RESTORATION OF INSTANTANEOUS AMPLITUDE AND PHASE OF SPEECH SIGNAL IN NOISY REVERBERANT ENVIRONMENTS

Yang Liu[1)], *Naushin Nower*[1)], *Yonghong Yan*[2)] *and Masashi Unoki*[1)]

[1)] School of Information Science, Japan Advanced Institute of Science and Technology
[2)] Institute of Acoustics, Chinese Academy of Sciences

ABSTRACT

We have proved that restoring the instantaneous amplitude as well as instantaneous phase on Gammatone filterbank plays a significant role for speech enhancement. However, it is still challenging topic with dereverberation since previously proposed scheme can only work in noisy environments. In this paper, we extend our previously proposed scheme to be general speech enhancement of removing the effects of noise and reverberation by restoring instantaneous amplitude and phase simultaneously. Objective and subjective experiments were conducted under various noisy reverberant conditions to evaluate the effectiveness of the extension of proposed scheme. The signal to error ratio (SER), correlation, PESQ, and SNR loss were used in objective evaluations. The normalized mean preference score was used in subjective evaluations. The results of both evaluations revealed that the proposed scheme could effectively improve quality and intelligibility of speech signals under noisy reverberant conditions.

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

1. INTRODUCTION

In real environments, the quality and intelligibility of speech are always degraded due to background noise and reverberation. Especially, the performance of applications such as hearing-aids and speech coders might be severely reduced in the presence of background noise and reverberation. Therefore, it is necessary to simultaneously remove these effects.

In the past a few decades, various methods have already been proposed to remove the effects of noise or reverberation in real environments. There are, for example, Wiener filtering [1], MMSE-STSA [2], Corpus-based approach [3], MINT approach [4], Kurtosis approach [5], and multiple-step linear prediction approach [6]. The first three methods can remove the effects of noise well in only noisy environments, meanwhile, the last three methods can reduce the effects of reverberation in only reverberant environments. It is obvious that all these methods cannot work well in noisy reverberant environments because a combination of different systems cannot simultaneously deal with the effects of additive noise and reverberation as convolved noise. Therefore, it is still a challenging problem to remove these effects simultaneously.

Recent studies have shown the importance of phase in speech enhancement [7,8]. We therefore have previously proposed a speech enhancement scheme motivated by the effectiveness of phase manipulation [9] which can significantly improve the quality and intelligibility of noisy speech. However, this scheme can only deal with additive noise but the convolved noise such as late reflection has not yet been modeled in this scheme, due to the properties of convolved noise.

This paper aims to extend the previous scheme to be a general speech enhancement of removing the effects of noise and reverberation simultaneously with consideration of phase information. A novel point is that effects of noise corresponding to additive and convolved noises (late reflection) on instantaneous amplitude and phase can be removed by Kalman filtering with efficient linear prediction (LP).

2. PREVIOUS SCHEME

Our previous scheme [9] intends to improve both instantaneous amplitudes and phases on output of the Gammatone filterbank (GTFB) which was designed by considering the properties of auditory system for noisy speech [10]. In this scheme, the noisy speech $y_N(t)$, where $y_N(t) = x(t) + n(t)$, is only observed. Here, $x(t)$ is the clean speech and $n(t)$ is background noise. The output of the kth sub-band, $Y_{N,k}(t)$, is represented as the analytical form by

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

where $Y_{N,1,k}(t)$ and $Y_{N,2,k}(t)$ are the sub-band components of $x(t)$ and $n(t)$, respectively. In addition, ω_k is the center frequency of the kth sub-band. $A_{N,k}(t)$ and $\phi_{N,k}(t)$ are the instantaneous amplitude and phase of the noisy speech. Then, Kalman filter with LP is applied to remove the effects of noise on the instantaneous amplitude and phase. Finally, the restored signal, $\hat{x}(t)$ is resynthesized from the restored sub-bands components by inverse GTFB.

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 a Grant-in-Aid for Scientific Research (A) (No. 25240026).

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

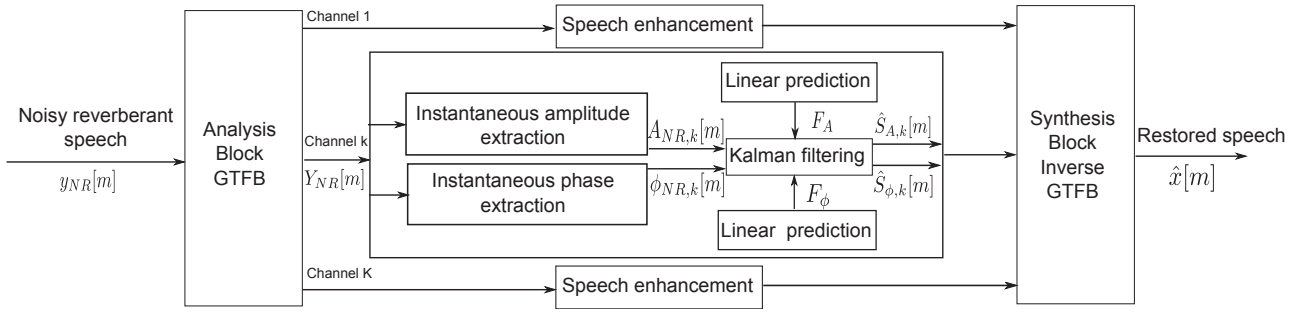

Fig. 1. Block diagram of proposed scheme for speech enhancement.

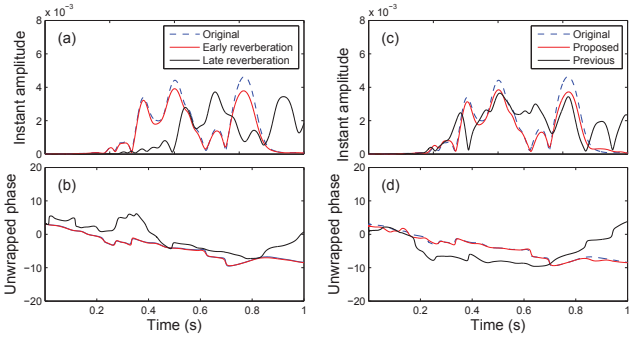

Fig. 2. Example of instantaneous amplitudes ((a) and (c)) and phases ((b) and (d)) under reverberant condition ($T_R = 2$ s) in 80th sub-band (left) and results of restoration by previous and proposed methods (right).

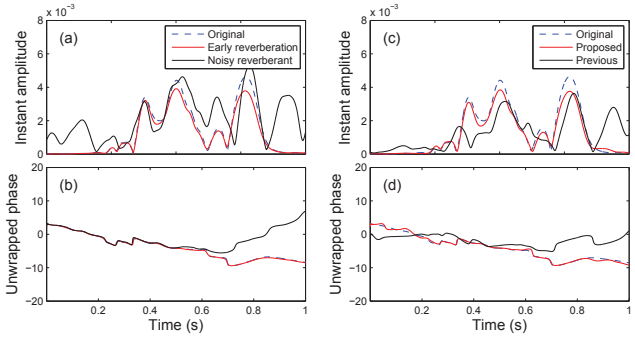

Fig. 3. Example of instantaneous amplitudes ((a) and (c)) and phases ((b) and (d)) under noisy reverberant condition ($T_R = 2$ s and SNR $= 0$ dB) in 80th sub-band (left) and results of restoration by previous and proposed methods (right).

3. PROPOSED SCHEME

The proposed scheme is an extension of the previous scheme and the block diagram of the proposed scheme is shown in Fig. 1. The proposed scheme consists of three stages: analysis stage, modification stage, and resynthesis stage.

The noisy reverberant speech, $y_{NR}(t) = x(t) * h(t) + n(t)$, is observed. Here, $h(t)$ is the room impulse response (RIR). The RIR, $h(t)$, contains both effects of early 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 reflection. Then we have $y_{NR}(t) = x(t) * h_E(t) + x(t) * h_L(t) + n(t)$, where $x(t) * h_E(t)$ is early reverberation and $x(t) * h_L(t)$ is late reverberation. Early reverberation may not significantly degrade the quality and intelligibility of speech because human beings cannot distinguish short echo and original speech while late reverberation is detrimental to the quality and intelligibility.

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

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

where $Y_{NR,1,k}(t)$ and $Y_{NR,2,k}(t)$ are the components of $x(t) * h_E(t)$ and $x(t) * h_L(t) + n(t)$, respectively. $A_{NR,k}(t)$ and $\phi_{NR,k}(t)$ are the instantaneous amplitude and phase of the noisy reverberant speech $Y_{NR,k}(t)$.

As shown in Figs. 2(a) and 2(b), the shapes of instantaneous amplitude and phase of early reverberation under reverberant conditions (reverberation time, $T_R = 2$ s) are similar to those of clean speech. From Figs. 3(a) and 3(b), these similarities can be observed under noisy reverberant conditions ($T_R = 2$ s and the signal to noise ratio SNR $= 0$ dB). The same trends were also observed in the other sub-bands and under the other noisy reverberant conditions. Therefore, in this paper, we only focus on dealing with the summation of late reverberation as convolved noise and additive noise.

3.1. Kalman filtering

The state and observation equations are defined in the Kalman filter. The state equations of kth sub-band for instantaneous amplitude and phase are defined as

$$\boldsymbol{S}_{A,k}[m] = \boldsymbol{F}_A \boldsymbol{S}_{A,k}[m-1] + \boldsymbol{W}_{A,k}[m], \quad (3)$$

$$\boldsymbol{S}_{\phi,k}[m] = \boldsymbol{F}_\phi \boldsymbol{S}_{\phi,k}[m-1] + \boldsymbol{W}_{\phi,k}[m], \quad (4)$$

where m is sample 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. $\boldsymbol{F}_A$ and $\boldsymbol{F}_\phi$ are the transition matrices that can be

obtained by the LP method. $\boldsymbol{W}_{A,k}[m]$ and $\boldsymbol{W}_{\phi,k}[m]$ are assumed to be Gaussian white noise of kth sub-band, and the variances of $\boldsymbol{W}_{A,k}[m]$ and $\boldsymbol{W}_{\phi,k}[m]$ are Q_A and Q_ϕ, respectively. $\boldsymbol{S}_{A,k}[m]$ and $\boldsymbol{S}_{\phi,k}[m]$ are the state vectors of instantaneous amplitude and phase of early reverberation at sampling point m in kth sub-band respectively.

The observation equations for the instantaneous amplitude and phase of kth sub-band are defined as

$$\boldsymbol{O}_{A,k}[m] = \boldsymbol{H}_A\boldsymbol{S}_{A,k}[m] + \boldsymbol{V}_{A,k}[m], \quad (5)$$
$$\boldsymbol{O}_{\phi,k}[m] = \boldsymbol{H}_\phi\boldsymbol{S}_{\phi,k}[m] + \boldsymbol{V}_{\phi,k}[m], \quad (6)$$

where $\boldsymbol{O}_{A,k}[m]$ and $\boldsymbol{O}_{\phi,k}[m]$ are the observed instantaneous amplitude and phase of the noisy reverberant speech at time m in kth sub-band. $\boldsymbol{H}_A$ and $\boldsymbol{H}_\phi$ are the observation matrices which are $[0, 0, \cdots, 1]$. $\boldsymbol{V}_{A,k}[m]$ 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.

We need five steps to calculate the optimal estimations for both instantaneous amplitude and phase.

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

$$\hat{\boldsymbol{S}}_{A,k}[m|m-1] = \boldsymbol{F}_A\hat{\boldsymbol{S}}_{A,k}[m-1|m-1], \quad (7)$$
$$\hat{\boldsymbol{S}}_{\phi,k}[m|m-1] = \boldsymbol{F}_\phi\hat{\boldsymbol{S}}_{\phi,k}[m-1|m-1]. \quad (8)$$

The state vector of m is estimated from the state vector of $m-1$ under the principle of minimum mean-square error.

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

$$\boldsymbol{P}_A[m|m-1] = \boldsymbol{F}_A\boldsymbol{P}_A[m-1|m-1]{\boldsymbol{F}_A}^{\mathrm{T}} + Q_A, \quad (9)$$
$$\boldsymbol{P}_\phi[m|m-1] = \boldsymbol{F}_\phi\boldsymbol{P}_\phi[m-1|m-1]{\boldsymbol{F}_\phi}^{\mathrm{T}} + Q_\phi. \quad (10)$$

Step 3: The current values are estimated as:

$$\hat{\boldsymbol{S}}_{A,k}[m|m] = \hat{\boldsymbol{S}}_{A,k}[m|m-1] + \boldsymbol{e}_A, \quad (11)$$
$$\hat{\boldsymbol{S}}_{\phi,k}[m|m] = \hat{\boldsymbol{S}}_{\phi,k}[m|m-1] + \boldsymbol{e}_\phi. \quad (12)$$

Here, $\boldsymbol{e}_A = \boldsymbol{G}_A[m](\boldsymbol{O}_{A,k}[m] - \boldsymbol{H}_A\hat{\boldsymbol{S}}_{A,k}[m|m-1])$ and $\boldsymbol{e}_\phi = \boldsymbol{G}_\phi[m](\boldsymbol{O}_{\phi,k}[m] - \boldsymbol{H}_\phi\hat{\boldsymbol{S}}_{\phi,k}[m|m-1])$ are called innovation, where $\boldsymbol{G}_A[m]$ and $\boldsymbol{G}_\phi[m]$ are the Kalman gains.

Step 4: We update the Kalman gains by:

$$\boldsymbol{G}_A[m] = \frac{\boldsymbol{P}_A[m|m-1]{\boldsymbol{H}_A}^{\mathrm{T}}}{(\boldsymbol{H}_A\boldsymbol{P}[m|m-1]{\boldsymbol{H}_A}^{\mathrm{T}} + R_A)}, \quad (13)$$

$$\boldsymbol{G}_\phi[m] = \frac{\boldsymbol{P}_\phi[m|m-1]{\boldsymbol{H}_\phi}^{\mathrm{T}}}{(\boldsymbol{H}_\phi\boldsymbol{P}_\phi[m|m-1]{\boldsymbol{H}_\phi}^{\mathrm{T}} + R_\phi)}. \quad (14)$$

Step 5: We update the error covariances matrices by:

$$\boldsymbol{P}_A[m|m] = (\boldsymbol{I} - \boldsymbol{G}_A[m]\boldsymbol{H}_A)\boldsymbol{P}_A[m|m-1], \quad (15)$$
$$\boldsymbol{P}_\phi[m|m] = (\boldsymbol{I} - \boldsymbol{G}_\phi[m]\boldsymbol{H}_\phi)\boldsymbol{P}_\phi[m|m-1], \quad (16)$$

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

3.2. Linear prediction

LP analysis was used to obtain transition matrices $\boldsymbol{F}_A$ and $\boldsymbol{F}_\phi$ in Eqs. (3) and (4). We extracted the LP coefficients for the instantaneous amplitude and phase which can be regarded as the output of a $\boldsymbol{p}$-th order auto-regressive process by autocorrelation method. The Kalman filtering with transition matrices obtained from clean speech is referred as an ideal scheme (IS) which could be used to check the upper limitation of the improvement for speech enhancement.

We studied the properties of LP coefficients for instantaneous amplitude and phase to estimate them under various conditions. This investigation revealed that these LP coefficients had similarities in modulation domain between speakers, gender, and contents. We calculated the LP coefficients of each sub-band for early reverberation generated from the closed clean dataset and converted them to line spectral frequencies (LSFs) to obtain $\boldsymbol{F}_A$ and $\boldsymbol{F}_\phi$. LSFs have a well behaved dynamic range while LP coefficients have a large dynamic range of values, therefore it is easier to guarantee the stability of the resulting synthesis filter in LSF domain. We averaged the computed LSFs and converted them to LP coefficients as trained LP coefficients in $\boldsymbol{F}_A$ and $\boldsymbol{F}_\phi$.

We incorporated an offline training phase with the proposed scheme to train the LP coefficients based on their characteristics in the modulation domain. This is referred as the proposed scheme (PS), to compare the PS with the IS.

4. EVALUATION

We used a closed dataset, containing four sentences from two males and two females from the TIMIT database [11] to determine $\boldsymbol{F}_A$ and $\boldsymbol{F}_\phi$. We then used ten different sentences uttered by five males and five females as an open dataset, to evaluate the proposed scheme. The signal to noise ratios (SNRs) between $x(t)$ and $n(t)$ were fixed at 10 and 0 dB. Reverberation times, T_Rs, were fixed at 0.5 and 2 s [12]. The sampling frequency, F_s, was set to be 20 kHz. We used a GTFB [10] to decompose the signal into 128 sub-bands ($K = 128$). The frame size was 25 ms. The LP order, p, was set to 22.

4.1. Objective evaluation

We evaluated the improvement of the restored speech by measuring correlation (Corr) and signal to error ratio (SER). Correlation shows the similarity between the shapes of clean instantaneous amplitude and phase and restored instantaneous amplitude and phase and SER shows the level of the error that we can reduce. Correlation and SER are defined as follows

$$\mathrm{Corr}(x_k, \hat{x}_k) = \frac{\int_0^T \left(x_k(t) - \overline{x_k}\right)\left(\hat{x}_k(t) - \overline{\hat{x}_k}\right)dt}{\sqrt{\left\{\int_0^T (x_k(t) - \overline{x_k})dt\right\}\left\{\int_0^T (\hat{x}_k(t) - \overline{\hat{x}_k})dt\right\}}}, \quad (17)$$

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

Table 1. Comparisons: PESQ and SNR loss (AVG.)

Method / SNR/T_R	Noisy reverberant		Ideal scheme (IS)		Ref (IS)		Proposed scheme (PS)		Ref (PS)	
	PESQ	SNR loss	PESQ	SNR loss	PESQ	SNR loss	PESQ	SNR loss	PESQ	SNR loss
10 dB/0.5 s	1.85	0.89	3.08	0.65	2.75	0.75	2.82	0.66	2.63	0.77
10 dB/2 s	1.36	0.92	2.84	0.68	2.55	0.77	2.55	0.70	2.32	0.80
0 dB/0.5 s	1.41	0.93	2.82	0.69	2.57	0.79	2.35	0.72	2.25	0.81
0 dB/2 s	1.11	0.94	2.69	0.71	2.43	0.80	2.20	0.74	2.01	0.95

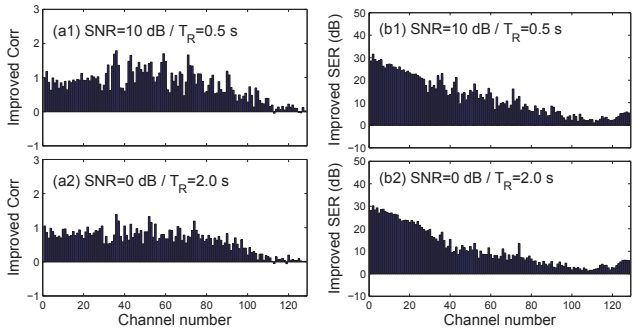

Fig. 4. Improvements in restoration accuracy of PS: (a) improved Corrs and (b) improved SERs.

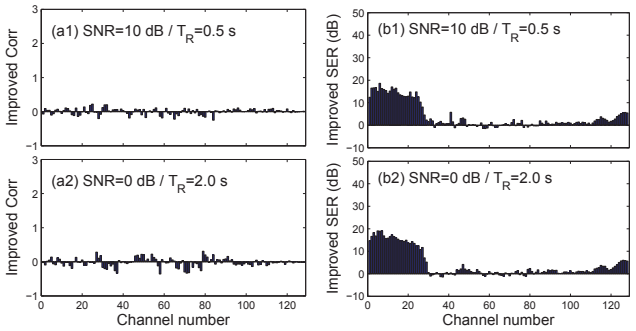

Fig. 5. Improvements in restoration accuracy of Ref (PS): (a) improved Corrs and (b) improved SERs.

where $x_k(t)$ and $\hat{x}_k(t)$ are clean and the restored speech in kth sub-band. These two measures were used to evaluate the reduction of effects of additive and convolved noise on the instantaneous amplitudes and phases in sub-bands. We defined the improved correlation as the subtraction of correlation between clean speech and restored speech from the correlation between clean speech and noisy reverberant speech in sub-band. The improved SER is also defined in the same way.

Figure 4 shows that the proposed scheme has large improvement in both Correlation and SER for restoring instantaneous amplitude and phase simultaneously. Figure 5 shows the improvement of only restoring instantaneously amplitude by our proposed scheme which indicates the importance of phase information. Figures 2 (c), 2(d), 3(c), and 3(d) show the comparison between the previous scheme and proposed scheme, it is easily observed that the proposed scheme provide better similarities in the instantaneous amplitude and phase in noisy reverberant environments.

Perceptual evaluation of sound quality (PESQ) [13] in the objective difference grades (ODGs) that covers from -0.5 (very annoying) to 4.5 (imperceptible) was used to evaluate subjective quality of the restored speech signals under noisy reverberant conditions. SNR loss [14] was also used to predict the improvement of speech intelligibility which ranges from 0 to 1.0, corresponding to the percent correctness (100% to 0%), under noisy reverberant conditions.

The results of objective evaluations are listed in Table 1. We made comparisons among noisy reverberant speech (NR), the restored speech by ideal scheme (IS), the restored speech by ideal scheme with only instantaneous amplitude (Ref (IS)), the restored speech by proposed scheme (PS), and the restored speech by proposed scheme with only instantaneous amplitude (Ref (PS)). From the results of PESQ, we found that the phase information can improve the quality of speech, comparing IS and PS with Ref (IS) and Ref (PS) respectively. From the results of SNR loss, it could be observed that phase information plays an important role for improving intelligibility of speech, comparing IS and PS with Ref (IS) and Ref (PS) respectively. From the results of evaluations, we could conclude that the proposed scheme can effectively reduce the effects of noise and reverberation by restoring the instantaneous amplitude simultaneously. Furthermore, phase information is quite important for improving the quality and intelligibility of speech under noisy reverberant conditions.

4.2. Subjective evaluation

Sentence-pair listening test was chosen for subjective evaluation. Noisy reverberant speech signals were generated under four noisy reverberant conditions: SNRs at 10 and 0 dB and reverberation times T_Rs at 0.5 and 2 s, for two male and two female speakers from TIMIT database.We made comparison for six categories of speech (clean (CL), IS, Ref (IS), PS, Ref (PS), and noisy reverberant (NR)), where CL is clean speech. Each of these six was compared with the other five categories. Therefore we have 30 sentence pairs 30 ($= 5 \times 6$) under each noisy reverberant condition. These sentence pairs were randomly shuffled and listeners were required to choose one of the three choices for each sentence pair: prefer the first one, prefer the second one, and no preference. Pairwise scoring

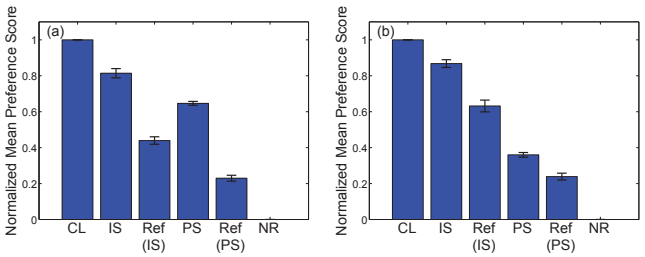

Fig. 6. Subjective evaluation in noisy reverberant environments: (a) $T_R = 0.5$ s and SNR=10 dB (b) $T_R = 2$ s and SNR=0 dB.

was employed: 1 point is added to the preferred speech and 0 to the other and 0.5 point is added for both ones with no preference. The experiment was conducted in sound-proof room and ten subjects with normal hearing were participated in this experiment. These participants were familiar with the task after a short practice session before formal test. Each listener listened 30 sentence pairs for each noisy reverberant condition and totally listened to 120 (30×4) sentence pairs.

Figure 6 shows the comparison of normalized mean preference score for the best and worst noisy reverberant conditions in our experiment. We found that IS was always better than PS and Ref (IS) was always better than Ref (PS), which means the trained $\boldsymbol{F}_A$ and $\boldsymbol{F}_\phi$ still have a little bit minor negative effects on speech restoration. However, if the training effects of $\boldsymbol{F}_A$ and $\boldsymbol{F}_\phi$ under various conditions will be improved, these gaps could be eliminated. We also found that IS was always better than Ref (IS) and PS was always better than Ref (PS). These indicated that the use of instantaneous phase plays an important role for speech enhancement in noisy reverberant environments.

5. CONCLUSION

We proposed a scheme for speech enhancement by using Kalman filter with a training phase in sub-bands on the Gammatone filterbank in noisy reverberant environments. The proposed scheme dealt with the temporal variations of instantaneous amplitudes and phases simultaneously. The results of objective evaluations revealed that the proposed scheme can improve much of the quality and intelligibility of speech. The results of subjective evaluations also indicated the importance of phase information for speech enhancement. We believe that the combination of the instantaneous amplitude and phase with accurate estimation of LP coefficients in sub-bands can contribute much to speech enhancement.

REFERENCES

[1] P. Scalart and J. V. Filho, "Speech enhancement based on a priori signal to noise estimation," *Proc. ICASSP1996*, pp. 629–633, 1996.

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

[3] 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.

[4] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. ASSP*, vol. 36, no. 2, pp. 145–152, 1988.

[5] O. Tanrikulu and A. G. Constantinides, "Least-mean Kurtosis: a novel higher-order statistics based adaptive filtering algorithms," *Electronic Letters*, vol. 30, no. 3, pp. 189–190, 1994.

[6] 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. ASSP*, vol. 17, no. 4, pp. 534–545, 2009.

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

[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] 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.

[10] 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.

[11] W. M. Fisher, G. R. Doddington, R. George, and M. Kathleen, "The DARPA speech recognition research database: specification and status," *Proc. DARPA workshop*, pp. 93–99, 1986. Report TR-I-0028, 2010.

[12] M. R. Schroeder, "Modulation transfer functions: definition and measurement," Acustica, vol. 49, pp. 179–182, 1981.

[13] 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.

[14] 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.