Applied Acoustics 112 (2016) 192–200

Contents lists available at ScienceDirect

Applied Acoustics

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

A new regularized forward blind source separation algorithm for automatic speech quality enhancement

Meriem Zoulikha, Mohamed Djendi *

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

ARTICLE INFO

Article history:
Received 13 July 2015
Received in revised form 1 May 2016
Accepted 2 May 2016
Available online 3 June 2016

Keywords:
Speech enhancement
Blind source separation
Adaptive filtering
Segmental SNR
VAD

ABSTRACT

This paper addresses the problem of speech enhancement and acoustic noise reduction by adaptive filtering algorithms in a moving car through blind source separation (BSS) structures. In this paper we propose a new regularized forward blind source separation (RFBSS) algorithm that does not need voice activity detection (VAD) systems, and allows getting efficient speech enhancement performances with low complexity.

The proposed RFBSS algorithm is compared with recent and classical speech enhancement algorithms in different noisy conditions. This comparison is evaluated in terms of Cepstral distance (CD), the system mismatch (SM) and the Segmental signal-to-noise ratio (SegSNR) criteria. The obtained results show the efficiency of the proposed algorithm and its superiority in comparison with competitive algorithms in speech enhancement applications.

1. Introduction

The problem of speech enhancement from very noisy observations has became in last year's one of the recent challenges that a big part of scientists and researchers in the world are investigating [1]. This application of speech enhancement is still taking an important place in our daily life because it is used and present in every new teleconferencing systems, such as hands-free telephony, hearing aid, and teleconferencing. Several techniques have been proposed to deal with the problem of acoustic noise reduction and speech enhancement applications [2–5]. For instance, in [6–10] several single and dual microphones based on adaptive techniques are proposed to correct the distortions of the speech signal. Moreover, several adaptive techniques and algorithms combined with single, two- and multi-channel techniques for noise reduction and speech enhancement applications have been recently proposed as a new insight for this problem [11–14]. Other techniques that are mainly proposed to overcome the need of a VAD system, are substantially used to estimate the noise components power spectral density (PSD) in many speech enhancement techniques employing one, two or multi sensors systems. However, different techniques have been proposed to deal with this VAD system. In [15], a dual-microphone subband-based VAD using higher-order cumulant is used for detecting speech arriving from random directions. Another technique for VAD is proposed in [16] and based on a dual-microphone adaptive noise cancelling algorithm which uses the VAD method in the adaptive noise cancellation to avoid voice leakage. With such a method, the noise cancellers can well asses and cancel the environmental noises from the noisy signals and restore the desired voices. In [17], the authors provide a method based on left–right hidden Markov model (HMM) to identify the start and end of the speech. The method proposes two models of non-speech and speech instead of existed two states, formally, each model could include several states, also, they analyze other features, such as pitch index, pitch magnitude and fractal dimension of speech and non-speech. In a car environment, where the driver makes a hands-free communication with a far-end talker, this distant-talking involves a decrease of the desired speech signal quality [18–21]. This distortion is originally caused by the high number of quasi-infinite noise sources since vibrations, road, fan and wind noise from open windows generate a continuous and distributed noise background. The uttered signal by the driver will thus be recorded by the microphones sensor as a noisy signals. One of the main challenges for noise suppression algorithms methods is the detection or the estimation of the signal of interest. For example, the BSS techniques have shown a good performances that makes the hands-free system in a car communication highly robust toward noise components [22–32]. It is well known that the BSS is a powerful technique for acoustic noise reduction and speech enhancement in many situation such as in a car configuration involving loosely spaced microphones and short impulse responses. Most speech enhancement algorithms

* Corresponding author at: Blida University, LATSI Laboratory, Algeria.
E-mail addresses: m_zoulikha@hotmail.fr (M. Zoulikha), m_djendi@yahoo.fr (M. Djendi).

http://dx.doi.org/10.1016/j.apacoust.2016.05.012
0003-682X/

which are based on the BSS structure use either a manual VAD system (MVAD) system, or an integrated bloc which realizes automatic VAD (AVAD) to control the adaptation of the cross-filters. Recently, a particular attention has been made to both forward and backward BSS (i.e. FBSS and BBSS) structures applied to enhance corrupted speech signals and to cancel the acoustic noise components. Several works have dealt with these two structures [18–23]. However, all of these techniques used a MVAD system that makes them inoperative in practice. The only work that proposed a VAD system with this FBSS and BBSS structure is given in [1]. In this paper, we focus our interest on the BBSS structure and propose a new automatic adaptive algorithm that enhances the speech signal and reduces the acoustic noise without need of any VAD system. The solution given in this paper is algorithmic and not structural as it is given in [1]. The organization of this paper is as follows: in Section 2, we present the principle of BSS. The Model description of FBSS is detailed in Section 3, and the forward symmetric adaptive decorrelating (FSAD) algorithm is presented in Sections 4. The new RFBSS is described in Sections 5. In Section 6, we show the simulation results of the proposed RFBSS algorithm and its performances in comparison with competitive and recent techniques. Finally, we conclude our work in Section 7.

2. Blind source separation (BSS)

In real environment such as in cars, the recorded signals by two microphones are a linear combination between the speech $s(n)$ and the noise $b(n)$ components. The latter depends on the microphone's positions, acoustic characteristics of the car's interior, the sources themselves, etc. Therefore, the main problem is to find, with the least *a priori* knowledge, useful signals which have been mixed. In order to overcome this problem, the BSS structure of [8] is used frequently to extract the sources signal from the only knowledge of noisy signals. In this paper, we focus on the convolutive noisy signals and the FBSS structure. The principle of the FBSS is shown in Fig. 1.

3. Model description of FBSS structure

In order to separate the noisy observation components from the given structure (Fig. 2a), we used the FBSS which is shown in Fig. 2b [20].

The noisy signals $p_1(n)$ and $p_2(n)$, which will be used as the FBSS inputs, are given by:

$$p_1(n) = s(n) + h_{21}(n) * b(n), \tag{1}$$

$$p_2(n) = b(n) + h_{12}(n) * s(n), \tag{2}$$

where $s(n)$ and $b(n)$ are two sources of speech and noise, respectively; $h_{12}(n)$ and $h_{21}(n)$ represent the cross-coupling effects between the two-channels. The symbol $(*)$ represents the linear convolution operator. For this model, both adaptive filters $w_{12}(n)$ and $w_{21}(n)$ are used to identify the cross-talk path $h_{21}(n)$ and $h_{21}(n)$, respectively. The output signals, $u_1(n)$ and $u_2(n)$, of the FBSS structure are given by the following formulas:

$$u_1(n) = p_1(n) - p_2(n) * w_{21}(n), \tag{3}$$

$$u_2(n) = p_2(n) - p_1(n) * w_{12}(n), \tag{4}$$

where $w_{12}(n)$ and $w_{21}(n)$ are the cross-adaptive filters. If we do further development of the above relations, we obtain:

$$u_1(n) = b(n) * [h_{21}(n) - w_{21}(n)] + s(n) * [\delta(n) - h_{12}(n) * w_{21}(n)], \tag{5}$$

$$u_2(n) = s(n) * [h_{12}(n) - w_{12}(n)] + b(n) * [\delta(n) - h_{21}(n) * w_{12}(n)], \tag{6}$$

if we use the optimal assumption for both adaptive filters $w_{12}(n)$ and $w_{21}(n)$,this means that the following solutions hold:

$$\begin{cases} w_{12}^{opt} = h_{12}, \\ w_{21}^{opt} = h_{21}. \end{cases} \tag{7}$$

Then the two outputs of the FBSS structure are given by [20]:

$$u_1(n) = s(n) * [\delta(n) - h_{12}(n) * h_{21}(n)], \tag{8}$$

$$u_2(n) = b(n) * [\delta(n) - h_{21}(n) * h_{12}(n)]. \tag{9}$$

4. FSAD algorithm

The symmetric adaptive decorrelating (SAD) algorithm combined with the FBSS and BBSS structures, were first proposed in [10,11]. In this study, we focus on SAD algorithm applied to the FBSS

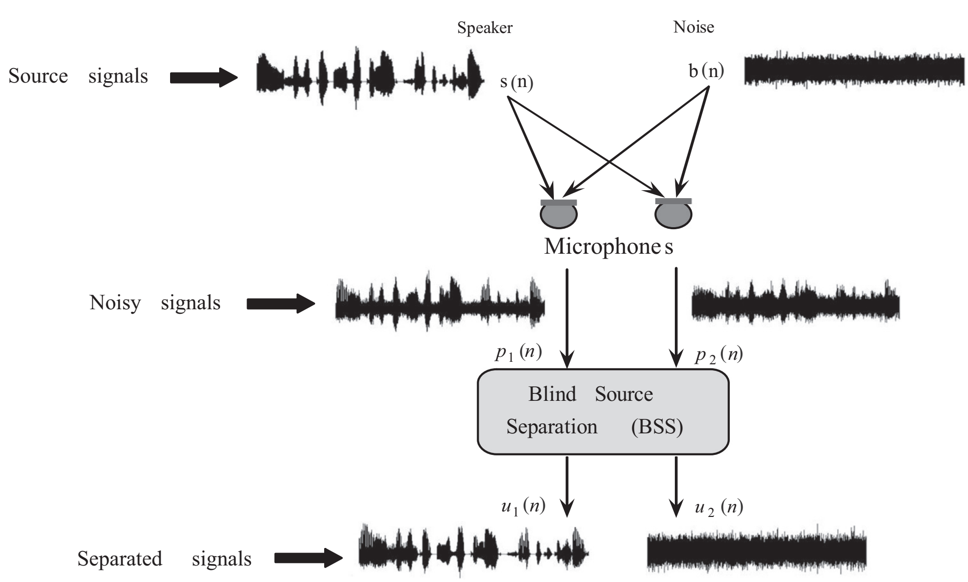

Fig. 1. Principal scheme of blind source separation (BSS).