ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/273388794

Restoration scheme of instantaneous amplitude and phase using Kalman filter with efficient linear prediction for speech...

Article *in* Speech Communication · March 2015

DOI: 10.1016/j.specom.2015.02.006

CITATIONS
2

READS
57

3 authors, including:

Naushin Nower
University of Dhaka
22 PUBLICATIONS **39** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

PACTMAN: Trust, Privacy and Consent in Future Pervasive Environments (funded by Engineering and Physical Sciences Research Council (EPSRC)) View project

All content following this page was uploaded by Naushin Nower on 07 June 2015.

The user has requested enhancement of the downloaded file.

Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 70 (2015) 13–27

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Restoration scheme of instantaneous amplitude and phase using Kalman filter with efficient linear prediction for speech enhancement

Naushin Nower *, Yang Liu, Masashi Unoki

School of Information Science, Japan Advanced Institute of Science and Technology (JAIST), 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan

Received 12 August 2014; received in revised form 18 February 2015; accepted 23 February 2015
Available online 5 March 2015

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 the phase spectrum in the short-time Fourier transform plays an important role in speech enhancement. Thus, the proposed scheme concentrates on simultaneously restoring both instantaneous amplitudes and phases. The Kalman filter, which is an optimal estimator in this scheme, is used for both instantaneous amplitudes and phases in the sub-band representation to remove the effect of noise. We found that the effectiveness of the Kalman filter depended on accurate estimates of LP coefficients. We propose an effective LP training phase to derive gender and content independent LP coefficients as central processing for Kalman filtering. We carried out objective and subjective tests under various noisy conditions to evaluate the effectiveness of the proposed scheme and compared it with typical methods. The signal to error ratio (SER), perceptual evaluation of speech quality (PESQ), and SNR loss were used as objective measures in these simulations. The mean preference score was used in subjective evaluations. The results revealed that the proposed scheme could effectively improve these objective and subjective measures more than those with typical methods.
© 2015 Elsevier B.V. All rights reserved.

Keywords: Speech enhancement; Instantaneous amplitude and phase; Kalman filter; Gammatone filterbank; Linear prediction

1. Introduction

The required speech signal in real world scenarios is frequently smeared by various kinds of noise. This noise not only degrades the perceptual aspects of speech quality and speech intelligibility but also reduces 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 signals in noisy environments have to be enhanced.

Speech enhancement is concerned with improving the quality and intelligibility of corrupted speech in the presence of noise. Various methods of speech enhancement have already been proposed during the past two decades to remove the effects of noise from noisy speech to improve its quality. Of these, classical methods of speech enhancement, such as spectral subtraction (SS) (Boll, 1979), the Ephraim–Malah algorithm (MMSE-STSA estimator) (Ephraim and Malah, 1984), and the Scalart–Filho algorithm (Wiener filtering) (Scalart and Filho, 1996), have attracted a great deal of attention because of their simplicity and efficiency in estimating spectral magnitude. The SS method (Boll, 1979) subtracts the estimated noise magnitude spectrum from the noisy speech magnitude spectrum, where the noise spectrum can be estimated and updated during periods when speech is absent. The Wiener filter

* Corresponding author.
E-mail addresses: naushin@jaist.ac.jp (N. Nower), yangliu@jaist.ac.jp (Y. Liu), unoki@jaist.ac.jp (M. Unoki).

http://dx.doi.org/10.1016/j.specom.2015.02.006
0167-6393/© 2015 Elsevier B.V. All rights reserved.

(Scalart and Filho, 1996) algorithm filters noisy speech signals by using a filter derived based on the minimum mean-square error (MMSE) criterion. These methods employ the short-time Fourier analysis-modification-synthesis (AMS) framework for speech enhancement.

There are various statistical methods of model-based speech enhancement in the literature in addition to these. Modeling in the model-based approaches is done using the statistical properties of the speech signal over multiple frames. This modeling is performed using the hidden Markov model (HMM) (Ephraim et al., 1989; Ephraim, 1992; Zhao and Kleijn, 2007), the Gaussian mixture model (GMM), or codebook-based methods (Sriram et al., 2007). HMM-based speech enhancement is a renowned model-based technique and resolves common problems with classical methods of speech enhancement in dealing with rapid variations in noise characteristics (Veisi and Hossein, 2013). Nishikawa et al. (2003) combined an independent component analysis (ICA) based noise estimator with multi-channel-wise non-linear signal processing to reduce noise further in their method of noise reduction. However, all of their improvements were limited because they only considered the spectral domain.

Recent research has investigated the importance of speech enhancement in the modulation domain, such as the modulation-domain Kalman filter (MDKF) (So and Paliwal, 2011; Paliwal et al., 2012). So and Paliwal's method modeled temporal changes in the magnitude spectrum for both speech and noise without taking into consideration noise in the phase component (So and Paliwal, 2011). Consequently, the corpus based approach (Ji et al., 2011), model based speech enhancement with spectral estimation (Ruofei et al., 2012), and speech enhancement using nonnegative matrix factorization (NMF) (Mohammadiha et al., 2013; Sawada et al., 2013) are included in modern methods of speech enhancement. All the existing methods process corrupted speech signals by modifying or correcting speech in either temporal or spectral magnitude only and keeping the phase component unchanged. This is because the phase spectrum that is conventionally considered is unimportant and has been demonstrated not to contribute much toward speech enhancement. Wang and Lim emphasized this (Wang and Lim, 1982), which is perhaps the most cited work to justify the unimportance of phase in speech enhancement.

However, recent studies have reported that the use of the phase spectrum in the short-time Fourier transform (STFT) can significantly improve speech enhancement (Shannon and Paliwal, 2006; Paliwal and Alsteris, 2005; Roux et al., 2008). Shannon and Paliwal (2006) reported that they had carried out magnitude only and phase only experiments to investigate the effects of phase in speech enhancement. Clean magnitude was used in the magnitude-only experiment and the phase was set to a random value. In contrast, a clean phase was used in the phase-only experiment and the magnitude was set to one. Their results indicated that the phase spectrum also contained useful and important information. After that, Paliwal and Alsteris (2005) investigated whether the shape and length of the window function used in STFT for phase manipulation were important factors for speech enhancement.

Fardkhaleghi and Savoji (2010) investigated the role of the phase spectrum in speech enhancement using Wiener filtering and Martin minimum statistics. To do this, the noisy phase of each processed frame was corrected by applying three proposed optimization algorithms, which demonstrated that better results were achieved using phase correction. Zhang and Zhao (2013) also subtracted real and imaginary spectra separately in the modulation frequency domain to enhance magnitude as well as phase through spectral subtraction for better enhancement. Kleinschmidt et al. (2011) proposed a method of utilizing phase information by complementing it with conventional magnitude-only spectral subtraction speech enhancement through complex spectrum subtraction (CSS). They demonstrated that an average improvements of 20% in relative word accuracy were possible for automatic speech recognition systems when accurate estimates of the phase spectrum were available.

Paliwal et al. (2011) demonstrated that modifying the phase spectrum could greatly improve speech enhancement. By investigating various cases where different combinations of noisy, clean (noiseless), and compensated amplitude and phase spectra were considered. This suggested that significant speech enhancement could be possible if the clean phase was known or the compensated phase spectrum was available. They also studied the effect of mismatched or matched windows for both amplitude and phase spectra estimates during AMS in STFT. Their results indicated that a proper choice of an analysis window and AMS setting in the phase spectrum could significantly improve speech enhancement. All existing research on the phase spectrum has either emphasized the importance of phase in speech enhancement or investigated a suitable size and shape for the window to enable phase manipulation. Thus, we obviously need to consider both the amplitude and phase of the noisy signal to enable better speech enhancement.

It is well-known that all existing speech enhancement algorithms based on STFT-AMS can improve speech quality but not speech intelligibility (Loizou and Kim, 2011). The reasons for this are still unclear so that many researchers have investigated probable strategies for reducing distortions and enhancing features related to speech intelligibility. However, it has been found from psychoacoustical studies that the temporal envelope (TE) and temporal fine structure (TFS) are important cues for speech perception (Drullman, 1995; Moore, 2008). It has also been revealed that TE and TFS play an important role in improving the intelligibility of noise-degraded speech (Swaminathan, 2010; Swaminathan and Heinz, 2012). Therefore, AMS in the filterbank is a suitable framework for speech enhancement, rather than AMS in STFT. Hence, it is expected that temporal amplitude and phase

manipulations acting as AMS in the filterbank can drastically improve the quality as well as the intelligibility of noise degraded speech.

Our main aim that was motivated by the effectiveness of phase manipulation from the existing literature was to propose a speech enhancement scheme acting as AMS on the filterbank to enhance both instantaneous amplitude and phase by using a recursive Kalman filter in a Gammatone filterbank (GTFB) (Unoki and Akagi, 1999). We dealt with instantaneous amplitude and phase on the GTFB because temporal smoothed information (amplitude and phase) are directly related in improving the quality and intelligibility of speech. The Kalman filter is of particular interest in smooth methods of prediction in dealing with instantaneous amplitude and phase in sub-bands. Moreover, it can be viewed as a joint estimator for both the magnitude and phase spectrum of speech, under non-stationary conditions (So and Paliwal, 2011). We believe that the ability of the Kalman filter to process non-stationary signals as well as estimate both the magnitude and phase spectrum make it preferable over STFT-based methods of enhancement. We propose a concept of speech enhancement based on these analyses by restoring both instantaneous amplitude and phase using Nower et al.'s Kalman filter (Nower et al., 2014) as a preliminary study. However, the main concept in the proposed scheme for speech enhancement without details on central processing cannot confirm the concept's validity.

We concentrated on deriving accurate transition matrices, which is the main core processing technique in Kalman filtering. This is because enhanced performance by the Kalman filter is dependent on the accuracy and reliability of transition matrices. As these transition matrices of the state equation of both instantaneous amplitude and phase are unknown, it is difficult to set these transition matrices in Kalman filtering to enable suitable speech enhancements.

We analyzed the speech signal as a highly temporal correlated time series signal and used linear prediction as a state of the art concept to derive these coefficients as modulation characteristics of amplitude and phase in each sub-band of Kalman filtering in this research. We studied the characteristics of LP in the modulation domain and utilized the properties of stability of line spectral frequencies (LSFs) to train the LP coefficients in an effective way that could be regarded as speaker, gender, and content independent LP coefficients for Kalman filtering. Thus, our main contribution in this paper is to concentrate on core processing of the Kalman filter to provide a state of the art concept of speech enhancement.

The rest of the paper is organized as follows. Section 2 presents the proposed scheme for speech enhancement. We describe the details on implementing the algorithm in Section 3. The results from evaluation and a discussion are presented in Sections 4 and 5. Section 6 ends with a conclusion and description of future work.

2. Proposed scheme

Our proposed method of speech enhancement was intended to improve both instantaneous amplitude and instantaneous phase acting as AMS on a Gammatone filterbank. The model consisted of three steps: (i) an analysis stage, where instantaneous amplitude and instantaneous phase were extracted from noisy speech by GTFB, (ii) a modification stage, where instantaneous amplitude and instantaneous phase were enhanced by a Kalman filter with LP (with/without a training phase), and (iii) re-synthesis by using the inverse Gammatone filterbank. There is a block diagram of the proposed method in Fig. 1.

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

$$Y_k(t) = Y_{1,k}(t) + Y_{2,k}(t) \tag{1}$$

$$= A_k(t)\exp\left(j\omega_k t + j\phi_k(t)\right) \tag{2}$$

where $Y_{1,k}(t)$ is the component of $x(t)$ and $Y_{2,k}(t)$ is that of $n(t)$ that have passed through the filterbank. 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 noisy speech.

2.1. Modeling of instantaneous amplitude and phase in Kalman filter

The Kalman filter which is 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 signals 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.

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

$$\boldsymbol{S}[m] = \boldsymbol{F}\boldsymbol{S}[m-1] + \boldsymbol{W}[m], \tag{3}$$

$$\boldsymbol{O}[m] = \boldsymbol{H}\boldsymbol{S}[m] + \boldsymbol{V}[m], \tag{4}$$

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 driven noise and observation noise that are assumed to be white Gaussian noise. The state equations of the kth channel for instantaneous amplitude and instantaneous phase are:

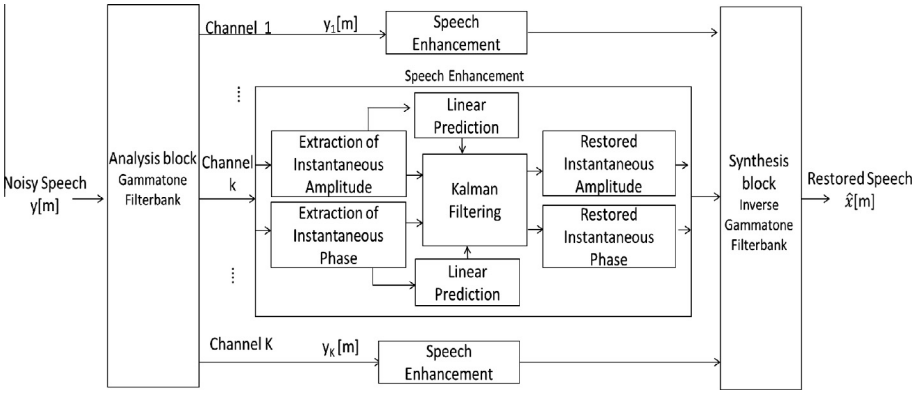

Fig. 1. Block diagram of proposed scheme (non-blind method) for speech enhancement.

$$\boldsymbol{S}_{A,k}[m] = \boldsymbol{F}_A \boldsymbol{S}_{A,k}[m-1] + \boldsymbol{W}_{A,k}[m], \tag{5}$$

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

where $\boldsymbol{S}_{A,k}[m]$ and $\boldsymbol{S}_{\phi,k}[m]$ correspond to the states of instantaneous amplitude and instantaneous phase of the kth channel. Since, instantaneous amplitude and phase can be modeled with an auto-regressive (AR) process of order p, the state vector can be represented as:

$$\boldsymbol{S}_{A,k}[m] = [S_{A,k}[m-p+1], \ldots, S_{A,k}[m]]^T, \tag{7}$$

$$\boldsymbol{S}_{\phi,k}[m] = \left[S_{\phi,k}[m-p+1], \ldots, S_{\phi,k}[m]\right]^T, \tag{8}$$

where $\boldsymbol{F}_A$ and $\boldsymbol{F}_\phi$ are the transition matrices that can be obtained with the method of linear prediction. $\boldsymbol{W}_{A,k}[m]$ and $\boldsymbol{W}_{\phi,k}[m]$ are assumed to be the white Gaussian noise of kth channel, and the variances of $\boldsymbol{W}_{A,k}[m]$ and $\boldsymbol{W}_{\phi,k}[m]$ correspond to Q_A and Q_ϕ.

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

$$\boldsymbol{O}_{A,k}[m] = \boldsymbol{H}_A \boldsymbol{S}_{A,k}[m] + \boldsymbol{V}_{A,k}[m], \tag{9}$$

$$\boldsymbol{O}_{\phi,k}[m] = \boldsymbol{H}_\phi \boldsymbol{S}_{\phi,k}[m] + \boldsymbol{V}_{\phi,k}[m], \tag{10}$$

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

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

Step 1: We set the initial state vectors $\boldsymbol{S}_{A,k}[1|1] = [10^{-12} \cdots 10^{-12}]$ and $\boldsymbol{S}_{\phi,k}[1|1] = [10^{-12} \cdots 10^{-12}]$. These values are used to initialize the state vectors only and will reach close to the original state vectors after a few iterations. We can then estimate the instantaneous amplitude and phase of clean speech in the 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 estimate of time $m-1$.

$$\boldsymbol{S}_{A,k}[m|m-1] = \boldsymbol{F}_A \hat{\boldsymbol{S}}_{A,k}[m-1|m-1], \tag{11}$$

$$\boldsymbol{S}_{\phi,k}[m|m-1] = \boldsymbol{F}_\phi \hat{\boldsymbol{S}}_{\phi,k}[m-1|m-1]. \tag{12}$$

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

$$\boldsymbol{P}_A[m|m-1] = \boldsymbol{F}_A \boldsymbol{P}_A[m-1|m-1]\boldsymbol{F}_A^{\mathrm{T}} + Q_A, \tag{13}$$

$$\boldsymbol{P}_\phi[m|m-1] = \boldsymbol{F}_\phi \boldsymbol{P}_\phi[m-1|m-1]\boldsymbol{F}_\phi^{\mathrm{T}} + Q_\phi. \tag{14}$$

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

$$\boldsymbol{S}_{A,k}[m|m] = \boldsymbol{S}_{A,k}[m|m-1] + \boldsymbol{e}_A, \tag{15}$$

$$\boldsymbol{S}_{\phi,k}[m|m] = \boldsymbol{S}_{\phi,k}[m|m-1] + \boldsymbol{e}_\phi. \tag{16}$$

Here, $\boldsymbol{e}_A = \boldsymbol{G}_A[m] \quad \times (O_{A,k}[m] - H_A S_{A,k}[m|m-1])$ and $e_\phi = G_\phi[m](O_{\phi,k}[m] - H_\phi S_{\phi,k}[m|m-1])$ are called innovation. Where $G_A[m]$ and $G_\phi[m]$ are the Kalman gains.

Step 4: We update the Kalman gains by:

$$G_A[m] = \frac{P_A[m|m-1]H_A^{\mathrm{T}}}{(H_A P[m|m-1]H_A^{\mathrm{T}} + R_A)}, \tag{17}$$

$$G_\phi[m] = \frac{P_\phi[m|m-1]H_\phi^{\mathrm{T}}}{(H_\phi P_\phi[m|m-1]H_\phi^{\mathrm{T}} + R_\phi)}. \tag{18}$$

Step 5: We update the error covariance matrices by:

$$P_A[m|m] = (I - G_A[m]H_A)P_A[m|m-1], \tag{19}$$

$$P_\phi[m|m] = (I - G_\phi[m]H_\phi)P_\phi[m|m-1]. \tag{20}$$

where I is the unit matrix.

2.2. Assumptions

The Kalman filter provides an optimal solution to the estimated parameters, whenever the observation noise and driven noise maintain white and Gaussian noise properties. We have assumed that driven noise $W_{A,k}[m]$ and

$W_{\phi,k}[m]$ in Eqs. (7) and (8) and observation noise $V_{A,k}[m]$ and $V_{\phi,k}[m]$ in Eqs. (9) and (10) are white Gaussian noise as random variables in this paper. The power spectrum density (PSD) plot for white Gaussian noise gives almost a fixed power in all the frequencies. The PSD indicates how much power is contained in each of the spectral components. In addition, the mean values of $W_{A,k}[m], W_{\phi,k}[m]$, $V_{A,k}[m]$, and $V_{\phi,k}[m]$ for white Gaussian noise are zero and variances denoted as Q_A, Q_ϕ, R_A, and R_ϕ have finite values. The distribution of noise should follow a Gaussian distribution for the Gaussian assumption. A Kolmogorov–Smirnov (K–S) test (Massey, 1951) and a Chi square test (Satorra and Bentler, 2001) are usually used in formal testing for Gaussianity. We will consider these assumptions in detail in Section 5.

3. Implementation of algorithm

3.1. Auditory-motivated filterbank

We used an auditory-motivated Gammatone filterbank, which was designed by considering the properties of the auditory system and the detection of a discontinuous point dealing with a complex spectrum. Please see Unoki and Akagi (1999) for details on the filterbank design.

3.2. Calculation of instantaneous amplitude and phase

We extract instantaneous amplitude $S_{A,k}[m]$ and instantaneous unwrapped phase $S_{\phi,k}[m]$ of Eqs. (9) and (10) (Unoki and Akagi, 1999) by using the Gammatone filterbank. Instantaneous amplitude $S_{A,k}[m]$ and unwrapped phase $S_{\phi,k}[m]$ are calculated as:

$$S_{A,k}[m] = |\tilde{f}(c,m)| \tag{21}$$

$$S_{\phi,k}[m] = \int_0^m \left(\frac{d}{d\tau}\arg(\tilde{f}(c,m) - \omega_k)\right)d\tau. \tag{22}$$

where $c = \alpha^{k-K/2}$ and α is the scale of the Gammatone filterbank. Here, $|\tilde{f}(c,m)|$ is the amplitude spectrum defined by the wavelet transform and arg $(\tilde{f}(c,m))$ is the unwrapped phase spectrum defined by the complex wavelet transform. There is an example of the instantaneous amplitude and instantaneous unwrapped phase extraction in a particular sub-band ($k = 28$) in Fig. 2.

3.3. Linear prediction

We utilized LP analysis to set the LP coefficients F_A, and F_ϕ in Eqs. (7) and (8) for Kalman filtering. To do this, we incorporated two methods, i.e., non-blind and blind, to extract the LP coefficients. The purpose of the non-blind method was to identify the possible upper limit of improvements in real enhancement applications. We assumed that clean speech was available for non-blind Kalman filtering to derive the LP coefficients. Accurate LP coefficients were

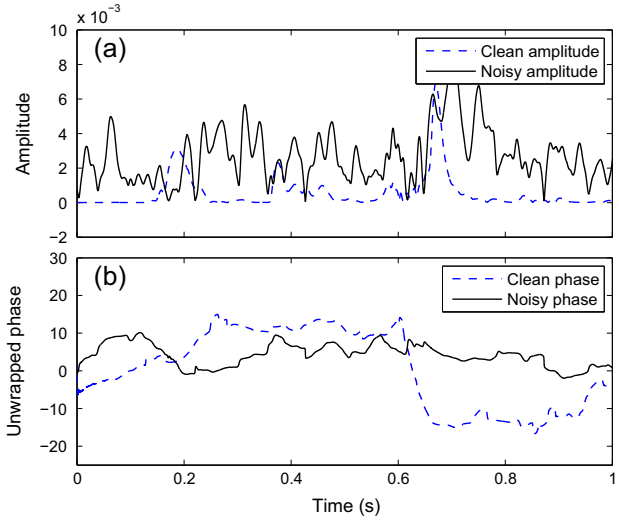

Fig. 2. Example of (a) instantaneous amplitude $S_{A,k}[m]$ and (b) instantaneous unwrapped phase $S_{\phi,k}[m]$ extraction in sub-band (channel $k = 28$) using Gammatone filterbank. SNR $= -10$ dB noise (white).

calculated from clean speech and we used these coefficients in Kalman filtering to restore noisy speech. However, clean speech may not be available in real life for LP analysis. Thus, our main aim is to propose a general method of speech enhancement that can be applied in real environments. To do this, we propose a blind method of deriving LP. A training phase was incorporated into the proposed blind method to set the proper LP coefficients from a closed data set. These trained LP coefficients were regarded as speaker and gender independent and could be utilized on an open data set, whenever clean speech was not available. There was no closed data set for non-blind Kalman filtering, and we assumed that we had LP coefficients from corresponding clean speech to restore speech. We will thus explain these two methods (non-blind and blind methods) to set F_A and F_ϕ.

3.3.1. Non-blind Kalman filter

We assumed that the sampling sequences of clean speech's amplitude and phase in the non-blind Kalman filtering were $S_{A,k}[m]$ and $S_{\phi,k}[m]$, where $m = 0, 1, 2, \ldots, M$. These can be regarded as the output of a p-th order AR process. The models of linear prediction can be represented as:

$$S_{A,k}[m] = \sum_{i=1}^{p} a_i S_{A,k}[m-i] \tag{23}$$

$$S_{\phi,k}[m] = \sum_{i=1}^{p} b_i S_{\phi,k}[m-i]. \tag{24}$$

Here, $S_{A,k}[m]$ and $S_{\phi,k}[m]$ are the optimal estimates 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 order of prediction.

There are two types of methods of calculating the LP coefficients, i.e., through auto-correlation (AC) and covariance. We chose the AC method to calculate the LP coefficients, and $\{a_i\}$ $(i = 1, 2, \ldots, p)$ and $\{b_i\}$ $(i = 1, 2, \ldots, p)$ could be obtained by solving the Yule–Walker equation as:

$$R[q_a] - \sum_{i=1}^{p} a_i R[q_a - i] = 0 \tag{25}$$

$$R[q_b] - \sum_{i=1}^{p} b_i R[q_b - i] = 0. \tag{26}$$

Here, $R[q_a]$ and $R[q_b]$ are the AC functions of the instantaneous amplitude and phase of 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. We can then obtain the transition matrices, F_A and F_ϕ, given in Eqs. (27) and (28) to estimate instantaneous amplitude and instantaneous phase by using the Kalman filter.

$$F_A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ 0 & 0 & 0 & \cdots & 1 \\ a_p & a_{(p-1)} & a_{(p-2)} & \cdots & a_1 \end{bmatrix} \tag{27}$$

$$F_\phi = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ 0 & 0 & 0 & \cdots & 1 \\ b_p & b_{(p-1)} & b_{(p-2)} & \cdots & b_1 \end{bmatrix}. \tag{28}$$

We extract the LP coefficients in this case to restore speech from corresponding clean speech.

3.3.2. Blind Kalman filter

The blind method is proposed for real environments, whenever clean speech is not available to calculate LP coefficients. There is a block diagram of the proposed blind approach in Fig. 3. As was previously stated, the effectiveness of Kalman filtering depends on proper estimates of LP coefficients. We studied the properties of LP coefficients for instantaneous amplitude and phase to estimate the LP coefficients. We found from this investigation that the LP coefficients of instantaneous amplitude and phase had some similarities in values. We also checked that these similarities were not sensitive to a specific gender, individual speaker, or the content of speech. These properties are quite different from the LP properties of the spectral envelope. There is an example of similarities in LP coefficients for four (two males and two females) different speakers and different types of content in Fig. 4. These were derived from Eqs. (27) and (28).

We incorporated an offline prior training phase into a closed data set to train the LP coefficients for blind Kalman filtering based on LP characteristics in the modulation domain. We calculated the LP coefficients of each sub-band from the closed clean data and converted them to LSFs, which are alternative representation of LP parameters. We averaged the computed LSFs and converted LP coefficients as training LP coefficients.

It is well known that LSFs have some useful properties over LP coefficients. LP parameters have a large dynamic range of values, which is not beneficial for quantization (Paliwal and Atal, 1993). LSFs, on the other hand, have a well behaved dynamic range. It is easier to guarantee the stability of the resulting synthesis filter when interpolation is done in the LSF domain. Whenever LP coefficients are encoded as LSFs, we do not need to spend the same number of bits for each LSF. This is because higher LSFs correspond to high frequency components and high frequency components have less effect in speech perception. Consequently, higher LSFs can be quantized using fewer

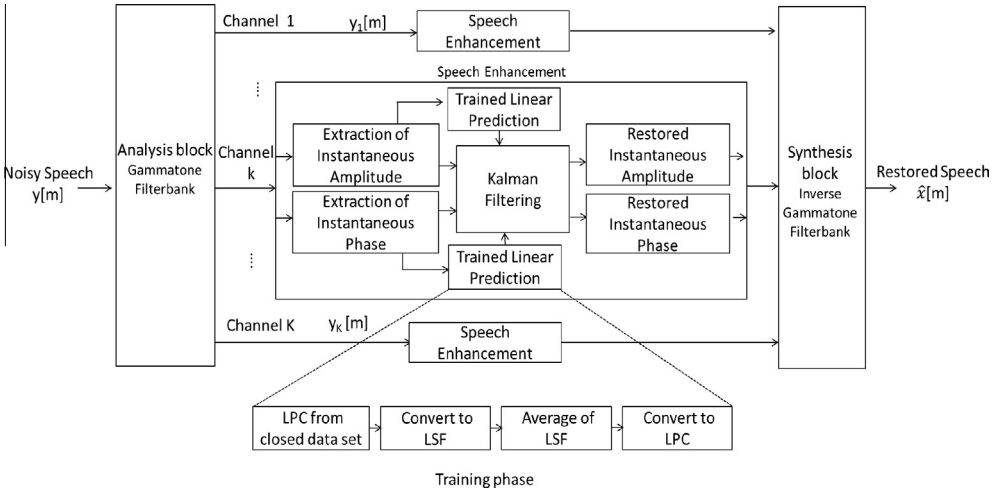

Fig. 3. Block diagram of proposed scheme (blind Kalman filter method) for speech enhancement.

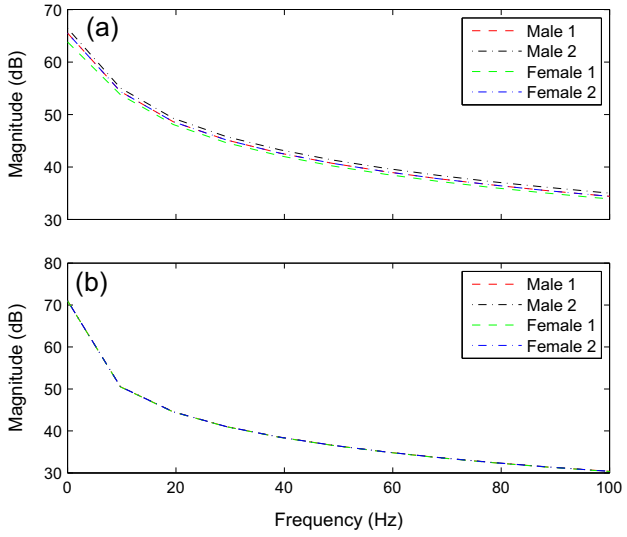

Fig. 4. Example of spectrum analysis of LP coefficient similarities in four different speakers and content: (a) instantaneous amplitude and (b) phase in sub-band (channel $k = 28$).

bits than lower LSFs. This reduces the bit rate while keeping the speech quality almost the same. A short segment of speech in the LP analysis of speech is assumed to be generated as the output of an all-pole filter, $H(z) = 1/A(z)$, where $A(z)$ is the inverse filter defined as:

$$A(z) = 1 + a_1 z^{-1} + \cdots + a_p z^{-p} \tag{29}$$

In order to define LSFs, the inverse filter polynomial is used to construct two polynomials:

$$P(z) = A(z) + z^{-(M+1)} A(z^{-1}) \tag{30}$$

$$Q(z) = A(z) - z^{-(M+1)} A(z^{-1}). \tag{31}$$

The roots of the polynomials are called LSFs. Thus, we calculated the LP coefficients of each sub-band from the closed clean dataset and converted them to LSFs. We averaged the computed LSFs and converted them to LP coefficients as trained LP coefficients.

4. Evaluation

We evaluated speech enhancements with the proposed method using both subjective and objective measures.

4.1. Objective evaluation

Objective measures included SER, Correlation, PESQ, and SNR losses. We carried out experiments using 50 different English sentences uttered by male and female speakers that were extracted from the Texas Instruments and Massachusetts Institute of Technology (TIMIT) database to evaluate the effectiveness of both the proposed methods. Half the utterances were selected from males and the other half was selected from females from the open data set. We created a closed data set, containing five sentences from

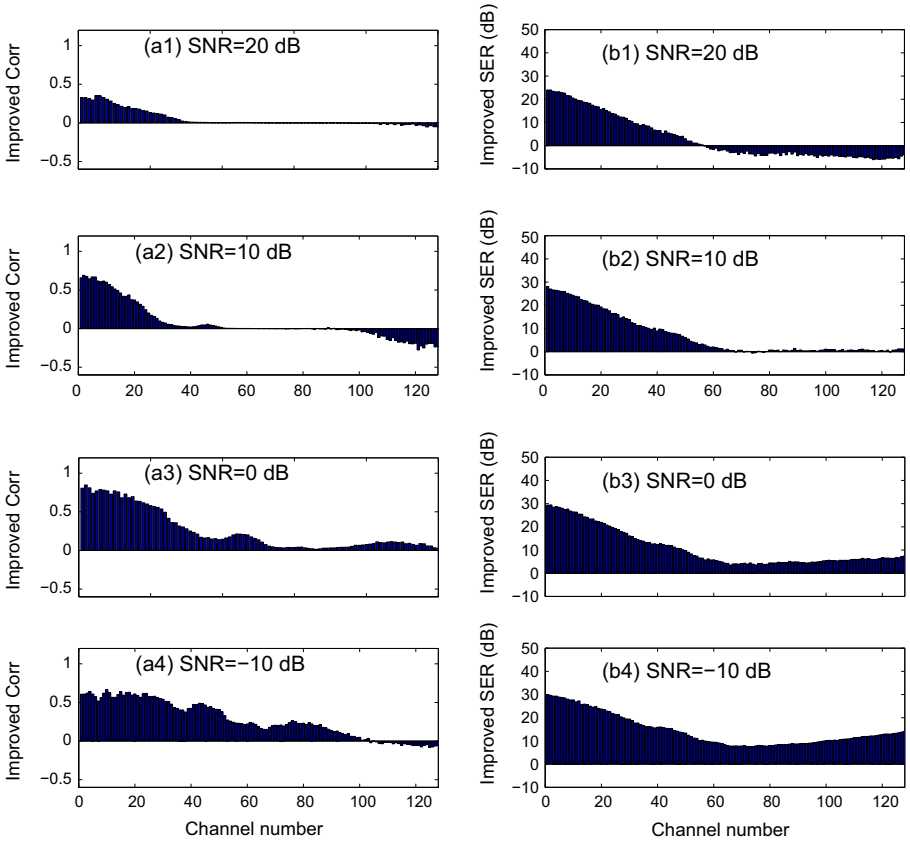

Fig. 5. Improvements in restoration accuracy with non-blind Kalman filter: (a) improved Corrs. and (b) improved SERs. SNR = 20 dB to −10 dB.

two male and three female speakers extracted from the TIMIT database to train the LP coefficients. We defined the other data set as an open data set. We used white, babble, and pink noise to evaluate the proposed scheme. The signal to noise ratios (SNRs) for white noise between $x(t)$ and $n(t)$ were fixed 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 (Unoki and Akagi, 1999) to divide the signal into 128 channels ($K = 128$). We used a sampling frequency (F_s) of 20 kHz. We utilized a 25-ms-long rectangular window. The LP order, p, was set to 12.

We evaluated improvements to restore speech by measuring the correlation and signal to error ratio (SER). The correlation indicated similarities between the shapes of clean instantaneous amplitude and phase and restored instantaneous amplitude and phase and SER indicated the amount of error that could be reduced. The correlation and SER are defined as:

$$\mathrm{Corr}(x_k,\hat{x}_k) = \frac{\int_0^T (x_k(t)-\overline{x_k})\left(\hat{x}_k(t)-\overline{\hat{x}_k}\right)dt}{\sqrt{\left\{\int_0^T (x_k(t)-\overline{x_k})dt\right\}\left\{\int_0^T (\hat{x}_k(t)-\overline{\hat{x}_k})dt\right\}}}, \tag{32}$$

$$\mathrm{SER}(x_k,\hat{x}_k) = 10\log_{10}\frac{\int_0^T (x_k(t))^2 dt}{\int_0^T (x_k(t)-\hat{x}_k(t))^2 dt}, \tag{33}$$

where $x_k(t)$ is the clean speech of the kth channel and $\hat{x}_k(t)$ is the restored speech of the kth channel.

Fig. 5 shows the improvement in correlation and SER in each channel using the non-blind method (non-blind Kalman filtering) under the given white noise conditions. The heights of the bars in the figure indicate the mean value of the improvements in SER. Most of the channels have positive improvements in SER under 20, 10, 0, and −10 dB noise conditions, except for some of the higher channels at 20-dB. The signal components in the higher channels under 20-dB conditions are very similar to those of the clean signal, and are thus almost much closer to those of clean speech. Thus, it is easy to see that the proposed method can effectively reduce noise in both instantaneous amplitude and phase.

The performance of the blind method (blind Kalman filtering) is shown in Fig. 6. The results prove that the blind Kalman filter with the trained LP coefficients also works well, and thus we obtained positive improvements in correlation and SER under all noise conditions, except for some of the higher channels at 20 dB and some middle channels under 10-dB and 0-dB conditions. It can be observed from a comparison of the results between non-blind and blind Kalman filtering that we achieved almost the same improvements in correlation and SER except for some channels. This is because our trained LP coefficients acted as clean LP coefficients and could be used as gender and content independent LP coefficients. Fig. 7 shows examples of restored instantaneous amplitude and

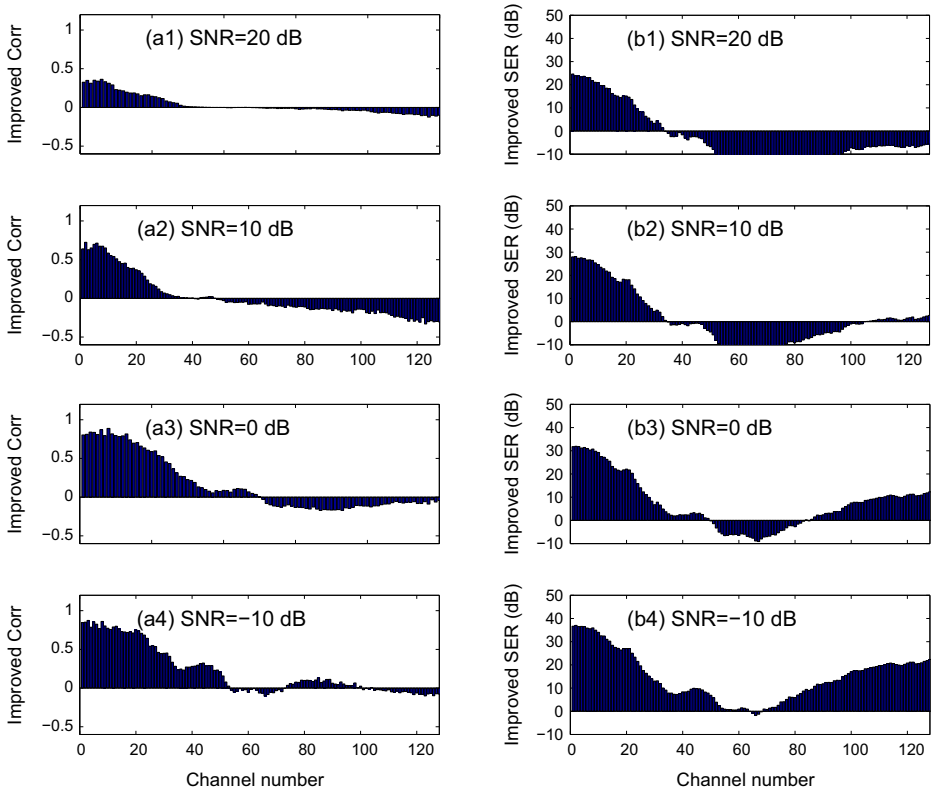

Fig. 6. Improvements in restoration accuracy with blind Kalman filter: (a) improved Corrs. and (b) improved SERs. SNR = 20 dB to −10 dB.

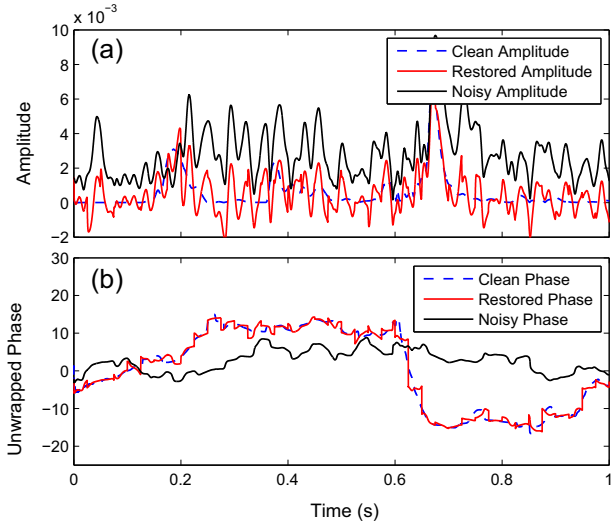

Fig. 7. Example of comparisons of (a) clean, restored, and noisy instantaneous amplitude and (b) clean, restored, and Noisy instantaneous unwrapped phase in sub-band (channel $k = 28$) with proposed blind Kalman filtering. SNR = −10 dB noise (white).

phase in a particular sub-band obtained with the proposed method (blind Kalman filtering). We can observe that the restored amplitude and phase match the clean amplitude and phase.

Moreover, we also chose the method of Wiener filtering (Scalart–Filho algorithm) under the same conditions to compare its effectiveness with that of our proposed methods. Based on the results in Fig. 8, we can see that both of our proposed methods could obviously improve SER and the correlation much more than Wiener filtering.

We calculated the perceptual evaluation of sound quality (PESQ) (Hu and Loizou, 2008) and SNR losses (Ma and Loizou, 2011) for all stimuli that we used in these evaluations to evaluate the quality and intelligibility of restored speech. PESQ in objective difference grades (ODGs) that covered from −0.5 (very annoying) to 4.5 (imperceptible) were used to evaluate subjective quality. SNR losses that ranged from 0.0 to 1.0 were used to evaluate the intelligibility of speech. SNR losses (0–1.0) corresponded to the percent of correctness (100–0%). The results from objective measures are listed in Table 1, which indicate that both of our proposed methods provided much better quality and improved intelligibility in restored speech than the existing methods of speech enhancement. We found from the results of evaluations that the proposed method could effectively reduce noise from both amplitude and phase and also improve quality and speech intelligibility.

We carried out our simulations under pink and babble noise conditions. The accuracy of restoration with the proposed blind Kalman filtering under pink and babble noise conditions are shown in Figs. 9 and 10. The proposed methods could reduce error and improve the correlations in pink and babble noise in all channels except for some

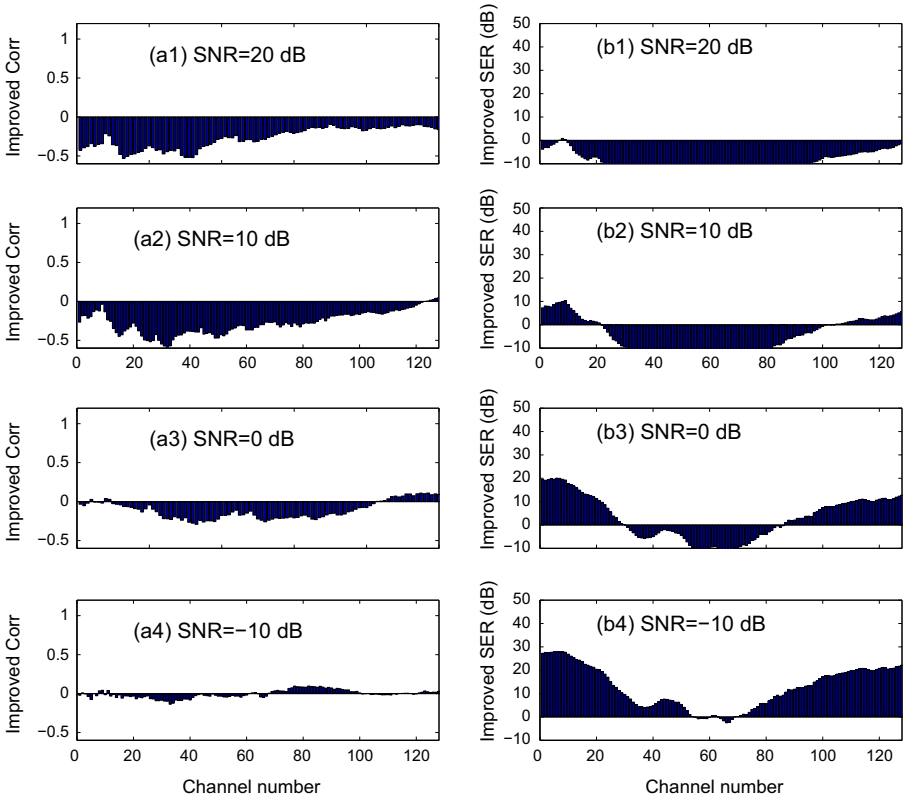

Fig. 8. Improvements in restoration accuracy with Wiener filter: (a) improved Corrs. and (b) improved SERs. SNR = 20 dB to −10 dB.

Table 1
Comparison of results for PESQ and SNR losses (averaged values).

SNR (dB)	Methods											
	Noisy speech		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	PESQ	SNR loss
20	2.99	0.57	3.61	0.53	3.50	0.54	3.11	0.83	2.73	0.84	2.72	0.88
10	2.31	0.74	3.59	0.62	3.06	0.62	2.42	0.78	2.11	0.93	1.94	0.94
0	1.65	0.87	3.10	0.67	2.47	0.71	1.75	0.91	1.71	0.94	1.76	0.97
−10	1.15	0.95	2.84	0.73	1.62	0.82	1.09	0.90	1.01	0.94	1.35	0.96

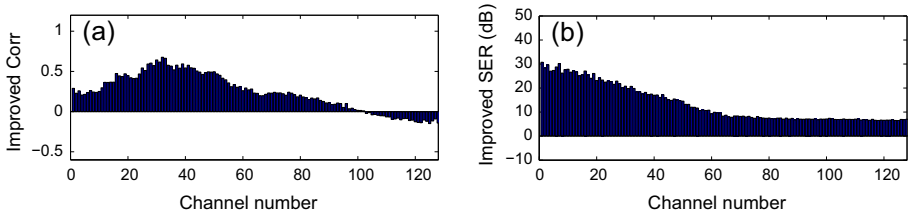

Fig. 9. Improvements in restoration accuracy with blind Kalman filter under pink noise conditions: (a) improved Corrs. and (b) improved SERs. SNR = −2.07 dB.

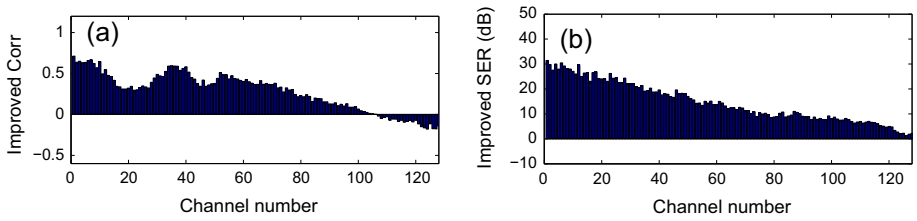

Fig. 10. Improvements in restoration accuracy with blind Kalman filter under babble noise conditions: (a) improved Corrs. and (b) improved SERs. SNR = −5.60 dB.

higher channels. We found that the signal components in higher channels were very similar to those of clean speech and had very high SNR. Thus, it was very difficult to reduce noise in higher channels since it was very similar to that in clean speech.

4.2. Subjective evaluation

The subjective evaluation was performed through a sentence-pair listening test. The listening materials included three noise types (white, pink, and babble) at an SNR of 0 dB for two male speakers and two female speakers from the TIMIT database. We restored noisy speech using four methods (MMSE, Wiener filter, the proposed non-blind method, and the proposed blind method). We compared them under six conditions (MMSE, Wiener filter, the proposed non-blind method, the proposed blind method, clean speech, and noisy speech). Each of these six was compared with the other five restored signals. Thus, we had 30 sentence pairs ($5 \times 6 = 30$) for each type of noise and we had a total of 90 ($30 \times 3 = 90$) sentence pairs. These sentence pairs were randomly shuffled and divided into three groups where each group contained 30 sentence pairs. Listeners were required to mark one of the three choices for each sentence pair as to whether they preferred the first one, preferred the second one, or had no preference.

Pairwise scoring was employed: a score of +1 was awarded to the preferred method and +0 to the other, and for the no preference response each method was awarded a score of +0.5. Fourteen subjects with normal hearing participated in the experiment. The listening test was conducted in a sound-proof room. The fourteen participants were familiarized with the task during a short practice session before the formal test. Each listener evaluated three groups of sentence pairs. The normalized mean preference score from the subjective evaluation experiment is summarized in Table 2. Figs. 11–13 compare the methods under white, pink, and babble noise conditions. We found from these comparisons that the proposed methods were preferred over the conventional methods. It can also clearly be seen that the proposed blind method had very similar performance to the proposed non-blind method for all noise types.

4.3. Effectiveness of phase

We restored instantaneous amplitude by applying blind Kalman filtering and combined it with noisy instantaneous

Table 2
Comparison of mean preference scores.

Methods	Clean	Noisy	Proposed (non-blind)	Proposed (blind)	MMSE	Wiener
White noise	8.78	0.64	6.64	6.24	3.28	2.82
Pink noise	9.64	0.71	6.44	6.17	4.21	2.64
Babble noise	9.21	0.28	6.28	6.18	4.07	2.21

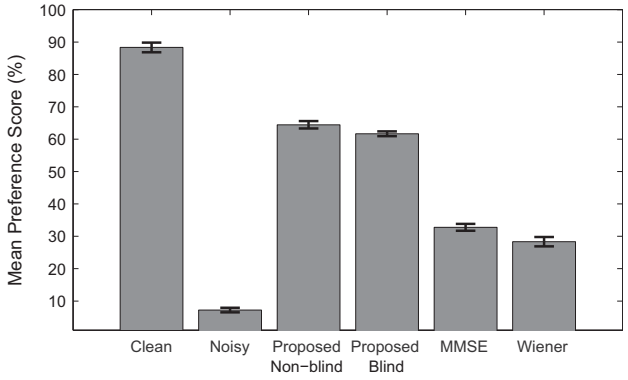

Fig. 11. Subjective evaluation under white noise conditions.

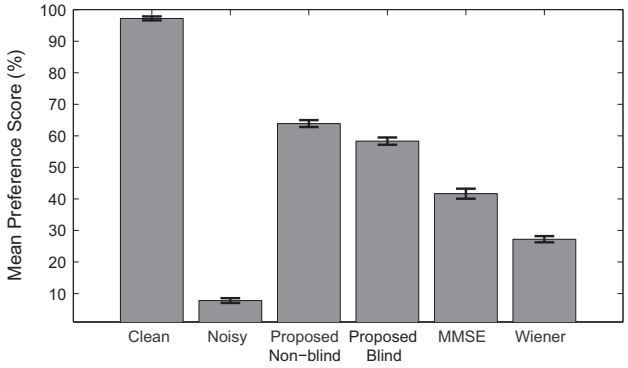

Fig. 12. Subjective evaluation under pink noise conditions.

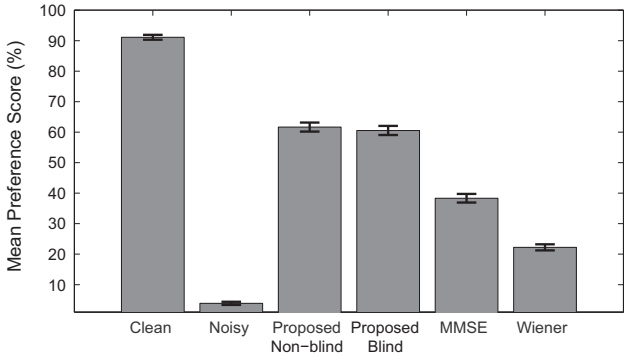

Fig. 13. Subjective evaluation under babble noise conditions.

phase to obtain enhanced speech to evaluate the effectiveness of phase in speech enhancement. The results obtained for the accuracy of amplitude restoration in terms of correlation and SER are shown in Fig. 14. It can easily be observed that for better speech enhancement we need to incorporate both amplitude and phase by comparing the results without the incorporating phase in Fig. 14 and those with the restoring phase in Fig. 6. We obtained better positive improvements by restoring both amplitude and phase under all SNR conditions, as shown in Fig. 6. The same results were also observed for improvements to quality and intelligibility. That is, we obtained restored speech with better quality and intelligibility, whenever restoration was applied both to amplitude and phase, as listed in Table 3. We compared the PESQ and SNR losses for restored speech (both amplitude and phase), amplitude only restored speech, and phase only restored speech. The results from the evaluations indicated that for better speech enhancements we need to consider both amplitude and phase to reduce noise and to effectively improve quality and intelligibility.

5. Discussion

The proposed method enhanced both instantaneous amplitudes and instantaneous phases for better speech improvements. Our main purpose in this paper is to point out the important role of phase for AMS-GTFB. We could ensure the important role of phase in speech enhancements from the objective and subjective evaluations. We evaluated the performance of the proposed method under various noise and SNRs conditions. We found from comparisons with typical methods that the proposed method provided better solutions in all cases. This is because the proposed method enhanced both amplitude and phase rather than only improved amplitude compared to the other methods. We used a Kalman filter in the proposed method for enhancement and it required noise to be white Gaussian noise for best optimal estimates. We assumed that the noise had white and Gaussian properties in the beginning.

We concentrated on the noise assumptions of Kalman filtering to test and verify these assumptions. We investigated the statistical properties of noise in each sub-band to check Gaussianity and whiteness. We used the K–S test to validate the Gaussian assumptions of noise to carry out a formal test of Gaussianity. We considered ten samples of noise as a statistical sample and applied the K–S test to verify the PSD and distribution in each sub-band. To obtain the observation noise of amplitude and phase in the sub-bands, we used the following equations.

$$V_{A,k} = S_{A,k} - A_{A,k} \tag{34}$$

$$V_{\phi,k} = S_{\phi,k} - A_{\phi,k}. \tag{35}$$

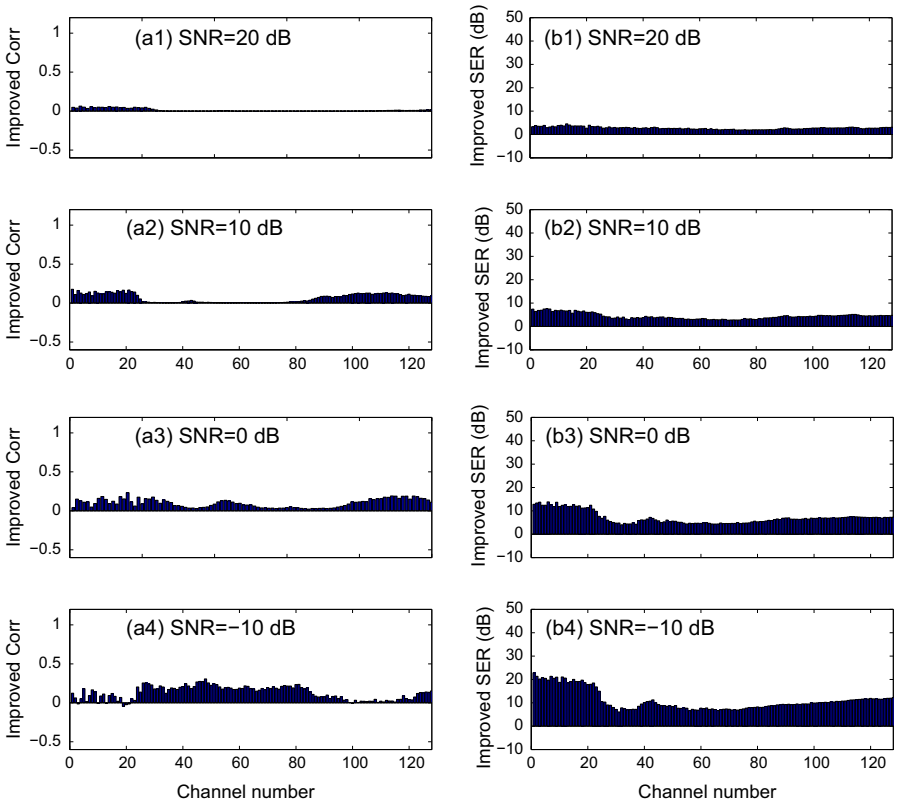

Fig. 14. Improvements in restoration accuracy of amplitude only using blind Kalman filter: (a) improved Corrs. and (b) improved SERs. SNR = 20 dB to −10 dB.

Table 3
Comparison of restored speech with amplitude only restoration and phase only restoration (averaged values).

SNR (dB)	Methods							
	Noisy speech		Restored speech (blind)		Amplitude only restoration (blind)		Phase only restoration (blind)	
	PESQ	SNR loss	PESQ	SNR loss	PESQ	SNR loss	PESQ	SNR loss
20	2.99	0.57	3.50	0.54	3.46	0.55	3.08	0.61
10	2.31	0.74	3.06	0.62	2.97	0.66	2.44	0.73
0	1.65	0.87	2.47	0.71	2.39	0.77	1.83	0.85
−10	1.15	0.95	1.62	0.82	1.47	0.88	1.30	0.94

where $A_{A,k}$ and $A_{\phi,k}$ are the clean amplitude and phase in the kth sub-band. Driven noise was considered to be an LP residue.

The noise assumption for driven noise in the sub-bands could be partially satisfied by using the K–S test; unfortunately, it could not be satisfied in all sub-bands. The values were always positive for observation noise in instantaneous amplitude, which may be a reason for the rejection of Gaussianity. We could see partial acceptance for only driven noise in both instantaneous amplitude and phase through segmental LP analysis by using the K–S test, but not for all.

We calculated the normalized PSD and distribution of noise in each sub-band. The PSD plots give almost a fixed power for all the frequencies for white noise. As an example, the results for the PSD of a particular sub-band ($k = 28$) for the observation noise of both instantaneous amplitude and phase are shown in Figs. 15(a) and 16(a). The results indicate that the PSD of observation noise, both in amplitude and phase have a flat constant power and we also observed flat PSD properties in other sub-bands. We can also obtain similar results for driven noise, which are shown in Figs. 17(a) and 18(a) as white noise.

We analyzed the distribution of noise by plotting histograms of them. Figs. 15(b) and 16(b) show the distribution for observation noise of the amplitude and phase in a particular sub-band ($k = 28$) as an example. We can see that the distribution for the observation noise of instantaneous amplitude seems to be a normal distribution. However, the observation noise in the phase, shown in Fig. 16(b) does not have Gaussianity. The distributions for the driving noise of instantaneous amplitude and phase

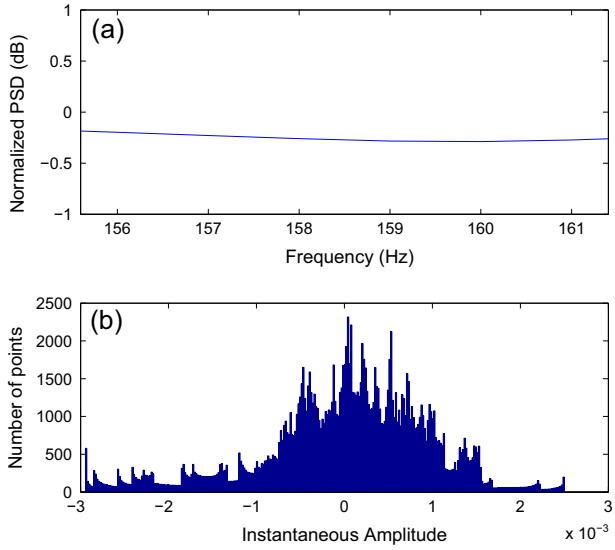

Fig. 15. Results of analysis for observation noise in Eq. (9): (a) normalized PSD of $V_{A,k}$ and (b) distribution of $V_{A,k}$ in 28th channel.

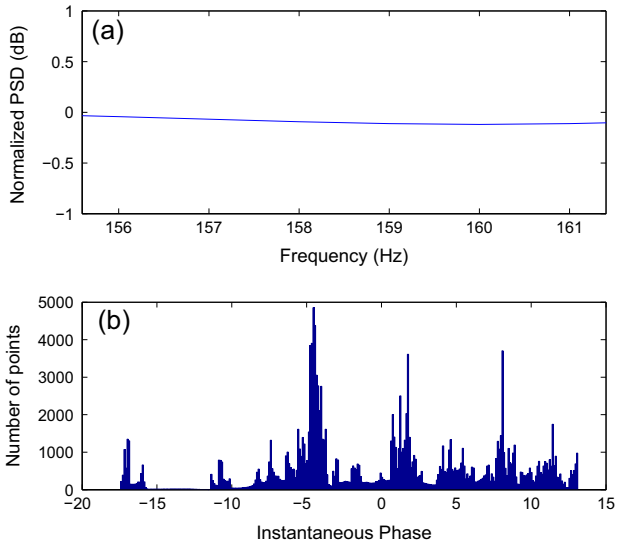

Fig. 16. Results of analysis for observation noise in Eq. (10): (a) normalized PSD of $V_{\phi,k}$ and (b) distribution of $V_{\phi,k}$ in 28th channel.

are shown in Figs. 17(b) and 18(b) in a specific sub-band ($k = 28$), which seem to be normal distributions.

We have presented the results obtained from verifying observation and driven noise in only a specific sub-band due to space limitations. However, we analyzed the results in each sub-band and found that in all channels the observation and driven noise in amplitude and phase was white noise since their PSDs maintained a flat constant power. The distribution of observation noise in amplitude, on the other hand, for most of the sub-bands, could not maintain a Gaussian distribution. Thus, we statistically failed to confirm that the distributions of driven noise in the amplitude and phase in all channels provided Gaussian distributions.

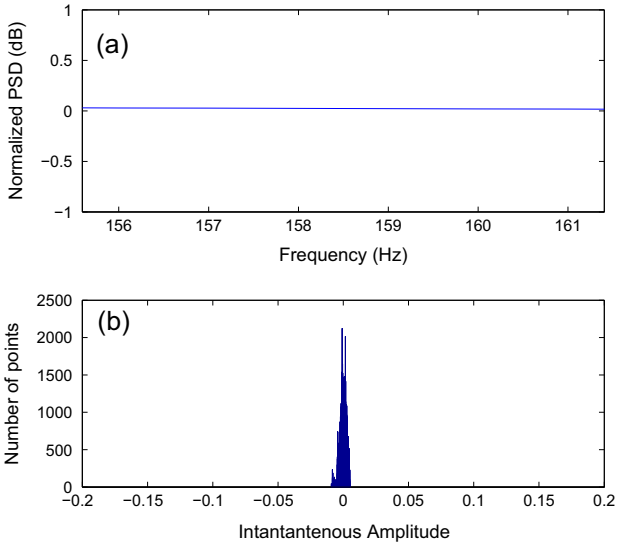

Fig. 17. Results of analysis for driving noise in Eq. (7): (a) normalized PSD of $W_{A,k}$ and (b) distribution of $W_{A,k}$ in 28th channel.

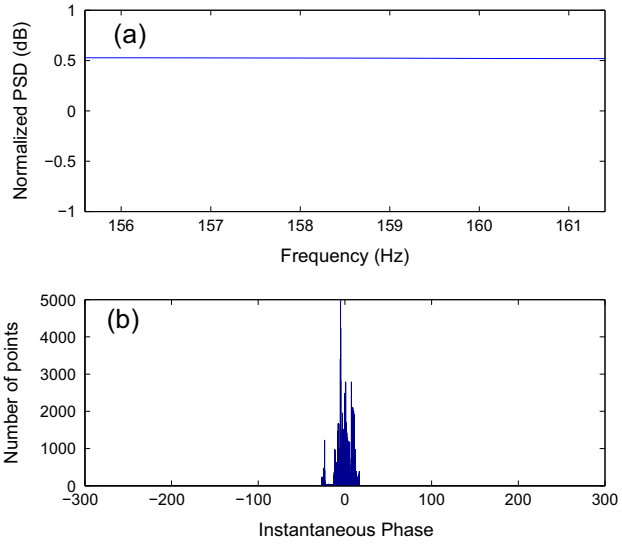

Fig. 18. Results of analysis for driving noise in Eq. (8): (a) normalized PSD of $W_{\phi,k}$ and (b) distribution of $W_{\phi,k}$ in 28th channel.

Hence, both driven noise in Eqs. (7) and (8) and observation noise in Eqs. (9) and (10) may generally be able to be regarded as white noise but not Gaussian noise.

We will consider the noise assumptions for Kalman filtering in this section. We believe that if the noise assumptions for Kalman filtering are satisfied, we can obtain more improvements in the evaluation results by using the proposed method. The extended non-linear Kalman filter had been considered for non-Gaussian noise in the literature (Cipra and Rubio, 1991). The current Kalman filter was considered under non-Gaussian conditions, so we need to investigate a non-linear Kalman filter to solve this issue in the future. We intend to extend our proposed scheme by using this approach in future work.

6. Conclusion

We proposed a speech enhancement scheme by using a Kalman filter with/without a training phase in the sub-bands as the AMS on the Gammatone filterbank. We presented a scheme that achieved greater speech enhancements by simultaneously improving the instantaneous amplitude and instantaneous phase. We concentrated on central processing by Kalman filtering by providing an effective way of training the LP coefficients for blind Kalman filtering. The results from our objective test revealed that the proposed scheme with the training phase performed almost like non-blind Kalman filtering and also outperformed the existing conventional speech enhancement algorithms in terms of improvements in speech quality, intelligibility, and preferences. The results from subjective evaluation also indicated listener's preferences for our proposed method. We believe that this resulted from the effect of combining phase enhancement with amplitude enhancement and accurate LP coefficients in the sub-band representation. Although the proposed method always yielded improvements, future research is required to ensure improvements in all channels under high SNR conditions. We plan to investigate whether the proposed scheme can be extended to enhance speech in noisy reverberant environments.

The proposed method concentrated on speech enhancements by modeling the Kalman filter to the updated equations of instantaneous amplitude and phase. It dealt with temporal variations in the instantaneous amplitude and phase of the speech signal rather than noise. However, although this method can also be used in speech enhancements under non-stationary conditions, future investigations are required to ensure this occurs.

Acknowledgments

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

References

Boll, S., 1979. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust. Speech Signal Process. ASSP-27 (2), 113–120.

Ephraim, Y., Malah, D., 1984. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Process. ASSP-32 (6), 1109–1211.

Scalart, P., Filho, J.V., 1996. Speech enhancement based on a priori signal to noise estimation. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processings (ICASSP), pp. 629–623.

Ephraim, Y., Malah, D., Juang, B.-H., 1989. On the application of hidden Markov models for enhancing noisy speech. IEEE Trans. Acoust. Speech Signal Process. ASSP-37 (12), 1846–1856.

Ephraim, Y., 1992. A Bayesian estimation approach for speech enhancement using hidden Markov models. IEEE Trans. Signal Process. 40 (4), 725–735.

Zhao, D.Y., Kleijn, B.W., 2007. HMM-based gain modeling for enhancement of speech in noise. IEEE Trans. Audio, Speech, Lang. Process. 15 (3), 882–892.

Sriram, S., Jonas, S., Kleijn, W.B., 2007. Codebook-based Bayesian speech enhancement for nonstationary environments. IEEE Trans. Audio, Speech, Lang. Process. 15 (2), 441–452.

Veisi, H., Hossein, S., 2013. Speech enhancement using hidden Markov models in Mel-frequency domain. Speech Commun. 55 (2), 205–220.

Nishikawa, T., Saruwatari, H., Shikano, K., 2003. Blind source separation of acoustic signals based on multistage ICA combining frequency-domain ICA and time-domain ICA. IEICE Trans. Fund. Electron., Commun. Comput. Sci. 86 (4), 846–858.

So, S., Paliwal, K.K., 2011. Modulation-domain Kalman filtering for single-channel speech enhancement. Speech Commun. 53 (6), 818–829.

Paliwal, K.K., Schwerin, B., Wójcicki, K., 2012. Speech enhancement using a minimum mean-square error short-time spectral modulation magnitude estimator. Speech Commun. 54 (2), 282–305.

Ji, M., Srinivasan, R., Crooks, D., 2011. A corpus-based approach to speech enhancement from nonstationary noise. IEEE Trans. Audio, Speech, Lang. Process. 19 (4), 822–836.

Ruofei, C., Cheung-Fat, C., Cheung, H.S., 2012. Model-based speech enhancement with improved spectral envelope estimation via dynamics tracking. IEEE Trans. Audio, Speech, Lang. Process. 20 (4), 1324–1336.

Mohammadiha, N., Smaragdis, P., Leijon, A., 2013. Supervised and unsupervised speech enhancement using nonnegative matrix factorization. IEEE Trans. Audio, Speech, Lang. Process. 21 (10), 2140–2151.

Sawada, H., Kameoka, H., Araki, S., Ueda, N., 2013. Multichannel extensions of non-negative matrix factorization with complex-valued data. IEEE Trans. Audio, Speech Lang. Process. 21 (5), 971–982.

Wang, D., Lim, J.S., 1982. The unimportance of phase in speech enhancement. IEEE Trans. Acoust. Speech Signal Process. ASSP-30 (4), 679–681.

Shannon, B.J., Paliwal, K.K., 2006. Role of phase estimation in speech enhancement. In: Proceedings of IEEE SAPA@ INTERSPEECH, pp. 1427–1430.

Paliwal, K.K., Alsteris, L.D., 2005. On the usefulness of STFT phase spectrum in human listening tests. Speech Commun. 45 (2), 153–170.

Roux, J.L., Ono, N., Sagayama, S., 2008. Explicit consistency constraints for STFT spectrograms and their application to phase reconstruction. In: Proceedings of IEEE SAPA@ INTERSPEECH, pp. 23–28.

Fardkhaleghi, P., Savoji, M.H., 2010. New Approaches to speech enhancement using phase correction in Wiener Filtering. 5th IEEE International Symposium on Telecommunications (IST), pp. 895–899.

Zhang, Y., Zhao, Y., 2013. Real and Imaginary modulation spectral subtraction for speech enhancement. Speech Commun. 55 (4), 509–522.

Kleinschmidt, T., Sridharan, S., Mason, M., 2011. The use of phase in complex spectrum subtraction for robust speech recognition. Comp. Speech Lang. 25 (3), 585–600.

Paliwal, K.K., Wójcicki, K., Shannon, B., 2011. The importance of phase in speech enhancement. Speech Commun. 53 (4), 465–494.

Loizou, P.C., Kim, G., 2011. Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions. IEEE Trans. Audio, Speech, Lang. Process. 19 (1), 47–56.

Drullman, R., 1995. Temporal envelope and fine structure cues for speech intelligibility. J. Acoust. Soc. Am. 97 (1), 585–592.

Moore, B.C.J., 2008. The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. J. Assoc. Res. Otolaryngol. 9 (4), 399–406.

Swaminathan, J., 2010. The Role of Envelope and Temporal Fine Structure in the Perception of Noise Degraded Speech, Ph.D Thesis. Purdue University.

Swaminathan, J., Heinz, M.G., 2012. psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise. J. Neurosci. 32 (5), 1747–1756.

Unoki, M., Akagi, M., 1999. A method of signal extraction from noisy signal based on auditory scene analysis. Speech Commun. 27 (3), 261–279.

Nower, N., Liu, Y., Unoki, M., 2014. Restoration of instantaneous amplitude and phase using Kalman filter for speech enhancement. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processings (ICASSP), pp. 4666–4670.

Massey Jr., F.J., 1951. The Kolmogorov–Smirnov test for goodness of fit. J. Am. Statist. Assoc. 46 (253), 68–78.

Satorra, A., Bentler, P.M., 2001. A scaled difference chi-square test statistic for moment structure analysis. Psychometrika 66 (4), 507–514.

Paliwal, K.K., Atal, B.S., 1993. Efficient vector quantization of LPC parameters at 24 bits/frame. IEEE Trans. Audio, Speech, Lang. Process. 1 (1), 3–14.

Hu, Y., Loizou, P.C., 2008. Evaluation of objective quality measures for speech enhancement. IEEE Trans. Audio, Speech, Lang. Process. 16 (1), 229–238.

Ma, J., Loizou, P.C., 2011. SNR loss: a new objective measure for predicting the intelligibility of noise-suppressed speech. Speech Commun. 53 (3), 340–354.

Cipra, T., Rubio, A., 1991. Kalman filter with a non-linear non-Gaussian observation relation. Trabajos de Estadistica 6 (2), 111–119.

View publication stats