第32回信号処理シンポジウム
2017年11月8日～10日（盛岡）

Analysis of Complex Kalman Filtering in Speech Enhancement

Nguyen Minh Dat　　Masashi Unoki
School of Information Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa, 923-1292 Japan
E-mail: {datnm, unoki}@jaist.ac.jp

Abstract This paper proposes a restoration scheme for the instantaneous amplitude and phase of speech signal by using the complex version of Kalman filtering in speech enhancement. The previous studies have proved that restoring the instantaneous amplitude as well as instantaneous phase by Kalman filtering with linear prediction (LP) on Gammatone filterbank plays a significant role in speech enhancement. However, the existing problem is the individual restoring for the instantaneous amplitude and phase. Thus, this paper aims to solve this problem by studying the feasibility of restoring both instantaneous amplitude and phase simultaneously base on the complex Kalman filtering. The proposed method concentrates on analyzing the separation of real and imaginary parts of the analytical speech signal simultaneously. The complex Kalman filtering with LP are applied to the analytical speech signal. The expected outcomes are improvements in the signal to error ratio, correlation and the quality as well as intelligibility of speech signals. Results of evaluations showed that the proposed scheme could effectively improve the previous one in noisy reverberant environments.

1 Introduction

In voice communication, the quality and intelligibility of speech are very important, which are degraded under the influence of background noise and reverberation. In particular, the performance of applications such as speech coders, hearing aids and automatic speech recognition (ASR) systems could be reduced. Therefore, these effects need to be removed simultaneously with more explicit consideration.

In previous studies, various methods which were concerned with speech enhancement have already been proposed to remove the influence of noise or reverberation to improve the quality or intelligibility of speech signals. Among the many developed methods, short-time Fourier transform based analysis-modification-synthesis (STFT-AMS) framework was widely used for speech enhancement [1]. Based on the survey of noise reduction methods, both single and multi channel approaches were considered. In these cases, the spectral subtraction (SS) has been shown to effectively suppress stationary noise [2]. This method performs subtraction of an estimated noise magnitude spectrum from a noisy speech magnitude spectrum. In addition, the methods which related to minimum mean-square error short-time spectral amplitude (MMSE-STSA) estimator [3] and the Wiener filtering algorithm [4] have drawn a great deal of attention.

Besides, the effectiveness of phase manipulation is of interest to researchers in speech enhancement. The importance of phase information in speech enhancement was shown in [5], [6]. There presented the modulation-phase-only experiments that proved the modulation phase spectrum gives an important role in the intelligibility of speech signal. Many researchers proposed the speech enhancement scheme on the filterbank to enhance both the instantaneous amplitude and phase by using recursive Kalman filter in a Gammatone filterbank (GTFB) [7].

Kalman filter is usually used in speech enhancement to improve and restore the instantaneous amplitude as well as instantaneous phase. Paliwal and Basu were the first researchers who applied the Kalman filter in speech enhancement [8]. This method is the most suitable for reduction of white noise with Kalman assumption. Kalman flter is a mathematical procedure which operates through a prediction and correction mechanism. Kalman flter combines all the available data measured, plus the knowledge of the system and the measurement devices, to produce an estimation of the desired variables in such a manner that the error is statistically minimized [9]. In addition, Kalman filter is of particular interest in smooth prediction method for dealing with the instantaneous amplitude and phase in sub-band. It can be viewed as a joint estimator for both the magnitude and phase spectrum of speech, under non-stationary condition [10].

Nower *et al.*'s methods [11], [12] dealt with the instantaneous amplitude and phase on the GTFB because temporal smoothed information (amplitude and phase) is directly related to improve the quality and intelligibility of speech. In addition, these research carefully analyzed the assumption of Kalman filter for speech enhancement in noisy environments. The modulation characteristics of amplitude and phase in each sub-band of Kalman filtering were considered by using the linear prediction to derive these coefficients.

Moreover, in reverberant environments, the dereverberation is quite important. A few solutions that have proposed related to STFT-AMS framework, cepstral mean normalization (CMN) [13], which could suppress the effect of early reverberation by normalizing cepstral features, has been shown to be the simplest and most effective method. However, this method still exist the problem, which is not effective with the pres-

ence of the late reverberation. Another method based on multiple-step linear prediction (MSLP) was proposed by Kinoshita *et al.* [14]. This method estimates the late reverberation by long-term MSLP and then these were suppressed by subsequent SS.

Almost previously studied methods cannot work well in noisy reverberant environments to completely remove the influence of noise and reverberation simultaneously. Liu *et al.*'s methods [15]–[17] tried to solve this problem completely in order to improve the quality as well as intelligibility of s peech signals. In [15], the derivation of the accurate transition matrices was deeply concentrated. This is because they are quite important parameters in Kalman filtering from noisy reverberant speech. Besides, the enhancement performance of Kalman filter depends on the accuracy and reliability of transition matrices. These matrices were applied linear prediction (LP) algorithm for instantaneous amplitude and instantaneous phase individually. In addition, the effects of noise and reverberation were removed simultaneously with consideration of phase information and the effects of noise corresponding to additive and convolved noises (late reverberant speech) on instantaneous amplitude and phase can be removed by Kalman filtering with efficient LP and the early reflection effect can be removed by CMN. Thus, the quality and intelligibility of speech were improved. However, this method is still remaining issue that is how to consider and restore both instantaneous amplitude and phase in noisy reverberant environments simultaneously.

Therefore, the objective of this paper is to design complex version of Kalman filtering for speech enhancement in order to improve previous research works. This paper will use [15] as a preliminary study. The novel point is applying both instantaneous amplitude and phase simultaneously by analyzing the complex Kalman filtering. The proposed scheme is expected to work well and improve the quality and intelligibility of speech signal.

The rest of this paper is organized as follows. Section 2 explains the details of the proposed scheme in noisy reverberant environments. The complex Kalman filtering that is applied for instantaneous amplitude and phase simultaneously. Section 3 presents the objective and evaluation of the achieved results. Section 4 describes the discussion of this research. Section 5 makes the conclusion and future works.

2 Proposed scheme

2.1 Signal definition

The noisy reverberant speech $y_{NR}(t)$, where $y_{NR}(t) = x(t) * h(t) + n(t)$, is observed in [15]. Here, $x(t)$ is the clean speech, $h(t)$ is the room impulse response (RIR) and $n(t)$ is background noise. The RIR, $h(t)$, contains both effects of early reflection and late reverberation so that this can be represented as $h(t) = h_E(t) + h_L(t)$, where $h_E(t)$ is early reflection and $h_L(t)$ is late reverberation.

The output of the k-th sub-band, $Y_{NR,k}(t)$, is represented as the analytical form by:

$$\begin{aligned} Y_{NR,k}(t) &= Y_{NR,1,k}(t) + Y_{NR,2,k}(t), \\ &= A_{NR,k}(t) \exp\left(j\omega_k t + j\phi_{NR,k}(t)\right), \end{aligned} \tag{1}$$

where $Y_{NR,1,k}(t)$ and $Y_{NR,2,k}(t)$ are the components of noisy reverberant speech, respectively. $A_{NR,k}(t)$ and $\phi_{NR,k}(t)$ are the instantaneous amplitude and phase of the noisy reverberant speech $Y_{NR,k}(t)$, which are calculated as follows:

$$A_{NR,k}(t) = |\tilde{f}(c,t)|, \tag{2}$$

$$\phi_{NR,k}(t) = \int_0^t \left(\frac{d}{d\tau} \arg(\tilde{f}(c,\tau) - \omega_k) \right) d\tau, \tag{3}$$

where, $c = \alpha^{k-K/2}$, α is the scale of GTFB. $|\tilde{f}(c,t)|$ is the amplitude spectrum defined by the wavelet transform and $\arg(\tilde{f}(c,t))$ is the unwrapped phase spectrum defined by the complex wavelet transform.

In previous research [15], Kalman filter is applied for instantaneous amplitude $A_{NR,k}$ and phase $\phi_{NR,k}$ individually. In this paper, we proposed a scheme as an extension of the previous scheme. The block diagram of the proposed scheme for speech enhancement is shown in Fig. 1. We concentrated on the analysis of complex Kalman filtering that considers both instantaneous amplitude and phase simultaneously. The noisy reverberant speech, $Y_{NR,k}(t)$ from Eq. (1),

$$\begin{aligned} Y_{NR,k}(t) &= A_{NR,k}(t) \exp\left(j\omega_k t + j\phi_{NR,k}(t)\right) \\ &= A_{NR,k}(t) \cos\left(\omega_k t + \phi_{NR,k}(t)\right) \\ &\quad + jA_{NR,k}(t) \sin\left(\omega_k t + \phi_{NR,k}(t)\right), \end{aligned} \tag{4}$$

Hence,

$$Y_{NR,k}(t) = Y_{r,NR,k}(t) + jY_{i,NR,k}(t), \tag{5}$$

where $Y_{r,NR,k}(t) = A_{NR,k}(t) \cos\left(\omega_k t + \phi_{NR,k}(t)\right)$, and $Y_{i,NR,k}(t) = A_{NR,k}(t) \sin\left(\omega_k t + \phi_{NR,k}(t)\right)$ are components of real and imaginary parts, respectively.

The real and imaginary parts of speech signal will be applied to Kalman filtering individually. In other words, the instantaneous amplitude and phase of speech signals are applied simultaneously. This method is called as Complex Kalman filtering.

2.2 Complex Kalman filtering

The complex version of Kalman filtering is an algorithm, which applied for real and imaginary parts of speech signal, individually. The state equations of k-th sub-band for real and imaginary parts are defined as:

$$\boldsymbol{S}_{Y_r,k}[m] = \boldsymbol{F}_{Y_r,k} \boldsymbol{S}_{Y_r,k}[m-1] + \boldsymbol{W}_{Y_r,k}[m], \tag{6}$$

$$\boldsymbol{S}_{Y_i,k}[m] = \boldsymbol{F}_{Y_i,k} \boldsymbol{S}_{Y_i,k}[m-1] + \boldsymbol{W}_{Y_i,k}[m], \tag{7}$$

where m is sampling number ($m = 0, 1, 2, \cdots, M$; $t = m/F_s$), M is the number of time samples and F_s is the sampling frequency. $\boldsymbol{F}_{Y_r,k}$ and $\boldsymbol{F}_{Y_i,k}$ are the transition

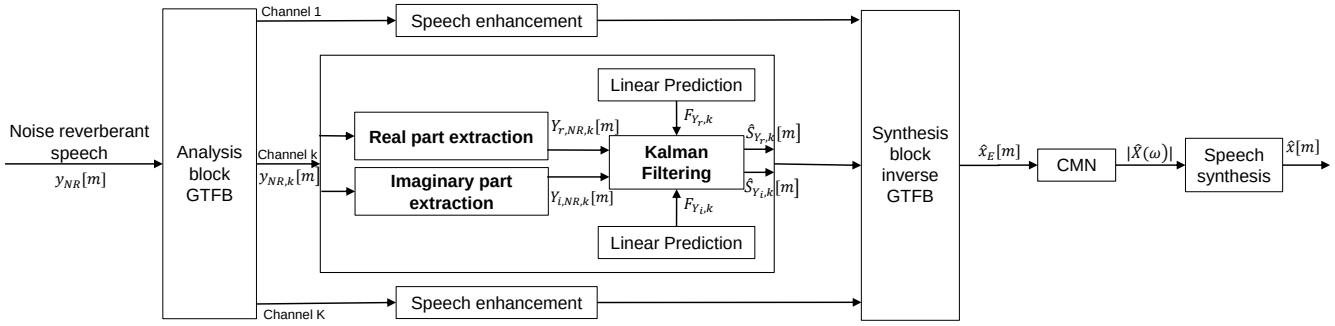

Figure 1: Block diagram of proposed scheme for speech enhancement.

matrices of k-th sub-band that can be obtained by the LP method. $\boldsymbol{W}_{Y_r,k}[m]$ and $\boldsymbol{W}_{Y_i,k}[m]$ are assumed to be Gaussian white noise of k-th sub-band, and the variances of $\boldsymbol{W}_{Y_r,k}[m]$ and $\boldsymbol{W}_{Y_i,k}[m]$ are $Q_{Y_r,k}$ and $Q_{Y_i,k}$, respectively. $\boldsymbol{S}_{Y_r,k}[m]$ and $\boldsymbol{S}_{Y_i,k}[m]$ are the state vectors of real and imaginary parts of early reverberant speech at sampling point m in k-th sub-band respectively.

The observation equations for the real and imaginary parts of k-th sub-band are defined as:

$$\boldsymbol{O}_{Y_r,k}[m] = \boldsymbol{H}_{Y_r}\boldsymbol{S}_{Y_r,k}[m] + \boldsymbol{V}_{Y_r,k}[m], \quad (8)$$

$$\boldsymbol{O}_{Y_i,k}[m] = \boldsymbol{H}_{Y_i}\boldsymbol{S}_{Y_i,k}[m] + \boldsymbol{V}_{Y_i,k}[m], \quad (9)$$

where $\boldsymbol{O}_{Y_r,k}[m]$ and $\boldsymbol{O}_{Y_i,k}[m]$ are the observed real and imaginary parts of the noisy reverberant speech at sampling point m in k-th sub-band. $\boldsymbol{H}_{Y_r}$ and $\boldsymbol{H}_{Y_i}$ are the observation matrices that are $[0, 0, \cdots, 1]$. $\boldsymbol{V}_{Y_r,k}[m]$ and $\boldsymbol{V}_{Y_i,k}[m]$ are observation noise (Gaussian white noise), and the variances of $\boldsymbol{V}_{Y_r,k}[m]$ and $\boldsymbol{V}_{Y_i,k}[m]$ are $R_{Y_r,k}$ and $R_{Y_i,k}$.

We calculate the optimal estimations for real and imaginary parts of the speech signal as follows **5** steps.

Step 1: Initial state vectors are set to be $\hat{\boldsymbol{S}}_{Y_r,k}[1|1] = [10^{-12}\cdots 10^{-12}]$ and $\hat{\boldsymbol{S}}_{Y_i,k}[1|1] = [10^{-12}\cdots 10^{-12}]$. These values are used to initialize the state vector only and will come close to the original state vector after a few iterations.

$$\hat{\boldsymbol{S}}_{Y_r,k}[m|m-1] = \boldsymbol{F}_{Y_r,k}\hat{\boldsymbol{S}}_{Y_r,k}[m-1|m-1], \quad (10)$$

$$\hat{\boldsymbol{S}}_{Y_i,k}[m|m-1] = \boldsymbol{F}_{Y_i,k}\hat{\boldsymbol{S}}_{Y_i,k}[m-1|m-1]. \quad (11)$$

The state vector of m is estimated from the state vector of $m-1$ under the principle of MMSE.

Step 2: The initial error covariance matrices $\boldsymbol{P}_{Y_r,k}[1|1] = \mathrm{diag}(R_{Y_r,k}[1]\cdots R_{Y_r,k}[1])$ and $\boldsymbol{P}_{Y_i,k}[1|1] = \mathrm{diag}(R_{Y_i,k}[1]\cdots R_{Y_i,,k}[1])$ are set as:

$$\boldsymbol{P}_{Y_r,k}[m|m-1] = \boldsymbol{F}_{Y_r,k}\boldsymbol{P}_{Y_r,k}[m-1|m-1]{\boldsymbol{F}_{Y_r,k}}^{\mathrm{T}} + Q_{Y_r,k}, \quad (12)$$

$$\boldsymbol{P}_{Y_i,k}[m|m-1] = \boldsymbol{F}_{Y_i,k}\boldsymbol{P}_{Y_i,k}[m-1|m-1]{\boldsymbol{F}_{Y_i,k}}^{\mathrm{T}} + Q_{Y_i,k}. \quad (13)$$

Step 3: The current values are estimated as:

$$\hat{\boldsymbol{S}}_{Y_r,k}[m|m] = \hat{\boldsymbol{S}}_{Y_r,k}[m|m-1] + \boldsymbol{e}_{Y_r,k}, \quad (14)$$

$$\hat{\boldsymbol{S}}_{Y_i,k}[m|m] = \hat{\boldsymbol{S}}_{Y_i,k}[m|m-1] + \boldsymbol{e}_{Y_i,k}. \quad (15)$$

Here, $\boldsymbol{e}_{Y_r,k} = \boldsymbol{K}_{Y_r,k}[m](\boldsymbol{O}_{Y_r,k}[m] - \boldsymbol{H}_{Y_r}\hat{\boldsymbol{S}}_{Y_r,k}[m|m-1])$ and $\boldsymbol{e}_{Y_i,k} = \boldsymbol{K}_{Y_i,k}[m](\boldsymbol{O}_{Y_i,k}[m] - \boldsymbol{H}_{Y_i}\hat{\boldsymbol{S}}_{Y_i,k}[m|m-1])$ are called innovation, where $\boldsymbol{K}_{Y_r,k}[m]$ and $\boldsymbol{K}_{Y_i,k}[m]$ are the Kalman gains of k-th sub-band.

Step 4: Update the Kalman gains by:

$$\boldsymbol{K}_{Y_r,k}[m] = \frac{\boldsymbol{P}_{Y_r,k}[m|m-1]{\boldsymbol{H}_{Y_r}}^{\mathrm{T}}}{(\boldsymbol{H}_{Y_r}\boldsymbol{P}_{Y_r,k}[m|m-1]{\boldsymbol{H}_{Y_r}}^{\mathrm{T}} + R_{Y_r,k})}, \quad (16)$$

$$\boldsymbol{K}_{Y_i,k}[m] = \frac{\boldsymbol{P}_{Y_i,k}[m|m-1]{\boldsymbol{H}_{Y_i}}^{\mathrm{T}}}{(\boldsymbol{H}_{Y_i}\boldsymbol{P}_{Y_i,k}[m|m-1]{\boldsymbol{H}_{Y_i}}^{\mathrm{T}} + R_{Y_i,k})}. \quad (17)$$

Step 5: Update the error covariance matrices by:

$$\boldsymbol{P}_{Y_r,k}[m|m] = (\boldsymbol{I} - \boldsymbol{K}_{Y_r,k}[m]\boldsymbol{H}_{Y_r})\boldsymbol{P}_{Y_r,k}[m|m-1], \quad (18)$$

$$\boldsymbol{P}_{Y_i,k}[m|m] = (\boldsymbol{I} - \boldsymbol{K}_{Y_i,k}[m]\boldsymbol{H}_{Y_i})\boldsymbol{P}_{Y_i,k}[m|m-1]. \quad (19)$$

where $\boldsymbol{I}$ is the unit matrix.

2.3 Linear prediction

In previous research [15], LP analysis was used to obtain transition matrices for instantaneous amplitude and phase individually. This paper still use this method to calculate transition matrices $\boldsymbol{F}_{Y_r,k}$ and $\boldsymbol{F}_{Y_i,k}$ in Eqs. (6) and (7) for real and imaginary parts respectively.

The LP coefficients for complex Kalman filtering are extracted from early reverberant speech that can be regarded as the output of a $\boldsymbol{p}$-th order auto-regressive process by an autocorrelation method as follows:

$$R[q_a] - \sum_{i=1}^{p} a_i R[q_a - i] = 0, \quad (20)$$

$$R[q_b] - \sum_{i=1}^{p} b_i R[q_b - i] = 0. \quad (21)$$

Here, $R[q_a]$ and $R[q_b]$ are the autocorrelation functions of the real and imaginary parts of early reverberant speech, $R[q_a] = E\{S_{Y_r,k}[m]S_{Y_r,k}[m-q_a]\}$ and $R[q_b] = E\{S_{Y_i,k}[m]S_{Y_i,k}[m-q_b]\}$, where $E\{\cdot\}$ is the expectation.

The transition matrices $\boldsymbol{F}_{Y_r,k}$ and $\boldsymbol{F}_{Y_i,k}$ for each k-th sub-band are as follows:

$$\boldsymbol{F}_{Y_r,k} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \hat{a}_p & \hat{a}_{p-1} & \hat{a}_{p-2} & \cdots & \hat{a}_1 \end{bmatrix}, \quad (22)$$

$$\boldsymbol{F}_{Y_i,k} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \hat{b}_p & \hat{b}_{p-1} & \hat{b}_{p-2} & \cdots & \hat{b}_1 \end{bmatrix}, \quad (23)$$

where $\hat{a}_p$ and $\hat{b}_p$ are the trained LP coefficients for real and imaginary parts, respectively.

3 Evaluation

3.1 Objective

In this section, we concentrate on the analysis and evaluation to show the effectiveness of the proposed scheme and then prove the proposed method improves the previous one in noisy reverberant environments.

First, we indicate the improvement in quality and intelligibility of restored speech signal by comparison between previous method and proposed one in terms of perceptual evaluation of sound quality (PESQ) [18] and SNR loss [19], respectively. PESQ in objective different grades (ODGs) that covers from -0.5 to 4.5 was used to evaluate subjective quality of the restored speech signals under noisy reverberation conditions. SER loss was also used to predict the improvement in speech intelligibility, which ranges from 0 to 1.0, corresponding to the percent correctness (100 to 0%), under noisy reverberant conditions. The evaluation results based on comparison of the noisy reverberant speech, the previous and proposed method.

In addition, the measurement and comparison of correlation (Corr.) and signal to error ratio (SER) were used to consider the restoration accuracy for instantaneous amplitude and phase. Correlation shows the similarity between the shapes of the clean instantaneous amplitude and phase and the restored instantaneous amplitude and phase. Meanwhile, SER shows the level of error that we can reduce in speech enhancement [15].

Correlation (Corr) and SER are defined as follows:

$$\mathrm{Corr}(x_k, \hat{x}_k) = \frac{\int_0^T (x_k(t) - \overline{x_k})\big(\hat{x}_k(t) - \overline{\hat{x}_k}\big)dt}{\sqrt{\big\{\int_0^T (x_k(t) - \overline{x_k})dt\big\}\big\{\int_0^T (\hat{x}_k(t) - \overline{\hat{x}_k})dt\big\}}}, \quad (24)$$

$$\mathrm{SER}(x_k, \hat{x}_k) = 10\log_{10}\frac{\int_0^T (x_k(t))^2 dt}{\int_0^T (x_k(t) - \hat{x}_k(t))^2 \mathrm{d}t}, \quad (25)$$

where $x_k(t)$ is the clean speech and $\hat{x}_k(t)$ is the restored speech of k-th sub-band.

3.2 Evaluations

First, Table 1 lists the results of comparison among the noisy reverberant speech, restored speech in the previous and proposed scheme in terms of PESQ and SNR loss under various reverberant conditions: SNR at 10 and 0 dB and reverberation times T_R at 0.5 and 2 s. From the results of PESQ and SNR loss, we can see the quality and intelligibility of speech signal are improved. These results also show that the phase information had an important role in the improvement on the intelligibility of speech signal. Furthermore, we easily recognize the results of proposed scheme are better than previous scheme. In the reverberant condition, with SNR = 0 dB and $T_R = 0.5$, the proposed method works most effectively.

Figures 2 to 5 illustrate the comparison of the improvements in Corr. and SER under various reverberant conditions: SNR at 10 and 0 dB and reverberation times T_R at 0.5 and 2 s in noisy reverberant environments.

For the improvements in Corr., although this method has not reduced so much the negative parts, but in the positive parts, the height of bar, which is the level of similarity between the clean and restored instantaneous amplitude and phase, is higher than previous method under the various reverberant conditions. For improvements in SER, the height of bar indicates the improvements in mean values and magnitude of estimated errors was reduced. The orange color parts of bars are the improvements in proposed method. However, this effectiveness probably occurs at the low frequency channels, especially the most effective at the 8th channel. We can explain that the channels in lower frequencies are not sensitive so much or do not fluctuate rapidly, so we can estimate the correct signal easier.

4 Discussion

In this paper, there are a few key points that need to be discussed.

There are two main differences between previous and proposed schemes. First, we designed and applied the complex Kalman filtering for both real and imaginary parts of speech signals. In other words, we applied both instantaneous amplitude and phase simultaneously. Second, we concentrated on the training LP coefficients from early reverberant speech to consider the accuracy of transition matrices for calculation in the complex Kalman filtering. The each sub-band has a different transition matrix.

In addition, the importance of phase information that directly affects to the quality and intelligibility of speech signal was discussed in previous method. However, in this paper, we only focus on the improvement in terms of signal to error ratio (SER), perceptual evaluation of sound quality (PESQ) and SNR loss for the quality and intelligibility of speech signal respectively. The results of evaluations showed that the proposed method works more effective than the previous one. Therefore, the ac-

Table 1: Comparisons: PESQ and SNR loss (average values).

SNR/ T_R	Methods					
	Noisy reverberant		Previous scheme		Proposed scheme	
	PESQ	SNR loss	PESQ	SNR loss	PESQ	SNR loss
10 dB/0.5 s	1.85	0.91	2.72	0.70	2.75	0.68
10 dB/2 s	1.38	0.93	2.51	0.73	2.58	0.71
0 dB/0.5 s	1.45	0.94	2.31	0.74	2.41	0.72
0 dB/2 s	1.09	0.95	2.19	0.75	2.22	0.73

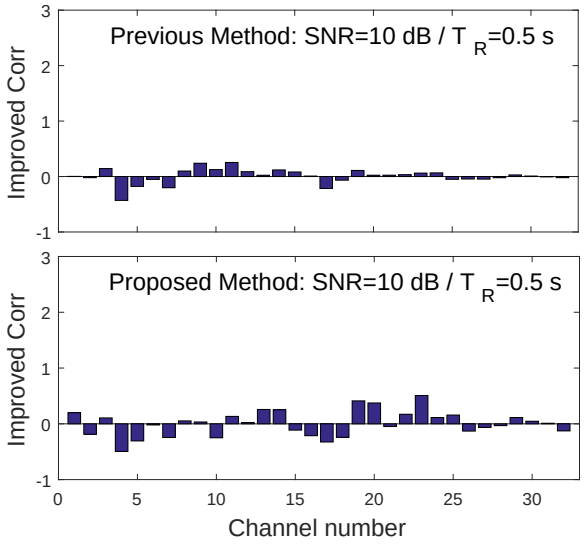

Figure 2: Improved Corrs in restoration accuracy in noisy reverberant environments under $T_R = 0.5$ s and SNR = 10 dB.

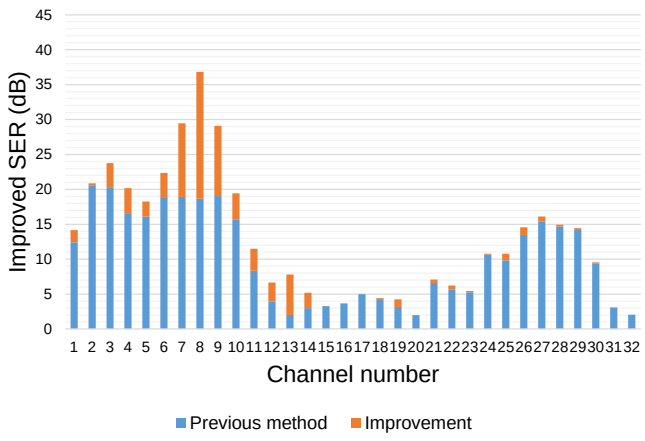

Figure 3: Improved SERs in restoration accuracy in noisy reverberant environments under $T_R = 0.5$ s and SNR = 10 dB.

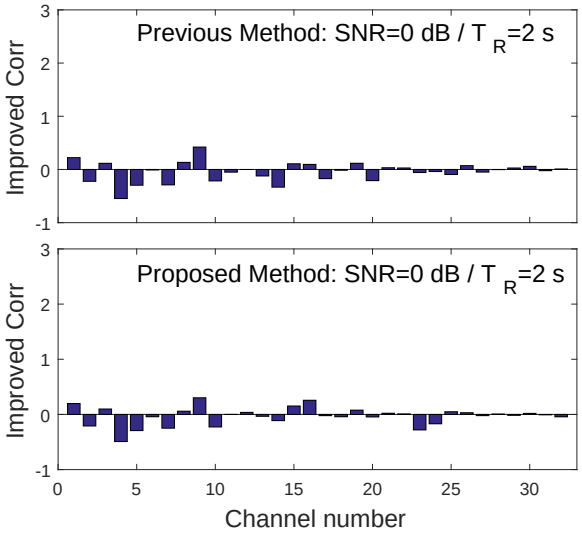

Figure 4: Improved Corrs in restoration accuracy in noisy reverberant environments under $T_R = 2$ s and SNR = 0 dB.

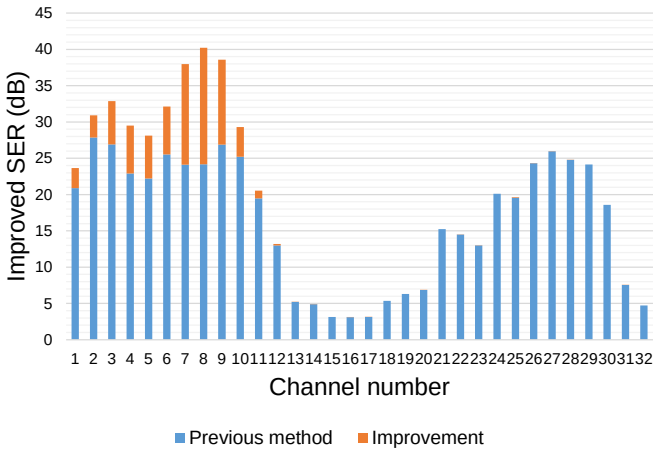

Figure 5: Improved SERs in restoration accuracy in noisy reverberant environments under $T_R = 2$ s and SNR = 0 dB.

curacy of instantaneous amplitude and phase information should be analyzed more explicitly.

Moreover, the effectiveness of the training LPC algorithm or other mechanisms as CMN, speech synthesis in the block scheme should be considered more explicitly with the input and output signal of each part. These results will help we can easily evaluate the impacts and effects in the principle of proposed scheme.

5 Conclusion

In this paper, we proposed a scheme for speech enhancement in noisy reverberant environments by using a complex Kalman filter with applying both the instantaneous amplitudes and phases simultaneously. The results of objective evaluations revealed that the proposed scheme can greatly improve the previous one in terms of PESQ and SNR loss, which directly related to the quality and intelligibility of speech signal in noisy reverberant environments. Besides, the results related to correlation and SER also are presented explicitly.

For future work, to improve the quality and intelligibility of speech signal in noisy reverberant environments more effectively, we can extend this scheme by combination of the real and imaginary parts simultaneously instead of applying the instantaneous amplitude and phase or real and imaginary parts in each sub-band, individually. Therefore, this research is as an open key to researcher in the near future to continue to contribute much to speech enhancement.

Acknowledgment

This research was supported by a Grant-in-Aid for challenging Exploratory Research (No. 16K12458) and Innovative Areas (No. 16H01669) from MEXT, Japan.

References

[1] M. Parchami, W. Zhu, B. Champagne, and E. Plourde, "Recent developments in speech enhancement in the short-time Fourier transform domain," *IEEE Circuits and Systems Magazine,* vol. 6, no. 3, pp. 45–77, 2016.

[2] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust. Speech Signal Process.,* vol. 27, no. 2, pp. 113–120, 1979.

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

[4] P. Scalart and J. V Filho, "Speech enhancement based on a priori signal to noise estimation," *IEEE Int. Conf.,* vol. 2, pp. 629–632, 1996.

[5] B. J. Shannon and K. K. Paliwal, "Role of phase estimation in speech enhancement," *Proc. INTERSPEECH 2006,* pp. 1427–1430, 2006.

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

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

[8] K. Paliwal and A. Basu, "A speech enhancement method based on Kalman filtering," *ICASSP 87,* vol. 12, pp. 177–180, 1987.

[9] M. Mathe, S. P. Nandyala, and T. Kishore Kumar, "Speech enhancement using Kalman filter for white, random and color noise," *Int. Conf. Devices, Circuits Syst. ICDCS,* pp. 195–198, 2012.

[10] S. So and K. K. Paliwal, "Modulation-domain Kalman filtering for single-channel speech enhancement," *Speech Commun.,* vol. 53, no. 6, pp. 818–829, 2011.

[11] N. Nower, Y. Liu, and M. Unoki, "Restoration of instantaneous amplitude and phase using Kalman filter for speech enhancement," *Proc. ICASSP2014*, pp. 4666–4670, 2014.

[12] N. Nower, Y. Liu, and M. Unoki, "Restoration scheme of instantaneous amplitude and phase using Kalman filter with efficient linear prediction for speech enhancement," *Speech Commun.,* vol. 70, pp. 13–27, June 2015.

[13] M. Wu and D. Wang, "A two-stage algorithm for one microphone reverberant speech enhancement," *IEEE Trans. Audio. Speech. Lang. Process.*, vol. 14, no. 3, pp. 774–784, 2006.

[14] K. Kinoshita, M. Delcroix, T. Nakatani, and M. Miyoshi, "Suppression of late reverberation effect on speech signal using long-term multiple-step linear prediction," *IEEE Trans. Audio. Speech. Lang. Process.,* vol. 17, no. 4, pp. 534–545, 2009.

[15] Y. Liu, N. Nower, S. Morita, and M. Unoki, "Speech enhancement of instantaneous amplitude and phase for applications in noisy reverberant environments," *Speech Commun.,* vol. 84, pp. 1–14, 2016.

[16] Y. Liu, N. Nower, S. Morita, and M. Unoki, "Robust front-end for speech recognition by human and machine in noisy reverberant environments: the effect of phase information," *Proc. ISCSLP2016,* pp. 1–5, 2016.

[17] Y. Liu, N. Nower, Y. Yan, and M. Unoki, "Restoration of instantaneous amplitude and phase of speech signal in noisy reverberant environments," *Proc. EUSIPCO2015,* pp. 879–883, 2015.

[18] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio. Speech. Lang. Process.,* 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 Commun.,* vol. 53, no. 3, pp. 340–354, 2011.