RESTORATION OF INSTANTANEOUS AMPLITUDE AND PHASE USING KALMAN FILTER FOR SPEECH ENHANCEMENT

Naushin Nower, Yang Liu, and Masashi Unoki

School of Information Science, Japan Advanced Institute of Science and Technology
Email: {naushin, yangliu, unoki}@jaist.ac.jp

ABSTRACT

This paper proposes a restoration scheme for the instantaneous amplitudes and phases in sub-bands by using a Kalman filter with linear prediction (LP). A few important studies have already proved that phase spectrum on the short term Fourier transform plays an important role in speech enhancement. Thus, the proposed scheme concentrates on restoring of both the instantaneous amplitudes and phases simultaneously. In this scheme, the Kalman filter is used for both instantaneous amplitudes and phases in the sub-band representation to remove the effect of noise. Thus, it can sufficiently reduce the noise effects in both. Simulations were carried out in various noisy environments to evaluate the effectiveness of the proposed scheme. The signal to error ratio (SER), perceptual evaluation of speech quality (PESQ), and SNR loss were used as objective measures. Results showed that the proposed scheme can effectively improve these objective measures more than conventional methods.

***Index Terms*—** Speech enhancement, Instantaneous amplitude and phase, Kalman filter, Gammatone filterbank, Linear prediction

1. INTRODUCTION

In real world scenarios, the desired speech signal is smeared by various kinds of interferences, such as background noise, reverberation, and competing speech. These interferences not only degrade the perceptual aspects of speech quality and intelligibility but also reduce the performance of various automated speech systems such as automatic speech recognition systems, speaker recognition systems, and hearing aids. Therefore, the quality and intelligibility of speech signal in the noisy environments have to be enhanced.

Various methods for speech enhancement have already been proposed to remove the effects of noise from the noise degraded speech to improve its quality. Well-known speech enhancement algorithms, such as Spectral Subtraction (SS) [1], the Ephraim-Malah algorithm (MMSE-STSA estimator) [2], the Scalart-Filho algorithm (Wiener filtering) [3], and the Corpus-Based approach [4] process the corrupted speech signals by modifying or correcting the spectral magnitude only and keeping the phase component unchanged. This is because the phase spectrum conventionally considered is unimportant and has been shown not to contribute much towards speech enhancement. Wang and Lim emphasized this point [5] and theirs is perhaps the most cited work to justify the unimportance of phase for speech enhancement.

However, recent studies have reported that the use of phase spectrum in the short-time Fourier transform (STFT) can significantly improve speech enhancement [6, 7, 8, 9]. Shannon and Paliwal [6] reported that magnitude-only and phase-only experiments have been carried out to investigate the effect of phase in speech enhancement. In the magnitude-only experiment, the clean magnitude was used and the phase was set to a random value. In contrast, in the phase-only experiment, the clean phase was used and the magnitude was set to one. Their results showed that the phase spectrum also contains useful and important information. After that, Paliwal and Alsteris [7] investigated whether the shape and length of the window function used in the STFT for phase manipulation are important factors for speech enhancement. Paliwal *et al.* [8] showed that modifying the phase spectrum can greatly improve speech enhancement. For this, they investigated various cases where the different combinations of noisy, clean (noiseless), and compensated amplitude and phase spectra are considered. This suggested that significant speech enhancement can be possible if the clean phase is known or the compensated phase spectrum is available. They also studied the effect of mismatched or matched windows for both amplitude and phase spectra estimation during analysis modification synthesis (AMS) in the STFT. The results show that the proper choice of an analysis window and AMS setting on the phase spectrum can significantly improve the speech enhancement. Thus, for better speech enhancement, we obviously need to consider both the amplitude and phase of the noisy signal.

It is well-known that all existing speech enhancement algorithms based on STFT-AMS can improve speech quality but not speech intelligibility [10]. The reasons for that are still unclear so that many researchers have investigated the expected strategy for reducing distortions and enhancing features related to speech intelligibility. On the other hand, from psycho acoustical studies, it is found that temporal envelope (TE) and temporal fine structure (TFS) are important cues for speech perception [11, 12]. It is also revealed that TE and TFS play an important role of improving intelligibility of noise-degraded speech [13, 14]. Therefore, AMS in the filterbank is suitable framework for speech enhancement, rather than AMS in the STFT. All existing researches on phase spectrum either emphasizes the importance of phase in speech enhancement or investigates the suitable size and shape of the window for phase manipulation. Hence, it is expected that temporal amplitude and phase manipulations as the ASM in the filterbank can drastically improve quality as well as intelligibility of noise degraded speech.

Motivated about the effectiveness of phase manipulation from the existing literature, the aim of this paper is to propose a speech enhancement scheme as the ASM on the filterbank. This scheme aims to enhance both the instantaneous amplitude and phase by using recursive Kalman filter in a Gammatone filterbank [15] as the AMS on the filterbank for speech enhancement. We deal with the instan-

This work was supported by an A3 foresight program made available by the Japan Society for the Promotion of Science. It was also partially supported by the Strategic Information and Communications R & D Promotion Programme (SCOPE; 131205001) of the Ministry of Internal Affairs and Communications (MIC), Japan.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

taneous amplitude and phase on the Gammatone filterbank because temporal smoothed information (amplitude and phase) are directly related to improve quality and intelligibility of speech. In addition, the Gammatone filterbank can be regarded as an approximation of human cochlear filterbank. The Kalman filter is of particular interest in smooth prediction method for dealing with the instantaneous amplitude and phase in sub-band. Since the transition matrices of the state-equation of both instantaneous amplitude and phase are unknown, it is difficult to set these transition matrices in the Kalman filtering for suitable speech enhancement.

In this paper, linear prediction (LP) analysis is used to set these transition matrices that are regarded as modulation characteristics of amplitude and phase in each sub-band. We incorporate the auto correlation function to derive transition matrices from the clean speech as a non-blind method. In addition, to make it more realistic model, we implement an off-line training phase without clean speech to set matrices as a blind method.

2. PROPOSED METHOD

Our proposed method for speech enhancement is intended to improve both instantaneous amplitude and instantaneous phase as the ASM on a Gammatone filterbank. The model consists of three steps: (i) Analysis stage, where instantaneous amplitude and instantaneous phase are extracted from the noisy speech by the Gammatone filterbank. (ii) Modification stage, where instantaneous amplitude and instantaneous phase are enhanced by a Kalman filter with linear prediction (LP) (with/without training phase). (iii) Re-synthesis by the inverse Gammatone filterbank.

First, only the noisy speech $y(t)$, where $y(t) = x(t) + n(t)$, is observed in the proposed model. Here, $x(t)$ indicates the clean speech and $n(t)$ represents a noise or the other signal. t is continuous time and m is sampling number ($m = 0, 1, 2, \cdots, M$; $t = m/F_s$) where M is the number of time samples and F_s is the sampling frequency. From the observed signal $y(t)$, instantaneous amplitude and instantaneous phase are extracted into the frequency components by the Gammatone filterbank (the number of channels is K). The output of the kth channel is represented as the analytical form by

$$Y_k(t) = Y_{1,k}(t) + Y_{2,k}(t) = A_k(t)\exp\left(j\omega_k t + j\phi_k(t)\right) \quad (1)$$

where $Y_{1,k}(t)$ and $Y_{2,k}(t)$ are the components of $x(t)$ and $n(t)$ that have passed through the filterbank, respectively. In addition, ω_k is the center frequency of the kth channel, $A_k(t)$ is the instantaneous amplitude, and $\phi_k(t)$ is the instantaneous phase of the noisy speech.

The Kalman filter, an efficient computational recursive solution for estimating a signal, is widely used in fields related to statistical processing. It not only exploits the statistical characteristics of signal and noise but also utilizes the speech production model based on the source-filter model. Therefore, we believe that the Kalman filter can be used to remove noise from both instantaneous amplitude and instantaneous phase.

The state and observation equations are the main equations in the Kalman filter, which are defined as:

$$\begin{aligned} \boldsymbol{S}[m] &= \boldsymbol{F}\boldsymbol{S}[m-1] + \boldsymbol{W}[m], & (2)\\ \boldsymbol{O}[m] &= \boldsymbol{H}\boldsymbol{S}[m] + \boldsymbol{V}[m], & (3)\end{aligned}$$

where $\boldsymbol{S}[m]$ is the state in discrete time m and $\boldsymbol{O}[m]$ is the observation in discrete time m. $\boldsymbol{W}[m]$ and $\boldsymbol{V}[m]$ are driving noise and observation noise that are assumed to be Gaussian white noise. The state equations of kth channel for instantaneous amplitude and instantaneous phase are as follows

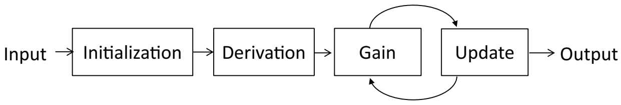

Fig. 1. Steps of Kalman filtering.

$$\begin{aligned} \boldsymbol{S}_{A,k}[m] &= \boldsymbol{F}_A\boldsymbol{S}_{A,k}[m-1] + \boldsymbol{W}_{A,k}[m], & (4)\\ \boldsymbol{S}_{\phi,k}[m] &= \boldsymbol{F}_\phi\boldsymbol{S}_{\phi,k}[m-1] + \boldsymbol{W}_{\phi,k}[m], & (5)\end{aligned}$$

where $\boldsymbol{S}_{A,k}[m]$ and $\boldsymbol{S}_{\phi,k}[m]$ are the states of instantaneous amplitude and instantaneous phase of kth channel respectively. Since, instantaneous amplitude and phase can be modeled with an auto regressive (AR) process of order p, the state vector can be represented as: $\boldsymbol{S}_{A,k}[m] = [S_{A,k}[m-p+1], S_{A,k}[m-p+2], \cdots, S_{A,k}[m]]^T$, and $\boldsymbol{S}_{\phi,k}[m] = [S_{\phi,k}[m-p+1], S_{\phi,k}[m-p+2], \cdots, S_{\phi,k}[m]]^T$, where $\boldsymbol{F}_A$ and $\boldsymbol{F}_\phi$ are the transition matrices that can be obtained by the linear prediction method. $\boldsymbol{W}_{A,k}[m]$ and $\boldsymbol{W}_{\phi,k}[m]$ are assumed to be Gaussian white noise of kth channel, and the variances of $\boldsymbol{W}_{A,k}[m]$ and $\boldsymbol{W}_{\phi,k}[m]$ are Q_A and Q_ϕ, respectively.

The observation equations for the instantaneous amplitude and instantaneous phase are defined as

$$\begin{aligned} \boldsymbol{O}_{A,k}[m] &= \boldsymbol{H}_A\boldsymbol{S}_{A,k}[m] + \boldsymbol{V}_{A,k}[m], & (6)\\ \boldsymbol{O}_{\phi,k}[m] &= \boldsymbol{H}_\phi\boldsymbol{S}_{\phi,k}[m] + \boldsymbol{V}_{\phi,k}[m], & (7)\end{aligned}$$

where $\boldsymbol{O}_{A,k}[m]$ and $\boldsymbol{O}_{\phi,k}[m]$ are the observed instantaneous amplitude and phase of the noisy speech at time m in kth channel, respectively. $\boldsymbol{H}_A$ and $\boldsymbol{H}_\phi$ are the observation matrices, which are $[0, 0, \cdots, 1]$ in this research. $\boldsymbol{V}_{A,k}[m]$ and $\boldsymbol{V}_{\phi,k}[m]$ are observation noise (Gaussian white noise) and the variances of $\boldsymbol{V}_{A,k}[m]$ and $\boldsymbol{V}_{\phi,k}[m]$ are R_A and R_ϕ, respectively.

We need five steps to calculate the optimal estimations for both instantaneous amplitude and instantaneous phase. The steps of Kalman filtering are shown in Fig. 1.

Step 1: We set the initial state vectors $\boldsymbol{S}_{A,k}[1|1] = [10^{-12} \cdots 10^{-12}]$ and $\boldsymbol{S}_{\phi,k}[1|1] = [10^{-12} \cdots 10^{-12}]$. Then, we can estimate the instantaneous amplitude and phase of clean speech at next time step from the initial state vector. Repeating this step, we can estimate the instantaneous amplitude and phase of clean speech of time m from the optimal estimation of time $m-1$.

$$\boldsymbol{S}_{A,k}[m|m-1] = \boldsymbol{F}_A\hat{\boldsymbol{S}}_{A,k}[m-1|m-1], \quad (8)$$

$$\boldsymbol{S}_{\phi,k}[m|m-1] = \boldsymbol{F}_\phi\hat{\boldsymbol{S}}_{\phi,k}[m-1|m-1]. \quad (9)$$

Step 2: We define the error covariance matrix $\boldsymbol{P}_A[m|m] = E[(\hat{\boldsymbol{S}}_{A,k}[m|m] - S_{A,k}[m|m])(\hat{\boldsymbol{S}}_{A,k}[m|m] - S_{A,k}[m|m])^T]$ and $\boldsymbol{P}_\phi[m|m] = E[(\hat{\boldsymbol{S}}_{\phi,k}[m|m] - S_{\phi,k}[m|m])(\hat{\boldsymbol{S}}_{\phi,k}[m|m] - S_{\phi,k}[m|m])^T]$. We update the covariance of $\boldsymbol{S}_{A,k}[m|m-1]$ and $\boldsymbol{S}_{\phi,k}[m|m-1]$ from the covariance of $\hat{\boldsymbol{S}}_{A,k}[m-1|m-1]$ and $\hat{\boldsymbol{S}}_{\phi,k}[m-1|m-1]$. We set the initial $\boldsymbol{P}_A[1|1] = \mathrm{diag}(R_A \cdots R_A)$ and $\boldsymbol{P}_\phi[1|1] = \mathrm{diag}(R_\phi \cdots R_\phi)$.

$$\begin{aligned} \boldsymbol{P}_A[m|m-1] &= \boldsymbol{F}_A\boldsymbol{P}_A[m-1|m-1]\boldsymbol{F}_A{}^T + Q_A, & (10)\\ \boldsymbol{P}_\phi[m|m-1] &= \boldsymbol{F}_\phi\boldsymbol{P}_\phi[m-1|m-1]\boldsymbol{F}_\phi{}^T + Q_\phi. & (11)\end{aligned}$$

Step 3: We estimate the current value and smooth the previous value as follows.

$$\begin{aligned} \boldsymbol{S}_{A,k}[m|m] &= \boldsymbol{S}_{A,k}[m|m-1] + \boldsymbol{e}_A, & (12)\\ \boldsymbol{S}_{\phi,k}[m|m] &= \boldsymbol{S}_{\phi,k}[m|m-1] + \boldsymbol{e}_\phi, & (13)\end{aligned}$$

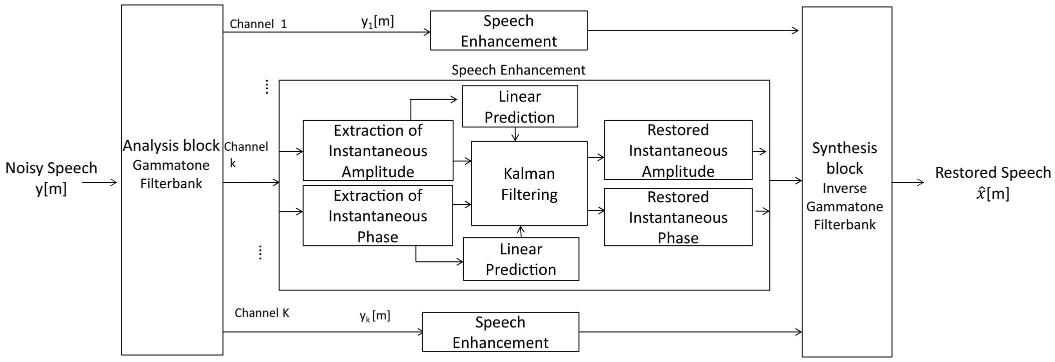

Fig. 2. Block diagram of proposed method for speech enhancement.

where $\boldsymbol{G}_A[m]$ and $\boldsymbol{G}_\phi[m]$ are the Kalman gains. Here, $\boldsymbol{e}_A = \boldsymbol{G}_A[m] \times(\boldsymbol{O}_{A,k}[m] - \boldsymbol{H}_A\boldsymbol{S}_{A,k}[m|m-1])$ and $\boldsymbol{e}_\phi = \boldsymbol{G}_\phi[m](\boldsymbol{O}_{\phi,k}[m] - \boldsymbol{H}_\phi\boldsymbol{S}_{\phi,k}[m|m-1])$ are called new informations. In this step, the value of the previous estimations is innovative.

Step 4: Updating the Kalman gains.

$$\boldsymbol{G}_A[m] = \boldsymbol{P}_A[m|m-1]{\boldsymbol{H}_A}^T/(\boldsymbol{H}_A\boldsymbol{P}[m|m-1]{\boldsymbol{H}_A}^T + R_A), \quad (14)$$

$$\boldsymbol{G}_\phi[m] = \boldsymbol{P}_\phi[m|m-1]{\boldsymbol{H}_\phi}^T/(\boldsymbol{H}_\phi\boldsymbol{P}_\phi[m|m-1]{\boldsymbol{H}_\phi}^T + R_\phi). \quad (15)$$

Step 5: Updating the covariances of $\boldsymbol{S}_{A,k}[m|m]$ and $\boldsymbol{S}_{\phi,k}[m|m]$.

$$\boldsymbol{P}_A[m|m] = (\boldsymbol{I} - \boldsymbol{G}_A[m]\boldsymbol{H}_A)\boldsymbol{P}_A[m|m-1], \quad (16)$$

$$\boldsymbol{P}_\phi[m|m] = (\boldsymbol{I} - \boldsymbol{G}_\phi[m]\boldsymbol{H}_\phi)\boldsymbol{P}_\phi[m|m-1], \quad (17)$$

where $\boldsymbol{I}$ is the unit matrix. A block diagram of the proposed method is shown in Fig. 2.

2.1. Linear Prediction Method

We developed a linear prediction method to extract the LP coefficients for Kalman filtering from the clean speech, which is referred as to non-blind Kalman filtering. In addition, we also developed a LP method that requires the training phase to estimate the LP coefficients, which is known as blind Kalman filtering. We can utilize blind Kalman filtering, whenever the clean speech is not available.

In the non-blind linear prediction method, we assume that the sampling sequences of clean speech's amplitude and phase are $S_{A,k}[m]$ and $S_{\phi,k}[m]$, where $m = 0, 1, 2, \cdots, M$. This can be regarded as the output of a p-th order AR process. The model of linear prediction can be represented as: $S_{A,k}[m] = \sum_{i=1}^{p} a_i S_{A,k}[m-i]$ and $S_{\phi,k}[m] = \sum_{i=1}^{p} b_i S_{\phi,k}[m-i]$. Here, $S_{A,k}[m]$ and $S_{\phi,k}[m]$ are the optimal estimation of $S_{A,k}[m]$ and $S_{\phi,k}[m]$ under the principle of minimum mean square error (MMSE), $\{a_1, a_2, \cdots, a_p\}$ and $\{b_1, b_2, \cdots, b_p\}$ are LP coefficients and p is the prediction order.

There are two types of methods for calculating the LP coefficients: an autocorrelation (AC) method and a covariance method. We chose the AC method to calculate the LP coefficients, and $\{a_i\}$ $(i = 1, 2, \cdots, p)$ and $\{b_i\}$ $(i = 1, 2, \cdots, p)$ could be obtained by solving the Yule-Walker equation as $R[q_a] - \sum_{i=1}^{p} a_i R[q_a - i] = 0$ and $R[q_b] - \sum_{i=1}^{p} b_i R[q_b - i] = 0$. Here, $R[q_a]$ and $R[q_b]$ are the AC functions of instantaneous amplitude and phase of the clean speech, $S_{A,k}[m]$ and $S_{\phi,k}[m]$, $R[q_a] = E\{S_{A,k}[m]S_{A,k}[m-q_a]\}$ and $R[q_b] = E\{S_{\phi,k}[m]S_{\phi,k}[m-q_b]\}$, where $E\{\cdot\}$ is the expectation. Then, we can obtain the transition matrix $\boldsymbol{F}_A$ and $\boldsymbol{F}_\phi$ for estimating instantaneous amplitude and instantaneous phase by using the Kalman filter.

For the blind LP prediction, we utilize a training phase for obtaining LP coefficients. We trained LP coefficients by using the autocorrelation method, prior to fitting them to the all subjective data. During the training phase, we found that the LP coefficients had some similarities. By utilizing this feature, we set LP coefficients to a single value for Kalman filtering. We also checked that these similarities were not sensitive to the specific gender and individuality of the speaker. Whenever, the clean speech was not available, we used the trained LP coefficients for Kalman filtering.

3. EVALUATIONS

To evaluate the effectiveness of both proposed methods, we carried out experiments using 12 different Japanese sentences uttered by male and female speakers from the ATR database [16]. The signal to noise ratios (SNRs) between $x(t)$ and $n(t)$ were fixed at from 20 dB to -10 dB at intervals of 10 dB. All noisy signals $y(t)$ were generated by adding $x(t)$ with $n(t)$. We used a Gammatone filterbank [15] to divide the signal into 128 channels ($K = 128$). We used the sampling frequency (F_s) of 20 kHz. We utilized a 25-ms-long rectangular window. The LP order, p, was set to 12.

We have evaluated the improvement of the restored speech by measuring the signal to error ratio (SER). SER shows the level of the error that we can reduce. SER is defined as follows

$$\mathrm{SER}(x_k, \hat{x}_k) = 10 \log_{10} \frac{\int_0^T (x_k(t))^2 dt}{\int_0^T (x_k(t) - \hat{x}_k(t))^2 dt}, \quad (18)$$

where $x_k(t)$ is the clean speech of kth channel and $\hat{x}_k(t)$ is the restored speech of kth channel. Figure 3 shows the improvement in SER in each channel using non-blind method (non-blind Kalman filtering) under the mentioned white noise conditions. In the figure, the height of the bar indicates the mean value of the improvement in SER. All the channels have positive improvement in SER in 20, 10, 0, and -10 dB noise conditions, excepting with case of higher channels in 20-dB condition. This is because signal components in higher channels in 20-dB condition are almost similar to those of clean signal. Thus, it is easy to see that the proposed method can effectively reduce the noise in both instantaneous amplitude and phase.

The performance of the blind method (blind Kalman filtering) is shown in Fig. 4. The results prove that blind Kalman filter with the trained LP coefficients also works well, thus we always obtain positive improvements in SER in all noise conditions, excepting with the same case mentioned in the above. Moreover, we also choose the Wiener filtering method (Scalart-Filho algorithm) under the same conditions to compare its effectiveness with that of our proposed

Table 1. Comparison of result of PESQ and SNR loss (averaged values).

SNR \ Method	Proposed (Non-blind)		Proposed (Blind)		SS [1]		MMSE [2]		Wiener filter [3]	
	PESQ	SNR loss	PESQ	SNR loss	PESQ	SNR loss	PESQ	SNR loss	PESQ	SNR loss
20 dB	3.27	0.531	3.23	0.536	2.40	0.802	2.74	0.848	2.46	0.882
10 dB	3.15	0.601	2.93	0.628	1.45	0.928	2.00	0.938	1.35	0.974
0 dB	2.92	0.665	2.31	0.714	1.28	0.947	1.30	0.948	1.25	0.968
−10 dB	2.58	0.720	1.54	0.788	1.01	0.949	1.15	0.950	0.96	0.953

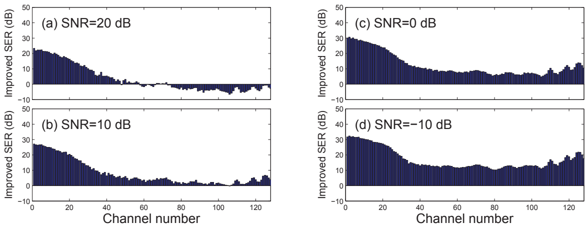

Fig. 3. Improvement in SER by using non-blind Kalman filtering.

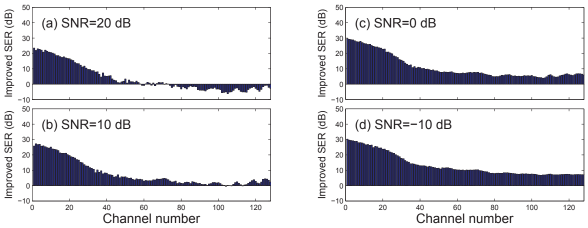

Fig. 4. Improvement in SER by using blind Kalman filtering.

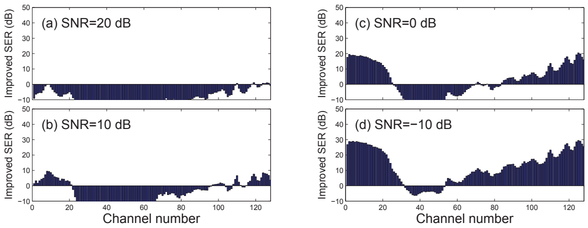

Fig. 5. Improvement in SER by using Wiener filter.

method. Based on the results in Fig. 5, we can see that both of our proposed methods can obviously improve the SER much more than the Wiener filtering method.

To evaluate the quality and intelligibility of the restored speech, we calculated the perceptual evaluation of sound quality (PESQ) [17, 18] and SNR loss [19] for all stimuli that we used the above evaluations. PESQ in the objective difference grades (ODGs) that covers from -0.5 (very annoying) to 4.5 (imperceptible) was used to evaluate subjective quality. SNR loss that ranges from 0.0 to 1.0 was used to evaluate intelligibility of speech. SNR losses (0 to 1.0) are corresponded to the percent correctness (100% to 0%). The results of that objective measures are listed in Table I. The results indicate that both of our proposed methods provide better quality and improved intelligibility in the restored speech much more than the existing speech enhancement methods. From the result of evaluations, we can say that the proposed method can effectively reduce the noise from both the amplitude and phase and also improve the quality and speech intelligibility.

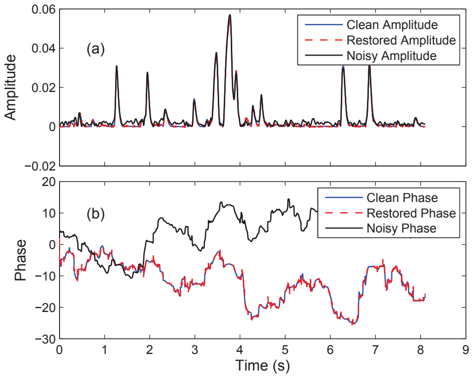

Fig. 6. Examples of (a) the instantaneous amplitude and (b) phase in a sub-band (channel $k = 43$) by the proposed scheme.

Figure 6 shows the restored instantaneous amplitude and phase in a particular sub-band by the proposed method (blind Kalman filtering). We can see that the restored amplitude and phase are perfectly matched with the clean amplitude and phase.

Discussion on relation to prior work: Many speech enhancement methods have already been proposed. However, all existing methods concentrate on the enhancement of amplitude only and left the phase part intact. Thus, noise is still present in the phase. Moreover, the importance of phase has been reported in several studies. Comparing with the other methods, in this study, we considered noise reduction from both instantaneous amplitude and phase from the sub-band representation for better speech enhancement. These guarantee better speech enhancement in the results.

4. CONCLUSION

In this paper, we proposed a speech enhancement scheme by using the Kalman filter with/without training phase in the sub-bands as the ASM on the Gammatone filterbank. We then presented greater speech enhancement scheme by improving the instantaneous amplitude and instantaneous phase simultaneously. Our simulation results revealed that the proposed scheme outperforms the existing conventional speech enhancement algorithms in terms of improvements for speech quality and intelligibility. We believe that this is the effect of combining phase enhancement with amplitude enhancement in the sub-band representation. Although the proposed method always shows improvement, but still, future research is required to ensure improvement in all channels for high SNR conditions. We plan to investigate whether the mask of data or speaker, or gender dependent trained LP coefficients can provide solutions. We will also plan to investigate whether the proposed scheme can be extended to be a speech enhancement in noisy reverberant environments.

5. REFERENCES

[1] S. F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," *IEEE Trans. ASSP*, vol. ASSP-27(2), pp. 113–120, 1979.

[2] Y. Ephraim and D. Mlah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. ASSP*, vol. ASSP-32, no. 6, pp. 1109–1211, Dec. 1984.

[3] P. Scalart and J. V. Filho, "Speech enhancement based on a priori signal to noise estimation," *Proc. ICASSP1996*, pp. 629-623, 1996.

[4] J. Ming, R. Srinivasan and D. Crookes, "A Corpus-Based Approach to Speech Enhancement From Nonstationary Noise," *IEEE Trans. ASSP*, vol. 19, no.4, pp. 822–836, 2011.

[5] D. L. Wang and J. S. Lim, "The unimportance of phase in speech enhancement," *IEEE Trans. ASSP*, vol. ASSP-30, no. 2, pp. 113–120, 1979.

[6] B. J. Shannon and K. K. Paliwal, "Role of Phase Estimation in Speech Enhancement," *Proc. INTERSPEECH 2006-ICSLP*, Pittsburgh, Pennsylvania, pp. 1427–1430, 2006.

[7] K. K. Paliwal and L. D. Alsteris, "On the usefulness of STFT phase spectrum in human stening test," *Speech Communication*, vol. 45, no. 2, pp. 153–170, 2005.

[8] K. K. Paliwal, K. Wojcicki and B. J. Shannon, "The importance of phase in speech enhancement," *Speech Communication*, vol. 53, no. 4, pp. 465–494, 2011.

[9] J. L. Roux, N. Ono and S. Sagayama, "Explicit consistency constraints for STFT spectrograms and their application to phase reconstruction," *Proc. Statistical and Perceptual Audition (SAPA)* , pp. 23–28, 2008.

[10] P. C. Loizou and G. Kim, "Reasons why Current Speech-Enhancement Algorithms do not Improve Speech Intelligibility and Suggested Solutions," *IEEE Trans. Audio, Speech, and Lang. Processing*, vol. 19, no. 1, pp. 47–56, 2011.

[11] R. Drullman, "Temporal envelope and fine structure cues for speech intelligibility," *J. Acoust. Soc. Am.*, vol. 97, no. 1, pp. 585–592, 1995.

[12] B. C. J. Moore, "The Role of Temporal Fine Structure Processing in Pitch Perception, Masking, and Speech Perception for Normal-Hearing and Hearing-Impaired People," *J. Association for Research in Otolaryngology*, vol. 9, pp. 399–406, 2008.

[13] J. Swaminathan, "The role of envelope and temporal fine structure in the perception of noise degraded speech," Ph.D Thesis, Purdue University, 2010.

[14] J. Swaminathan and M. G. Heinz. "Psychophysiological Analyses Demonstrate the Importance of Neural Envelope Coding for Speech Perception in Noise." *Journal of Neuroscience*, vol. 32, no. 5, pp. 1747–1756, 2012.

[15] M. Unoki and M. Akagi, "A Method of signal extraction from noisy signal based on auditory scene analysis," *Speech Communication*, vol. 27, pp. 261–279, 1999.

[16] K. Takeda et al, "Speech database user's manual," in ATR Technical Report TR-I-0028, 2010.

[17] Y. Hu and P. C. Loizou, "Evaluation of objective measures for speech enhancement," *Proc. Interspeech 2006*, pp. 1447–1450, 2006.

[18] Y. Hu and P. C. Loizou, "Evaluation of Objective Quality Measures for Speech Enhancement," *IEEE Trans. Audio, Speech, and Lang. Processing*, vol. 16, no. 1, pp. 229–238, 2008.

[19] J. Ma and P. C. Loizou, "SNR loss: A new objective measure for predicting the intelligibility of noise-suppressed speech," *Speech Communication*, vol. 53, pp. 340–359, 2011.