2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

A Fast Discrete-time Learning Algorithm for Speech Enhancement Using Noise Constrained Parameter Estimation

Youshen Xia, Guiliang Lin and Wei Xing Zheng

***Abstract*—This paper proposes a fast discrete-time learning algorithm for speech enhancement of single-channel noisy speech signal, based on a noise constrained least squares estimate. Unlike existing learning algorithms for the noise constrained estimate, the proposed discrete-time learning algorithm has a low complexity and fast speed. Simulation results show that the proposed discrete-time learning algorithm has a faster speed than the existing learning algorithms for speech enhancement. Moreover, the proposed discrete-time learning algorithm has a good performance in having a significant gain in SNR at colored noise.**

Keywords:Noise constrained estimation, discrete-time learning algorithm, speech enhancement, colored noise

I. 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.The noisy signal space is separated into two orthogonal subspaces: the noisy subspace and the signal subspace. 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 [6-10]. Speech enhancement algorithms may be divided into single-channel algorithms and multi-channel algorithms. Compared with other speech enhancement methods, the parametric method does not require stationary of additive noise and the speech signal. In this paper, we focus on the parametric method for single-channel speech enhancement.

A difficulty of the parametric method is that the Kalman filtering algorithm includes the unknown AR model parameters and the unknown noise variance. Thus the quality of the speech signal recovery based on the Kalman filtering algorithm greatly depend on how these unknown parameters are estimated in advance. So, it is important to develop a good parameter estimation method for good speech enhancement performance. The least squares (LS) method is the most basic and common estimation method for AR model parameters. The LS method is appropriate for on line identification and is asymptotically unbiased when the noise distribution is white. In practice, however, the measured AR signal is usually corrupted with colored noise. As a result, the LS method often gives a biased estimation of the true parameters and will be very poor in the worst case. To improve the accuracy of the LS estimation, many significant methods, such as Yule-Walker equations, the maximum likelihood method, the instrumental variable method and the improved least-square method have been developed [11-15]. In order to deal with non-gaussian noise environments, the high-order statistic method were developed . A generalized least absolute deviation (GLAD) method for AR parameter estimation was developed under non-Gaussian noise environments [16]. It was shown the GLAD method can obtain a good AR parameter estimate with a smaller mean square error in the presence of non-Gaussian measurement noise than the conventional LAD method. As a result, a the GLAD estimation-based algorithm for speech enhancement was developed in paper [18]. However, since the cost function of the GLAD method is non smooth, the resulting algorithm will have a very slow convergence rate. Recently, a speech enhancement algorithm [19] for the removal of noise from speech signal was presented by using a novel noise constrained least-squares (NCLS) method [17]. The NCLS estimation-based Kalman filtering algorithm is based on a discrete-time learning algorithm.

To increase computational efficiency, this paper proposes a fast discrete-time learning algorithm for speech enhancement of single-channel noisy speech signal, based on a noise constrained least squares estimate. Unlike existing learning algorithms for the noise constrained estimate, the proposed discrete-time learning algorithm has a low complexity and fast speed. Simulation results show that the proposed discrete-time learning algorithm has a faster speed than the existing learning algorithms for speech enhancement. Moreover, the proposed discrete-time learning algorithm has a good performance in having a significant gain in SNR at colored noise.

Youshen Xia and Guiliang Lin are with College of Mathematics and Computer Science, Fuzhou University, China (ysxia@fzu.edu.cn)

Wei Xing Zheng is with the School of Computing, Engineering and Mathematics, University of Western Sydney, Sydney, NSW 2751, Australia

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

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

II. Speech model and estimation

A. 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 σ_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

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

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:

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

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.

It is seen that the the Kalman filtering includes three unknown parameters to be estimated: the AR model parameters $\{a_i\}$ in the transition matrix F, the derive noise variance σ_w^2, and the observed noise variance σ_v^2. While the two variance estimates may be computed by the AR model parameters. So, the quality of the speech signal recovery based on the Kalman filtering algorithm greatly depend on how the model parameters are estimated in advance. More exactly, speech enhancement performance of the parametric method is basically determined by employing the AR model parameter estimation.

B. Noise constrained estimation

It is well known that the noise corrupted in noisy speech is usually non-Gaussian. To deal with non-Gaussian noise, recently two noise constrained estimation methods [16,17] were developed. Let observed vector $\mathbf{y}(t) = [y(1), ..., y(N)]^T$ and noise vector $\mathbf{n}(t) = [n(1), ..., n(N)]^T$, and let

$$B = \begin{pmatrix} y(0) & y(-1) & \dots & y(1-p) \\ y(1) & y(0) & \dots & y(2-p) \\ \vdots & \vdots & \ddots & \vdots \\ y(N-1) & y(N-2) & \dots & y(N-p) \end{pmatrix}.$$

Then the noisy speech model can be written as a linear equation in a matrix and vector form

$$B\mathbf{a}^* - \mathbf{y} - \mathbf{n} = \mathbf{0}. \tag{5}$$

An l_1 norm-based noise constrained estimation method (called the GLAD estimation method) was proposed. It is to find an optimal solution, $(\mathbf{a}_\gamma^*, \mathbf{z}^*)$, of the following optimization problem

$$\begin{aligned} \min \quad & \|B\mathbf{a} - \mathbf{y} - \mathbf{z}\|_1 \\ \text{s.t} \quad & \mathbf{a} \in R^p,\ \mathbf{z} \in \Omega_\gamma, \end{aligned} \tag{6}$$

where $\|\cdot\|_1$ denotes l_1 norm, $\Omega_\gamma = \{\mathbf{z} \in R^N \mid \gamma_1 \mathbf{m} \le \mathbf{z} \le \gamma_2 \mathbf{m}\}$, $\mathbf{m} = E[y(t)]\mathbf{e}$, and γ_1 and γ_2 are design parameters which is determined by a sequential cross-validation technique [17,18]. To overcome the non-smooth cost function in the GLAD estimation, an l_2 norm-based noise constrained estimation method was presented. The noise constrained least-squares (NCLS) estimate is obtained by solving the following quadratic convex optimization problem

$$\begin{aligned} \min \quad & \|Y\mathbf{a} - \mathbf{y} - \mathbf{z}\|_2^2 \\ \text{s.t} \quad & \mathbf{a} \in R^p,\ \mathbf{z} \in \Omega_\gamma, \end{aligned} \tag{7}$$

where $\|\cdot\|_2$ denotes l_2 norm.

III. Learning algorithm for speech enhancement

A. Existing learning algorithms

A neural network can operate in either continuous time or discrete time form. A continuous-time neural network described by a set of ordinary differential equations enables us to solve optimization problems in real time due to the massively parallel operations of the computing units and due to its real-time convergence rate. In comparison, discrete-time

models can be considered as special cases of discretization of continuous-time models.

To solve the noise constrained estimation problem (6), one continuous-time cooperative learning algorithm [16] was proposed as follows

State equation

$$\frac{d\mathbf{x}}{dt} = -\mu B^T g^0(\mathbf{w} + B\mathbf{x} - \mathbf{y} - \mathbf{z}), \tag{8a}$$

$$\frac{d\mathbf{w}}{dt} = -\mu\{\mathbf{w}+BB^T\mathbf{w}-g^0(\mathbf{w}+B\mathbf{x}-\mathbf{y}-\mathbf{z})-(\mathbf{z}-g^1(\mathbf{z}+\mathbf{w}))\}, \tag{8b}$$

$$\frac{d\mathbf{z}}{dt} = -\mu\{\mathbf{z}-g^1(\mathbf{z}+\mathbf{w})+\mathbf{e}+g^0(\mathbf{w}+B\mathbf{x}-\mathbf{y}-\mathbf{z})\}. \tag{8c}$$

Output equation

$$\mathbf{a}(t) = \mathbf{x}(t), \tag{8d}$$

where $\mathbf{x} \in R^p$, $\mathbf{w} \in R^N$, $\mathbf{z} \in R^N$, $g^0(\mathbf{w})$ is the projection on the set Ω_γ and $g^1(\mathbf{z})$ is the projection on the set $X_1 = \{\mathbf{z} \in R^N \mid max_j|z_j| \leq 1\}$. Based on (8), a speech enhancement algorithm was developed in paper [18].

To solve the noise constrained estimation problem (7), another discrete-time learning algorithm [17] to solve (6) as follows:

State equation

$$\mathbf{x}(k+1) = (I - \beta\hat{B}^T\hat{B})\mathbf{x}(k) + \beta\hat{B}^T\mathbf{z}(k) + q, \tag{9a}$$

and

$$\mathbf{z}(k+1) = (1-\beta)\mathbf{z}(k) + \beta g(\hat{B}\mathbf{x}(k) - \hat{\mathbf{y}}), \tag{9b}$$

Output equation

$$\mathbf{a}(k+1) = \mathbf{x}(k+1) \tag{9c}$$

where $I \in R^{p\times p}$ is an unit matrix, $\hat{B} = B/\alpha$,$\hat{\mathbf{y}} = \mathbf{y}/\alpha$, $\alpha = \|B\|_2^2$, $q = \beta\hat{B}^T\mathbf{y}$, $\beta > 0$ is a given step length, and $g(\mathbf{z})$ is the projection on the set Ω_γ/α. Based on (9), another speech enhancement algorithm was developed in paper [19]. Although the l_2 norm noise constrained estimation-based speech enhancement algorithm can speed up the l_1 norm noise constrained estimation-based speech enhancement algorithm, its computation rate does not satisfy the requirement of real-time computation.

It is seen that the l_1 norm-based learning algorithm has the total number of neurons is equal to $p + 2N$ and the l_2 norm-based learning algorithm has the total number of neurons is equal to $p + N$. Therefore, the two learning algorithms have a model complexity being $O(N)$.

B. Proposed learning algorithm

To reduce model complexity and increase computation rate, in this paper we propose the following discrete-time learning algorithm for solving (7):

State equation

$$\mathbf{x}(k+1) = (I - \beta\hat{B}^T\hat{B})\mathbf{x}(k) + \beta\hat{B}^T g(\hat{B}\mathbf{x}(k) - \hat{\mathbf{y}}) + q \tag{10}$$

TABLE I
COMPLEXITY COMPARISON OF THREE LEARNING ALGORITHMS

Algorithm	Computational complexity	Asymptotic complexity
New algorithm	$2Np + p^2$	$O(2Np))$
Algorithm (8)	$2N^2 + N(3p + 11) + p^2$	$O(2N^2 + 3Np)$
Algorithm (9)	$2Np + N + p^2$	$O(2N(p + 1))$

Output equation

$$\mathbf{a}(k+1) = \mathbf{x}(k+1)$$

where $\beta > 0$ is a given step length and $\hat{B}$, q, and $g(\mathbf{z})$ are defined in (8), respectively.

It can be seen that the proposed learning algorithm has the total number of neurons is equal to p and thus has a model complexity being $O(1)$ since $N >> p$. Furthermore, Table I also show that the proposed learning algorithm has a lower computational complexity and asymptotic complexity [21].

Based on (10), we now propose a fast learning-based speech enhancement algorithm as follows:

Step 1: From the input noisy speech signal $\{y(n)\}$, compute matrix Y defined in (6). Compute the autocorrelation matrix R_y and vector r_y using the input noisy speech signal $\{y(n)\}$.

Step 2: Based on the proposed learning algorithm, compute the optimal solution, $(\mathbf{a}^*, \mathbf{z}^*)$ to the constrained optimization problem in (7), and let the AR model estimate be $\hat{\mathbf{a}} = \mathbf{a}^*$.

Step 3: Compute both the observation noise variance the deriving noise variance σ_w^2 by using $\hat{\mathbf{a}}$ based on the formulation given in paper [19].

Step 4: Compute the state matrix F using the obtained NCLS estimate and perform the Kalman filtering algorithm defined in (4) to obtain $\mathbf{x}(n)$.

Step 5. Output speech signal estimate: $z(n) = C^T\mathbf{x}(n)$.

IV. COMPUTATIONAL EXAMPLES

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 quality of enhanced speech components are evaluated in the time domain by means of the spectrogram. 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. It is easy to know that the higher SNR is, the better the performance is. The simulation is conducted in MATLAB.

All testing speech data were chosen from the NOIZEUS speech corpus. The clean speech data, a male signal called "sp01" and a female signal called "sp30," are collected. In our experiments,the frame size was 256 samples with 50% overlap.

Consider the male speech corrupted by colored observation noise modeled as

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

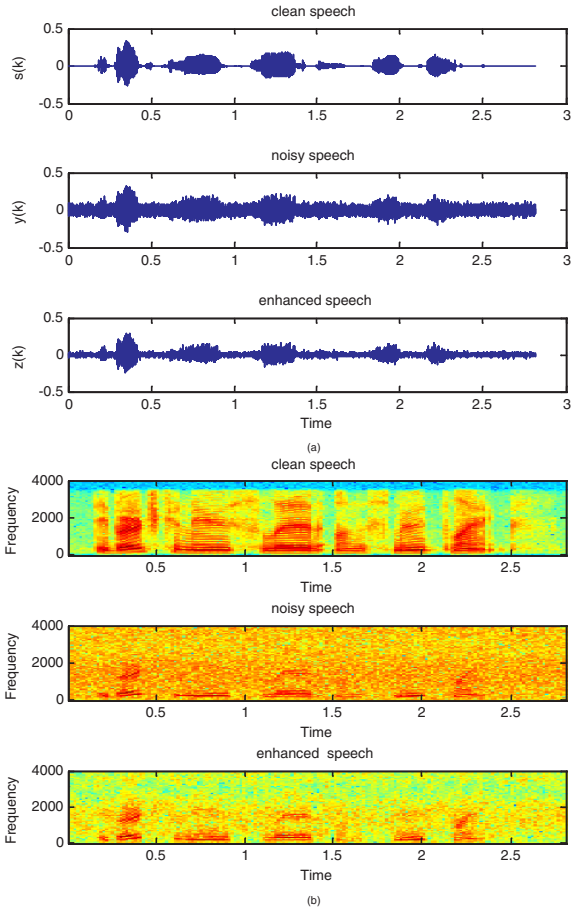

Fig. 1. Spectrogram and waveform results of the clean, noisy speech, and enhanced speech by the proposed algorithm in noisy speech sp01 (0dB)

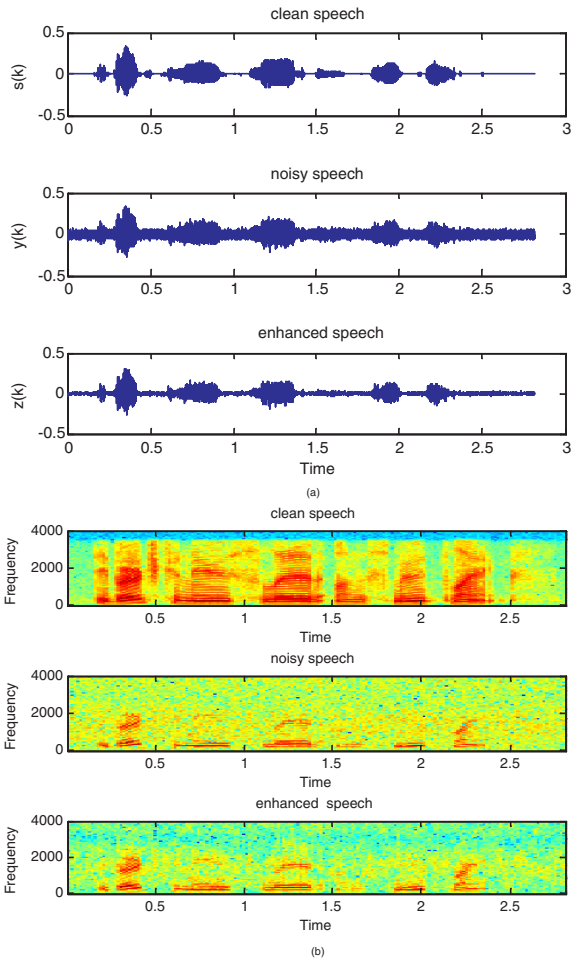

Fig. 2. Spectrogram and waveform results of the clean, noisy speech, and enhanced speech by the proposed algorithm in noisy speech sp01 (5dB)

where $u(k)$ is white Gaussian noise with variance 0.018. It results in input SNR being 0dB. The noisy speech signal has the sampling frequency of 8000 Hz. 256 samples are used for each frame. We perform the proposed algorithm with a 8th order speech AR filter. The waveform and spectrogram results of the clean speech (sp01), its noisy speech, and restored speech by the proposed algorithm are depicted in Fig. 1. It is seen that the proposed algorithm can suppresses high-frequency noise. The enhanced speech has SNR being 6.634. The waveform and spectrogram results of the clean speech (sp30), its noisy speech, and restored speech by the proposed algorithm are depicted in Fig. 3. It is seen that the proposed algorithm can suppresses high-frequency noise. The enhanced speech has SNR being 5.064. Furthermore, let $u(k)$ is white Gaussian noise with variance 0.012. It results in input SNR being 5dB. The waveform and spectrogram results of the clean speech (sp01), its noisy speech, and restored speech by the proposed algorithm are depicted in Fig. 2. It is seen that the proposed algorithm can suppresses high-frequency noise. The enhanced speech has SNR being 8.32. The waveform and spectrogram results of the clean speech (sp30), its noisy speech, and restored speech by the proposed algorithm are depicted in Fig. 4. It is seen that the proposed algorithm can suppresses high-frequency noise. The enhanced speech has SNR being 7.43.

TABLE II
RESULTS OF COMPUTATION TIME BY THREE ALGORITHMS

Algorithm	algorithm (9)	algorithm (8)	proposed algorithm
CPU(s)	0.047×256	6.31×256	0.031×256

Finally, for a comparison of computation time, we perform the proposed algorithm and the existing speech enhancement based the learning algorithms defined in (8) and (9), respectively. Table I displays computed results of the computation time by the three algorithms. Obliviously, the proposed learning algorithm has a very faster speed than the other two learning algorithms.

V. CONCLUSION

This paper proposes a novel discrete-time learning algorithm for speech enhancement of single-channel noisy speech

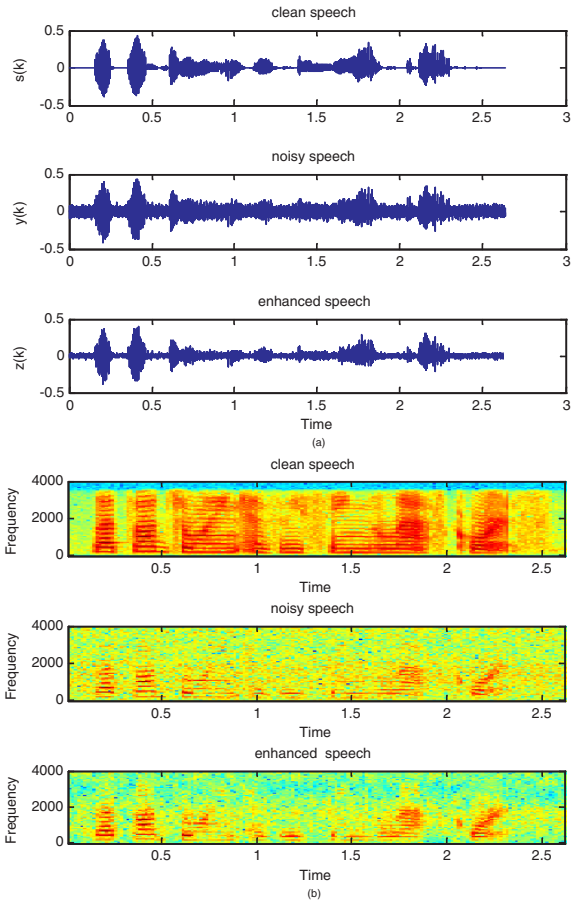

Fig. 3. Spectrogram and waveform results of the clean, noisy speech, and enhanced speech by the proposed algorithm in noisy speech sp30 (0dB)

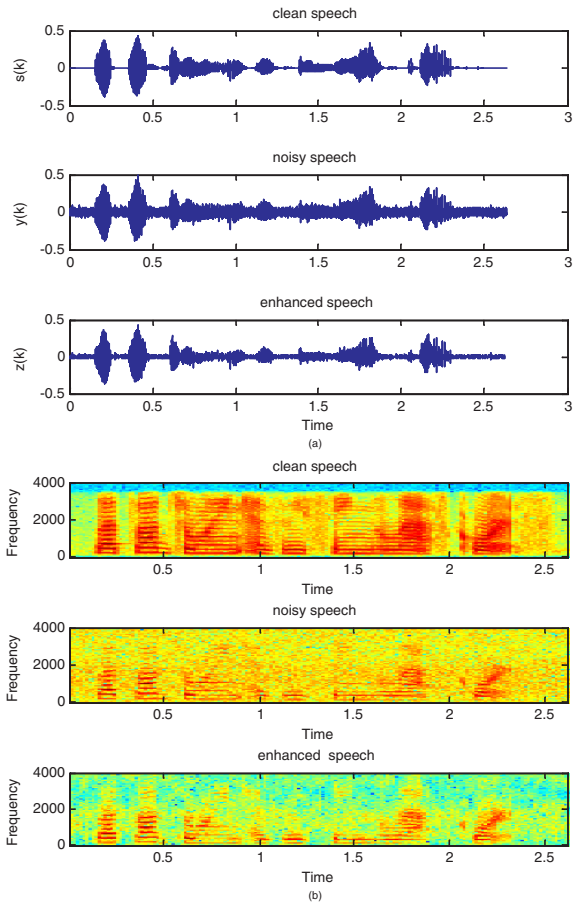

Fig. 4. Spectrogram and waveform results of the clean, noisy speech, and enhanced speech by the proposed algorithm in noisy speech sp30 (5dB)

signal, based on a novel noise constrained parameter estimation. Unlike existing learning algorithms for novel noise constrained parameter estimation, the proposed discrete-time learning algorithm has a low computation complexity and fast speed. Simulation results show that the proposed discrete-time learning algorithm has a faster speed than the existing learning algorithms for AR parameter estimation and speech enhancement. Moreover, the proposed discrete-time learning algorithm has a better performance in having a significant gain in SNR than related methods at different noise.

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

[9] 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.

[10] 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.

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

[12] S. M. Kay *Fundamentals of Statistical Signal Processing: Estimation Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1993.

[13] T. Soderstrom and P. Stoica,"Comparison of some instrumental variable methods-consistency and accuracy aspects", *Automatica*, vol. 17, pp. 101-115, 1981.

[14] W. X. Zheng, "Autoregressive parameter estimation from noise data," *IEEE Transactions on Circuits and systems*, Part II, vol. 47, no. 1, pp. 71-75, 2000.

[15] S. Alliney and S. A. Ruzinsky, "An algorithm for the minimization of mixed L_1 and L_2 norms with application to bayesian- estimation," *IEEE Transactions on Signal Processing*, vol. 42, pp. 618-627, 1994.

[16] G. B. Giannakis and J. M. Mendel, "Cumulant-Based Order Determination Of Non-Gaussian ARMA Models," *IEEE Transactions On Acoustics Speech And Signal Processing*, vol. 38, pp. 1411-1423, 1990.

[17] Y. S. Xia and M. S Kamel, "A generalized least absolute deviation method for parameter estimation of autoregressive signals," *IEEE Transactions Neural Networks*, vol 19, no.1, pp. 107-118, 2008.

[18] Y. S. Xia, M. S. Kamel, and L. Henry, "A Fast Algorithm for AR Parameter Estimation Using A Novel Noise-Constrained Least Squares method," *Neural Networks*, vol. 33, pp. 396-405, 2010.

[19] 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.

[20] Y. S. Xia, Fast speech enhancement using a novel noise constrained least square estimation*International Conference on Audio, Language and Image Processing*, 2012, 16/7/2012-18/7/2012, pp 980-985, shang hai, China.

[21] T. H. Corman, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*, Chapter 2, McGraw-Hill, New York, 1990.