Pink Noise Whitening Method for Pitch Synchronous LPC Analysis

Liqing LIU* and Tetsuya SHIMAMURA†
* Saitama University, Saitama, Japan
E-mail: leisei@sie.ics.saitama-u.ac.jp Tel/Fax: +81-048-8583776
† Saitama University, Saitama, Japan
E-mail: shima@sie.ics.saitama-u.ac.jp Tel/Fax: +81-048-8583496

***Abstract*—We present a new noise whitening method for pitch synchronous LPC analysis under pink noise circumstances. First, we utilize a rectangular window to extract two frames whose shifting interval is a full pitch period. Then we perform a subtraction operation between the two frames to obtain a new noise signal which is considered to be not corrupted by the voiced speech signal. The obtained new noise signal can be used to design a new prediction whitening filter. The new whitening filter not only whitens the pink noise signal, but also can keep the vocal tract and formant natures of voiced speech signal. Utilizing the whitened signal, we can improve the pitch synchronous addition and subtraction (PSAS) method under pink noise circumstances. We discuss the properties of the whitened signal and PSAS method. Experimental results indicate the effectiveness of the proposed method.**

I. Introduction

In recent years, noise reduction plays a very significant role in speech analysis systems. Depending on the frequency domain properties, adverse noise is classified into white noise, colored noise, impulsive noise and so on. As well known, white noise is defined as a random signal process and its frequency spectrum is flat which means that it has equal power in all frequencies. On the other hand, colored noise refers to any broadband noise with a non-flat spectrum. Pink noise is a representation of colored noise, which has a predominantly low frequency spectrum.

As a widely used technique for speech analysis, linear predictive coding (LPC)[1] can estimate some basic parameters like formant, vocal tract function, spectrum, pitch and so on. This technique represents the voiced speech signal by a set of predictive parameters. Hence the LPC technique lies on the accuracy of the estimated predictive coefficients. The standard LPC technique like the autocorrelation method[2] assumes that the predictive coefficients should be estimated from clean voiced speech. Under noisy circumstances, therefore, the inaccuracy of the estimated predictive parameters will lead to degradation in speech analysis.

Several methods have been proposed to reduce the effects of noise on predictive parameters. Most of them aimed at white noise as noise. For example, a high-order Yule-Walker estimator[3] takes advantage of the property of the contaminated noise and does not involve the zero-th lag autocorrelation of voiced speech signal. This method is effective for white noise, however, it is invalid for pink noise because the autocorrelation function of pink noise will not concentrate only on the zero-th lag. And the non-stability of the resulting all-pole filter is also its shortcoming. These are also true for noise compensation[4]. The noise compensation method utilizes the property of white noise whose power is assumed to concentrate on the zero-th lag. Thus this method can not also adapt to pink noise.

Recently pitch synchronous analysis[5][6] has been applied to LPC analysis[7]. For pitch synchronous analysis of voiced speech, the analysis segment position coincides with the two pitch pulses and the number of samples in analysis-segment duration is less than or equal to the number of samples in a pitch period. Combining the pitch synchronous technique with LPC technique, the pitch synchronous LPC technique can become a suitable tool for vocal tract and speech source analysis. Thus Paliwal et al. proposed a pitch synchronous LPC analysis method by modifying the autocorrelation method in [7]. The method guarantees the stability of the estimated all-pole filter and is shown to provide a more accurate estimate than the autocorrelation and covariance methods of linear prediction for clean voiced speech case. However, it can not solve the problem under noisy circumstances. Shimamura et al.[8] proposed a noise reduction method based on pitch synchronous addition (PSA) for pitch synchronous LPC analysis. The PSA method is shown to provide a superior performance in white noise because of the distribution property of random noise process. Nevertheless it is weak for pink noise. Furthermore the performance will be easily affected by the length of pitch. It will provide a good performance in the case of high-pitched female and children speech. However, in the case of low-pitched male speech, it will not provide a desirable performance. Based on the PSA method, Liu et al.[9] proposed a pitch synchronous addition and subtraction (PSAS) method. The PSAS method is an iterative noise compensated method based on pitch synchronous LPC analysis. It can improve the performance of the PSA method in white noise. Unlike most of the noise compensation methods which utilize the non-speech durations to estimate the noise power, the PSAS method can estimate the noise power in every current frame. However, it still can not settle pink noise. A serious spectrum distortion will arise under pink noise circumstances.

In this paper, we present a new prediction whitening method for pitch synchronous LPC analysis. The proposed whitening method just whitens the adverse noise and could keep the formant properties of voiced speech signal. Utilizing the proposed whitening method, we can improve the performance of the PSAS method under pink noise circumstances.

II. Proposed Method

A. Proposed Prediction Whitening Filter

A linear prediction model is an all-pole model and the future value of a signal $s(n)$ is forecasted by a linear combination of its past values $s(n-i)$ and a certain input $u(n)$:

$$s(n) = -\sum_{i=1}^{p} a_i s(n-i) + Gu(n) \tag{1}$$

where a_i are the predictive coefficients, G is the gain function and $u(n)$ is a driving noise which is a zero-mean white Gaussian noise. We transform (1) into

$$e(n) = s(n) + \sum_{i=1}^{p} a_i s(n-i) \tag{2}$$

where $e(n)$ is the prediction error.

Comparing (1) with (2), we see that they have some similarities in an equation form. However, there is a totally different point. In (1), $u(n)$ is a driving noise which behaves as an input to an autoregressive filter and $s(n)$ is an output of the autoregressive filter, while in (2) $s(n)$ is treated as an input to a linear prediction filter and $e(n)$ is an output of the linear prediction process. In general, the prediction error $e(n)$ can be regarded as a white noise process and the linear prediction filter also can be considered to be a whitening filter. Certainly the coefficients of linear prediction filter in (2) is equivalent to the coefficients of the autoregressive filter in (1).

We apply the prediction whitening filter in (2) to pitch synchronous LPC analysis. Based on the pitch synchronous LPC analysis, we develop a new prediction whitening filter which just whitens the noise signal and maintains the frequency properties of voiced speech signal. Let us assume that an observed noisy speech signal can be expressed by

$$x(n) = s(n) + w(n) \tag{3}$$

where $s(n)$ denotes the clean voiced speech and $w(n)$ denotes the adverse noise. Then we utilize a rectangular window with a length of 20-25 ms to extract two frames whose shifting interval is one pitch period, T, as shown in Fig. 1. In Fig. 1, a sampling rate of 10 kHz is assumed. We assume that $x^1(l)$ and $x^2(l)$ represent Frame 1 and Frame 2, respectively, where framing is represented commonly for $l = 1, 2, ..., L$. L is the length of the frame. According to (3), $x^1(l) = s^1(l) + w^1(l)$ and $x^2(l) = s^2(l) + w^2(l)$. A new subtraction signal, $y(l)$, is obtained through the subtraction operation between Frame 1 and Frame 2 as follows:

$$\begin{aligned} y(l) &= x^1(l) - x^2(l) \\ &= s^1(l) + w^1(l) - s^2(l) - w^2(l) \end{aligned} \tag{4}$$

where $l = 1, 2, ..., L$.

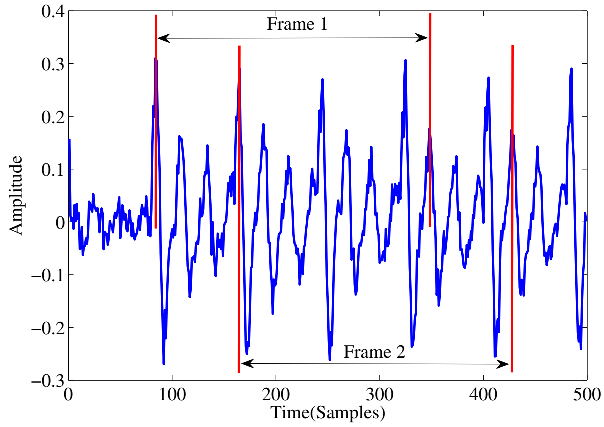

Fig. 1. Waveform of voiced speech synthetic vowel /o/ synthesized by [7]

Since a clean voiced speech signal has a clear periodicity which corresponds to the pitch period, $s^1(l)$ is assumed to be identical to $s^2(l)$. Hence the new signal $y(l)$ can result in

$$y(l) = w^1(l) - w^2(l). \tag{5}$$

The subtraction signal $y(l)$ can be a new noise signal without corruption by voiced speech signal. Then we substitute $y(l)$ into $s(n)$ in (2), resulting in

$$e(l) = y(l) + \sum_{i=1}^{q} b_i y(l-i) \tag{6}$$

with different predictive coefficients b_i of order q.

In (6), the obtained new noise signal $y(l)$ is an input signal to linear prediction whitening filter. The parameters of the prediction whitening filter can be calculated by the autocorrelation method of LPC technique. The resulting prediction whitening filter is determined by the noise signal and is uncorrelated with the voiced speech signal. Thus the new prediction whitening filter can whiten the pink noise. On the other hand, it is unable to whiten the voiced speech signal. In other words, the vocal tract and formant natures of voiced speech signal will not be transformed by the whitening filter. Simply saying, the prediction whitening filter estimated from (6) is a whitening filter for pink noise. For a voiced speech signal it is merely a common filter whose frequency characteristics are shifted by the pink noise. In the case of pink noise, the prediction whitening filter will behave as an approximated high-pass filter for the voiced speech signal.

For the purpose showing the behavior of the new prediction whitening filter, we corrupt a synthetic vowel /o/[7] with pink noise at SNR=10dB. The reason for selecting a synthetic vowel /o/ is that the true values of spectral parameters are known in advance. First we compare the original adverse pink noise with the whitened pink noise. The results in Fig. 2 show the whitening effect clearly. Fig. 2(e) shows that the frequency characteristics of adverse pink noise get close to a flat spectrum of ideal white noise after whitening. Furthermore the autocorrelation function of adverse pink noise

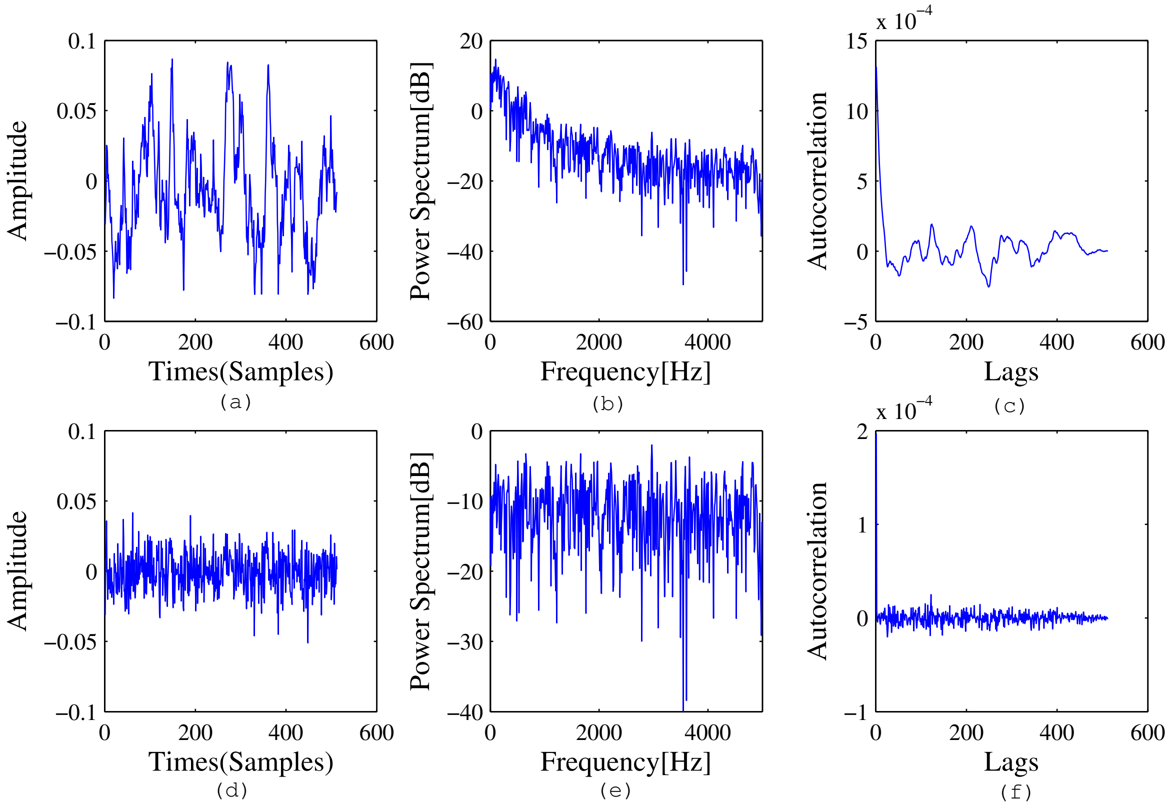

Fig. 2. (a)Original adverse pink noise. (b)Frequency characteristics of original adverse pink noise. (c)Autocorrelation of original adverse pink noise. (d)Adverse pink noise after whitening. (e)Frequency characteristics of adverse pink noise after whitening. (f)Autocorrelation of adverse pink noise after whitening.

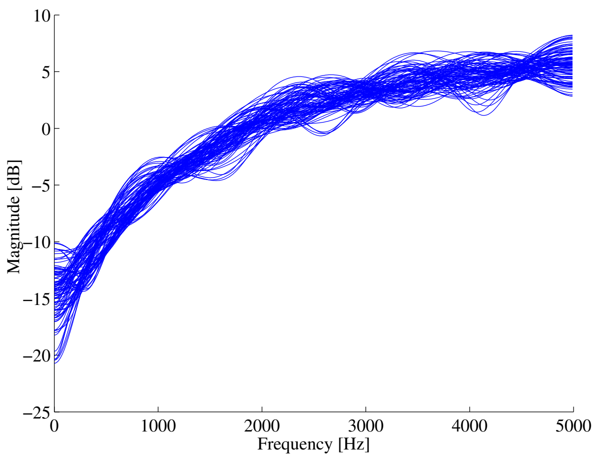

Fig. 3. Frequency characteristics of prediction whitening filter

after whitening can be assumed to be zero except for the zero-th lag in Fig. 2(f).

Fig. 3 shows the frequency characteristics of the resulting prediction whitening filter for continuous 100 frames. Fig. 3 suggests that in the case of pink noise, the frequency characteristics of the prediction whitening filter is approximately a high-pass filter.

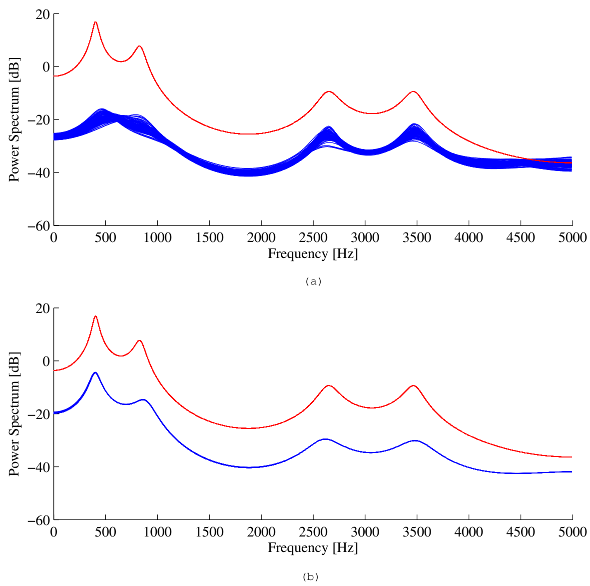

Fig. 4. Spectra of clean voiced speech after whitening

Next we check the clean voiced speech signal after whitening. Fig. 4(a) shows the spectra for continuous 100 frames of clean voiced speech after passing through the prediction whitening filter. Here the red line represents the true spectrum of synthetic vowel /o/. Fig. 4(a) indicates that the spectrum in low-frequency regions is restrained, while the spectrum in high-frequency regions is strengthened. However, the vocal tract properties of voiced speech are almost not altered. In order to eliminate the effect of the prediction whitening filter, we need to add an inverse filter of the prediction whitening filter further. In other words, we need to divide the estimated spectrum in Fig. 4(a) by the squared amplitude response of the prediction whitening filter. Then we can obtain a new spectrum. Compensating for the squared spectrum produces a close shape to the true one without influence of the prediction whitening filter as shown in Fig. 4(b).

In the next subsection, we utilize the new prediction whitening filter to ameliorate the PSAS method under pink noise. As mentioned earlier, the PSAS method is an iterative noise compensated method based on PSAS for pitch synchronous LPC analysis. Like most of the noise compensation methods, the PSAS method could not produce a desirable performance under pink noise circumstances. Unlike the white noise whose autocorrelation function is assumed to be zero except for the zero-th lag, the autocorrelation function of pink noise is not a pulse function, which has the maximum value at zero-th lag, and will decrease at increasing lags. Hence the noise compensation method can not provide a good noise reduction under pink noise circumstances. To make the PSAS method adapt to pink noise, the prewhitening procedure is requisite. We discuss the properties of the PSAS method after whitening in the next subsection.

B. PSAS method

First we introduce the PSAS method briefly. Pitch synchronization is very significant for pitch synchronous LPC analysis. Here the speech signal sample with the maximum value in a period is taken as the first sample in one frame[7][10]. Then according to the pitch period, one frame noisy speech signal, $x(n)$, is divided into K blocks such as

$$x_i(j) \quad i = 1, 2...K \quad j = 1, 2...P \tag{7}$$

where K is the number of pitch period and P is the number of samples in each pitch period.

Depending on the clear periodicity of clean voiced speech signal, an enhanced speech signal, $x_{ave}(j)$, is derived from the average operation of PSA as

$$x_{ave}(j) = \frac{1}{K}\sum_{i=1}^{K} x_i(j), \tag{8}$$

while the modified noise signal, $w_{as}(j)$, is obtained by the average operation of PSAS as

$$w_{as}(j) = \frac{1}{K}\sum_{i=1}^{K}(-1)^{i+1}x_i(j). \tag{9}$$

In (9), the value of K is set to be even. It has been shown in [9] that the noise power of the modified noise signal $w_{as}(j)$ is equivalent to the noise power in the enhanced speech signal under white noise circumstances. That is, the autocorrelation function of the clean voiced speech, $R_{ss}(k)$, can be approximately obtained by subtracting the autocorrelation of the modified signal, $R_{w_{as}w_{as}}(k)$, from the autocorrelation of the enhanced speech signal, $R_{x_{ave}x_{ave}}(k)$ as:

$$R_{ss}(k) = R_{x_{ave}x_{ave}}(k) - \lambda R_{w_{as}w_{as}} \tag{10}$$

where $0 \le \lambda \le 1$ and λ is gradually decreased by a rate of 0.1 until to ensure the stability of the LPC filter.

For the calculation of a biased autocorrelation function, the length of one pitch period is so short that it will induce severe distortion. To avoid such autocorrelation distortion, the computing of autocorrelation function in our work makes use of the modified autocorrelation method proposed by Paliwal in which instead of assuming zero extension of the signal, a periodic extension of the pitch period signal is assumed[7].

C. Improved PSAS Method

To overcome the drawback of the PSAS method under pink noise circumstances, we should whiten the noisy speech by the proposed prediction whitening filter first. The improved method is summarized as the below procedure:

- Step 1: Utilize the rectangular window with a length of 20-25 ms to extract the analysis frame and auxiliary frame whose shifting interval is a full pitch period. The two frames are applied to (4) and (6) and the proposed whitening filter is obtained.
- Step 2: Whiten the noisy speech signal of the analysis frame by the whitening filter obtained in Step 1.
- Step 3: Divide the whitened speech signal into K blocks, according to pitch period. Then apply the PSAS method to them.
- Step 4: Estimate the predictive coefficients by the Levinson-Durbin recursion.
- Step 5: Calculate the power spectrum from the resulting predictive coefficients and then divide it by the squared amplitude spectrum of the whitening filter in Step 1.

III. Experiments

A. Results on Synthetic Vowel

The performance of the improved PSAS method was investigated using a synthetic vowel /o/. The synthetic vowel /o/ is contaminated by adverse pink noise. Table I is the parameter specification for experiments. Figs. 5 and 6 show the LPC power spectra estimated by the PSAS method and improved PSAS method, respectively, in continuous frames at SNR=10dB. It can be seen from these two figures that the results estimated by the improved PSAS method in Fig. 6 provide more stable spectral sharpness and get closer to the true one than the ones in Fig. 5. Especially in the third and forth formants, the improved PSAS method provides more better sharpness.

Here we introduce the measurement of the cepstrum distance to show the improvement of the improved PSAS method.

TABLE I
EXPERIMENTAL PARAMETER SPECIFICATION FOR SYNTHETIC VOWEL /o/

Sampling frequency	10 kHz
Pitch period	8 ms
Analysis window	Rectangular
LPC order	10
Additive noise	pink
Number of frames	100
Frame shifting	T

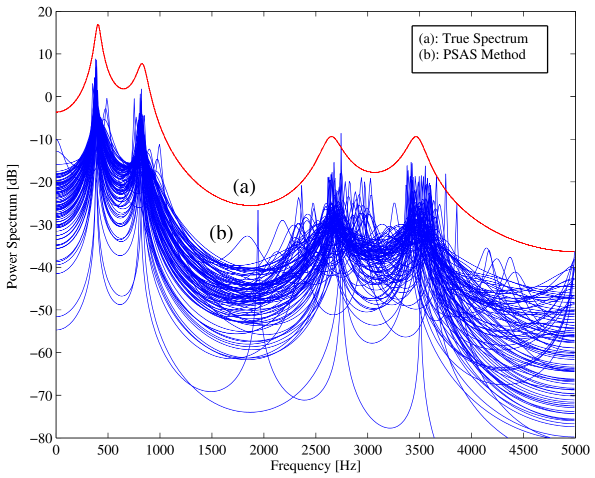

Fig. 5. LPC power spectra of PSAS method for synthetic vowel /o/ contaminated by pink noise at SNR=10dB

The cepstrum distance is calculated by

$$CD = \frac{10}{ln10}\sqrt{2\sum_{i=1}^{M}(c_i - \tilde{c}_i)^2} \tag{11}$$

where c_i are true cepstrum coefficients and $\tilde{c}_i$ are the estimated

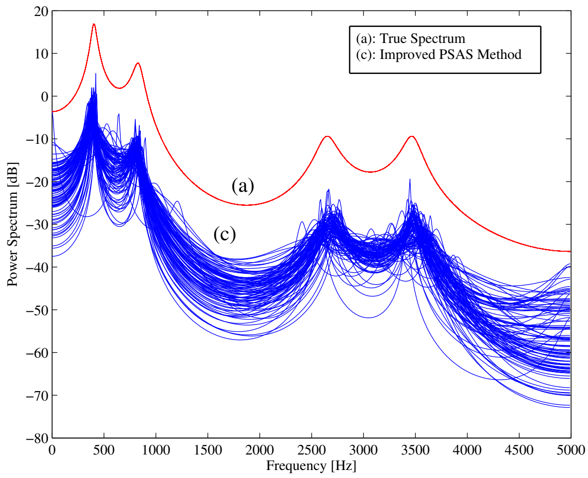

Fig. 6. LPC power spectra of improved PSAS method for synthetic vowel /o/ contaminated by pink noise at SNR=10dB

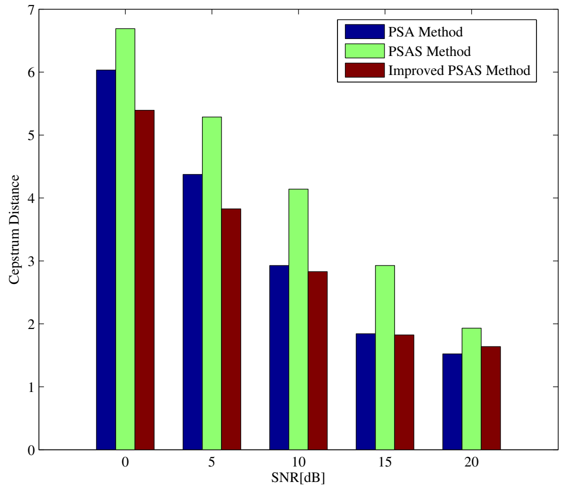

Fig. 7. Comparison of cepstrum distance for synthetic vowel /o/

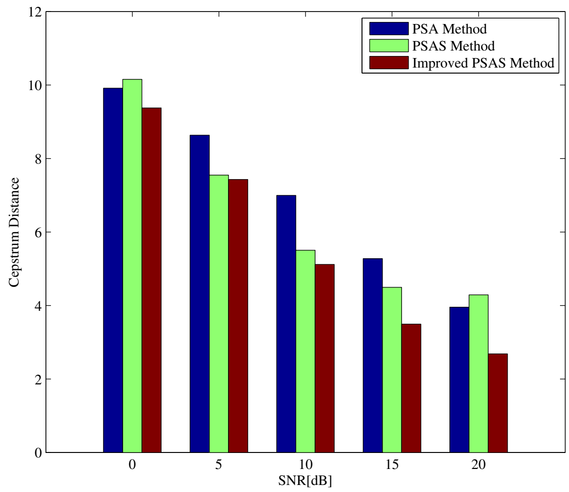

Fig. 8. Comparison of cepstrum distance for real vowel /a/

cepstrum coefficients calculated from the noisy speech signal[1]. We compare experimentally the improved PSAS method with the methods of PSA and PSAS. The input SNR was varied from 0 dB to 20 dB. The results are shown in Fig. 7. Fig. 7 shows that the PSAS method has the poorest performance under pink noise and the improved PSAS method provides a significant improvement.

B. *Results on Real Vowel*

Experiments have been also carried out on a real vowel /a/. Table II is the experimental parameter specification for the real vowel /a/. Unlike the synthetic vowel signal, the

[1]In order to calculate the $\tilde{c}_i$ of the improved PSAS method, we need to estimate the predictive coefficients of the resulting all-pole filter again. We utilize the compensated power spectrum in Step 5 to obtain the autocorrelation function. Then the new predictive coefficients are estimated by the Levinson-Durbin method

TABLE II

EXPERIMENTAL PARAMETER SPECIFICATION FOR REAL VOWEL /a/

Sampling frequency	10 kHz
Pitch period	7 ms
Analysis window	Rectangular
LPC order	10
Additive noise	pink
Number of frames	100
Frame shifting	T

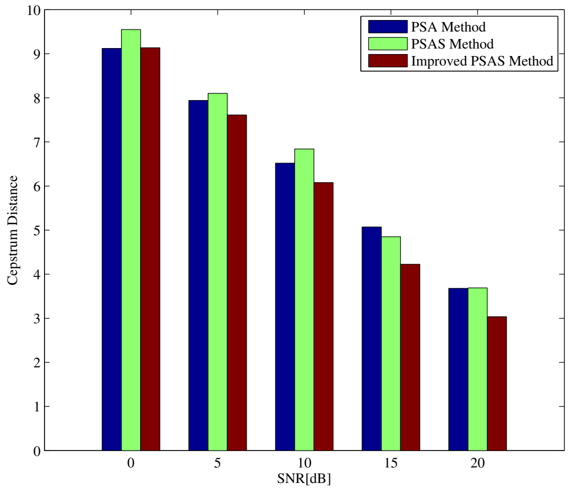

Fig. 9. Comparison of cepstrum distance for synthetic vowel /a/ under a mixed noisy condition

real vowel signal is a non-stationary speech waveform and has time-varying amplitude. Hence in (4), $s^1(l) - s^2(l)$ will not be identical to zero. However when the length of frame is limited to the duration about 20-25ms, the voiced speech can be assumed stationary. Therefore, even for this case, the prediction whitening filter is expected to have the capability to whiten the noisy signal. As shown in Fig. 8, actually the improved PSAS method provides the best improvement regardless to SNR under pink noise circumstances.

IV. Discussion

So far the experiments have been carried out under pink noise circumstances. However the improved PSAS method can also be adapted to a mixed environment of white noise and pink noise. The reason is that the proposed whitening method will not influence seriously the white noise. Thus here we discuss the improved PSAS method in a mixed environment. Figs. 9 and 10 show comparison results of cepstrum distance for synthetic vowel and real vowel regardless to SNR, respectively. For example, a mixed noise at SNR=0dB means that a pink noise at SNR=0dB plus white noise at SNR=0dB. Both of the results show that the improved PSAS method provide a better performance than the others.

V. Conclusions

A new prediction whitening method has been proposed. Based on the proposed whitening method, an improved PSAS method has been derived. From experimental results, the new method can improve the PSAS method and provide better performance than the other methods (PSA and PSAS methods) under pink noise. It also can be adapted to a mixed noisy environment.

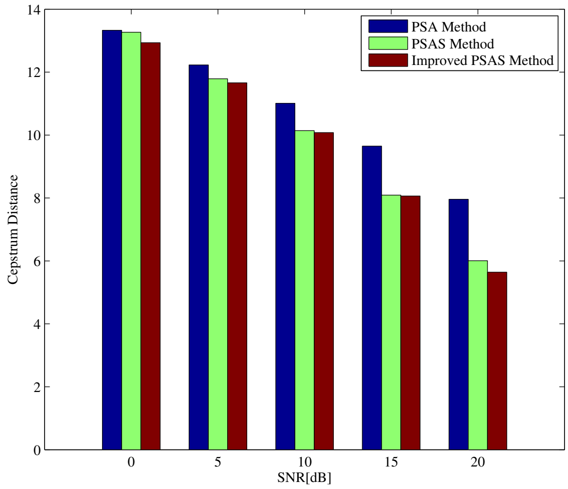

Fig. 10. Comparison of cepstrum distance for real vowel /a/ under a mixed noisy condition

References

[1] B. S. Atal, and S. Hanauer: " Speech analysis and synthesis by linear prediction of the speech wave", J. Acoust. Soc. Am., Vol. 50, No. 2, pp. 637-655, Aug. 1971.

[2] J. Markel: "Digital inverse filtering - A new tool for formant trajectory estimation", IEEE Trans. Audio and Electroacoust., Vol. AU-20, No. 2, pp. 129–137, June 1972.

[3] S. M. Kay, "Noise compensation for autoregressive spectral estimates", IEEE Trans. Acoust., Speech, and Signal Process., Vol. ASSP-28, No. 3, pp. 292-303, 1980.

[4] Q. F. Zhao, T.Shimamura and J.Suzuki, "Improvement of LPC analysis of speech by noise compensation", IEICE Trans, Vol. J81-A, No. 11, pp. 1583-1591, Nov 1998.

[5] M. V. Mathews, J. E. Miller, and E. E. David, "Pitch synchronous analysis of voiced sounds", J. Acoust. Soc. Amer, vol. 33, pp. 179-186, 1961.

[6] Y. Medan and E. Yair, "Pitch synchronous spectral analysis scheme for voiced speech", IEEE Trans. Acoust., Speech, and Signal Process., Vol. 37, No. 9, pp. 1321-1328, 1989.

[7] K. K. Paliwal and P.V.S. Rao, "A modified autocorrelation method of linear prediction for pitch-synchronous analysis of voiced speech", Signal Processing, Vol. 3, No. 2, pp. 181–185, 1981.

[8] T. Shimamura and Y. Kuroiwa, "LPC analysis based on noise reduction using pitch synchronous addition", IEICE Trans, Vol. J87, No. 4, pp. 458–469, April 2004.

[9] L. Q. Liu and T. Shimamura, "A new noise reduction method for pitch synchronous LPC analysis" 2012 Spring Meeting, Acoustical Society of Japan, pp. 377-380, March 2012.

[10] K. K. Paliwal, "Performance of the weighted Burg methods of AR spectral estimation for pitch synchronous analysis of voiced speech", Speech Commun., Vol. 3, No. 3, pp. 221-231, Dec 1984.