Speech Enhancement Algorithm Based on Improved Spectral Subtraction

Liuyang Gao Yunfei Guo Shaomei Li Fucai Chen
(National Digital Switching System Engineering & Technological R&D Center)
Zhengzhou 450002 China
E-mail:gly03040104@163.com

***Abstract*-To improve the enhancement effect of noisy speech signals, this paper presents and analyses a new speech enhancement algorithm based on improved spectral subtraction. In contrast to the standard spectral subtraction algorithm, the new algorithm accurately estimates the noise according to that the amplitude spectral of narrowband white Gaussian noise obeys Rayleigh distribution, based on that all noise can be changed into Additive White Gaussian Noise (AWGN). This algorithm also adopts a new speech activity detection technology based on frequency band variance to detect speech activity. The emulational analyse indicates that the algorithm in this paper is better suit for noise elimination than standard spectral subtraction.**

Key words- speech enhancement, spectral subtraction, AWGN, Rayleigh distribution, speech activity detection

I. INTRODUCTION

Environment noise inevitably influences our speech communication quality. Speech enhancement technology is an available approach to resolve the influence of noise, while engineers are in favor of single microphone speech enhancement technology based on short time spectral estimation, such as algorithms of spectral subtraction, Wiener filter, minimum mean square error(MMSE) estimation as well as masking properties of human auditory system, etc[1~3]. Above all, spectral subtraction is a well known speech enhancement technology with lots of advantages. But there are some problems, such as "Musical Noise". The "Musical Noise" influences enhancement effect severely. Lots of improved algorithms have been proposed, such as algorithms of amplitude spectral average, factional spectral subtraction [4], and algorithm based on masking properties of human auditory system[5]. These algorithms improve the enhancement effect on a certain extent, but the enhancement effect is still to be improved. That is because these algorithms have not concerned with the ultimate property of noise spectral. So new algorithm must be exploited to improve enhancement effect farther, according to the property of noise spectral.

II. THE PRINCIPLE AND MAIN PROBLEM OF SPECTRAL SUBTRACTION

A. The principle of spectral subtraction

Actual noise may be color noise, or non-additive noise, but after being filtered by a whitening filter and homostasis system, all noise can be changed into Additive White Gaussian Noise (AWGN). Formula (2-1) displays the model in time domain of a frame of short time speech signal with AWGN.

$$y(n)=x(n)+d(n),0 \le n \le N-1 \qquad (2\text{-}1)$$

In formula (2-1) n stands for the n-th point, while N stands for frame length. x(n) refers to speech without noise, while d(n) refers to AWGN and y(n) is speech with noise.

The corresponding model in frequency domain can be displayed as follow:

$$Y(k)=X(k)+D(k),0 \le k \le N-1 \qquad (2\text{-}2)$$

Based on supposition that d(n) is additive and irrelevant to x(n),the expectation of power spectral is:

$$E[|Y(k)|^2]=E[|X(k)|^2]+E[|D(k)|^2],0 \le k \le N-1 \qquad (2\text{-}3)$$

The standard spectral subtraction takes the average signal of frames without speech (just noise) as the estimation of $E[|D(k)|^2]$, which is supposed to λ_k^2. For short time process, amplitude spectral estimation of original speech can be presented as follow:

$$|\hat{X}(k)|=\begin{cases}\{|Y(k)|^2-\lambda_k^2]\}^{1/2}, & |Y(k)|^2 \ge \lambda_k^2 \\ 0, & \text{else}\end{cases} \qquad (2\text{-}4)$$

The standard spectral subtraction takes the phase of speech as the phase of enhanced speech directly, based on that human auditory system perceives speech by the amplitude spectral of signal and has a thick skin to

Sponsored by The 863 National High Technology Research and Development Program of China Fund (No. 2008AA011002)

978-1-4244-4738-1/09/$25.00 ©2009 IEEE

phase spectral. Then the spectral estimation of original speech can be obtained.

B. The main problem of spectral subtraction

Main problem of standard spectral subtraction in speech enhancement is the problem of remaining noise. In formula (2-4), noise amplitude spectral is estimated from statistical average value of frames without speech, while actually noise amplitude spectral fluctuates during each frame. When estimating the amplitude spectral of pure speech signal, part of noise has been left in the speech signal. That is called "Musical Noise", which brings perceptual quality of enhanced speech to serious decline.

To solve this problem, this paper proposes another method of noise estimation according to the property of noise. The emulational experiment indicates that this method estimates noise spectral more accurately and wipes off musical noise in effect.

III. IMPROVED SPECTRAL SUBTRACTION BASED SPEECH ENHANCEMENT ALGORITHM

A. Speech activity detection

Usually, people use speech activity detection algorithms based on power, such as short time power algorithm, short time average amplitude algorithm, double threshold algorithm, etc. The basic problem of these algorithms occurs just because voting threshold is determined by experiential value [6].

This paper adopts the speech activity detection algorithm based on frequency band variance [7].This algorithm gets speech endpoint by using short time frequency band variance as detection parameter.

Firstly, define vector:

$$X = \{x(w_0), x(w_1), \ldots, x(w_n)\}$$

The average value is:

$$E = \frac{1}{n+1}\sum_{i=0}^{n} x(w_i) \qquad (3\text{- }1)$$

Then frequency band variance is calculated as follow：

$$D = \frac{1}{n+1}\sum_{i=0}^{n} [x(w_i) - E] \qquad (3\text{- }2)$$

The frequency band variance of speech signal is much larger than the one of environment noise. That is to say, D>>Dr. We take M as (3~5) Dr. If D $\geq$ M, we believe there is speech signal in this frame, otherwise, there is only noise.

B. Noise amplitude spectral estimation

The amplitude spectral D(k) of narrowband white Gaussian noise obeys to Rayleigh distribution[8]:

$$f(x) = \begin{cases} \dfrac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, x \geq 0 \\ 0, x < 0 \end{cases} \qquad (3\text{- }3)$$

According to full expectation formula,

$$\begin{aligned} |\hat{D}_i(k)| &= E[|D_i(k)|\,||D_i(k)|<|Y_i(k)|] \\ &= \int E[x \,|\, x<|Y_i(k)|]dF_{D_i(k)}(x) \\ &= \int_0^{|Y_i(k)|} xf(x)dx \Big/ \int_0^{|Y_i(k)|} f(x)dx \\ &= \int_0^{|Y_i(k)|} \frac{x^2}{\sigma_i^2} e^{-\frac{x^2}{2\sigma^2}} dx \Big/ \int_0^{|Y_i(k)|} \frac{x}{\sigma_i^2} e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \frac{-e^{-\frac{|Y_i(k)|^2}{2\sigma_i^2}} + \frac{\sqrt{2\pi}\sigma_i}{2|Y_i(k)|} erf(\frac{|Y_i(k)|}{\sqrt{2}\sigma_i})}{1 - e^{-\frac{|Y_i(k)|^2}{2\sigma_i^2}}} |Y_i(k)| \end{aligned} \qquad (3\text{- }4)$$

In formula (3- 4), i stands for the i-th frame, As we know, $erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2} dt$ .Following property of Rayleigh distribution, $E[|D_i|] = \sqrt{\pi\sigma_i^2/2}$, then

$$\sigma_i^2 = 2[E[|D_i|]]^2 / \pi \qquad (3\text{- }5)$$

Considering recorded speech, we can take several beginning frames for noise signal. Practically, this paper substitutes the first frame for noise, that's to say, $|D_1(k)| = |Y_1(k)|$, according to formula (3-5),

$$\sigma_1^2 = 2[E(|Y_1|)]^2 / \pi \,. \qquad (3\text{- }6)$$

Then detect speech activity of next frame by using speech activity detection algorithm described in section A of this chapter, in order to judge whether there is speech in next frame. If speech exists, subtract noise spectral; otherwise, update the noise amplitude spectral estimation. Taking short time stationarity of white Gaussian noise into account, we consider that

$E[|D_i|] = E[|\hat{D}_{i-1}|]$, then get recursion of $E[|D_i|]$. As the amplitude spectral of the former frame $|\hat{D}_{i-1}|$ has already been calculate, according to formula (3- 5), σ_i^2 can be expressed as follow:

$$\sigma_i^2 = 2[E(|\hat{D}_{i-1}|)]^2 / \pi \qquad (3\text{- }7)$$

At last, according to formula (3- 7) and formula (3- 4), we gets $|\hat{D}_i|$.

IV. EXPERIMENTAL RESULTS AND ANALYSE

This paper takes use of male's speech "I like nature", recorded in the laboratory with little noise and the sampling frequency is 22050Hz. This signal is enframed with frame length 256(2.54ms). Since there is little noise in this signal, the enhancement is not necessary. Add AWGN to the signal before processing. Figure 4- 1 presents the result of algorithm proposed in this paper, compared with standard spectral subtraction, when SNR=10dB. In figure 4-1, picture (a) stands for recorded laboratory speech, while picture (b) for speech with AWGN (SNR=10dB). Picture (d) presents the result of spectral subtraction improved with arithmetic proposed in this paper, compared with the standard spectral subtraction (picture (c)).

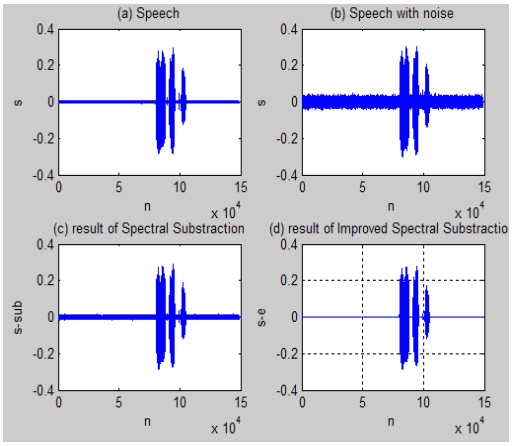

Figure 4-1:The enhancement effect (SNR=10dB)

After the first experiments it becomes clear that the algorithm proposed in this paper does better in enhancement effect than standard spectral subtraction does. Then adds heavier noise. Figure 4-2 presents the result of enhancement effect when SNR=0dB.

Just as the figure 4-2 indicates, the noise elimination performance of the algorithm in this paper is better suited for musical noise restrain than the one of standard spectral subtraction. There are two reasons for that. One is that preferably speech activity detection technology is used for filtering the noise of period without speech downright, the other is to improve accuracy of noise amplitude spectral estimation during period with speech, according to that noise amplitude spectral obeys Rayleigh distribution.

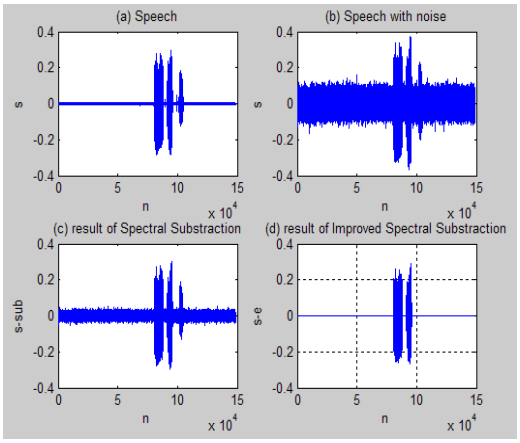

Figure 4-2:The enhancement effect (SNR=0dB)

V. CONCLUSION

This paper presents and analyses a new speech enhancement algorithm based on improved spectral subtraction, to improve the enhancement effect of noisy speech signals. The theoretical analysis indicates that the new algorithm estimates the noise more accurately than standard spectral subtraction algorithm does. There are two reasons, one is that the new algorithm estimates the noise according to that the amplitude spectral of narrowband white Gaussian noise obeys Rayleigh distribution, the other is that this algorithm adopts a new speech activity detection technology based on frequency band variance to detect speech activity. The emulational experiment shows that the performance of this improved algorithm is much better than that of standard spectral subtraction.

REFERENCES

[1] BOLL S F. Suppression of acoustic noise in speech using spectral subtraction[J].IEEE Trans Acoust Speech Signal Process,1979,27(2):113-120.

[2] SOON I Y,KOH S N. Speech enhancement using 2-D Fourier transform[J].IEEE Trans Speech Audio Process,2003,1l(6):717—724.

[3] EPHRAIM Y,MALAH D. Speech enhancement using a minimum mean-square error short-rime spectral amplitude estimator[J] .IEEE Trans Acoust Speech Signal Process,1984,32(6):1109-1121.

[4] Wang Zhenli,Zhang Xiongwei. A Method Based on Fractional Spectral Subtraction for Speech Enhancement. China Journal of

Electronics & Information Technology[J], 2007,29(5):1096-1020.

[5] Cai Hantian,Yuan Botao. A speech enhancement algorithm based on masking properties of human autlitory system. Journal of China Institude of Communications[J], 2002,23(8):93-98

[6] Wang Bingxi,Qu Dan,Peng Xuan. Practical Speech Recognition Foundation[M].Beijing: National Defense Industry Press,2005:106~113

[7] Li Zupeng,Yao Peiyang. A new method of speech activity detecting. China Telecommunication Technology [J],2000,3 : 68~71

[8] Zhu Hua,Huang Huining,Li Yongqing. Stochastic Signal Analyze [M]. Beijing: Beijing University of Technology Press,1990:317~322