

A New Improved Algorithm of Speech Enhancement Based on MCRA and Noncausal a Priori SNR Estimator

Na Ying

Institute of Communication Engineering
Hangzhou Dianzi University
Hangzhou, China
e-mail: name@xyz.com

Jianwei Liu

Institute of Communication Engineering
Hangzhou Dianzi University
Hangzhou, China
e-mail: name@xyz.com

Xuzhen Wang

Institute of Communication Engineering
Hangzhou Dianzi University
Hangzhou, China
e-mail: name@xyz.com

Qinfang Wang

Institute of Communication Engineering
Hangzhou Dianzi University
Hangzhou, China
e-mail: 1028504704@qq.

Jianzhi Hua

Institute of Communication Engineering
Hangzhou Dianzi University
Hangzhou, China
e-mail: jianzh@hdu.edu.cn279481626

Qingbiao Yang

Institute of Communication Engineering
Hangzhou Dianzi University
Hangzhou, China
e-mail: 279481626@qq.com

Abstract—In view of the Minimum Mean Square Error of Log-Spectral Amplitude estimator (MMSE-LSA) noise power spectrum estimation algorithm does not adapt to the actual non-stationary noise and the accuracy of the estimation of prior SNR is not high on the traditional logarithmic domain. However, most of the nonstationary noises are encountered in real world. This paper puts forward a new improved speech enhancement algorithm of MMSE-LSA. First, it uses Minima Controlled Recursive Averaging(MCRA) to estimate noise which can track non-stationary effectively. Then combined with non-causal prior SNR estimation, it has higher accuracy of estimation of actual speech power spectrum, eliminating noise further. The simulation results show that under the non-stationary noise, the effect of restrain background noise and music noise residual has large improvement by the improved algorithm, the signal-to-noise ratio and the voice quality have been improved. In particular it is clear that the proposed algorithm has an improved capability to retain low amplitude voiced speech components in low SNR conditions.

Keywords—Speech enhancement; noncausal a prior SNR; BR-MCRA; MMSE-LSA; SNR

I. INTRODUCTION

With the rapid development of communication technology, the function of speech enhancement in speech signal processing is more and more important. However, there is a sharp drop in the quality of speech when the noise level is higher in real environment. The easiest way

to alleviate the speech quality degradation is to employ a speech enhancement technique in which the background noise is reduced before the noisy speech is sent into various processing systems.

The minimum mean square error of log-spectral amplitude estimator has been proposed in [1], which can suppress background noise effectively. It shows estimation of the spectral speech power can be done by employing the decision-directed approach, the paper [2] for detailed analyses. However, the enhanced speech obtained by the algorithms is accompanied by the unpleasant musical noise artifact which is characterized by tones with random frequencies. In order to reduce the musical noise artifact, several solutions have been proposed.

In this paper, we present a new improved MMSE-LSA algorithm that combine BR-MCRA algorithm with noncausal a prior SNR algorithm based approaches [3] and [4]. The noncausal estimator employs future spectral measurements to predict the spectral variances of the clean speech better.

II. SPEECH SPECTRAL MODEL

Let $x(t)$ and $d(t)$ denote speech and uncorrelated additive noise signal, the observed signal is

$$y(t) = x(t) + d(t) \quad (1)$$

Applying the Short-Time Fourier Transform (STFT) to the observed signal, we have in the time-frequency

domain signal $D(k, l)$, $X(k, l) = A(k, l) \exp[j \nu(k, l)]$, $Y(k, l) = R(k, l) \exp[j \theta(k, l)]$, which is the l th frequency spectrum component of the first frame noise signals, clean speech signal, noise speech signal respectively. $A(k, l)$, $R(k, l)$, $\nu(k, l)$, $\theta(k, l)$ are their corresponding amplitude and phase.

We assume that clean speech signal and noise signal are zero-mean random variables, and denote by $\lambda_x(k, l) = \{E[X(k, l)]^2\}$, $\lambda_d(k, l) = E\{|D(k, l)|^2\}$, the speech and noise spectral variances, respectively.

Because human is sensitive to the amplitude of speech signal primarily, thus it is simply to obtain the valuation issue:

$$\hat{A} = E\left\{\left(\log A(k, l) - \log \hat{A}(k, l)\right)^2\right\} \quad (2)$$

The estimation of \hat{A} can be rewritten:

$$\hat{A}_k = \exp\left\{E\left[\ln A_k | Y_K\right]\right\} \quad (3)$$

The paper [1] proposed:

$$\hat{A}_k = \frac{\xi_k}{1 + \xi_k} \exp\left\{\frac{1}{2} \int_{\nu_k}^{\infty} \frac{e^{-t}}{t} dt\right\} R_k \quad (4)$$

Where $\nu(k, l) = \frac{\xi(k, l)}{1 + \xi(k, l)} \gamma(k, l)$, $\xi(k, l)$ and $\gamma(k, l)$

denote the prior SNR and posteriori SNR respectively.

$$\xi_k = \frac{\lambda_x(k)}{\lambda_d(k)} \quad (5)$$

$$\gamma_k = \frac{R_k^2}{\lambda_d(k)} \quad (6)$$

Based on spectrum estimation, the estimated amplitude can be expressed in terms of a gain function,

$$G(\xi(k, l), \gamma(k, l)) = \frac{\hat{A}(k, l)}{R(k, l)} = \frac{\xi(k, l)}{1 + \xi(k, l)} \exp\left\{\frac{1}{2} \int_{\nu(k, l)}^{\infty} \frac{e^{-t}}{t} dt\right\} \quad (7)$$

The decision-directed a priori SNR estimate is proposed by Ephraim and Malah in [1], which is given by

$$\xi_k(l) = \alpha \frac{\hat{A}_k^2(l-1)}{\lambda_d(k, l-1)} + (1 - \alpha) \max[\gamma_k(l) - 1, 0] \quad (8)$$

Where l denotes the current frame, $l-1$ denotes the previous frame. α denotes the smooth coefficient.

III. THE PROPOSED ALGORITHM

Speech enhancement algorithm in the frequency domain, there are three main factors presents [5]: the selection of the gain function, the accuracy of the noise power spectrum and a prior SNR. Therefore, based on the statistical model, this paper presents a new improved MMSE-LSA algorithm, estimating and updating the noise power spectrum through BR-MCRA algorithm, then deriving a recursive estimator for the a priori SNR.

A. Noise Estimation

Noise contamination is inevitable which severely degrade the performance of communication system[6]. The traditional SNR based on voice activity detector (VAD) is difficult to tune and its application in low SNR speech signal often shows in clipped speech[7].

In this paper, we chose the optimized noise estimate algorithm with magnitude compensation. The modified noise estimate algorithm is not only more accurate, but also can achieve less speech distortion and residual noise.

By feeding these continuous statistics estimates to noise reduction filters, non-stationary noise can be reduced without causing significant speech distortion[9].

Thus, according to signal power spectrum based on IMCRA algorithm of minimum bidirectional search algorithm proposed by Yu-min Zeng and Peng Wang in [10], we present a new improved algorithm.

Firstly, let $S_{\min 1}(k, l)$, $S_{\min 2}(k, l)$, $S_{\min}(k, l)$ denote the minima values of the forward search frame, the backward search frame and the eventually determine speech frame. Then calculate the final minimum frame of determine speech.

$$S_{\min}(k, l) = \max\{S_{\min 1}(k, l), S_{\min 2}(k, l)\} \quad (9)$$

Lastly, get BR-MCRA of the noise spectrum by taken advantage of a Minima Controlled Recursive Averaging (MCRA) algorithm presented by Israel Cohen in [3].

Let $S_r(k, l) = S(k, l) / S_{\min}(k, l)$, and

$$\begin{cases} I(k, l) = 1 & S_r(k, l) > \sigma, \text{ speech presence} \\ I(k, l) = 0 & S_r(k, l) < \sigma, \text{ speech absence} \end{cases}$$

Where $S(k, l)$ is the smooth spectrum of the speech with noise. We choose threshold $\sigma=5$.

Let $p(k, l)$ the probability of speech presence. According to the speech signal correlation between adjacent frames in [11], we obtain

$$p(k, l) = \alpha_d p(k-1, l) + (1 - \alpha_d) I(k, l) \quad (10)$$

$$\hat{\alpha}_d(k, l) = \alpha_p + (1 - \alpha_p) p(k, l) \quad (11)$$

Where $\alpha_d=0.2$ is the updating constant of $p(k, l)$.

$\hat{\alpha}_d(k, l)$ is smoothing parameter. α_p is empirical constant.

Then the improved estimate of the noise is

$$\begin{aligned} \hat{\lambda}_d(k, l) &= \hat{\lambda}_d(k-1, l) p(k, l) \\ &+ \left[\alpha_p \hat{\lambda}_d(k-1, l) + (1 - \alpha_p) |Y(k, l)|^2 \right] \\ &\times (1 - p(k, l)) \\ &= \hat{\alpha}_d \hat{\lambda}_d(k-1, l) + (1 - \hat{\alpha}_d) |Y(k, l)|^2 \end{aligned} \quad (12)$$

B. Noncausal Spectral Enhancement

The noncausal estimator employs future spectral measurements to predict the spectral variances of the clean speech better. Reference to the literature [4], the current l th frame of clean speech frequency component $\lambda_x(k, l)$

can be obtained by estimating the speech signal $\hat{\lambda}_{x|l+L}(k, l)$.

Noncausal prior SNR estimation method use the mixed speech signal's frequency spectrum information to estimate the first frame of the speech, which is closer to the actual speech power spectrum.

Let $\hat{\lambda}_{x|l+L}(k, l)$ denote the conditional variance of X given y_0^{l+L} excluding the noisy measurement $Y(k, l)$. Let $\hat{\lambda}'_{x|(l, l+L)}$ denote the conditional variance of given the noisy measurements $y_0^{l+L}/\{Y\}$. Then, an estimate for $\lambda_x(k, l)$ can be obtained by computing the conditional variance of X given $Y(k, l)$ and $\hat{\lambda}_{x|l+L}(k, l)$.

$$\begin{aligned}\lambda_x(k, l) &= \hat{\lambda}_{x|l+L}(k, l) \\ &= \frac{\hat{\lambda}'_{x|l+L}}{\lambda_D + \hat{\lambda}'_{x|l+L}} \left(\lambda_D + \frac{\hat{\lambda}'_{x|l+L} |Y|^2}{\lambda_D + \hat{\lambda}'_{x|l+L}} \right)\end{aligned}\quad (13)$$

$$\begin{aligned}\hat{\lambda}'_{x|l+L}(k, l) &= \max \left\{ \mu \hat{\lambda}^2(k, l-1) \right. \\ &\quad \times \left[\mu' \sum_{i=-w}^w b(i) \hat{\lambda}_{x|l+L-1}(k-i, l-1) \right. \\ &\quad \left. \left. + (1-\mu') \hat{\lambda}_{x|l+L}(k, l) \right], \lambda_{\min} \right\}\end{aligned}\quad (14)$$

where $\mu (0 \leq \mu \leq 1)$, $\mu' (0 \leq \mu' \leq 1)$ denote Smoothing coefficients. λ_{\min} is a lower bound on the variance of X . Let $b(i)$ denote a normalized window function of length $2w+1$, and $\sum_{(i) \in \Gamma} b(i) = 1$.

$$\hat{\lambda}_{x|l+L}(k, l) = \max \left\{ \frac{\sum_{(n,i) \in \Gamma} b(i) |Y(k-i, l+n)|^2}{\sum_{(n,i) \in \Gamma} b(i)} - \beta \lambda_D, 0 \right\} \quad (15)$$

And $\beta (\beta \geq 1)$ is an over-subtraction factor to compensate for a sudden increase in the noise level.

For comparison, the new improved a priori SNR estimate is given by

$$\xi_k(l) = \max \left[\alpha \frac{\lambda_x(k, l)}{\lambda_d(k, l)} + (1-\alpha) \max[\gamma_k(l) - 1, 0], \xi_{\min} \right] \quad (16)$$

The parameter α controls the trade-off between speech distortion and noise fluctuations, and $\alpha = 0.95$. The minimum value of ξ_k is $\xi_{\min} = -15\text{dB}$, which is suggested to reduce the level of musical noise.

IV. EXPERIMENTAL RESULTS

The clean speech signals in the experiment are recorded by English language in a quiet environment. The additive noises are selected from volvo noise. The sample rate of all the test data is 8 kHz. Three different SNR conditions, -5dB, 0dB, 5dB, are used in the test. The frame size of the input signal is 256 sample points and every

frame uses the Hamming window with 50% overlap between adjacent frames.

Fig .1 and Fig .2 show the waveforms of the clean speech and the noisy speech signal corrupted by volvo noise with the 5 dB SNR. Fig .3 shows the enhanced speech signal using traditional MMSE-LSA algorithm. Fig .4 shows the enhanced speech signal using proposed method.

From these figures, it is clear that the proposed method can reduce background noise effectively and the residual noise is stationary.

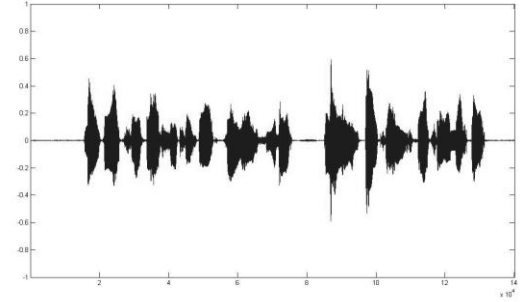


Figure 1. Waveforms of clean speech

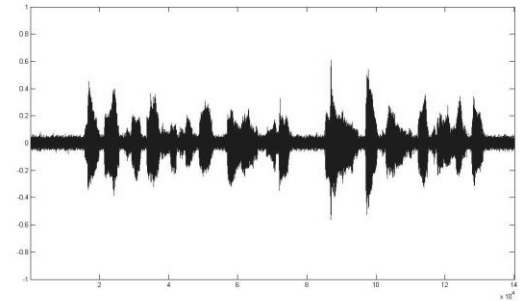


Figure 2. Waveforms of the noisy speech by the proposed method in volvo noise (SNR = 5dB)

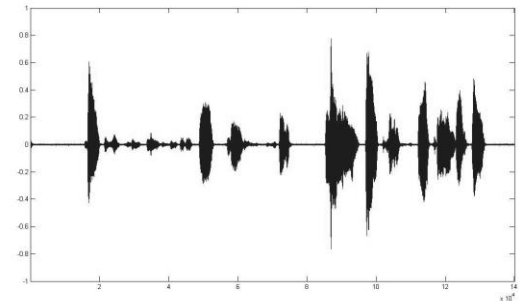


Figure 3. Waveforms of the traditional enhanced speech proposed method in volvo noise (SNR = 5dB)

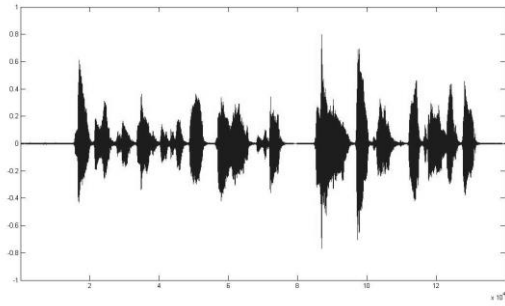


Figure 4. Waveforms of the improved enhanced speech proposed method in volvo noise (SNR = 5dB)

The Samsung Super Noise Reduction is a normal method of measuring speech enhancement algorithm. Table 1 lists the different output SSNR of the traditional algorithm and the improved algorithm with different input SNR. Table 1 shows the SNR improvement and noise reduction of the proposed method is greater than the traditional algorithms. With the decrease of input SNR, the effect of noise reduction is more distinctly.

TABLE I. COMPARISON OF SSNR

Input SNR(dB)	White noise		Volvo noise	
	<i>traditional algorithm</i>	<i>improved algorithm</i>	<i>traditional algorithm</i>	<i>improved algorithm</i>
-5	0.86	2.34	2.34	8.08
0	1.70	4.95	6.66	8.58
5	1.91	6.93	8.94	8.95

V. CONCLUSION

This paper presents a new improved MMSE-LSA speech enhancement algorithm. This algorithm based on traditional MMSE-LSA algorithm, using noncausal estimation for a prior SNR.

An important aspect of speech enhancement algorithms is the estimation of the spectral noise power. Therefore the non-stationary noise is estimated and updated by BR-MCRA, eliminating the musical noise further.

The results show that the improved algorithm can reduce the residual noise and musical noise effectively, improving SNR and keeping a good quality of voice. For non-stationary noise the method is at a clear advantage.

ACKNOWLEDGMENT

The Project Supported by Zhejiang Provincial Natural Science Foundation of China (LY12F01009) ; National Natural Science Foundation of China (61201301);Research on Application of technology for public welfare of Zhejiang Province (2014C31056) .

REFERENCES

- [1] Ephraim Y, Malah D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator[J]. IEEE Trans on Acoustics Speech and Signal Processing, 1985, 33(2):443-445.
- [2] C. Breithaupt and R. Martin, "Analysis of the decision-directed SNR estimator for speech enhancement with respect to low-SNR and transient conditions," IEEE Trans. Audio, Speech, Language Process., vol. 19, no. 2, pp. 277 - 289, Feb. 2011.
- [3] I. Cohen, B. Berdug. Noise estimation by minima controlled recursive averaging for robust speech enhancement [J]. Signal Processing, 2002, 9(1):12-15.
- [4] Cohen I. Speech enhancement using a non-causal a priori SNR estimator[J]. Signal Processing, 2004, 11(9): 725-728.
- [5] Plapous C, Marro C. Improved signal-to-noise ratio estimation for speech enhancement[J]. IEEE Trans Audio, Speech, Language Process, 2006, 14(6):2098-2108.
- [6] Kit Yan Chan, Cedric K. F. Yiu, Tharam S. Dillon, Sven Nordholm, and Sai Ho Ling, "Enhancement of speech recognitions for control automation using an intelligent particle swarm optimization," IEEE Transatcions on Industrial Informatics, VOL.8, NO.4, Nov. 2012, pp.869-879.
- [7] Anuradha R.Fukane, Shashikant L.Sahare, "Noise estimation Algorithms for Speech Enhancement in highly non-stationary Environments." IJCSI International Journal of Computer Science Issues, Vol. 8, pp. 39-44, 2011.
- [8] Yu-min zeng, Peng Wang. Based on the bidirectional search method of the minimum control recursive average speech enhancement algorithm [J]. Acta acoustica (Chinese version), 2010, 35(1):81-87.
- [9] Shan An, Changchun Bao, Bingyin Xia. An adaptive β -order MMSE estimator for speech enhancement using super-Gaussian speech model[J]. Signal and Information Processing (China SIP), 2013 6(10): 327-331.
- [10] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction," Speech Commun., vol. 49, pp. 636 - 656, 2007.
- [11] R. Martin. Spectral subtraction based on minimum statistics[J]. Seventh European Signal Processing Conference, 1999, 9: 1182-1185