

KALMAN FILTER WITH PHASE SPECTRUM COMPENSATION ALGORITHM FOR SPEECH ENHANCEMENT

Stephen So, Kamil K. Wójcicki, James G. Lyons, Anthony P. Stark, and Kuldip K. Paliwal

Signal Processing Laboratory, Griffith School of Engineering
Griffith University, Brisbane, QLD, Australia, 4111.
{s.so, k.wojcicki, j.lyons, a.stark, k.paliwal}@griffith.edu.au

ABSTRACT

In this paper, we propose to combine the Kalman filter with a recent speech enhancement technique, called the phase spectrum compensation procedure, or PSC. More specifically, we apply the PSC technique to initialise the Kalman filter, whereby PSC is used to clean the noisy speech prior to LPC estimation for the Kalman recursion. We refer to the combined technique as the Kalman-PSC filter. Using an objective speech quality measure, formal subjective listening tests and spectrogram analysis, we show that the proposed method results in improved speech quality.

Index Terms— Kalman filter, phase spectrum compensation (PSC), speech enhancement

1. INTRODUCTION

The *Kalman filter* is an unbiased, time-domain, linear MMSE estimator that was first proposed for speech enhancement by Paliwal and Basu [1]. In speech enhancement we are primarily interested in suppression of noise from noise corrupted speech so that the speech quality and intelligibility are improved. The Kalman filter has been of particular interest in speech enhancement, due to several advantages it has over other spectral domain-based enhancement methods: (1) the speech production model is made inherent in the Kalman recursion equations; (2) enhanced speech from the Kalman filter contains no ‘musical noise’ [2];¹ and (3) the Kalman filter can process non-stationary speech signals. The enhancement performance of the Kalman filter is somewhat dependent on the accuracy and reliability of the linear prediction coefficient (LPC) estimates. Ideally, LPC coefficients estimated from the clean speech should be used for enhancement of the noisy speech using the Kalman filter (as was done in [1]). However, in practice this is not possible as the clean speech is not known *a priori* and the LPC coefficients have to be estimated from the noisy speech. Depending on the noise characteristics and the signal-to-noise ratio (SNR), the LPCs obtained from the noisy speech using the conventional autoregressive (AR) spectral estimation methods (such as the auto-correlation method) may be inaccurate and unreliable, leading to a suboptimal Kalman filter.

One method of overcoming this problem is to use an iterative LPC estimation method [3], where in each iteration, the Kalman filter enhances the speech prior to LPC estimation. While this method generally improves the SNR after a few iterations, the enhanced speech is often distorted and accompanied by some residual noise. The limitation of the iterative method may stem from the fact that the Kalman filter, in the first iteration, is itself dependent on the availability of accurate LPC estimates.

¹Assuming that accurate linear prediction coefficients are available.

In this paper, we propose to combine the Kalman filter with a recent speech enhancement technique called the *phase spectrum compensation* (PSC) [4, 5]. More specifically, we use the PSC technique to enhance the noisy speech prior to LPC estimation for the Kalman recursion. We refer to the proposed approach as the Kalman-PSC filter. We evaluate the performance of the combined technique using both subjective and objective measures, as well as spectrogram analysis. Our experimental study, on enhancing speech corrupted by additive white noise, shows that the Kalman-PSC filter improves speech quality and compares favourably with other speech enhancement methods.

2. KALMAN FILTER FOR SPEECH ENHANCEMENT

2.1. Conventional Kalman filter

If the clean speech is represented as $x(n)$ and the additive noise signal as $v(n)$, then the noise-corrupted speech $y(n)$, which is the only observable signal in practice, is expressed as:

$$y(n) = x(n) + v(n). \quad (1)$$

In the Kalman filter, that is used for speech enhancement (see Fig. 1), $v(n)$ is a zero-mean white Gaussian noise that is uncorrelated with $x(n)$.² A p th order linear predictor is used to model the speech signal:

$$x(n) = -\sum_{k=1}^p a_k x(n-k) + w(n), \quad (2)$$

where $\{a_k, k = 1, 2, \dots, p\}$ are the LPCs and $w(n)$ is the white Gaussian excitation with zero mean and a variance of σ_w^2 . Rewriting Eqs. (1) and (2) using state vector representation gives:

$$\mathbf{x}(n) = \mathbf{A}\mathbf{x}(n-1) + \mathbf{d}w(n) \quad (3)$$

$$y(n) = \mathbf{c}^T \mathbf{x}(n) + v(n), \quad (4)$$

where $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-p+1)]^T$ is the ‘hidden’ state vector, \mathbf{A} is the linear prediction state transition matrix, while $\mathbf{d} = [1, 0, \dots, 0]^T$ and $\mathbf{c} = [1, 0, \dots, 0]^T$ are the measurement vectors for the excitation noise and observation, respectively. The linear prediction state transition matrix is given by:

$$\mathbf{A} = \begin{bmatrix} -a_1 & -a_2 & \dots & -a_{p-1} & -a_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}. \quad (5)$$

²Coloured noise can be modelled by another linear predictor model, which can be augmented into the Kalman state vector equations [3].

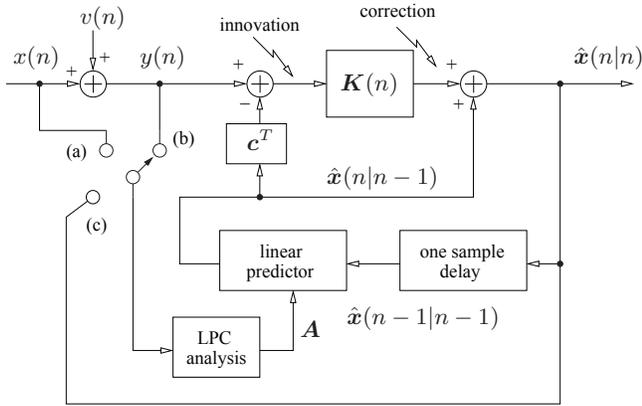


Fig. 1. Block diagram of the Kalman filter showing different LPC estimation methods (bolded variables are vectors). LPCs are estimated from: (a) clean speech $x(n)$; (b) noise-corrupted speech $y(n)$; (c) Kalman-enhanced speech (iterative LPC estimation).

The Kalman filter calculates $\hat{x}(n|n)$, which is an unbiased and linear MMSE estimate of the state vector $x(n)$, given the current sample of corrupted speech $y(n)$, by using the following recursive equations:

$$\hat{x}(n|n-1) = A\hat{x}(n-1|n-1) \quad (6)$$

$$P(n|n-1) = AP(n-1|n-1)A^T + \sigma_w^2 dd^T \quad (7)$$

$$K(n) = P(n|n-1)c[\sigma_v^2 + c^T P(n|n-1)c]^{-1} \quad (8)$$

$$\hat{x}(n|n) = \hat{x}(n|n-1) + K(n)[y(n) - c^T \hat{x}(n|n-1)] \quad (9)$$

$$P(n|n) = [I - K(n)c^T]P(n|n-1). \quad (10)$$

2.2. Accurate estimation of linear prediction coefficients

Ideally, the LPCs (to which we will refer to as ‘clean LPCs’) should be estimated from the clean speech in order for the Kalman filter to attain its full potential (see Fig. 1, method (a)). However, in practice only the noise-corrupted speech $y(n)$ is available, so the LPCs (to which we will refer to as ‘noisy LPCs’) need to be estimated from the noisy speech (see Fig. 1, method (b)). As has been reported numerous times in the literature (recently in [6]), the performance of the Kalman filter based on the noisy LPCs is suboptimal, where the enhanced speech is always accompanied by a large amount of wideband residual noise.

Gibson *et al.* [3] reported an iterative LPC estimation method, where each frame of speech is first enhanced by the Kalman filter initialised with the noisy LPCs (Fig. 1, method (b)). Then a new set of LPCs are estimated from the Kalman-enhanced speech (Fig. 1, method (c)). Convergence is usually achieved after three iterations, along with a large SNR improvement. The enhanced speech from the iterative Kalman filter is free from the wideband residual noise that plagued the non-iterative method, however some ‘musical-like’ artifacts do remain. The enhanced speech also suffers from distortion which can degrade intelligibility. A weakness of this method is the use of the noisy LPCs in the first iteration of the Kalman filter. The lack of accurate LPCs for the Kalman filter during this stage means that it will provide a poor initial estimate of the clean speech.

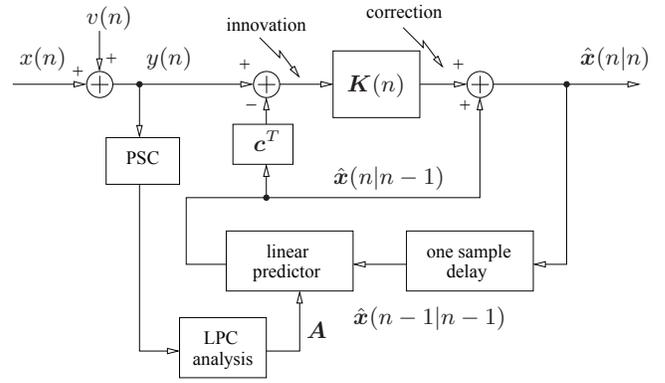


Fig. 2. Block diagram of the proposed Kalman-PSC filter.

2.3. Proposed approach

Recently, a novel approach to speech enhancement called the *phase spectrum compensation* (PSC) has been proposed in [4, 5]. PSC is based on the analysis-modification-synthesis (AMS) framework. In the PSC method, the noisy phase spectrum gets compensated for additive noise distortion. The PSC method uses a synthesis-based cancellation to attenuate lower energy signal components much more than higher energy components, resulting in background noise reduction. Both short-time magnitude and phase spectra play important roles in the speech enhancement process. This is unlike traditional AMS-based speech enhancement techniques. Importantly, PSC manages to reduce background noise without significantly distorting high-energy components belonging to the speech signal. The above properties of the PSC method make it of interest to us in the present work. We propose to employ the PSC procedure to enhance noisy speech prior to LPC computation within the Kalman filter framework. We refer to the proposed approach as Kalman-PSC filter. A block diagram of the proposed method is shown in Fig. 2. For comparison, we also investigate the use of the popular MMSE algorithm [7] in-place of the PSC method within the Kalman filter. We refer to this approach as the Kalman-MMSE filter.

3. EXPERIMENTS

3.1. Speech corpus

In our experiments we use the NOIZEUS speech corpus, which is composed of 30 phonetically-balanced sentences belonging to six speakers [8]. The corpus is sampled at 8 kHz and filtered to simulate receiving frequency characteristics of telephone handsets. For our experiments we generate a stimuli set corrupted by additive white Gaussian noise at four SNR levels: 0, 5, 10 and 15 dB.

3.2. Evaluation methods

In this study, we conducted both subjective and objective evaluations as well as spectrogram analysis. The subjective evaluation was in a form of AB listening tests to determine subjective method preference [9]. Two NOIZEUS sentences, belonging to different speakers, were included in these tests. White noise case (10 dB SNR) was investigated. Nine treatment types were used, seven corresponding to speech enhancement methods and two to clean and noisy speech. Stimuli pairs were played back to the listeners. The listeners were asked to make a subjective preference for each stimuli pair. Fourteen

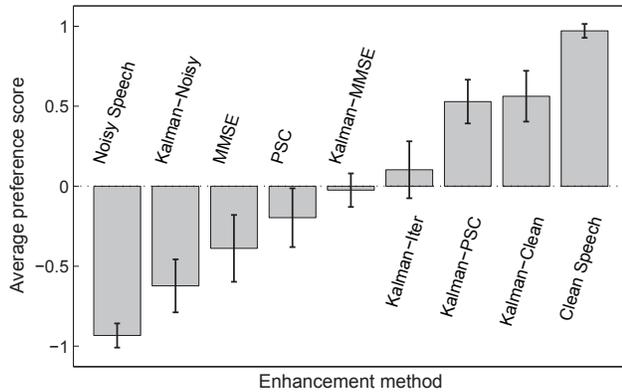


Fig. 3. Subjective results in terms of average preference scores for additive white Gaussian noise at 10 dB SNR.

English speaking listeners participated in the listening tests. The objective evaluation was carried out on the NOIZEUS corpus using an objective speech quality measure, namely PESQ.

3.3. Speech enhancement methods

In addition to the proposed method, various forms of the Kalman filter used for speech enhancement were included in our experiments. For completeness, the PSC and MMSE methods were also added. A total of seven speech enhancement methods, as listed below, were considered in our evaluations.

- Kalman-Clean – Kalman filter with clean LPCs (*oracle* case) [1]
- Kalman-Noisy – Kalman filter with noisy LPCs [6]
- Kalman-Iter – Kalman filter with iterative LPC estimation (three iterations) [3]
- Kalman-PSC – Kalman filter with LPCs estimated from noisy speech enhanced using the PSC algorithm (*proposed*)
- Kalman-MMSE – Kalman filter with LPCs estimated from noisy speech enhanced using the MMSE algorithm
- PSC – the phase spectrum compensation procedure [5]
- MMSE – the MMSE algorithm [7]

4. RESULTS AND DISCUSSION

The results of the subjective tests, along with their standard error bars, are shown in Fig. 3. The subjective results are in terms of average (method) preference scores. A score of one for a particular method, means that the method was preferred over all other methods every time. Conversely, a score of minus one means that the method was never preferred. As can be seen from the subjective results, all enhancement methods were preferred over the noisy speech. The most preferred method was the Kalman-Clean method. This is not surprising as Kalman-Clean is an *oracle* case method, where clean LPCs are made available during the enhancement process. At the other end of the scale, the least preferred method was Kalman-Noisy. This is also to be expected, as this method uses the noisy LPCs, which is suboptimal as discussed in Section 2.2. The

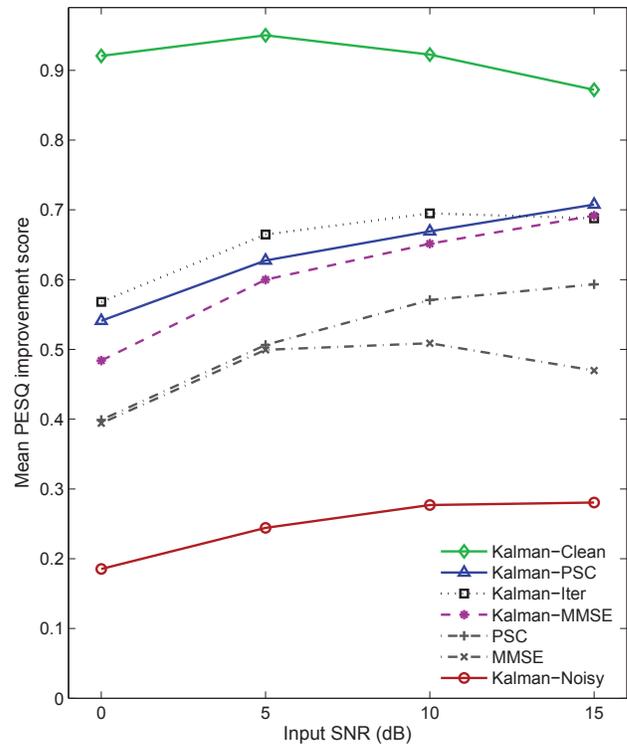


Fig. 4. Objective results in terms of mean PESQ improvement scores for additive white Gaussian noise at various input SNRs.

non-Kalman methods, namely MMSE and PSC, performed somewhat worse than the remaining Kalman methods. Overall, the listener preference for the Kalman-PSC method was found to be significantly higher than the preference for the Kalman-Noisy, MMSE, PSC, Kalman-MMSE and Kalman-Iter methods. Notably, the subjective preference for Kalman-PSC was found to be comparable to the preference for Kalman-Clean. These results suggest that employing the PSC procedure on noisy speech prior to LPC estimation leads to improved LPC estimates. This can be attributed to the fact that PSC attenuates lower energy signal components much more than higher energy components, thus reducing background noise without significantly distorting the speech signal.

The results of the objective experiment, in terms of mean PESQ improvement scores, are shown in Fig. 4. The objective results show a trend similar to the subjective results discussed in the preceding paragraph. In terms of objective results, Kalman-PSC also works better than most of the other methods. Only exception is Kalman-Iter, which seems to perform slightly better than Kalman-PSC. As we know the objective measures do not always reflect the subjective findings. The reason why subjective preference for Kalman-Iter is somewhat low can be explained by the presence of distracting residual noise that can be seen in the Kalman-Iter enhancement spectrogram in Fig. 5(e).

From the spectrograms in Fig. 5 it can also be seen that the noisy speech enhanced using Kalman-PSC and Kalman-Clean (Fig. 5(f,g), respectively) contains significantly less residual noise than the speech enhanced using Kalman-Noisy, Kalman-MMSE and Kalman-Iter methods (Fig. 5(c,d,e), respectively).

5. CONCLUSION

In this paper, we have proposed the Kalman-PSC filter. In the proposed method, the phase spectrum compensation (PSC) procedure is employed to enhance the noisy speech prior to LPC estimation for the Kalman recursion. Using an objective speech quality measure, spectrogram analysis, as well as formal subjective listening tests, we showed that the proposed method results in improved speech quality.

6. REFERENCES

- [1] K. Paliwal and A. Basu, "A speech enhancement method based on Kalman filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, vol. 12, Apr. 1987, pp. 177–180.
- [2] N. Ma, M. Bouchard, and R. Goubran, "Speech enhancement using a masking threshold constrained Kalman filter and its heuristic implementations," *IEEE Trans. Speech Audio Process.*, vol. 14, no. 1, pp. 19–32, Jan. 2006.
- [3] J. Gibson, B. Koo, and S. Gray, "Filtering of colored noise for speech enhancement and coding," *IEEE Trans. Signal Process.*, vol. 39, no. 8, pp. 1732–1742, Aug. 1991.
- [4] K. Wójcicki, M. Milacic, A. Stark, J. Lyons, and K. Paliwal, "Exploiting conjugate symmetry of the short-time Fourier spectrum for speech enhancement," *IEEE Signal Process. Lett.*, vol. 15, pp. 461–464, 2008.
- [5] A. Stark, K. Wojcicki, J. Lyons, and K. Paliwal, "Noise driven short time phase spectrum compensation procedure for speech enhancement," in *Proc. Int. Conf. Spoken Language Processing (INTERSPEECH-ICSLP)*, Sep. 2008.
- [6] S. So and K. Paliwal, "A long state vector Kalman filter for speech enhancement," in *Proc. Int. Conf. Spoken Language Processing (INTERSPEECH-ICSLP)*, Sep. 2008.
- [7] 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, pp. 1109–1121, Dec. 1984.
- [8] Y. Hu and P. Loizou, "Subjective comparison of speech enhancement algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, 2006, pp. 153–156.
- [9] P. Sorqvist, P. Handel, and B. Ottersten, "Kalman filtering for low distortion speech enhancement in mobile communication," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, vol. 2, 1997, pp. 1219–1222.

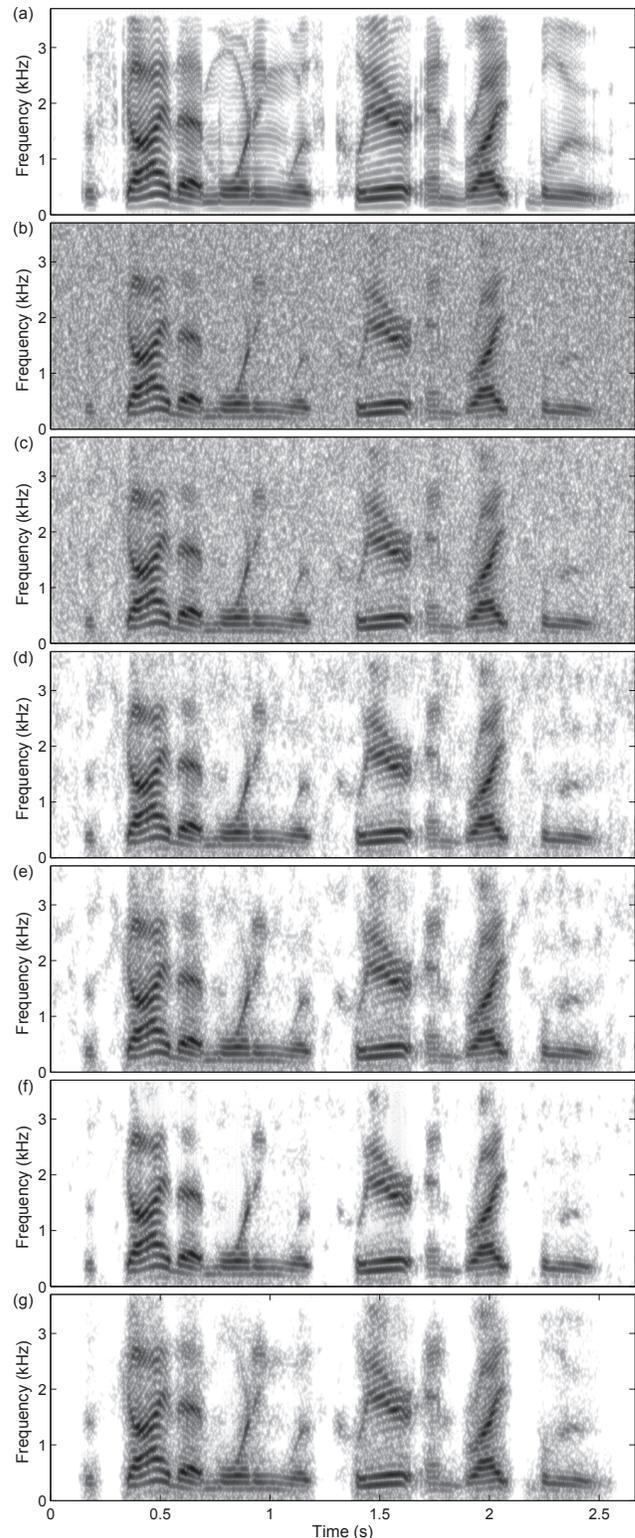


Fig. 5. Spectrograms of sp10.wav utterance, ‘The sky that morning was clear and bright blue,’ by a male speaker from the NOIZEUS corpus: (a) clean speech; (b) speech degraded by white noise (10 dB SNR); as well as noisy speech enhanced using: (c) Kalman-Noisey; (d) Kalman-MMSE; (e) Kalman-Iter; (f) Kalman-PSC; and (g) Kalman-Clean.