Neural Networks 67 (2015) 131–139

Contents lists available at ScienceDirect

Neural Networks

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

Low-dimensional recurrent neural network-based Kalman filter for speech enhancement☆

Youshen Xia [a,*], Jun Wang [b]

[a] College of Mathematics and Computer Science, Fuzhou University, China

[b] Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 25 May 2014
Received in revised form 1 March 2015
Accepted 19 March 2015
Available online 7 April 2015

Keywords:
Recurrent neural network
Speech enhancement
Non-Gaussian noise
Noise-constrained estimation

ABSTRACT

This paper proposes a new recurrent neural network-based Kalman filter for speech enhancement, based on a noise-constrained least squares estimate. The parameters of speech signal modeled as autoregressive process are first estimated by using the proposed recurrent neural network and the speech signal is then recovered from Kalman filtering. The proposed recurrent neural network is globally asymptomatically stable to the noise-constrained estimate. Because the noise-constrained estimate has a robust performance against non-Gaussian noise, the proposed recurrent neural network-based speech enhancement algorithm can minimize the estimation error of Kalman filter parameters in non-Gaussian noise. Furthermore, having a low-dimensional model feature, the proposed neural network-based speech enhancement algorithm has a much faster speed than two existing recurrent neural networks-based speech enhancement algorithms. Simulation results show that the proposed recurrent neural network-based speech enhancement algorithm can produce a good performance with fast computation and noise reduction.

1. Introduction

Speech enhancement techniques have been successfully used in many areas such as mobile communication systems, speech recognition systems, and hearing aid devices, where received speech signals are corrupted by white or colored noise (Kay, 1993; Loizou, 2007). The main objective of speech enhancement is to improve the performance of speech communication in noise environments. Over the past decades, much research has focused on this area. Speech enhancement techniques may be divided into single-channel speech enhancement and multi-channel speech enhancement (Bobillet et al., 2007; Boll, 1979; Doclo & Moonen, 2002, 2005; Ephraim & Malah, 1984; Epharim & Van Trees, 1995a; Ephraim & Van Trees, 1995b; Gabrea, 2005; Gabrea, Grivel, & Najim, 1999; Gannot, Burshtein, & Weinstein, 1998; Gerkmann & Hendriks, 2012; Gibson, Koo, & Gray, 1991; Kay, 1993; Labarre, Grivel, Najim, & Todini, 2004; Lee & Jung, 2000; Ning, Bouchard, & Goubran, 2006; Roberto & Guidorzi, 2007; Wang, Li, & Dong, 2010; Xia & Yu, 2010; Xia, 2012; Xia & Wang, 2013). In this paper we focus on the single-channel speech enhancement.

There are mainly three types of single-channel speech enhancement algorithms. The first type is called the frequency domain method, including the Wiener filter algorithm and the MMSE amplitude spectrum estimation algorithm (Doclo & Moonen, 2005; Ephraim & Malah, 1984; Gerkmann & Hendriks, 2012; Wang et al., 2010). The Wiener algorithm requires estimating the power spectra of speech and noise and its performance depends on the estimation of the speech and noise spectra. The Wiener algorithm has a good noise reduction effect but could muffle speech. The MMSE amplitude spectrum estimation algorithm consists of two parts: a priori SNR estimate and an MMSE spectral amplitude estimate. This algorithm has a better performance than the conventional spectral subtraction algorithm (Boll, 1979), however, it needs an assumption that an estimate of the speech spectrum is available and white noise is Gaussian. The second type is the subspace method. Signal enhancement is to remove the noise subspace and to estimate the clean speech signal from the noisy speech subspace. Traditional subspace methods are suitable for white noise environments (Epharim & Van Trees, 1995a; Ephraim & Van Trees, 1995b). Several improved subspace methods were presented to deal with

☆ This work is supported by the National Natural Science Foundation of China under Grant No. 61179037 and 61473330, and in part from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project no. CUHK416811E).

* Corresponding author.
E-mail addresses: ysxia@fzu.edu.cn (Y. Xia), jwang@mae.cuhk.edu.hk (J. Wang).

http://dx.doi.org/10.1016/j.neunet.2015.03.008
0893-6080/

colored noise by adding the computational task for eigendecomposition of a non-symmetric matrix (Doclo & Moonen, 2002; Wei & Xia, 2013). The third type is called the parameter estimation-based Kalman filtering method in which the speech signal is modeled as autoregressive process and the speech signal is then recovered from Kalman filtering (Bobillet et al., 2007; Gabrea, 2005; Gabrea et al., 1999; Gannot et al., 1998; Gibson et al., 1991; Labarre et al., 2004; Lee & Jung, 2000; Ning et al., 2006; Park & Choi, 2008). Compared with other two type methods, the Kalman filtering method has no assumption of stationary speech signals.

Traditional parameter estimation-based Kalman filtering algorithms differ only by the choice of the algorithm used to estimate model parameters and the choice of the models adopted for the speech signal and additive noise. For example, Gibson et al. (1991) proposed a method that provides a sub-optimal solution, using the estimate-maximize algorithm based on the maximum likelihood argument. Gannot et al. (1998) proposed the use of the EM algorithm to iteratively estimate the spectral parameters of speech and noise parameters. Lee and Jung (2000) have developed a time-domain approach, without a priory information, to enhance speech signals. Gabrea presented (Gabrea, 2005) an adaptive parameter estimation method. Bobillet et al. presented (Labarre et al., 2004) an optimal smoothing and parameter identification algorithm. These parameter identification algorithms have a standard Gaussian noise assumption (Alimorad & Mahmood, 2011). To deal with the situation in non-Gaussian white noise environments, the Bayesian estimation-based methods were developed (Alliney & Ruzinsky, 1994; Christmas & Everson, 2011; Giannakis & Mendel, 1990; Smidl & Quinn, 2005). Park and Choi presented (Park & Choi, 2008) a neural network method for speech enhancement. Among these methods, the noise statistical distribution is required to be known and there is also a slow speed for parameter learning. Recently, to avoid the requirement of a priori statistical information, two noise constrained estimation-based methods for robust parameter identification were presented by minimizing a generalized least absolute deviation cost function and a quadratic cost function (Xia & Kamel, 2008; Xia, Kamel, & Henry, 2010), respectively. For their implementation, two recurrent neural networks were presented in Xia (2012) and Xia and Yu (2010), respectively. Because the two neural network methods have the total number of neurons which is larger than the sample length of the speech signal, their order of complexity is usually depends on the sample length of the speech signal. So, resulting neural network-based speech enhancement algorithms have a very slower speed.

To increase computational efficiency, we propose a low-dimensional recurrent neural network for fast speech enhancement by using a noise-constrained least squares estimate for Kalman filter parameters. It is shown that the proposed neural network is globally asymptomatically stable to the optimal solution of a noise constrained estimation problem. Because the noise-constrained estimate has a robust performance against non-Gaussian noise, the proposed recurrent neural network-based speech enhancement algorithm can minimize the estimation error of Kalman filtering parameters in non-Gaussian noise. Furthermore, having the low order of complexity, the proposed neural network-based speech enhancement algorithm has a much faster speed than two existing recurrent neural networks-based speech enhancement algorithms. Simulation results show that the proposed recurrent neural network-based speech enhancement algorithm produces a good performance in fast computation and noise reduction.

The paper is organized as follows. In Section 2, autoregressive (AR) model and its noise-constrained estimation are introduced. In Section 3, two existing recurrent neural networks for estimating the AR model parameter are discussed, and a new recurrent neural network with global convergence is proposed. In Section 4, the speech model and Kalman filter are described, and a new recurrent neural network-based speech enhancement algorithm is presented. In Section 5, computed examples are reported. Section 6 gives the concluding remarks of this paper.

2. AR model and estimation

Consider the following pth-order AR signal system:

$$x(t) = \sum_{i=1}^{p} a_i^* x(t-i) + v(t), \tag{1}$$

where p is the known order of the system, $\mathbf{a}^* = [a_1^*, \ldots, a_p^*]^T$ is the unknown AR parameter vector, $v(t)$ is the driving noise, $x(t)$ is an AR signal process with $x(t) = 0$ for $t \leq 0$, and $x(t)$ is observed in additive measurement noise $w(t)$:

$$y(t) = x(t) + w(t), \tag{2}$$

and $w(t)$ is assumed to be uncorrelated with $v(t)$. For simplicity, we denote the noisy signal vector by $\mathbf{y}_t = [y(t-1), \ldots, y(t-p)]^T$, and the measurement noise vector by $\mathbf{w}_t = [w(t-1), \ldots, w(t-p)]^T$. Then the AR signal observation model can be written as

$$y(t) = \mathbf{y}_t{}^T \mathbf{a}^* - n(t), \tag{3}$$

where $n(t) = \mathbf{w}_t{}^T \mathbf{a}^* - w(t) - v(t)$. The problem under study is to estimate AR parameter vector $\mathbf{a}^*$ from noisy observations $\{y(t)\}_1^N$ where N is the number of observations. The most basic approach to estimate the AR parameter vector is the least square (LS) method. The LS estimation minimizers

$$E(\mathbf{a}) = \frac{1}{N} \sum_{t=1}^{N} (y(t) - \mathbf{y}_t{}^T \mathbf{a})^2$$

and is given by

$$\mathbf{a}_{LS} = \left(\frac{1}{N} \sum_{t=1}^{N} \mathbf{y}_t \mathbf{y}_t{}^T \right)^{-1} \left(\frac{1}{N} \sum_{t=1}^{N} \mathbf{y}_t y(t) \right),$$

where $\mathbf{a} = [a_1, \ldots, a_p]^T$. In addition, there is an error between the LS estimate $\mathbf{a}_{LS}$ and the true AR parameter vector $\mathbf{a}^*$:

$$\mathbf{a}_{LS} \approx \mathbf{a}^* - \sigma^2 \hat{R}^{-1} \mathbf{a}^*,$$

where $\hat{R} = \frac{1}{N} \sum_{t=1}^{N} \mathbf{y}_t \mathbf{y}_t{}^T$ and σ^2 is the variance of the measurement noise.

Many AR parameter estimation methods have been developed to improve the LS estimation. For example, one is called the instrumental variable (IV) method (Bobillet et al., 2007; Labarre et al., 2004). Most of the IV algorithms is used for solving a set of high-order Yule–Walker equations. Another is called the bias correction method (Alimorad & Mahmood, 2011) where the AR model parameters, the observation noise variance, and the driving noise variance are estimated in an alternating iteration. The main feature among them is that the estimate of the AR model parameters is usually dependent on the estimate of observation noise variance with an assumption of Gaussian white noise. In practice, the noise corrupted in noisy speech is usually non-Gaussian and colored. Although the Bayesian estimation method developed can handle non-Gaussian noise cases (Christmas & Everson, 2011; Smidl & Quinn, 2005), it requires a priori statistical information.

Recently, to avoid a priori statistical information, two noise-constrained estimation methods were developed in Xia and