Applied Acoustics 73 (2012) 12–20

Contents lists available at ScienceDirect

Applied Acoustics

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

Dual-channel spectral subtraction algorithms based speech enhancement dedicated to a bilateral cochlear implant

Fathi Kallel [a,b,*], Mondher Frikha [a], Mohamed Ghorbel [a], Ahmed Ben Hamida [a], Christian Berger-Vachon [b]

[a] Research Unit in Advances Technologies for Medical and Signals (ATMS), Laboratory of Electronics and Information Technologies (LETI), National Engineering School of Sfax, University of Sfax, Route Soukra km 3, Sfax, B.P.W, 3038, Tunisia

[b] PACS Team, INSERM Unit 1028: "Cognition and Brain Dynamics", Lyon Neurosciences Centre, EPU-ISTIL, Claude Bernard University, Boulevard du 11 Novembre 1918, 69622 Villeurbanne, France

ARTICLE INFO

Article history:
Received 11 October 2010
Received in revised form 19 May 2011
Accepted 23 June 2011
Available online 22 July 2011

Keywords:
Dual-channel noise PSD estimation
DC-NLSS
DC-MBSS
BCI

ABSTRACT

In this paper, two speech enhancement algorithms (SEAs) based on spectral subtraction (SS) principle have been evaluated for bilateral cochlear implant (BCI) users. Specifically, dual-channel noise power spectral estimation algorithm using power spectral densities (PSD) and cross power spectral density (CPSD) of the observed signals was studied. The enhanced speech signals were obtained using either Dual Channel Non Linear Spectral Subtraction 'DC-NLSS' or Dual-Channel Multi-Band Spectral Subtraction 'DC-MBSS' algorithms. For performance evaluation, some objective speech assessment tests relying on Perceptual Evaluation of Speech Quality (PESQ) score and speech Itakura-Saito (IS) distortion measurement were performed to fix the optimal number of frequency band needed in DC-MBSS algorithm. In order to evaluate the speech intelligibility, subjective listening tests were assessed with 50 normal hearing listeners using a specific BCI simulator and with three deafened BCI patients. Experimental results, obtained using French Lafon database corrupted by an additive babble noise at different Signal-to-Noise Ratios (SNR), showed that DC-MBSS algorithm improves speech understanding better than DC-NLSS algorithm for single and multiple interfering noise sources.

1. Introduction

Most cochlear implant (CI) users perform well in quiet listening conditions and many users can now achieve even more than 80% word recognition scores regardless the used device [28]. However, speech recognition scores are enormously degraded in noisy environments [25]. Furthermore, as mentioned by CI users, better and comfortable speech recognition in noisy environments would be considered as one of their most significant challenges [23]. As a first trial to improve speech intelligibility in noisy environments, individuals with severe to profound hearing loss can now be implanted with two cochlear implants, one in each ear. In fact, a bilateral cochlear implantation provides patients the advantages of bilateral information. Bilateral hearing permits optimal performance of the auditory system, with a better understanding of speech in quiet and even better understanding in noisy environments [22]. Recent works compared also speech performance in noisy environment with matched bilateral CI with respect to unilateral CI users. BCI group showed significantly better performance on speech perception in noisy environments compared to the unilateral CI subjects [26,7,8]. Different other clinical studies have demonstrated a substantial increase in speech intelligibility with bilateral cochlear implants compared to monaural listening configurations in noise [30,33,18].

To reduce background effects of noise, some speech enhancement algorithms originally developed for normal hearing listeners have been applied to CI speech processing [17,36,14,11]. These algorithms were able to somewhat improve CI users' performance in noisy listening conditions. Considerably, larger benefits in speech intelligibility could be obtained when resorting to multi-microphone adaptive signal processing strategies, instead. Such strategies make use of spatial information due to the relative position of the emanating sounds, and could therefore better exploit situations in which the target and masker are spatially separated [13,34,5]. Several noise-reduction algorithms using two or more microphones were also available, and most of these proposed algorithms were based on beamformer techniques, especially, the adaptive beamformer algorithms. The performance of adaptive beamforming with two microphones with bilateral cochlear implant was assessed by different studies [29]. In the study of Chung et al. [5], authors conducted experiments to investigate whether directional microphones and adaptive multi-channel noise reduction algorithms could enhance overall CI performance. Results indicated that directional microphones could provide an average improvement of around 3.5 dB. Spriet et al. [29] investigated the

* Corresponding author.
E-mail address: fathikallel@yahoo.fr (F. Kallel).

0003-682X/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apacoust.2011.06.010

performance of the beam pre-processing strategy in the Nucleus Freedom speech processor with five CI users. The performance with the beam strategy was evaluated at two noise levels and with two types of noise, speech-weighted noise and multi-talker babble. The tested algorithm improved the speech reception threshold by approximately 5 to 8 dB. Kokkinakis and Loizou [16] proposed a multi-microphone based adaptive noise reduction strategy exploiting information simultaneously collected by two behind-the-ear processors (BTE) microphones. Four microphones were employed (two omni-directional and two directional) in each of the two BTE processors (one per ear). Results indicated that the proposed multi-microphone strategies improved speech understanding in single and multi-noise source scenarios.

We note that mush of the focus of the previously published studies has been to investigate, in the mono-channel case, the pre-processing noisy speech signal by noise reduction algorithms in order to feed enhanced signals to CI listeners. Only a small number of SEAs have been evaluated to the bilateral case. As one contribution of this work is the application of SEAs for the BCI users in an uncorrelated additive babble noise environments. We propose then new SEAs built upon series of previously published works on spectral subtraction algorithms. The first is based on a non linear spectral subtraction approach according to the work of Berouti et al. [1]. Whereas the second is built on a multi-band spectral subtraction algorithm proposed by Kamath and Loizou [15] and Udrea et al. [32].

Our proposed SEAs namely DC-NLSS and DC-MBSS are an extension of the previously mentioned mono-channel spectral subtraction algorithms to the dual-channel conception case. However, spectral subtraction principles are combined together with a noise PSD estimation technique based on the use of two omni-directional microphones. This noise PSD estimator takes into account the coherence between both received noisy speech signals.

The paper is outlined as follows. Section 2 provides theoretical overview of SEAs. Section 3 derives bilateral cochlear implant simulator principle. Section 4 evaluates the experimental results. Section 5 gives an overall discussion of all obtained results. Finally, Section 6 devotes to the conclusion.

2. Speech enhancement algorithms

The bilateral cochlear implant configuration is as illustrated in Fig. 1. Both left and right cochlear implants are fitted with one microphone. The two received noisy speech signals ($y_1(n)$ and $y_2(n)$) are processed together with the proposed SEAs to generate enhanced speech signals ($\hat{s}_1(n)$ and $\hat{s}_2(n)$) which are then used for stimulation.

We suppose that the noise received by the microphones can be represented by two additive uncorrelated babble noise signals so that the picked up noisy speech signals can be expressed in temporal domain as follows:

$$y_i(k) = s_i(k) + d_i(k) \quad i = \{1,2\} \tag{1}$$

where 'i' is the microphone index, $y_i(k)$, $s_i(k)$ and $d_i(k)$ represent respectively noisy speech, clean speech and noise signals.

Note that $i = 1$ corresponds to the signal picked up by the microphone placed in the right ear and $i = 2$ corresponds to the signal picked up by the microphone placed in the left ear. The short-time Fast Fourier Transforms (FFT) of the received noisy signals is formulated as follows:

$$Y_{i,N}(f,n) = S_{i,N}(f,n) + D_{i,N}(f,n) \quad i = \{1,2\} \tag{2}$$

where $Y_{i,N}(f,n)$, $S_{i,N}(f,n)$ and $D_{i,N}(f,n)$ denote respectively the N-point FFTs of the $y_i(k)$, $s_i(k)$, and $d_i(k)$ for the frame n and the fth frequency bin. The parameter 'N' is left out for simplicity.

The proposed dual-channel speech enhancement algorithm contains then two major parts:

– Noise PSD estimation based on PSD and CPSD computation of received noisy speech signals.
– Enhanced speech signal estimation using spectral subtraction approach.

2.1. PSD and CPSD estimation

The noise PSD estimation needs a PSD and a CPSD estimation of the noisy received speech signals. The PSD of the noisy signal at the first channel '$P_{y1y1}(f,n)$', and at the second channel '$P_{y2y2}(f,n)$', and the CPSD '$P_{y1y2}(f,n)$' can be estimated as follows [19]:

$$\begin{aligned} P_{YiYi}(f,n) &= \lambda \cdot P_{YiYi}(f,n-1) + (1-\lambda)\cdot Y_i(f,n)\cdot Y_i^*(f,n) \quad i = \{1,2\} \\ P_{Y1Y2}(f,n) &= \lambda \cdot P_{Y1Y2}(f,n-1) + (1-\lambda)\cdot Y_1(f,n)\cdot Y_2^*(f,n) \end{aligned} \tag{3}$$

where $*$ is the complex conjugate operator, λ is a smoothing factor usually close to 1. This factor should satisfy the two following constraints:

– For lower values, the estimation takes into account the speech short term stationarity.
– For higher values, this factor serves to minimize the estimator variance.

A previous study [9] showed that for 16 kHz sampling frequency with 256 samples per frame and a 50% overlap, the upper limit values of λ were around 0.6–0.8. In this study, we choose $\lambda = 0.8$.

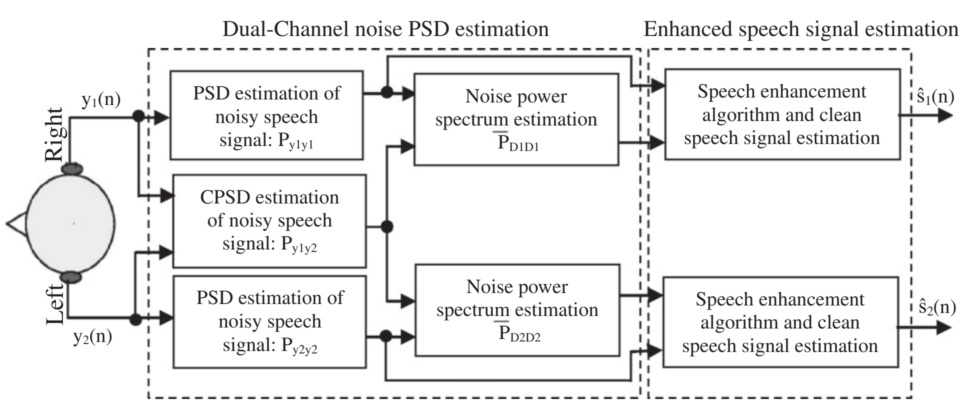

Fig. 1. Block diagram of the considered dual-channel speech enhancement algorithm.