Int. J. Electron. Commun. (AEÜ) 69 (2015) 545–554

Contents lists available at ScienceDirect

International Journal of Electronics and Communications (AEÜ)

journal homepage: www.elsevier.com/locate/aeue

LPC-based formant enhancement method in Kalman filtering for speech enhancement

Tarek Mellahi*, Rachid Hamdi

Department of Electronics, Badji Mokhtar University, BP12, Sidi Amar, Annaba, Algeria

ARTICLE INFO

Article history:
Received 21 February 2014
Accepted 12 November 2014

Keywords:
Kalman filtering
Speech enhancement
Linear prediction coefficients
Formant enhancement method

ABSTRACT

In this work, we are concerned by a new iterative Kalman filtering scheme where a linear predictor model parameters are estimated from noisy speech. However, when only noise-corrupted speech is available, the enhancement performance of the Kalman filter is somewhat dependent on the accuracy of the linear prediction coefficients (LPCs) and excitation variance estimates. Nevertheless, linear prediction based speech (LPC) analysis is known to be sensitive to the presence of additive noise. To overcome this problem we present in this paper an analysis and application of the LPC-based formant enhancement method by modifying the log magnitude spectrum of the LPC model and then re-evaluating new LPCs to be applied on the Kalman filter. These enhanced LPCs are useful indicator of Kalman filter performance. Our enhancement experiments use a NOIZEUS speech corpus where the proposed method achieves higher objective and subjective results compared with other enhancement methods.

1. Introduction

Speech enhancement in noisy environments, such as in streets, cars, trains, aircrafts, and noisy public venues, improves the quality and intelligibility of speech and reduces communication fatigue. Noise reduction benefits a wide range of applications as mobile phones, hands-free phones, teleconferencing, in-car cabin communication, hearing aids, automated voice services based on speech recognition and synthesis. Various speech enhancement methods have been reported in the literature and these include spectral subtraction [1], MMSE estimation methods [2], Wiener filtering [3], subspace methods [4], and Kalman filtering [5].

The Kalman filter is an unbiased, time-domain, linear minimum mean squared error (MMSE) estimator that originated from control systems theory [6]. Its role is to estimate the unknown states of a dynamic system, using a linear combination of noise-corrupted observations and predicted states. The Kalman filter has been of particular interest in speech enhancement because of several advantages it has over other spectral domain-based enhancement methods: (1) the speech production model is made inherent in the Kalman recursion equations by using a linear predictor as the dynamic model; (2) when accurate LPCs are available, the enhanced speech from the ideal Kalman filter contains no random frequency tones; (3) the Kalman filter makes no stationary assumptions, unlike the Wiener filter; (4) the Kalman filter can be turned-on at the first sample $n = 0$, where the recursion parameters are initialized with their expected values; and (5) the non-stationary Kalman filter can be viewed as a joint estimator for both the magnitude and phase spectrum of speech [7].

A widely used source-filter model of speech is the linear prediction (LP) model. Which is used for speech coding [8], recognition [9] and enhancement [10]. The LPCs are found by LPC estimation, which describes the inverse transfer function of the human vocal tract.

The enhancement performance of the Kalman filter is somewhat dependent on the accuracy of the LPC and excitation variance estimates. Ideally, these coefficients should be obtained from the clean speech, as was done by [5]. However, in practice, the LPCs and variances are generally not known a priori, so they must be estimated from the noise-corrupted speech. Depending on the noise characteristics and signal-to-noise ratio (SNR), the LPCs and excitation variance obtained from the noise-corrupted speech will be poor. For this reason, most of the methods that have been proposed focus on the way to estimate the LPCs and excitation variance. Thus, in [11] the authors develop an alternative suboptimal iterative speech enhancement method using expectation-maximization (EM) algorithm. In [12] the author proposes an estimation method of LPCs by using the robust recursive LS algorithm with variable forgetting factor. In addition, LPC estimates are calculated using the enhanced

* Corresponding author.
E-mail addresses: tarekmellahi@gmail.com (T. Mellahi), rachid.hamdi@univ-annaba.dz (R. Hamdi).

http://dx.doi.org/10.1016/j.aeue.2014.11.007
1434-8411/

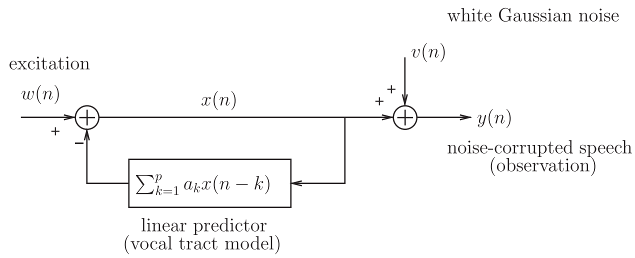

Fig. 1. Block diagram showing the speech production model (linear predictor) and white Gaussian noise that is added to the speech.

output of the Kalman filter from previous iteration [13,14]. However, these methods achieve lower SNRs. In this paper, we propose a new technique, referred to as the LPC-based formant enhancement method (LPC-FEM), for enhancing the structure of noise-corrupted speech spectrum in iterative Kalman filter.

The proposed method is based on modifying the log magnitude spectrum of the LPC model and then re-evaluating new LPCs to reduce the presence of background noise in iterative Kalman filter. The improvements provided by this method will be verified in objective and subjective speech enhancement experiments performed on the NOIZEUS speech corpus. We show that the proposed enhancement method (Kalman LPC-FEM) gives better performance than conventional iterative Kalman filtering schemes [13,14].

This paper is organized as follows. In Section 2, we describe the Kalman filtering for Speech enhancement algorithm is presented. Section 3 the Kalman LPC-FEM algorithm. Section 4, presents the objective measures and subjective scores being evaluated. The conclusions are given in Section 5.

2. Kalman filtering for speech enhancement

2.1. Kalman state-space equations of signal and noise models

The noise corrupted speech is expressed as:

$$y(n) = x(n) + v(n) \tag{1}$$

where $x(n)$ is the speech signal and $v(n)$ is the noise.

In the Kalman filter that is used for speech enhancement [5], $v(n)$ is a zero-mean, white Gaussian noise uncorrelated with $x(n)$. A p^{th} order linear predictor is used to model the speech signal:

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

where a_k are the LPCs of p^{th} order models and $w(n)$ is a white Gaussian excitation with zero mean and a variance of σ_w^2 (for more detail see Section 3.2).

Fig. 1 shows a block diagram of the speech production model and additive noise. Rewriting Eqs. (1) and (2) can be expressed in a state-space form for Kalman filtering as:

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

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

where **A** is the state transition matrix (containing the model parameters)

$$\mathbf{A} = \begin{bmatrix} -a_1 & -a_2 & \cdots & -a_{p-1} & -a_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \tag{5}$$

the hidden state vector is $\mathbf{x}(n) = [x(n), x(n-1), \ldots, x(n-p+1)]^T$ and $\mathbf{d} = \mathbf{c} = [1, 0, \ldots, 0]^T$ are p-dimensional vectors for respectively the excitation noise and observation.

2.2. Process equations

The Kalman filter recursively calculates an unbiased and linear MMSE estimate $\hat{\mathbf{x}}(n \mid n)$ of the state vector $\mathbf{x}(n)$ at time n, given the corrupted speech $y(n)$, by using the following recursive equations:

$$\hat{\mathbf{x}}(n \mid n-1) = \mathbf{A}\hat{\mathbf{x}}(n-1 \mid n-1) \tag{6}$$

$$\mathbf{P}(n \mid n-1) = \mathrm{AP}(n-1 \mid n-1)\mathbf{A}^T + \sigma_w^2 \mathrm{dd}^T \tag{7}$$

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

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

$$\mathbf{P}(n \mid n) = [\mathbf{I} - \mathbf{K}(n)\mathbf{c}^T]\mathbf{P}(n \mid n-1) \tag{10}$$

where $\hat{\mathbf{x}}(n \mid n-1)$ is a priori estimate of the current state vector at time n, given the observations up to $n-1$;

$\mathbf{P}(n \mid n-1)$ is the error covariance matrix of the a priori estimate, $\hat{\mathbf{x}}(n \mid n-1)$;

$\mathbf{K}(n)$ is the Kalman gain;

$\hat{\mathbf{x}}(n \mid n)$ is the a posteriori estimate of $\mathbf{x}(n)$, given the observations up to n;

$\mathbf{P}(n \mid n)$ is the error covariance matrix of the a posteriori estimate, $\hat{\mathbf{x}}(n \mid n)$.

During the operation of the Kalman filter, the corrupted speech $y(n)$ is windowed into non-overlapped short frames (e.g. 20 ms) then the LPCs and excitation variance σ_w^2 are estimated. These LPCs remain constant during the Kalman filtering of speech samples in the frame, while the Kalman gain $\mathbf{K}(n)$, error covariance $\mathbf{P}(n \mid n)$) and state vector estimate $\hat{\mathbf{x}}(n \mid n)$ are continually updated on a sample-by-sample basis (regardless of whichever frame we are in). When the LPC parameters are estimated from clean speech, the Kalman filter performs remarkably well [5]. However, when applied in practice, the clean speech is not available, the LPC parameters are estimated from noise-corrupted speech, which is the only observable signal in practice. In this case the performance of the Kalman filter degrades rapidly. In the field of the LPCs and excitation variance estimation from noise-corrupted speech, several methods are developed [11–14]. Therefore these LPCs estimation methods do not adequately address the problem of poor LPC estimates because the enhanced speech signal was found to suffer from annoying musical tones.

3. Proposed Kalman filter method (Kalman LPC-FEM) for speech enhancement

In this work, we present a re-evaluate LPCs with formant enhancement method (FEM). These new LPCs are used for enhancing the structure of noise-corrupted speech in the iterative Kalman filter. The advantages of this method over others are: (1) elimination of poor estimates LPCs due to additive noise by applying