Applied Acoustics 76 (2014) 209–222

Contents lists available at ScienceDirect

Applied Acoustics

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

A new efficient two-channel backward algorithm for speech intelligibility enhancement: A subband approach

Mohamed Djendi *, Rédha Bendoumia

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

ARTICLE INFO

Article history:
Received 12 May 2013
Received in revised form 13 August 2013
Accepted 14 August 2013
Available online 8 September 2013

Keywords:
Backward structures
LMS algorithm
Speech intelligibility
Noise reduction
Speech distortion
Subband technique

ABSTRACT

This paper addresses the problem of speech intelligibility enhancement by adaptive filtering algorithms employed with subband techniques. The two structures named the forward and backward blind source separation structures are extensively used in the speech enhancement and source separation areas, and largely studied in the literature with convolutive and non-convolutive mixtures. These two structures use two-microphones to generate the convolutive/non-convolutive mixing signal, and provide at the outputs the target and the jammer signal components. In this paper, we focus our interest on the backward structure employed to enhance the speech signal from a convolutive mixture. Furthermore, we propose a subband implementation of this structure to improve its behavior with speech signal. The new proposed subband-Backward BSS (SBBSS) structure allows a very important improvement of the convergence speed of the adaptive filtering algorithms when the subband-number is selected high. In order to improve the robustness of the proposed subband structure, we have adapted then applied a new criterion that combines the System Mismatch and the Mean-Errors criterion minimization. The proposed subband backward structure, when it is combined with this new criterion minimization, allows to enhance the output speech signal by reducing the distortion and the noise components. The performance of the proposed subband backward structure is validated through several objective criteria which are given and described in this paper.

1. Introduction

Significant progress has been made in both research and product-related evolution of hands-free communication devices. Promising applications of human–machine interfaces easing our daily life and calling for natural voice interaction include, for example, game stations, interactive television sets, humanoid robots, smart homes, or smart phones. Consequently, the user habitually desires to be free, desires to move freely and close-talking sensors become intolerable. However, the use of distant-talking and cross-talking sensors involves a decrease of the desired speech signal quality. This distortion is originally caused by additive undesired noise components. These noise components are resulting from interfering point sources, such as competing speakers, and from coherent and incoherent noise components which are emitted, for example, by household appliances and environments. [1–5].

In noisy environments, acoustic reflections of both the desired signal and other jammers extra-degrade the quality of the sensor signals [6–8]. For noise reduction (NR) and speech enhancement (SE) applications these impairments turn out to be a main challenge if they are non-stationary and irregular. However, and in order to improve the robustness of SE and NR systems in such noisy environments, we can procedure in two different approaches to overcome this constrained problem: the first approach is based on the identification then the adjustment of the acoustic model of the NR and SE systems to allows to quantify then correct the distortion caused by the environments (cross-talk, competing speakers, reverberation and coherent and incoherent noise components). In this way, we usually use adaptive techniques and algorithms for the identification of the acoustic model of NR and SE systems in single and multi-sensors approaches [9–13]. In the second approach, we use techniques and methods to enhance the desired signal and cancel the acoustic noise and unusual signal components. A several one-, two-and multichannel sensors techniques are proposed to deal with this problem. For example in [14–17], several single and two-sensor techniques are proposed to correct these distortions [18–20]. More advanced techniques are then proposed recently in [21,22]. Furthermore, several algorithms have been proposed, in combination with single, two-and multichannel techniques for NR and SE applications, are recently proposed as a new countermeasure for the presented problem [23–26]. Recently, a particular consideration has been made for

* Corresponding author.
E-mail addresses: m_djendi@yahoo.fr (M. Djendi), r.bendoumia@yahoo.fr (R. Bendoumia).

0003-682X/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.apacoust.2013.08.013

the two forward and backward blind source separation structures to be applied to enhance corrupted speech signals and cancel the acoustic noise components. Several works have dealt with these two structures as in [27–30]. In order to improve the NR and SE technique behavior with stationary and non-stationary noise components, the subband approach has been largely used for single, two-and multichannel implementation [31–34]. The subband adaptive filtering is an important approach that is usually used to overcome the problem of NR and SE techniques. The subband filter banks have been introduced in the area of NR and SE adaptive filtering methods in order to improve the performance of time domain adaptive filters. The main improvements brought by the subband approach are the faster convergence speed and the reduction of computational complexity of the used techniques due to the shorter adaptive filters in the filter bank subbands [35–37]. These good subband approach features are due to the fact that the fullband input signal is divided into a number of subband signals [38,39]. The subband adaptive filtering technique leads to the manipulation of each subband signal, and allows for each subband to converge almost separately for various modes, and thus improving the overall convergence behavior [40,41].

In this paper, we propose a new efficient two-channel backward blind source separation (SBBSS) algorithm which is implemented in subbands. This new algorithm is proposed to reduce the noise components and to enhance the speech signal when the observations are highly corrupted by the noise components. In the proposed algorithm, the output signals are estimated in subbands, whereas the coefficients of the two adaptive filters are explicitly adapted in the fullband form. This adaptive control mechanism is different compared with other classical subband BSS structures (where each sub-filter is adapted separately). The experimental results show the superiority of the proposed SBBSS algorithm in comparison with the fullband backward and the backward symmetric adaptive decorrelating (BSAD) ones.

This paper is organized and presented as follows: Section 2 describes the mixing process with two sources of speech and noise signals. In Section 3, a full description of the backward structure is presented. In Section 4, the BSAD algorithm is presented. The proposed subband backward blind source separation (SBBSS) algorithm and its full theoretical analysis are presented in Section 5 and its subsections. The simulations results of the proposed SBBSS algorithm in comparison with its fullband versions and the BSAD algorithm are presented in Section 6. Finally, we conclude our work in Section 7.

2. Description of the mixing process

In the general case of a convolutive mixture model, we assume that we have q source signals $s(n) = \lfloor s_1(n), s_2(n), \ldots, s_q(n) \rfloor$ which are real and statistically independent. These sources are convoluted with q impulse responses channels $h_{im}(k)$ and generate Ω sensors signals $p(n) = [p_1(n), p_2(n), \ldots, p_\Omega(n)]$. This general mixing model is given by the following Fig. 1. The relations between the source signals and the observations are given by the following relation [13]:

$$p_m(n) = \sum_{i=1}^{q}\sum_{j=0}^{M-1} h_{im}(j) s_i(n-j) \quad m = 1, 2, \ldots \Omega. \tag{1}$$

where Ω, q and M represent respectively, the number of sensors, the number of sources and the impulse responses length.

In this paper we consider the particular case of two sources and two sensors ($\Omega = q = 2$). The convolutive mixture model of two uncorrelated sources is shown in Fig. 2, where $s(n)$ and $b(n)$ are the speech signal and the noise respectively.

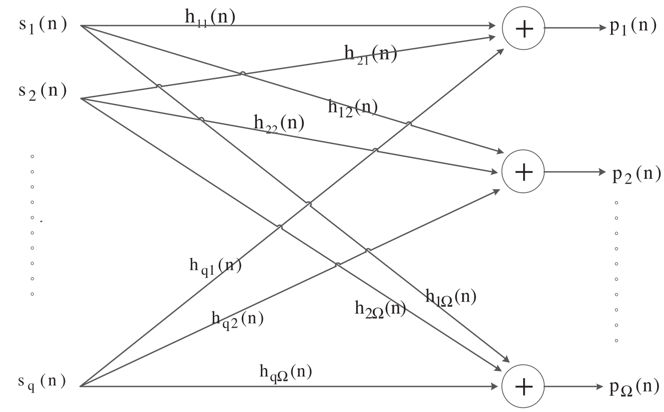

Fig. 1. The convolutive mixture model, $s_i(n)$ are the source signals, h_{im} are the impulse response of the channels throughout the sources are transformed and mixed, and $p_m(n)$ are the observed signals. (The index i and m are defined as follows $i = 1, 2, \ldots, q$ and $m = 1, 2, \ldots, \Omega$).

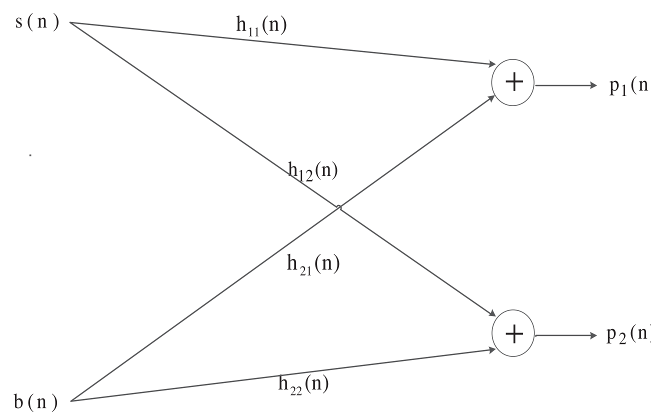

Fig. 2. The simplified mixture model, $s(n)$ and $b(n)$ are the speech signal and the noise respectively. $h_{11}(n)$, $h_{22}(n)$, $h_{12}(n)$ and $h_{21}(n)$ represent the impulse responses between the channels.

From this Fig. 2, the two observed signals at the sensor outputs of this model can be written as follows:

$$p_1(n) = \sum_{k=0}^{M-1} h_{11}(k) s(n-k) + \sum_{k=0}^{M-1} h_{21}(k) b(n-k) \tag{2}$$

$$p_2(n) = \sum_{k=0}^{M-1} h_{22}(k) b(n-k) + \sum_{k=0}^{M-1} h_{12}(n) s(n-k) \tag{3}$$

where $h_{11}(n)$ and $h_{22}(n)$ represent the direct acoustic path of each direct channel separately, $h_{12}(n)$ and $h_{21}(n)$ represent the cross-coupling effects between the channels. To simplify the problem of the mixing signals, we consider $h_{11}(n)$ and $h_{22}(n)$ the Kronecker impulse response, i.e. $h_{11}(n) = h_{22}(n) = \delta(n)$ [5,13]. According to this assumption and if the input signals are real, the two relations (2) and (3) can be rewritten as follows:

$$p_1(n) = s(n) + \sum_{k=0}^{M-1} h_{21}(k) b(n-k) \tag{4}$$

$$p_2(n) = b(n) + \sum_{k=0}^{M-1} h_{12}(k) s(n-k) \tag{5}$$

3. Description of the Backward BSS structure

In this section, we present the backward blind source separation (BBSS) structure and we give its full formulation and optimal