A Modified Speech Enhancement Algorithm Using A Universal Speaker Model

Li Guo, Wenbin Jiang, Rendong Ying and Peilin Liu
School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China
guoli0104@hotmail.com

***Abstract*—In this paper, we propose a statistical model-based speech enhancement algorithm using an improved minima controlled recursive averaging (IMCRA) noise estimation and a decision-directed (DD) priori SNR estimation. In the training stage, the Gaussian mixture model (GMM) of the Mel-frequency cepstral coefficients (MFCCs) of universal speaker is obtained. In speech enhancement stage, minima tracking process of IMCRA noise estimation is adjusted with the noisy power spectrum of current frame and an adjustment weighting factor. In addition, based on the universal GMM, some significant constant parameters are replaced by frequency-varying parameters, such as the weighting parameter in the DD priori SNR estimation and the adjustment weighting factor in the modified minima tracking process of IMCRA. The performance of proposed speech enhancement is evaluated by objective tests under various stationary and non-stationary noise environments. From experimental results, compared to the conventional approaches, the proposed scheme performs better and is suitable for being used as the pre-processing of speech processing systems.**

Keywords-GMMs, IMCRA, MFCCs, speech enhancement

I. INTRODUCTION

As a fundamental part of speech processing system, speech enhancement is a necessary pre-processing for background noise reduction. Various approaches for speech enhancement of noisy speech signals have been introduced and applied in recent 30 years. Many speech enhancement algorithms [1]-[10] operate in frequency domain and are usually composed of noise power estimation and speech estimation. The speech is estimated by multiplying the noisy speech magnitude spectrum by a gain function, which is on the basis of the estimated noise, the statistical model of speech and certain distortion criterion.

The noise power estimation is a crucial and difficult component in speech enhancement systems. As one of the efficient noise power estimation techniques, minimum statistics (MS) [1] obtains the noise power estimate by searching the minima values of a smoothed power estimate of the noisy signal within a finite window. To improve the accuracy of estimate during speech absence periods, MS is combined with the well-known soft decision scheme based on the speech absence probability (SAP) in [2]. However, only the smoothing of noisy power spectrum in time is taken into account in MS and there is a strong correlation of speech presence in neighboring frequency bins of the consecutive frames. Therefore, the improved minima controlled recursive averaging (IMCRA) technique [3] makes use of the smoothing in both time and frequency, which comprises two iterations of smoothing and minima tracking.

In most speech estimators, the estimated speech spectrum is obtained by multiplying the noisy spectrum by a gain function. The distribution of both the noisy speech and noise spectrum is assumed to follow one statistical model, including Gaussian and Laplacian. For weighting the spectral, several best known fidelity criteria are presented, such as the minimum mean-squared error (MMSE) [4], MMSE of the log-spectral amplitude (MMSE-LSA) [5][6] and the Wiener filter [7]. To further enhance these speech estimators, the gain function is then modified according to SAP [8].

Both the gain function and SAP are computed from a priori signal-to-noise ratio (SNR) and a posteriori SNR. The decision-directed (DD) approach [9] derived by Ephraim and Malah is widely used for estimating a priori SNR. A weighting factor, which is a fixed weight for the current short-time frame and the processing output of previous frames, is applied to control the tradeoff between noise reduction and transient distortion. In addition, there are some modifications, including an adaptive weighting factor determined by the deviation of the posteriori SNR [10] and a modified DD priori SNR estimator in [11] to reduce the degradation of speech quality caused by one-frame delay in traditional DD approach.

Moreover, some algorithms [12][13] improve the enhanced speech quality based on the off-line or on-line knowledge of noise type. Some parameters, such as the searching window size of MS [12], the weighting factor in DD approach [13], and the control parameter of minimum gain value [13], are variable according to the determined noise type.

In this paper, we propose a novel speech enhancement based on a universal GMM of speakers' MFCCs. The widely used IMCRA noise estimation and DD a priori SNR estimation are modified according to the feature of different frequency bins. For the IMCRA noise estimation, the noisy power spectrum of current frame is added to the tracked minima value of smoothed noisy power spectrum with adjustment weighting factors and a higher-bound limitation. The weighting factors of both this minimum noise power spectrum tracking and DD a priori SNR estimator are variable for different frequency bins according to the universal GMM.

The rest of this paper is organized as follows. In section II, we briefly introduce the IMCRA estimation and DD a priori SNR estimation. The training of universal GMM and the proposed modifications of IMCRA and DD are explained in section III. Section IV presents the experimental results, and finally, the conclusions are drawn in Section V.

978-1-4799-2186-7/14/$31.00 ©2014 IEEE

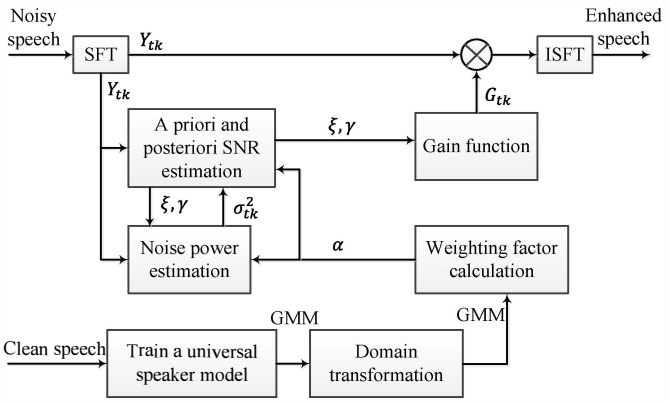


Figure 1. Overall block diagram of speech enhancement algorihtm based on a universal speaker model

II. REVIEW OF PREVIOUS SPEECH ENHANCEMENT

In this section, we briefly introduce the IMCRA noise estimation and decision-directed a priori SNR estimation. Considering only the additive noise, the observed noisy signal $y(n)$ is assumed as:

$$y(n)=x(n)+d(n) \tag{1}$$

where $x(n)$ and $d(n)$ denote clean speech and uncorrelated additive noise signals. After applying the short-time Fourier transform (SFT), the noisy signal is transformed into Y_{tk}.

A. IMCRA noise power estimation

There are two iterations of smoothing and minima tracking in IMCRA noise estimation. In the first iteration, a rough voice activity detection is provided. The noisy power spectrum is smoothed in frequency and time successively:

$$S^f_{tk}=\sum_{i=-w}^{w} b_i\left|Y_{t,k-i}\right|^2 \tag{2}$$

$$S_{tk}=\alpha_s S_{t-1,k}+(1-\alpha_s)S^f_{tk} \tag{3}$$

where b denotes a normalized window function of length $2w+1$ and α_s is a smoothing factor. The same as MS method, the minima of S_{tk} are picked within a finite window of length D.

$$S^{min}_{tk}\triangleq min\{S_{\tau k}|t-D+1\le\tau\le t\} \tag{4}$$

Based on above smoothing and minima tracking result, a rough decision about speech presence or absence is given by

$$I_{tk}=\begin{cases}1, & if\ \gamma^{min}_{tk}<\gamma_0\ and\ \zeta_{tk}<\zeta_0\ (speech\ absence)\\ 0, & if\ otherwise\quad (speech\ presence)\end{cases} \tag{5}$$

Based on the bias factor B_{min}, γ^{min}_{tk} and ζ_{tk} are defined by

$$\gamma^{min}_{tk}\triangleq\frac{|Y_{tk}|^2}{B_{min}S^{min}_{tk}},\quad \zeta_{tk}\triangleq\frac{S_{tk}}{B_{min}S^{min}_{tk}} \tag{6,7}$$

Then, in the second iteration, only the components identified as containing noise are smoothed as follows:

$$\tilde{S}^f_{tk}=\begin{cases}\dfrac{\sum_{i=-w}^{w} b_i I_{t,k-i}\left|Y_{t,k-i}\right|^2}{\sum_{i=-w}^{w} b_i I_{t,k-i}}, & if\ \sum_{i=-w}^{w} I_{t,k-i}\neq 0\\ \tilde{S}^f_{t-1,k}, & otherwise\end{cases} \tag{8}$$

After smoothing in frequency, smoothing in time and minima tracking are performed the same as the first iteration.

Under speech presence uncertainty, noise power is estimated with conditional speech presence probability.

$$\bar{\sigma}^2_{t+1,k}=\tilde{\alpha}_{tk}\bar{\sigma}^2_{t,k}+(1-\tilde{\alpha}_{tk})|Y_{tk}|^2 \tag{9}$$

$$\tilde{\alpha}_{tk}\triangleq\alpha_d+(1-\alpha_d)\tilde{p}_{tk} \tag{10}$$

where the smoothing factor $\tilde{\alpha}_{tk}$ is adjusted by the speech presence probability $\tilde{p}_{tk}$, which is estimated by the tracked minimum value of smoothed noise power spectrum $\tilde{S}^f_{tk}$.

B. Decision-directed a priori SNR estimation

The decision-directed approach, derived by Ephraim and Malah [9], provides a very useful estimation method for a priori SNR. When speech presence is assumed, the expression is simplified as:

$$\xi^{DD}_{tk}=max\left\{\alpha_\xi\frac{\left|\hat{X}_{t-1,k|H_1}\right|^2}{\hat{\sigma}^2_{t-1,k}}+(1-\alpha_\xi)C(\gamma_{tk}-1),\xi_{min}\right\} \tag{11}$$

where $C(x)=x$ if $x\ge 0$, and $C(x)=0$ otherwise. $\hat{X}_{t-1,k|H_1}$ is the estimated clean speech of previous frame and $\hat{\sigma}^2_{t-1,k}$ is the noise power spectrum of previous frame. The weighting factor α_ξ controls the trade-off between the noise reduction and transient signal distortion.

III. MODIFIED IMCRA NOISE ESTIMATION AND DD A PRIORI SNR ESTIMATION APPROACH

We propose modifications of IMCRA noise estimation and DD a priori SNR estimation. As described in section II, constant weighting parameters are utilized to control the tradeoff between noise reduction and signal distortion. Since the presence probability for different frequency bins are inconsistent, we can modify these parameters according to a universal model of speakers, which can also reduce the complexity increase caused by speaker recognition. In addition, the minima tracking process of IMCRA is also modified to get a more accurate estimated noise power spectrum. The overall process is shown in Figure 1, which composes of an offline training stage and a speech enhancement stage.

A. Featrue extraction and universal GMM training

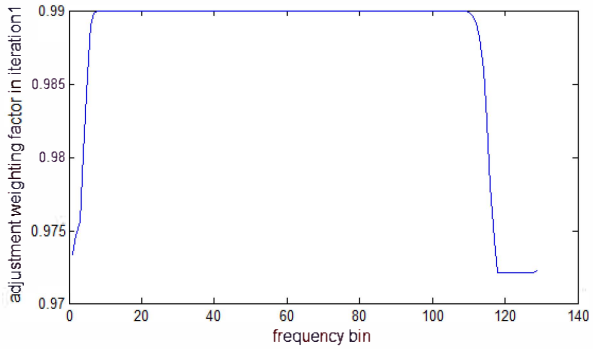
The training stage contains three steps, including speech feature extraction, universal GMM training, and domain transformation. To extract feature, the Mel-frequency cepstral coefficients (MFCCs) of clean speech are calculated, which are widely used for speech and speaker recognitions.

After that, a universal Gaussian mixture model of extracted MFCCs is trained. The GMM is described as:

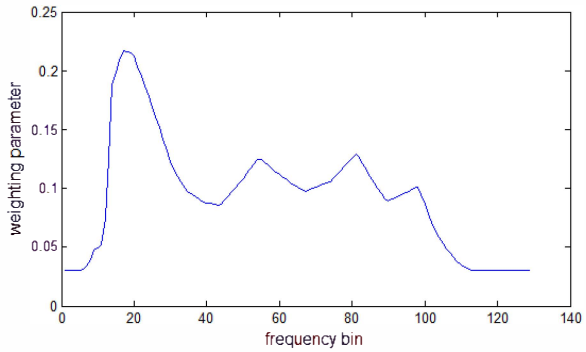
$$\lambda=\{\omega_i,\mu_i,\Sigma_i\},\quad 1\le i\le M \tag{12}$$

$$\mu^{mfcc} \xrightarrow{DCT^{-1}} \mu^{mfs} \xrightarrow{EXP} \mu^{mel} \xrightarrow{MEL^{-1}} \mu^{mag}$$

Figure 2. The operations needed to transform a Mel-frequency-cepstral domain vector to frequency domain



(a) adjustment weighting factor in first iteration of IMCRA



(b) weighting factor in DD a priori SNR estimation

Figure 3. Frequency-varying weighitng parameters: (a) the adjustment weighting factor in the first iteration of IMCRA noise estimation, and (b) the weighting parameter in DD a priori SNR estimation.

where M is the number of mixture, ω_i is mixture weight, μ_i is mean vector and Σ_i is covariance matrix.

The obtained universal GMM is in Mel-frequency domain, while most speech enhancement algorithms are processed in frequency domain. Therefore, the mean vector of GMM should be transformed to frequency domain according to the operating process shown in Figure 2. After domain transformation, the mixture weights ω_i and the mean vectors μ_i are combined as a mixture as below, which is utilized as feature of universal speakers in the following speech enhancement stage.

$$\mu^{sp} = \sum_{i=1}^{M} \omega_i \mu_i \tag{13}$$

B. Modified minima tracking process of IMCRA

IMCRA noise estimation contains two iterations of smoothing and minima tracking, as described in section II. In each iteration, after smoothing the power spectrum, the minima are tracked within a finite window length D. A larger window length is desirable to determine the minimum power in the case of stationary noises for stability [4], while a shorter window length is able to reduce the variance of the minima and shorten the delay when responding to a rising noise power [3]. Therefore, in the case of a fixed window length, in order to improve the tracking capability of noise estimator, the minima tracking process is adjusted with the noisy power spectrum of current frame and an adjustment weighting factor.

TABLE I. VALUES OF THE PARAMETERS OF ADJUSTMENT WEIGHTING FACTOR USED IN MODIFIED IMCRA, FOR A SAMPLING RATE OF 8 KHZ

Iterations		Parameters			
		a	*b*	*c*	*d*
Iteration 1	-	0.02	2	13	0.97
Iteration 2	HI_0^{tk}	0.02	2	13	0.96
	HI_1^{tk}	0.02	2	13	0.98

For the first iteration, the modified minima tracking process is shown in equation (14).

$$S_{tk}^{min} \triangleq \alpha_k^{Smin} \cdot min\{S_{tk} | t-D+1 \le \tau \le t\} + \left(1-\alpha_k^{Smin}\right) |Y_{tk}|^2 \tag{14}$$

where α_k^{Smin} is a frequency-varying adjustment weighting factor according to speakers' universal GMM.

In order to adjust the minima tracking process accurately, it is obvious that the adjustment weighting factor α_k^{Smin} should be larger for the case of speech presence, and be smaller on the contrary case. Thus, on the basis of this relationship, the adjustment weighting factor is defined as follows, using a flexible sigmoid-shape function.

$$\alpha_k^{Smin} = \frac{a}{exp\left(b-c\mu_k\right)+1} + d \tag{15}$$

$$\mu_k = \mu_k^{sp} \Big/ \left(\frac{1}{N}\sum_{i=1}^{N-1} \mu_i^{sp}\right) \tag{16}$$

where a, b, c and d are parameters that control the slope and the mean of a weighting factor. The parameter μ_k is modified mean vector of speakers' universal GMM. As an example, the adjustment weighting factor is shown in Figure 3(a).

In addition, a higher-bound constraint S_{max}^{min} is added to further reduce the excessive regulation of S_{tk}^{min}, as below.

$$S_{tk}^{min} = min\left\{S_{tk}^{min},\ S_{maxtk}^{min}\right\} \tag{17}$$

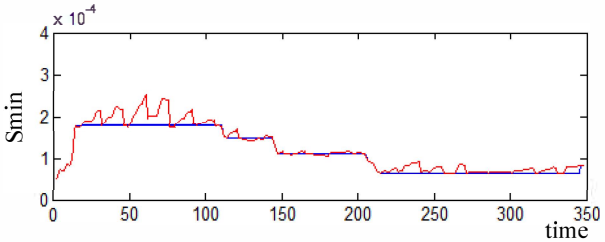
$$S_{maxtk}^{min} = T \cdot max\left\{S_{tk}^{min} \middle| t-D_M+1 \le \tau \le t\right\} \tag{18}$$

where T is a constant parameter of 1.1, and D_M is a finite search window length of 12.

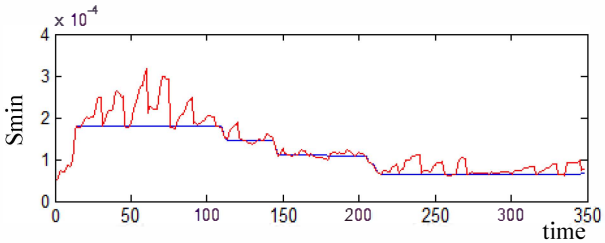
In the second iteration, hypotheses of smoothed signal absence is defined based on the rough decision I_{tk} of the first iteration. Otherwise it is the case of speech presence HI_1^{tk}.

$$HI_0^{tk}: \sum_{i=-w}^{w} I_{t,k-i} \neq 0 \tag{19}$$

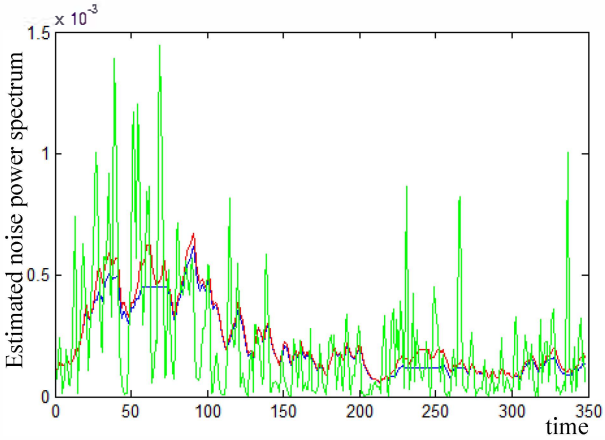
Considering the differences between speech presence and absence, the parameter d of the adjustment weighting factor α_k^{Smin} under HI_0^{tk} is smaller than the case of HI_1^{tk}. So the detailed parameters for a sampling rate of 8 kHz are presented in Table I. Higher-bound constraint is identical with the first iteration.



(a) Smin in the first iteration



(b) Smin in the second iteration



(c) Estimated noise power spectrum

Figure 4. Minima of smoothed noisy power spectrum and estimated noise power spectrum: (a)(b) the picked minima of noisy power using conventional (blue) and modified (red) IMCRA in the first and second iteration respectively, and (c) estimated by IMCRA (blue and red) and exact (green) noise power spectrum

The performance of conventional and modified IMCRA minima tracking processes is illustrated in the example of Figure 4. The noisy speech sequence is taken from NOIZEUS database with airport noise and SNR = 0dB. From the presented results, modified IMCRA noise estimator has better tracking capability, especially for a rising noise power case.

C. *Modified priori SNR estimation*

As a modification of decision-directed approach, a causal conditional estimator of priori SNR is presented in [14], which contains a "propagation" step and an "update" step. A special case of this causal recursive estimator degenerates to a DD estimator with a time-varying frequency-dependent weighting factor α_{tk} as follows:

$$\xi_{tk|t}=max\left\{\alpha_{tk}\frac{\left|\widehat{X}_{t-1,k|H_1}\right|^2}{\widehat{\sigma}^2_{t-1,k}}+(1-\alpha_{tk})C(\gamma_{tk}-1),\xi_{min}\right\} \quad (20)$$

$$\alpha_{tk}=1-\frac{\left|\widehat{X}_{t-1,k|H_1}\right|^4}{\left(\widehat{\sigma}^2_{t-1,k}+\left|\widehat{X}_{t-1,k|H_1}\right|^2\right)^2} \quad (21)$$

A larger weighting factor in DD estimator is to reduce musical noise phenomena during speech absence, while a smaller weighting factor is to reduce signal distortion during speech presence. So the above variable weighting factor helps reducing the level of musical noise without much audible distortion. But for some transients, the performance is not so ideal. Therefore, we consider a more generally suitable case based on the universal speaker's feature in this paper.

The estimated noise power spectrum of previous frame $\widehat{\sigma}^2_{t-1,k}$ in (25) is replaced by a related constant. Instead of the estimated output spectrum $\widehat{X}_{t-1,k|H_1}$, frequency-domain mean vector of speakers' universal GMM is utilized. So time-varying frequency-dependent weighting factor α_{tk} is simplified to a frequency-dependent weighting factor as below:

$$\alpha_k=1-b-\frac{\mu_k^{sp4}}{\left(a\left(\sum_{i=1}^{N}\mu_k^{sp2}/N\right)+\mu_k^{sp2}\right)^2} \quad (22)$$

where a is the constant estimate of noise power spectrum and b is a constant offset to weighting factor. For a sampling rate of 8 kHz case, a is set to 3.2 and b is set to 0.03. The frequency-varying weighting parameter is shown in Figure 4(b).

IV. EXPERIMENTAL RESULTS

The performance of speech enhancement based on the proposed modification of IMCRA noise estimation and DD a priori SNR estimation is demonstrated in this section. To evaluate the performance, the widely used optimal modified minimum mean-square error log-spectral amplitude (OMLSA) in [6] is chosen as speech estimation, which combines the MMSE-LSA gain function with the probability of speech presence.

The test speeches are taken from TIMIT and NOIZEUS database. Clean speeches used in this experiment are taken from TIMIT database, including 258 male and female speakers from 4 dialect regions. The clean speech signal is sampled at 8 kHz and degraded by seven stationary or non-stationary noises at SNRs of -5, 0, 5 and 10 dB, including airport, babble, car, exhibition, station, street and train from the NOIZEUS database. For each speaker, only one clean speech is used for training the speakers' universal GMM and other two noisy speeches are for testing. In the training stage, the number of components densities M in GMM is set to 16. The number of Mel-filters and Mel-cepstral coefficients is 24.

The performance is evaluated by three objective quality measures of the global SNR in dB, the perceptual evaluation of speech quality (PESQ) in the ITU-T P.862 and a composite measure for the overall quality, expressed by C_{ovl} [15].

$$C_{ovl}=1.594+0.805\cdot S_{PESQ}-0.512\cdot S_{LLR}-0.007\cdot S_{WSS} \quad (23)$$

where S_{PESQ}, S_{LLR} and S_{WSS} respectively represent PESQ, the log-likelihood ratio (LLR), and the weighted- slope spectral (WSS) distance, which are defined in [15].

TABLE II. PESQ RESULTS OBTAINED FROM THE CONVENTIONAL AND MODIFIED METHODS. (IM-OM IS SHORT FOR IMCRA-OMLSA AND MODIFIED IS THE PROPOSED MODIFICATION OF IMCRA AND DD)

Noise type	Method	Input SNR			
		0 dB	5 dB	10 dB	15 dB
Airport	IM-OM	1.91	2.29	2.65	3.03
	Modified	1.95	2.31	2.66	3.04
Babble	IM-OM	1.85	2.25	2.61	2.95
	Modified	1.88	2.26	2.61	2.96
Car	IM-OM	1.98	2.40	2.74	3.07
	Modified	2.03	2.44	2.78	3.11
Exhibition	IM-OM	1.73	2.20	2.56	2.92
	Modified	1.78	2.23	2.58	2.94
Station	IM-OM	1.90	2.40	2.69	3.03
	Modified	1.95	2.43	2.72	3.05
Street	IM-OM	1.83	2.23	2.58	2.92
	Modified	1.89	2.27	2.61	2.95
Train	IM-OM	1.75	2.17	2.53	2.89
	Modified	1.78	2.20	2.56	2.91

TABLE III. GLOBAL SNR RESULTS COMPARISON.

Noise type	Method	Input SNR			
		0 dB	5 dB	10 dB	15 dB
Airport	IM-OM	3.66	7.92	12.37	16.94
	Modified	3.69	7.94	12.47	17.17
Babble	IM-OM	3.86	8.15	12.49	16.79
	Modified	3.70	8.07	12.52	16.91
Car	IM-OM	5.80	9.57	13.32	17.41
	Modified	6.15	9.95	13.78	17.87
Exhibition	IM-OM	4.45	8.52	12.65	17.02
	Modified	4.50	8.60	12.82	17.29
Station	IM-OM	4.42	8.75	12.81	16.85
	Modified	4.65	8.95	13.06	17.16
Street	IM-OM	4.38	8.52	12.43	16.82
	Modified	4.56	8.74	12.68	17.08
Train	IM-OM	4.62	9.17	13.25	17.46
	Modified	4.67	9.26	13.44	17.65

TABLE IV. COMPOSITE MEASURE (C_{OVL}) RESULTS COMPARISON.

Noise type	Method	Input SNR			
		0 dB	5 dB	10 dB	15 dB
Airport	IM-OM	2.01	2.53	3.06	3.54
	Modified	2.03	2.53	3.07	3.55
Babble	IM-OM	1.87	2.46	2.99	3.44
	Modified	1.87	2.45	2.97	3.44
Car	IM-OM	2.20	2.75	3.18	3.60
	Modified	2.22	2.77	3.21	3.63
Exhibition	IM-OM	1.83	2.43	2.91	3.35
	Modified	1.86	2.45	2.91	3.36
Station	IM-OM	2.04	2.73	3.14	3.55
	Modified	2.07	2.73	3.15	3.57
Street	IM-OM	1.94	2.51	2.96	3.39
	Modified	1.97	2.53	2.98	3.41
Train	IM-OM	1.90	2.46	2.92	3.36
	Modified	1.92	2.48	2.94	3.37

The objective quality measures results of the conventional and proposed modified approaches are evaluated and averaged for each SNR. Compared with previous IMCRA and DD approach, the results of PESQ are presented in TABLE II. TABLE III and TABLE IV show the global SNR and the composite measure C_{ovl} results respectively. As seen in the tables of objective quality results, better PESQ, global SNR and composite measure C_{ovl} are achieved by the proposed modification of IMCRA noise estimation and DD a priori SNR estimation. For most tested noise environments, the performance improvement is consistent. From a viewpoint of average, the proposed approach improves the PESQ by more than 1.17%, which is achieved by more accurate noise power estimation and a priori SNR estimation.

TABLE V. PESQ RESULTS OF LOW-BIT SPEECH CODING SYSTEM. (USING A PRE-PROCESSING OF ORIGINAL NOISE PRE-PROCESSING IN MELPE, IMCRA-OMLSA, AND THE PROPOSED MODIFIED METHOD)

Noise type	Method	Input SNR			
		0 dB	5 dB	10 dB	15 dB
Airport	Melpe_npp	1.75	2.14	2.47	2.75
	IM-OM	1.78	2.15	2.46	2.75
	Modified	1.8	2.15	2.47	2.77
Babble	Melpe_npp	1.76	2.16	2.49	2.74
	IM-OM	1.73	2.14	2.49	2.73
	Modified	1.74	2.15	2.48	2.74
Car	Melpe_npp	1.88	2.26	2.55	2.77
	IM-OM	1.91	2.28	2.54	2.76
	Modified	1.93	2.31	2.57	2.79
Exhibition	Melpe_npp	1.68	2.11	2.42	2.67
	IM-OM	1.69	2.11	2.41	2.67
	Modified	1.73	2.14	2.44	2.7
Station	Melpe_npp	1.79	2.24	2.51	2.74
	IM-OM	1.79	2.26	2.51	2.74
	Modified	1.82	2.27	2.52	2.76
Street	Melpe_npp	1.73	2.13	2.44	2.68
	IM-OM	1.76	2.13	2.43	2.67
	Modified	1.8	2.17	2.45	2.7
Train	Melpe_npp	1.7	2.08	2.38	2.65
	IM-OM	1.66	2.07	2.36	2.65
	Modified	1.68	2.08	2.38	2.67

As a pre-processing unit, speech enhancement algorithm is widely used in speech processing system to reduce background noise. Although some speech enhancement methods are able to achieve higher PESQ scores, they may be not suitable for being used as a pre-processing unit. Therefore, the proposed modification is utilized in low-bit speech coding system Melpe [16]. PESQ scores of Melpe decoded speeches are evaluated and presented in TABLE V. Compared to the noise pre-processing used in Melpe codec and IMCRA-OMLSA, better performance of speech coding system is achieved by applying the modified speech enhancement. So the proposed method is effective and suitable for being a pre-processing.

V. CONCLUSIONS

This paper proposes a modified speech enhancement algorithm based on well-known IMCRA noise estimation and DD a priori SNR estimation. A universal GMM of speakers' MFCCs is trained first, and then transformed to frequency domain. In speech enhancement step, minima tracking process of IMCRA noise estimation is adjusted by the current frame's noisy power spectrum and adjustment weighting factors. According to the speaker universal GMM, the frequency-varying weighting parameters in DD priori SNR estimation and modified IMCRA are utilized instead of constant parameters. Compared to the conventional IMCRA and DD approach, the proposed modified method provides better performance of noise power estimation and a priori SNR estimation. From the results of objective quality tests, the proposed modification is able to improve

the enhanced speech quality and the performance of speech processing system as a pre-processing unit.

ACKNOWLEDGMENT

This paper was supported by National Natural Science Founding of China (NSFC) (61171171).

REFERENCES

[1] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Processing*, vol. 9, no. 5, pp. 504-512, 2001.

[2] J.-H. Chang and N.S. Kim, "Speech enhancement: new approaches to soft decision," *IEICE Trans. Inform. Systems*, vol. E84-D, no. 9, pp. 1231-1240, 2001.

[3] I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," *IEEE Trans. on speech and audio processing*, vol. 11, no.5, pp. 466-474, Sep. 2003.

[4] J.H. Chang, "Noisy speech enhancement based on improved minimum statistics incorporating acoustic environment- awareness," *Digital Signal Processing*, vol. 23, no. 4, pp. 1233-1238, Feb. 2013.

[5] J.-H. Choi and J.-H. Chang, "On using acoustic environment classification for statistical model-based speech enhancement," *Speech Communication*, vol. 54, no. 3, pp. 477-490, 2012.

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

[7] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, no.2, pp. 443-445, Apr. 1985.

[8] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal processing*, 81, pp. 2403-2418, 2001.

[9] P. Scalart and J. Filho, "Speech enhancement based on a priori signal to noise estimation," in *Proc. ICASSP*, pp. 629-632, 1996.

[10] I. Cohen, "Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator," *IEEE Signal Processing Letters*, vol. 9, no. 4, pp. 113-116, Apr. 2002.

[11] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.32, no.6, pp. 1109-1121, Dec. 1984.

[12] Y.S. Park and J.-H. Chang, "A novel approach to a robust a priori SNR estimator in speech enhancement," *IEICE Trans. Communication*, vol. E90-B, no. 8, pp. 2182-2185, 2007.

[13] P.C. Yong, S. Nordholm, and H.H. Dam, "Optimization and evaluation of sigmoid function with a priori SNR estimate for real-time speech enhancement," *Speech Communication*, vol.55, pp. 358-376, 2013.

[14] I. Cohen, "Relaxed statistical model for speech enhancement and a priori SNR estimation," *IEEE Trans. on Speech and Audio Processing*, vol. 13, no. 5, pp. 870-881, Sep. 2005.

[15] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 16, no. 1, pp. 229-238, Jan. 2008.

[16] A. V. McCree and T. P. Barnwell, "A mixed excitation LPC vocoder model for low bit rate speech coding," *IEEE Trans. on Speech and Audio Processing*, vol. 3, no. 4, pp. 242-250, Jul. 1995