Applied Acoustics 96 (2015) 39–52

Contents lists available at ScienceDirect

Applied Acoustics

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

A two-sensor Gauss–Seidel fast affine projection algorithm for speech enhancement and acoustic noise reduction

Ilyes Darazirar, Mohamed Djendi *

University of Blida 1, Signal Processing and Image Laboratory (LATSI), Route de Soumaa, B.P. 270, Blida 09000, Algeria

ARTICLE INFO

Article history:
Received 24 June 2014
Received in revised form 13 February 2015
Accepted 16 February 2015
Available online 30 March 2015

Keywords:
Speech enhancement
Two-sensor
Forward structure
LMS
BSS
Gauss–Seidel

ABSTRACT

This paper deals with the problem of noise reduction and speech enhancement by adaptive filtering using two-sensor configuration. The well-known Gauss–Seidel fast affine projection (GSFAP) algorithm is used in various applications such as acoustic echo cancelation (AEC), and adaptive noise control. In this paper, we use the GSFAP algorithm in noise reduction and speech enhancement application. Moreover, we derive a new two-sensor GSFAP (TS-GSFAP) algorithm for speech enhancement and acoustic noise reduction application. The application of the proposed TS-GSFAP algorithm to the forward blind source separation (FBSS) structure leads to a significant improvement of the output speech signals quality even in noisy conditions. A fair comparison of the proposed TS-GSFAP algorithm with other standard two-sensor type algorithms is presented. This comparison is based on the evaluation of several objective and subjective criteria such as: the segmental signal-to-noise ratio (SegSNR), the Cepstral distance (CD) and the system mismatch (SM), the perceptual evaluation of speech quality (PESQ), and the mean opinion score (MOS) criteria. The obtained results show the good performances of the proposed TS-GSFAP algorithm.

1. Introduction

In telecommunication systems such as teleconferencing applications and hand free systems, the target signal can be highly distorted by the acoustic noise components that are generated by the surrounding environment [1]. Therefore, it is desirable to be able to eliminate or to reduce these undesirable noise components at reception. However, acoustic noise cancelation and speech enhancement require the knowledge of the channel model to be equalized [2,3]. Moreover, the impulse responses that model the transmission channels can be time varying. In these particular situation, one of the suitable solution to this problem is to use adaptive filtering approach for a good estimation of the time varying parameters of the channel [4,5].

In adaptive filtering techniques, several types of algorithms are used. One of the most popular adaptive algorithms is the normalized least mean squares (NLMS) [6,7]. This algorithm has been used in different applications such as acoustic echo and noise cancelation (AEC) systems, and adaptive equalization [8–10]. The NLMS algorithm is simple and easy to implement, but on the other hand, it has a bad asymptotic convergence speed with correlated signals as in AEC where the input signal is a speech [8,11,12].

To improve the behavior of the NLMS, the affine projection-type (AP) algorithm is proposed [13]. The AP algorithm family is the generalization of the NLMS algorithm and employs many projection dimensions to converge toward the optimal solution. However, when the projection dimension increase, the convergence speed and the computational complexity of the AP algorithm increase [14,15]. For this reason, many fast AP (FAP) algorithms have been proposed to overcome this problem of complexity [18,20]. The most important drawback of these FAP algorithms is their instability with correlated signals, which results from the propagation and accumulation of numerical errors in a matrix inversion process. Multiple methods were proposed to do the inversion matrix process while minimizing errors accumulation like the conjugate gradient (CG) FAP [16–18] and the Gauss–Seidel FAP (GSFAP) algorithms [19–22]. The Gauss–Seidel method is often used to compute the inversion process as explained in [20]. The GSFAP algorithm were used in AEC [17,18] and multichannel applications [18] and had shown good behavior and performances.

Moreover, a modified version of the original AP algorithm was proposed in [19], which is called the pseudo affine projection (PAP) algorithm. This PAP algorithm is based on the use of the same projection principle, as in the original AP algorithm of [19].

* Corresponding author.
E-mail addresses: ilyes.darazirar@gmail.com (I. Darazirar), m_djendi@yahoo.fr (M. Djendi).

http://dx.doi.org/10.1016/j.apacoust.2015.02.010

Further improvement of the PAP algorithm is allowed by using prediction techniques of the input signal or the filtering errors [25]. In [22], the authors proposed to simplify the original version of the PAP algorithm by employing Gauss–Seidel method to compute the matrix inversion process; this new algorithm is called the Gauss–Seidel pseudo affine projection (GSPAP) algorithm [20–23].

The speech enhancement and acoustic noise cancelation has been widely studied with different configurations [24,25]. The one sensor scheme is the most cited in literature [26,27]. Several multi-sensor configuration and combination techniques were proposed in [28,29]. Recently, a particular attention have been focused on the two-sensor configuration especially the two-sensor forward and backward Blind Source Separation structures (BSS) that are used to enhance corrupted speech signals and cancel the acoustic noise components. Several works have dealt with these two structures as in [30–36] by using subband approach and also variable step-size techniques. Many other advanced adaptive techniques are proposed to correct the distortions of the output speech signal [12,25,30–32].

Several two-sensor forward and backward BSS algorithms have been proposed recently to reduce the acoustic noise and to enhance the speech signal. In this paper, we propose a new two-sensor forward BSS Gauss–Seidel Fast Pseudo Affine Projection (TS-GSFAP) algorithm that allows a high enhancement of a corrupted speech signals.

The presentation of this paper is organized as follows: Section 2 describes the adopted mixing process with two source signals i.e. speech and noise signals. In Section 3, a full description of the forward BSS (FBSS) structure is presented. In Section 4, the description of the two-sensor algorithms is given. Section 5 is dedicated to the computational complexity study of the two-sensor algorithms. The simulations results and a comparative study between the two-sensor algorithms are presented in Section 6. Finally, in Section 7 we conclude our work.

2. The two-sensor mixing model

In this paper, we focus our work on two-sensor case in presence of two source signals. First, we introduce the general case of the two-sensor mixing process [1]. We suppose two-sensor that provide two noisy observations that are noted $p_1(n)$ and $p_2(n)$ as it is shown in Fig. 1. The mixing signals received by the two-sensor are given by the formulas:

$$p_m(n) = \sum_{i=1}^{2}\sum_{j=0}^{L-1} \boldsymbol{h}_{im}(j)s_i(n-j) \quad m = 1,2. \tag{1}$$

where $s_i(n)$ represents two source signals that are real and statistically independent. L represents the impulse responses length, and $h_{im}(n)$ represents four impulse response (IR) channels that model the paths between these signal sources and the two-sensor. IR of each path is considered stationary, but these IRs may also be time varying. The two sources are a speech signal $s(n)$ and

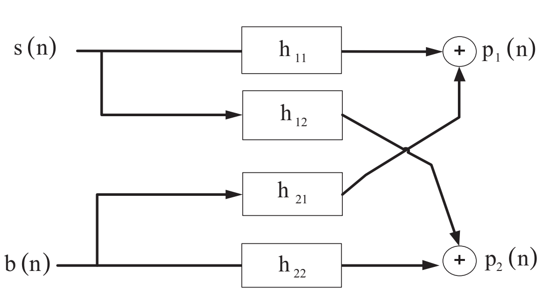

Fig. 1. General scheme of two-sensor mixing model.

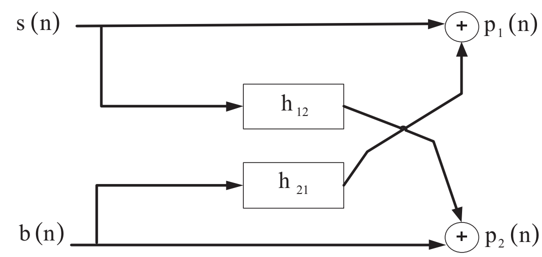

Fig. 2. A simplified scheme of two-sensor mixing model.

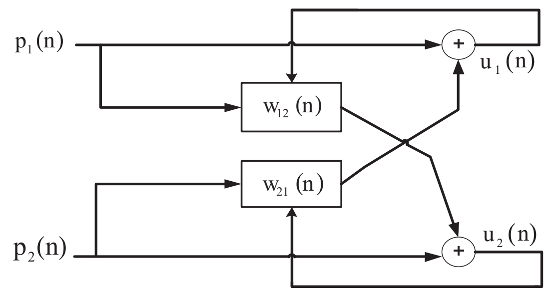

Fig. 3. The FBSS structure.

a noise $b(n)$. We assume that the two-sensor mixing model is stationary and the source signals $s(n)$ and $b(n)$ are mutually independent, i.e. $E[s(n)b(n-m)] = 0$, $\forall m$, where E represent the mathematical expectation, this implies that they are uncorrelated. We obtain the general scheme of the two-sensor mixing model, shown in Fig. 2.

$$p_1(n) = s(n) * \boldsymbol{h}_{11} + b(n) * \boldsymbol{h}_{21} \tag{2}$$

$$p_2(n) = s(n) * \boldsymbol{h}_{12} + b(n) * \boldsymbol{h}_{22} \tag{3}$$

where "$*$" represent the convolution operator. $\boldsymbol{h}_{11}, \boldsymbol{h}_{22}$ represent, separately, the direct acoustic path of each direct channel. $\boldsymbol{h}_{12}, \boldsymbol{h}_{21}$ represent the cross-coupling effects between the two channels. This mixing model can be simplified by taking $\boldsymbol{h}_{11} = \boldsymbol{h}_{22} = \delta(n)$ (the Kronecker impulse response) then the outputs (2) and (3) equation becomes as follow [1]:

$$p_1(n) = s(n) + b(n) * \boldsymbol{h}_{21} \tag{4}$$

$$p_2(n) = s(n) * \boldsymbol{h}_{12} + b(n) \tag{5}$$

We recall here that a full description of this simplified two-sensor mixing signals is given in our recently published work in [1].

3. Two-sensor FBSS structure

To recover the original components of the two noisy observations $p_1(n)$ and $p_2(n)$ we use a FBSS structure shown in Fig.3. The two-sensor FBSS (TS-FBSS) allows to retrieve the original components of two mixing signal by using two symmetric adaptive filters $\boldsymbol{w}_{12}(n)$ and $\boldsymbol{w}_{21}(n)$ that identify the cross-talk paths $\boldsymbol{h}_{12}$ and $\boldsymbol{h}_{21}$ [1].

The two output signals $u_1(n)$ and $u_2(n)$ of this TS-FBSS structure are obtained thorough the following relations:

$$u_1(n) = p_1(n) - p_2(n) * \boldsymbol{w}_{21}(n) \tag{6}$$

$$u_2(n) = p_2(n) - p_1(n) * \boldsymbol{w}_{12}(n) \tag{7}$$

where "$*$" represent the convolution operator and $\boldsymbol{w}_{21}(n), \boldsymbol{w}_{12}(n)$ are the adaptive filters that aims to suppress noise from the first output, and speech from the second output, respectively.

Inserting relations (4) and (5) in (6) and (7), respectively, we get the following relation of the two separated $u_1(n)$ and $u_2(n)$