MULTIPLICATIVE UPDATE OF AR GAINS IN CODEBOOK-DRIVEN SPEECH ENHANCEMENT

Qi He[1], *Chang-chun Bao*[1] *and Feng Bao*[2]

[1]Speech and Audio Signal Processing Laboratory, School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China
[2]Department of Electrical and Computer Engineering, The University of Auckland, Auckland 1142, New Zealand
very1990@emails.bjut.edu.cn, chchbao@bjut.edu.cn , fbao026@aucklanduni.ac.nz

ABSTRACT

This paper presents a novel technique for estimating auto-regressive (AR) parameters of speech and noise in the codebook-driven Wiener filtering speech enhancement method. We only train the shape codebook of speech spectrum offline, and the shape of noise spectrum is estimated online for solving the problem of noise classification. Unlike conventional codebook-driven methods, we exploit a multiplicative update rule to estimate the AR gains of speech and noise more accurately. Meanwhile, the Bayesian parameter-estimator without the noise codebook is also developed. Moreover, we achieve the goal of removing the residual noise between the harmonics of noisy speech by utilizing a very simple method, i.e., combining the codebook-driven Wiener filter with the speech-presence probability (SPP). The test results confirm the superiority of our method.

***Index Terms*—** Speech enhancement, Codebook-driven, Noise classification, SPP, Wiener filter

1. INTRODUCTION

Speech enhancement aims at removing the noise while guaranteeing the perceived quality and intelligibility of speech. Traditional speech enhancement methods, such as spectral subtraction [1], Wiener filter [2] and statistical-model-based methods [3] perform well for the stationary noises, but their performance becomes unsatisfactory for the non-stationary noises. It is mainly because these methods could not follow the changes of non-stationary noise energy and produce some unexpected noise.

To solve the problem aforementioned, some methods using the shape codebooks of speech and noise spectra have been proposed [4][5][6]. In these works, the shape codebooks of speech and noise spectra are trained offline. And the parameter-estimator based on the ML method [4][5] or the Bayesian MMSE method [6] is employed to estimate the AR parameters (AR coefficients and gains) of speech and noise. For the ML scheme, optimal speech and noise codebook entries are selected from respective codebooks, and the corresponding AR gains of speech and noise are estimated online. For the Bayesian MMSE scheme, the AR parameters of speech and noise are estimated as a weighted sum of all codebook entries of spectral shapes of speech and noise and the corresponding AR gains are estimated by ML method. The final Wiener filter constructed from the estimated AR parameters of speech and noise is used to enhance noisy speech. Because the AR gains are estimated on a short frame basis, even the rapid changes of the signal energy levels can be tracked accurately.

Although the codebook-driven Wiener filtering methods are more suitable for eliminating non-stationary noise than traditional methods, there are still many problems that need to be addressed. One is the inaccurate estimation of speech and noise AR parameters, especially the AR gains. For the ML and Bayesian MMSE codebook methods, the AR gains of speech and noise should be determined by minimizing the Itakura-Saito (IS) distortion between the modeled and observed noisy spectra. Since there is no closed-form solution for optimal speech and noise AR gain estimation, the conventional codebook-driven methods indirectly obtain the AR gain estimation based on the log-spectral (LS) distortion, which has a closed-form solution by applying the series expansion [4]. Nevertheless, we find that the accuracy of AR gains estimation is unsatisfactory. The enhanced speech contains some spectrum distortions and residual noises. Another is the problem of noise classification. The codebook for each kind of noise needs to be trained offline in conventional codebook-driven methods, which is obviously impractical since the noise is ever-changing in an actual environment. In addition, the conventional codebook-driven methods cannot suppress the residual noise between the harmonics of noisy speech, which will greatly affect the quality and intelligibility of the enhanced speech.

In this paper, the accuracy of speech and noise AR gains estimation is increased by using a multiplicative update rule [7][8]. And the problem of noise classification is also solved successfully by exploiting the Minima Controlled Recursive Averaging (MCRA) algorithm [9] to estimate the spectral shape of noise online. Considering the SPP almost equals to zero in some frequency bins, such as the frequency bins between the harmonics of noisy speech, we apply the SPP to modify the codebook-driven Wiener filter for eliminating the residual noise between the harmonics of noisy speech.

2. THE PROPOSED METHOD

In this section, we use the following noisy speech model where speech and noise are assumed independently.

$$\mathbf{y}_n = \mathbf{x}_n + \mathbf{w}_n \tag{1}$$

The $\mathbf{y}_n$, $\mathbf{x}_n$ and $\mathbf{w}_n$ denote the noisy speech, clean speech and noise signal, respectively.

2.1. The estimation of spectral shape of noise

To solve the problem of noise classification, the spectral shape of noise is estimated online by the MCRA algorithm instead of training the shape codebook of noise spectrum offline.

First, we use the MCRA algorithm to obtain the power spectrum of noise $P_w^{mcra}(k)$. Second, the auto-correlation coefficients $r_w(i)$ of noise can be calculated by Wiener-Khintchine theorem as follows:

$$r_w(m) = \frac{1}{K}\sum_{k=0}^{K-1} P_w^{mcra}(k)\cos(\frac{2\pi}{K}mk) \quad , \quad 0 \le m \le q \tag{2}$$

where q is the order of noise AR-model, K is the Fast Fourier Transform (FFT) size. Then the Levinson-Durbin recursive algorithm [10] is utilized to obtain the AR coefficients of noise $[\alpha_{w,0}^{mcra}, \alpha_{w,1}^{mcra}, ..., \alpha_{w,q}^{mcra}]$. Finally, the spectral shape of noise can be obtained by:

$$H_w^{mcra}(k) = \frac{1}{\left|A_w^{mcra}(k)\right|^2} = \frac{1}{\left|\sum_{m=0}^{q} \alpha_{w,m}^{mcra} \exp(-\frac{2\pi mk}{K})\right|^2} \tag{3}$$

2.2. The estimation of AR gains

Firstly, we provide a brief overview of the traditional method for the estimation of speech and noise AR gains.

For each pair of code-words $\{\boldsymbol{\theta}_x^i, \boldsymbol{\theta}_w^j\}$, where the *ith* speech code-word $\boldsymbol{\theta}_x^i = [\alpha_{x,0}^i, ..., \alpha_{x,p}^i]$ and *jth* noise code-word $\boldsymbol{\theta}_w^j = [\alpha_{w,0}^j, ..., \alpha_{w,q}^j]$, the modeled noisy spectrum is written as:

$$\hat{P}_y^{ij}(k) = \frac{\hat{g}_x^{ij}}{\left|A_x^i(k)\right|^2} + \frac{\hat{g}_w^{ij}}{\left|A_w^j(k)\right|^2} \tag{4}$$

where $A_x^i(k)$ and $A_w^j(k)$ are given by

$$\begin{aligned} A_x^i(k) &= \sum_{m=0}^{p} \alpha_{x,m}^i \exp(-\frac{2\pi m}{K}k) \\ A_w^j(k) &= \sum_{m=0}^{q} \alpha_{\mathrm{w},m}^j \exp(-\frac{2\pi m}{K}k) \end{aligned} \tag{5}$$

The $\hat{g}_x^{ij}$ and $\hat{g}_w^{ij}$ in Eq.4 are estimated by minimizing the following IS distortion between the modeled and observed noisy spectra [4].

$$d_{IS}(\boldsymbol{P}_y, \hat{\boldsymbol{P}}_y^{ij}) = \sum_{k=0}^{K-1}\left(\frac{P_y(k)}{\hat{P}_y^{ij}(k)} - \ln\left(\frac{P_y(k)}{\hat{P}_y^{ij}(k)}\right) - 1\right) \tag{6}$$

where $\boldsymbol{P}_y = [P_y(0), ..., P_y(K-1)]^T$ and $\hat{\boldsymbol{P}}_y^{ij} = [\hat{P}_y^{ij}(0), ..., \hat{P}_y^{ij}(K-1)]^T$. Using a series expansion for $\ln(x)$, the final AR gains is estimated by minimizing the following LS distortion [4].

$$d_{LS}(\boldsymbol{P}_y, \hat{\boldsymbol{P}}_y^{ij}) = \sum_{k=0}^{K-1}\left(\frac{\hat{g}_x^{ij}\Big/\left|A_x^i(k)\right|^2 + \hat{g}_w^{ij}\Big/\left|A_w^j(k)\right|^2}{P_y(k)} - 1\right)^2 \tag{7}$$

By differentiating Eq.7 with respect to $\hat{g}_x^{ij}$ and $\hat{g}_w^{ij}$ and setting the results to zero, the AR gains are determined by:

$$\boldsymbol{C}[\hat{g}_x^{ij} \quad \hat{g}_w^{ij}]^T = \boldsymbol{D} \tag{8}$$

where the matrices $\boldsymbol{C}$ and $\boldsymbol{D}$ are given in [4]. Obviously, the estimated AR gains are the closed-form solution of LS distortion but not the IS distortion.

In this paper, we use a multiplicative update rule [7][8] to obtain approximately closed-form solution of IS distortion. Since we train the shape codebook of speech spectrum offline and only the spectral shape of noise is estimated online, for each speech code-word $\boldsymbol{\theta}_x^i$, we can rewrite the modeled noisy spectrum as follows:

$$\hat{P}_y^i(k) = \frac{\hat{g}_x^i}{\left|A_x^i(k)\right|^2} + \frac{\hat{g}_w^i}{\left|A_w^{mcra}(k)\right|^2} \tag{9}$$

By expressing the Eq.9 in matrix form, we can get:

$$\begin{aligned} \hat{\boldsymbol{P}}_y^i &= \boldsymbol{H}_x^i \boldsymbol{W}_x^i + \boldsymbol{H}_w^{mcra} \boldsymbol{W}_w^i \\ &= \begin{bmatrix} \boldsymbol{H}_x^i & \boldsymbol{H}_w^{mcra} \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_x^i \\ \boldsymbol{W}_w^i \end{bmatrix} \\ &= \boldsymbol{H}_y^i \boldsymbol{W}_y^i \end{aligned} \tag{10}$$

with

$$\begin{aligned} \hat{\boldsymbol{P}}_y^i &= [\hat{P}_y^i(0), ..., \hat{P}_y^i(K-1)]^T \\ \boldsymbol{H}_x^i &= [\frac{1}{|A_x^i(0)|^2}, ..., \frac{1}{|A_x^i(K-1)|^2}]^T \\ \boldsymbol{H}_w^{mcra} &= [\frac{1}{|A_w^{mcra}(0)|^2}, ..., \frac{1}{|A_w^{mcra}(K-1)|^2}]^T \\ \boldsymbol{W}_w^i &= [\hat{g}_w^i] \ , \ \boldsymbol{W}_x^i = [\hat{g}_x^i] \end{aligned} \tag{11}$$

The $\boldsymbol{W}_x^i$ and $\boldsymbol{W}_w^i$ are named as the AR gain matrices of speech and noise, respectively. The IS distortion is rewritten as:

$$d_{IS}(\boldsymbol{P}_y, \hat{\boldsymbol{P}}_y^i) = \sum_{k=0}^{K-1}\left(\frac{P_y(k)}{\hat{P}_y^i(k)} - \ln\left(\frac{P_y(k)}{\hat{P}_y^i(k)}\right) - 1\right) \tag{12}$$

By differentiating Eq.12 with respect to gain matrices, we have [8]:

$$\begin{aligned} \frac{\partial d_{IS}(\boldsymbol{P}_y, \hat{\boldsymbol{P}}_y^i)}{\partial \boldsymbol{W}_x^i} &= (\boldsymbol{H}_x^i)^T[(\boldsymbol{H}_y^i\boldsymbol{W}_y^i)^{-2} \bullet (\boldsymbol{H}_y^i\boldsymbol{W}_y^i - \boldsymbol{P}_y)] \\ \frac{\partial d_{IS}(\boldsymbol{P}_y, \hat{\boldsymbol{P}}_y^i)}{\partial \boldsymbol{W}_w^i} &= (\boldsymbol{H}_w^{mcra})^T[(\boldsymbol{H}_y^i\boldsymbol{W}_y^i)^{-2} \bullet (\boldsymbol{H}_y^i\boldsymbol{W}_y^i - \boldsymbol{P}_y)] \end{aligned} \tag{13}$$

The symbol '$\bullet$' indicates the point-wise multiplication. By simplifying the above formula, we can get:

$$\frac{\partial d_{IS}(\boldsymbol{P}_y, \hat{\boldsymbol{P}}_y^i)}{\partial \boldsymbol{W}_x^i} = (\boldsymbol{H}_x^i)^T (\boldsymbol{H}_y^i \boldsymbol{W}_y^i)^{-1} - (\boldsymbol{H}_x^i)^T [(\boldsymbol{H}_y^i \boldsymbol{W}_y^i)^{-2} \bullet \boldsymbol{P}_y]$$
$$\frac{\partial d_{IS}(\boldsymbol{P}_y, \hat{\boldsymbol{P}}_y^i)}{\partial \boldsymbol{W}_w^i} = (\boldsymbol{H}_w^{mcra})^T (\boldsymbol{H}_y^i \boldsymbol{W}_y^i)^{-1} - (\boldsymbol{H}_w^{mcra})^T [(\boldsymbol{H}_y^i \boldsymbol{W}_y^i)^{-2} \bullet \boldsymbol{P}_y] \quad (14)$$

The $\boldsymbol{W}_x^i$ and $\boldsymbol{W}_w^i$ are obtained by iterating the following multiplicative rules to minimize Eq.12:

$$\boldsymbol{W}_x^i \leftarrow \boldsymbol{W}_x^i \bullet \frac{(\boldsymbol{H}_x^i)^T [(\boldsymbol{H}_y^i \boldsymbol{W}_y^i)^{-2} \bullet \boldsymbol{P}_y]}{(\boldsymbol{H}_x^i)^T (\boldsymbol{H}_y^i \boldsymbol{W}_y^i)^{-1}}$$
$$\boldsymbol{W}_w^i \leftarrow \boldsymbol{W}_w^i \bullet \frac{(\boldsymbol{H}_w^{mcra})^T [(\boldsymbol{H}_y^i \boldsymbol{W}_y^i)^{-2} \bullet \boldsymbol{P}_y]}{(\boldsymbol{H}_w^{mcra})^T (\boldsymbol{H}_y^i \boldsymbol{W}_y^i)^{-1}} \quad (15)$$

Substituting $\boldsymbol{W}_x^i = [\hat{g}_x^i]$ and $\boldsymbol{W}_w^i = [\hat{g}_w^i]$ into Eq.15, we have

$$\hat{g}_x^i \leftarrow \hat{g}_x^i \bullet \frac{(\boldsymbol{H}_x^i)^T [(\boldsymbol{H}_y^i \boldsymbol{W}_y^i)^{-2} \bullet \boldsymbol{P}_y]}{(\boldsymbol{H}_x^i)^T (\boldsymbol{H}_y^i \boldsymbol{W}_y^i)^{-1}}$$
$$\hat{g}_w^i \leftarrow Max\left\{ g_w^{mcra} \;,\; \hat{g}_w^i \bullet \frac{(\boldsymbol{H}_w^{mcra})^T [(\boldsymbol{H}_y^i \boldsymbol{W}_y^i)^{-2} \bullet \boldsymbol{P}_y]}{(\boldsymbol{H}_w^{mcra})^T (\boldsymbol{H}_y^i \boldsymbol{W}_y^i)^{-1}} \right\} \quad (16)$$

where g_w^{mcra} is the noise AR gain obtained by the MCRA algorithm in section **2.1**. The '*Max*' is used to avoid the underestimate of noise AR gain.

An example of average IS distortion is illustrated in Fig.1. The average IS distortion $\bar{D}_{IS} = \sum_{i=1}^{N_x} d_{IS}(\boldsymbol{P}_y, \hat{\boldsymbol{P}}_y^i) \big/ N_x$. The N_x is the size of speech codebook. The AR gains are estimated by the conventional and proposed methods, respectively. The speech material is corrupted by white noise with the SNR of 5dB.

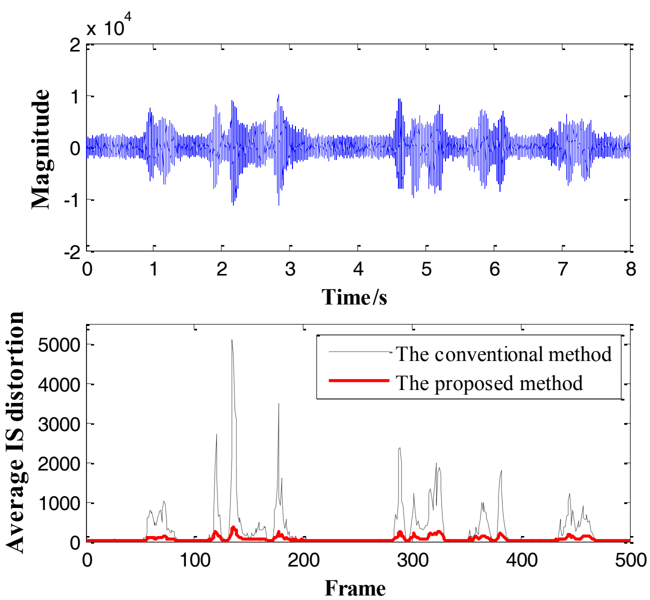

Fig.1 The average IS distortion comparison

From the Fig.1, we can find that the average IS distortion of the proposed method is obviously lower than that of conventional method, especially in the voiced segments. It means that the proposed method can obtain more accurate estimation of speech and noise AR gains.

Let $\boldsymbol{\theta}_x$ denote the random variable corresponding to the speech AR coefficients. And let g_x and g_w denote the random variables corresponding to the speech and noise AR gains, respectively. Let $\boldsymbol{\theta}$=[$\boldsymbol{\theta}_x$, g_x, g_w] denote the set of random variables. After getting each $\boldsymbol{\theta}^i = [\boldsymbol{\theta}_x^i, \hat{g}_x^i, \hat{g}_w^i]$, the desired Bayesian MMSE estimate $\hat{\boldsymbol{\theta}}$={$\hat{\boldsymbol{\theta}}_x$, $\hat{g}_x$, $\hat{g}_w$} can be written as follows

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= E(\boldsymbol{\theta} \mid \mathbf{y}) \\ &= \int \boldsymbol{\theta} \frac{p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{y})} d\boldsymbol{\theta} \\ &= \frac{1}{N_x} \sum_{i=1}^{N_x} \boldsymbol{\theta}^i \frac{p(\mathbf{y} \mid \boldsymbol{\theta}^i) p(\hat{g}_x^i) p(\hat{g}_w^i)}{p(\mathbf{y})}\end{aligned} \quad (17)$$

with $p(\mathbf{y}) = \frac{1}{N_x} \sum_{i=1}^{N_x} p(\mathbf{y} \mid \boldsymbol{\theta}^i) p(\hat{g}_x^i) p(\hat{g}_w^i)$ and $p(\mathbf{y} \mid \boldsymbol{\theta}^i) = \mathrm{C}\exp\left(-d_{IS}(\boldsymbol{P}_y, \hat{\boldsymbol{P}}_y^i)\right)$

where C is a constant.

The estimate $\hat{\boldsymbol{\theta}}$ is used to construct the following Wiener filter.

$$WF_0(k) = \frac{\hat{g}_x}{\left|\hat{A}_x(k)\right|^2} \Bigg/ \left(\frac{\hat{g}_x}{\left|\hat{A}_x(k)\right|^2} + \frac{\hat{g}_w}{\left|A_w^{mcra}(k)\right|^2} \right) \quad (18)$$

where $1 \big/ \left|\hat{A}_x(k)\right|^2$ denotes the estimate of spectral shape of speech.

2.3. Modified codebook-driven Wiener filter

Conventional codebook-driven Wiener filter is constructed by the estimated spectral envelopes of speech and noise, which usually causes an inaccurate fitting for the spectra between the harmonics of speech. Consequently, the residual noise still remains between the harmonics of the enhanced speech. In this section, we introduce the SPP to modify the codebook-driven Wiener filter Eq.18 for suppressing the residual noise.

We use $H_1(k)$ to denote the event that speech is present in frequency bin k. And let $p(H_1(k))$ denotes the priori SPP in frequency bin k. Then we can estimate the posteriori SPP in frequency bin k by [11]:

$$p\left(H_1(k) \mid \mathbf{y}\right) = \frac{1}{1 + (1/\,p(H_1(k)) - 1)(1 + \xi'(k)) \exp(-v'(k))} \quad (19)$$

where

$$\xi'(k) = \frac{\hat{g}_x \big/ \left|\hat{A}_x(k)\right|^2}{p(H_1(k)) P_w^{mcra}(k)}$$
$$v'(\mathrm{k}) = \frac{\xi'(k) P_y(k)}{(1 + \xi'(k)) P_w^{mcra}(k)} \quad (20)$$

Multiplying $WF_0(k)$ by the posteriori SPP, we have:

$$WF_1(k) = p\left(H_1(k) \mid \mathbf{y}\right) \frac{\hat{g}_x}{\left|\hat{A}_x(k)\right|^2} \Bigg/ \left(\frac{\hat{g}_x}{\left|\hat{A}_x(k)\right|^2} + \frac{\hat{g}_w}{\left|A_w^{mcra}(k)\right|^2} \right) \quad (21)$$

As known, the noisy signal only contains the noise in some frames. To remove these noise-only frames more efficiently, we construct the final Wiener filter as follows:

$$WF_2(k) = p\left(H_1(k)\,|\,\mathbf{y}\right)\frac{\rho \hat{g}_x}{\left|\hat{A}_x(k)\right|^2} \Bigg/ \left(\frac{\rho \hat{g}_x}{\left|\hat{A}_x(k)\right|^2} + \frac{(1-\rho)\hat{g}_w}{\left|A_w^{mcra}(k)\right|^2}\right) \quad (22)$$

where $\rho = \sum_{k=0}^{K-1} p_{mcra}(H_1(k)) \big/ K$, the $p_{mcra}(H_1(k))$ is the smoothed SPP in MCRA algorithm.

3. PERFORMANCE AND EVALUATIONS

In this section, we compare the performance of our method with that of two reference methods. The Ref. A [4] and Ref. B [6] denote the traditional ML and Bayesian MMSE codebook-driven methods, respectively. In our experiments, the test set which contains five male and four female speakers is selected from the NTT database and down-sampled to 8*kHz*. The speech is corrupted by white, babble, office, and street noise from the Noisex-92 database at 0dB, 5dB, and 10dB signal-to-noise ratio (SNR). The frame size is 32ms (256 samples), windowed using normalized Hamming window with 50% overlap between the adjacent frames. The FFT size is 512. The order of AR-models for speech and noise is 10. A 6 bit speech codebook is trained with one and half hours of clean speech. The performance tests are carried out under the evaluation measurements including the segment SNR (SSNR) [12], log-spectral distortion (LSD) [13], and the perceptual evaluation of speech quality (PESQ) [14]. Moreover, we give the spectrogram of enhanced speech.

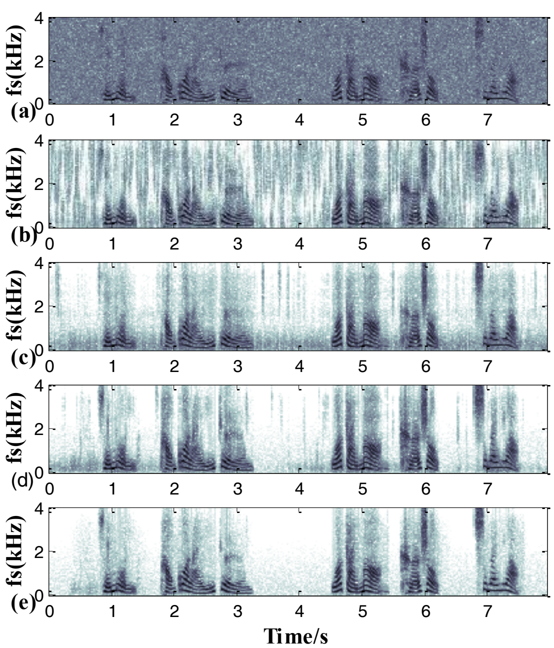

Fig.2. Power spectrum comparison

(a) Noisy speech (white noise, SNR=10dB), (b) enhanced speech using Ref. A, (c) enhanced speech using Ref. B, (d) enhanced speech using our method without SPP, (e) enhanced speech using our method with SPP

The Fig.2 shows the spectrogram of enhanced speech. By comparing the (d) with the (b) and (c) in Fig.2, we can see that the enhanced speech in (d) contains fewer residual noise and lower speech distortion. This is mainly because the proposed method has obtained more accurate estimation of AR gains. The comparison result between (e) and (d) shows that the residual noise between harmonics of noisy speech can be removed efficiently by the introduced SPP.

From the Table.1, we can find that the proposed method leads to the better perceived quality of speech than the reference methods in various input SNRs.

The test results in Table.2 and Table.3 show that our method can remove more noise than reference methods without producing extra speech distortion. Thus, it is further confirmed that the proposed method precedes the reference approaches.

Table.1. Test Results of PESQ

Enhancement methods	Average PESQ		
	0dB	5dB	10dB
Noisy speech	1.87	2.20	2.55
Ref. A	2.05	2.45	2.71
Ref. B	2.33	2.64	2.90
Proposed	2.40	2.76	3.06

Table.2. Test Results of SSNR

Enhancement methods	Average SSNR Improvement		
	0dB	5dB	10dB
Noisy speech	--	--	--
Ref. A	9.74	8.71	7.59
Ref. B	13.23	11.97	10.69
Proposed	16.42	15.46	14.16

Table.3. Test Results of LSD

Enhancement methods	Average LSD		
	0dB	5dB	10dB
Noisy speech	14.57	12.66	10.88
Ref. A	10.68	8.99	7.92
Ref. B	9.09	7.73	6.46
Proposed	7.41	6.11	5.15

4. CONCLUSIONS

In this paper, a novel speech and noise AR parameters estimation method is proposed for codebook-driven speech enhancement. The multiplicative update rule is applied to estimate the AR gains of speech and noise, which greatly increases the estimation accuracy. Using the MCRA algorithm to obtain the spectral shape of noise, the problem of noise classification is also solved. Furthermore, we introduce the SPP to combine with the codebook-driven Wiener filter for suppressing the residual noise between harmonics of noisy speech efficiently. The test results demonstrate that the proposed method outperforms the reference methods.

5. ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 61471014).

REFERENCES

[1] S. Boll, “Suppression of acoustic noise in speech using spectral subtraction”, *IEEE Transaction on Acoustic, Speech and Signal Processing*, vol. 27, no. 2, pp. 113-120, Apr. 1979.

[2] S. V. Vaseghi, Advanced Signal Processing and Digital Noise Reduction. New York: Wiley, 1998, ch.6.

[3] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error log-spectral amplitude estimator,” *IEEE Trans. Acoust. Speech Signal Process*., vol. 33, pp. 443-445, 1985.

[4] S. Srinivasan, J. Samuelsson, and W. B. Kleijn, “Codebook driven short-term predictor parameter estimation for speech enhancement”, *IEEE Trans. Audio, Speech, Lang. Process*., vol. 14, no. 1, pp. 163–176, Jan.2006.

[5] Feng Bao, Hui-jing Dou, Mao-shen Jia and Chang-chun Bao, Speech Enhancement Based on a Few Shapes of Speech Spectrum, *ChinaSIP 2014*, pp. 90-94.

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

[7] Daniel D. Lee and H. Sebastian Seung, “Algorithms for non-negative matrix factorization,” in *NIPS*, 2000, pp. 556–562.

[8] C. Févotte, N. Bertin, and J. L. Durrieu, “Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis,” *Neural Comput*., vol. 21, pp. 793–830, 2009.

[9] Cohen, I. “Noise estimation by minima controlled recursive averaging for robust speech enhancement”, *IEEE Signal Process. Lett.*, 9(1), 12-15. 2002.

[10] J. Markhoul. Linear Prediction:A Tutorial Review. *Prov. IEEE*, 1975, 63: 561 – 580

[11] P. C. Loizou, Speech enhancement: theory and practice. Boca Raton, FL, USA: CRC Press, 2007.

[12] S. R. Quackenbush, T. P. Barnwell, and M. A. Clements, “Objective measures of speech quality,” Englewood Cliffs, NJ: Prentice Hall, 1988.

[13] Abramson, A., Cohen, I., “Simultaneous Detection and Estimation Approach for Speech Enhancement”, *IEEE Trans. Speech Audio Process.*, 15(8), pp. 2348-2359, 2007.

[14] “Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-End Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs,” *ITU-T Recommendation*, P.862, Feb, 2001.