...,0,u(n),0,...,0,w(n)]^T \in R^{p+q}$, $H = [0,...,1,0,...,1]^T \in R^{p+q}$, $F = diag(F_a, F_b)$, $F_b \in R^{q\times q}$ is of the form defined in (3) with replacing elements $\{b_j\}$, and $\mathbf{x}(n) = [s(n-p+1),...,s(n-1),s(n),v(n-p+1),...,s(n-1),v(n)]^T$. After parameters $\{a_j\},\{b_j\},\sigma_u^2,\sigma_w^2$ are estimated, the speech signal is estimated by using the following argument Kalman filtering:

$$\begin{cases} \mathbf{k}(n) = P(n|n-1)H^T/(\sigma_v^2 + H^T P(n|n-1)H) \\ P(n|n-1) = FP(n-1)F^T + \sigma_w^2 GG^T \\ \hat{\mathbf{x}}(n) = F\hat{\mathbf{x}}(n-1) + \mathbf{k}(n)e(n) \\ P(n) = (I - \mathbf{k}(n)H^T)P(n|n-1) \end{cases} \tag{7}$$

where $e(n) = y(n) - H^T F\hat{\mathbf{x}}(n-1)$ and $k(n)$ is the Kalman gain vector. The speech signal estimate is then obtained by the $p$th componet of the state-vector estimate $\hat{\mathbf{x}}(n)$.

The adaptive recursive estimation algorithm is briefly described as follows:

Step 1. Perform the argument Kalman filtering (7) for state estimate:$\hat{\mathbf{x}}(k)$.

Step 2. Compute the noise AR parameters $\{b_j\}$ during silence period.

Step 3. Compute the speech AR parameters $\{a_j\}$ using the recursive least square estimation [14]:

$$\theta(k) = \theta(k-1) + Q(k)\mathbf{x}(k-1)\psi'(e(k))/\lambda_0$$

where $\theta(k) = [a_p(k),...,a_1(k)]^T$ and $Q(k) = Q(k-1) - h(k)\mathbf{x}(k-1)^T Q(k-1)\psi(e(k))$, $\lambda_0$ is the forgetting factor, $h(k)$ is the design step length, and $\psi(e(k))$ is the Huber influence function.

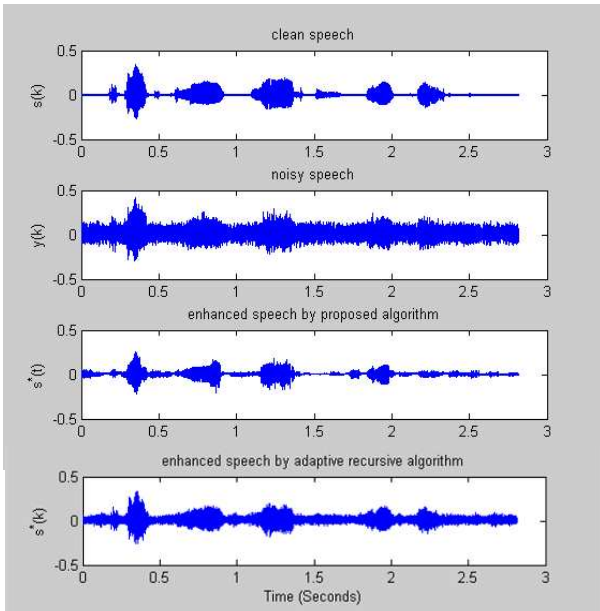


Fig. 1 Waveform of the male clean, noisy male speech (additive colored noise), and enhanced speech by the two algorithms in example 1.

Step 4. Compute the driving variance estimate by using an adaptive method [13].

Step 5. Compute the noise variance estimate by using

$$\sigma_w^2 \approx E[\beta^2(k)]$$

where $\beta^2(k) = H_1^T[\hat{\mathbf{x}}(k) - F\hat{\mathbf{x}}(k)]$ and $H_1 = [0,...,0,1]^T$.

## 4 Our parametric method

Different from the adaptive recursive estimation method, the speech AR parameter estimation is first obtained by using an improved least square method. The improved least square method for AR parameter estimation in colored noise was early introduced by Zheng [15]. Recently, Mahmoudi and Karimi [16] presented a further improvement on AR parameter estimation by using a joint technique between low-order equations and high-order equations given by

$$H_y\mathbf{a} = H_v\mathbf{a} + h_y - h_v \tag{8}$$

where $\mathbf{a} = [a_1,...,a_p]^T$, $H_y = [R_y, R_{ym}]^T$, $H_v = [R_v, R_{vm}]^T$, $h_y = [r_y, r_{ym}]^T$, $h_v = [r_v, r_{ym}]^T$, $r_y = E[y(t)\mathbf{y}(t)]$ and $r_v = E[v(t)\mathbf{v}(t)]$, $R_y = E[\mathbf{y}(t)\mathbf{y}^T(y)]$, $R_v = E[\mathbf{v}(t)\mathbf{v}^T(y)]$, and others are defined in a similar way. Based on (8), an iterative algorithm can be used to find the speech AR parameter estimate. Next, let $\hat{\mathbf{a}}$ be the speech AR parameter estimate. The drive noise variance estimate can be computed by

$$\sigma_w^2 \approx E[(y(t) - \mathbf{y_t}^T\hat{\mathbf{a}})^2]. \tag{9}$$

where $\mathbf{y_t} = [y(t-1),...,y(t-p)]^T$. Finally, we estimate the covariance matrix, $R_v$, of the measured colored noise during the silence period. Note that the main diagonal elements of $R_v$ is far larger than other elements. Then

$$R_v \approx \Lambda_v \approx \hat{\sigma}_v^2 I \tag{10}$$

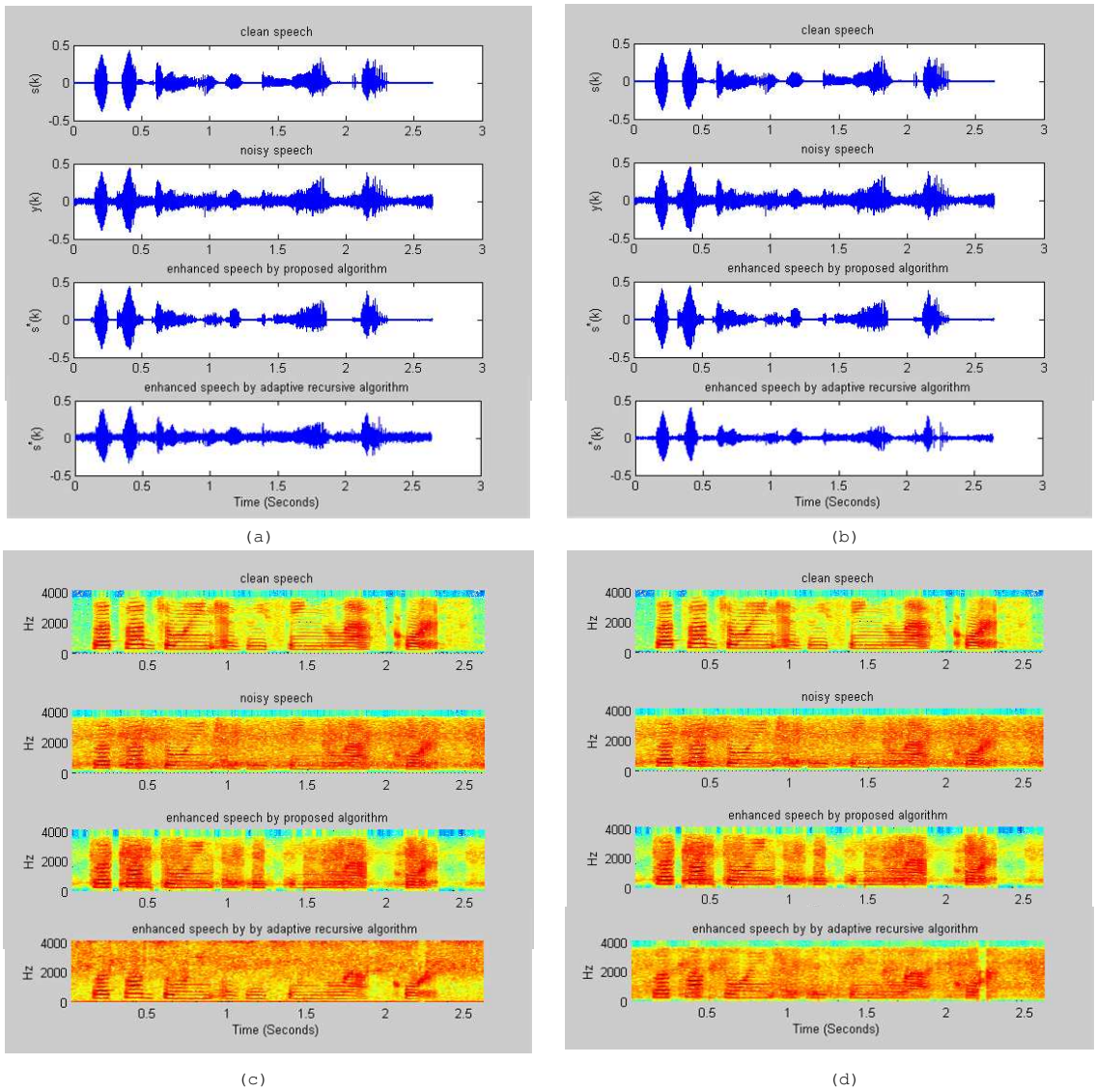


Fig. 2 (a) Waveform of the male clean, noisy male speech (train noise, SNR=5dB), and enhanced speech by the two algorithms; (b) Spectrogram results of the male clean, noisy male speech, and enhanced speech by the two algorithms in example 2.

where $\hat{\sigma}_v^2$ is the average of the main diagonal elements. Therefore, we summary our algorithm as follows:

Step 1. Compute auto-covariance estimates by using given observation $\{y(k)\}$

$$\hat{r}_k(N) = \frac{1}{N}\sum_{t=1}^{N} y(t)y(t-k),\ k = 0, 1, \cdots, p+m. \tag{11}$$

Step 2. Compute the initial estimates

$$\hat{\mathbf{a}}^{(0)} = \hat{\mathbf{a}}_H = (\hat{H}_y^T \hat{H}_y)^{-1} \hat{H}_y^T \hat{\mathbf{h}}_y,\ \mathbf{g} = \hat{\mathbf{r}}_y - \hat{R}_y \hat{\mathbf{a}}_H \tag{12}$$

Step 3. Calculate the estimate of the measurement noise covariance vector and use its elements to construct $\hat{H}_v^{(i)}$ and $\mathbf{h}_v^{(i)}$.

Step 4. Perform the bias correction

$$\hat{\mathbf{a}}^{(i)} = \hat{\mathbf{a}}_H + (H_y^T H_y)^{-1} H_y^T (\hat{H}_v^{(i)} \hat{\mathbf{a}}^{(i-1)} - \hat{\mathbf{h}}_v^{(i)}). \tag{13}$$

Step 5. If $\|\hat{\mathbf{a}}^{(i)} - \hat{\mathbf{a}}^{(i-1)}\|_2 / \|\hat{\mathbf{a}}^{(i)}\|_2 \leqslant \delta$ where $\delta$ is an appropriate small positive number, compute the drive noise variance estimate by using (9).

Step 6. Compute the covariance matrix of the measured colored noise during the silence period by using (10).

Step 7. Perform the Kalman filtering (4).

It is easy to see that the adaptive estimation algorithm performs an argument Kalman filtering and thus has a higher computation complexity than the proposed algorithm.

## 5 Simulation results

In this section, we give illustrative examples to demonstrate the effectiveness of the proposed algorithm. We evaluate the algorithm performance by using the signal to noise radio(SNR) and the quality of enhanced speech components. The SNR is defined by

$$SNR = 10log\frac{\sum_{n=1}^{N} x(n)^2}{\sum_{n=1}^{N} [x(n) - \hat{x}(n)]^2}$$

where $\hat{x}(n)$ is the estimated speech signal and $N$ is the total sample length. The quality of enhanced speech components are evaluated in the time domain and the frequency domain by means of the spectrogram. The clean speech data, a male signal called "sp01" and a female signal called "sp30," are collected from noisy speech corpus (NOIZEUS)[9]. The simulation is conducted in MATLAB.

**Example 1**. Consider the male speech corrupted by colored observation noise defined as

$$v(k) = 1.1v(k-1) + 0.9559v(k-2) + 0.5727v(k-3) + u(k)$$

where $u(k)$ is white noise and uniformly distributed with variance 0.1 on the interval $(0, 1)$. The noisy speech signal has the sampling frequency of 8000 Hz. 512 samples are used for each frame. For a comparison, we perform the proposed algorithm with a 12th order speech AR filter and the adaptive recursive estimation algorithm with a 4th order speech AR filter. The waveform results of the clean, noisy speech, and restored speech by the two algorithms are depicted in Fig. 1. It is seen that the proposed algorithm can suppresses more high-frequency noise than the adaptive estimation parametric method.

**Example 2** Consider the male speech and female speech corrupted by real train noise. The noisy speech signal has the sampling frequency of 8000 Hz. 512 samples are used for each frame. For a

| Noisy speech (male) | -5 dB | 0 dB | 5 dB |
|---|---|---|---|
| adaptive algorithm | $-0.338$ | 3.3651 | 5.3849 |
| our algorithm | 0.4893 | 4.366 | 8.366 |
| Noisy speech (female) | -5 dB | 0 dB | 5 dB |
| adaptive algorithm | 0.1078 | 3.5998 | 5.692 |
| our algorithm | 2.1495 | 4.1275 | 7.0954 |

**Table I Performance comparison of two algorithms in example 2.**

comparison, we perform the proposed algorithm with a 12th order speech AR filter and the adaptive estimation algorithm with a 4th order speech AR filter. The waveform of the male clean, noisy male speech with SNR=5dB, and enhanced speech by the two algorithms are depicted in Fig. 2 (a). The spectrogram results of the male clean, noisy male speech with SNR=5dB, and enhanced speech by the two algorithms are shown in Fig. 2(e). The waveform results of the female clean, noisy female speech with SNR=5dB, and enhanced speech by the two algorithms are depicted in Fig. 2 (b). The spectrogram results of the female clean, noisy female speech with SNR=5dB, and enhanced speech by the two algorithms are shown in Fig. 2(d). It is seen that the proposed algorithm preserves more harmonics and suppresses more high-frequency noise than the adaptive estimation method. Furthermore, Table I displays computed results of the SNR values by the two algorithms, respectively at -5,0, and 5dB. We see that the proposed algorithm has better performance in having a significant gain in SNR than the adaptive recursive estimation-based method at different colored noise.

## 6 Conclusion

This paper presents a novel parametric method for speech enhancement. Parameters of speech signal modeled as autoregressive process are estimated by using an improved least square estimation and then the speech signal is recovered from the Kalman filtering. Compared with the adaptive recursive estimation-based parametric method, the present method has a low computation complexity. Moreover, computed results shows that the presented parametric method has a better performance in having a significant gain in SNR than the adaptive recursive estimation-based parametric method at different colored noise.

## Acknowledgements

This work is supported by the National Natural Science Foundation of China under Grant No. 61179037.

## References

[1] L. R. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, Inc.,New Jersey, 1993

[2] S.F.Boll, “Suppression of acoustic noise in speech using spectral subtraction,” *IEEE transations on acoustics, Speech and Signal processing*, vol. 27, pp. 113-120, 1979.

[3] Y.Ephraim,D.Malah, “Speech enhancement and using minimum mean square error short-time spectral amplitude estimator,” *IEEE transaction on acoustics speech and signal processing*, vol. 32, pp. 1109-1121, 1984.

[4] C. E. Davila, “A Subspace approach to estimation of autoregressive parameters from noisy measurements,” *IEEE Transaction on Signal processing*, vol. 46, pp. 531-534,1984.

[5] S. Doclo and Marc Moonen, “GSVD-based optimal filtering for signal and multi-microphone speech enhancement,” *IEEE Transaction on Signal processing*, vol. 50, pp. 2230-2244,2002.

[6] W. Bobillet, R. Diversi, E. Grivel, et al. “Speech enhancement combining optimal smoothing and errors-in-variables identification of noisy AR processes,” *IEEE Transaction on Signal processing*, vol.55, pp.5564-5578, 2007.

[7] J. Dong, X.P. Wei, Q. Zhang. et al., “Speech enhancement algorithm based on high-order Cumulant parameter estimation,” *International Journal of Innovative Computing information and Control*, vol.5, pp. 2725-2733, 2009 .

[8] Y. S. Xia and Y. Yu, “Speech Enhancement Using Generalized Least Absolute Deviation Estimation,” *International Conference on Audio, Language and Image Processing*, Shanghai, China, Oct., 2010.

[9] G. Wang, C. G. Li, and L. Dong, “Noise Estimation Using Mean Square Cross Prediction Error for Speech Enhancement,” IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 57, pp. 1489-1498, 2010.

[10] J. D.Gibson,Boneung Koo,and Steven D. Gray , “Filtering of colored noise for speech enhancement and coding,” *IEEE Transaction on Signal Processing*, vol. 39, pp.1732-1742, 1991.

[11] M. Gabrea, E. Mandridake, and M. Najim, Adaptive Kalman Filter for Speech Enhancement from Colored Noise, in Proc. EUSIPCO 98, pp. 1457 1460.

[12] M. Gabrea, “robust adaptive Kalman filter for speech signal recovery in colored noise,” Proc. ICASSP, 2004.

[13] K. A. Myers and B. D. Tapley, Adaptive Sequential Estimation with Unknown Noise Statistics, IEEE Tran. Automatic Control, vol. AC-21, pp. 520 523, Aug. 1976.

[14] B. D. Kovacevic, M. M. Milosavljevic, and M. Dj. Veinovic, “Robust Recursive AR Speech Analysis,” Signal Processing, vol. 44, pp. 125 138, 1995.

[15] W. X. Zheng, “Estimation of the parameters of autoregressive signals from coloed noise-corrupted measurements”, *IEEE Signal Processing Letters*,vol. 7, pp. 201-204,2000.

[16] A. Mahmoudi and M. Karimi, “Parameter estimation of autoregressive signals from observations corrupted with colored noise,” *Signal Processing* vol. 90, pp. 157-164, 2010.