Biomedical Signal Processing and Control 39 (2018) 263–270

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

EEG-assisted modulation of sound sources in the auditory scene

Marzieh Haghighi [a,*], Mohammad Moghadamfalahi [a], Murat Akcakaya [b], Deniz Erdogmus [a]

[a] Northeastern University, 360 Huntington Avenue, Boston, MA 02115, United States
[b] University of Pittsburgh, 4200 Fifth Avenue, Pittsburgh, PA 15260, United States

ARTICLE INFO

Article history:
Received 2 March 2017
Received in revised form 19 July 2017
Accepted 1 August 2017
Available online 16 August 2017

Keywords:
Auditory BCI
Cocktail party problem
Auditory attention classification

ABSTRACT

Noninvasive EEG (electroencephalography) based auditory attention detection could be useful for improved hearing aids in the future. This work is a novel attempt to investigate the feasibility of online modulation of sound sources by probabilistic detection of auditory attention, using a noninvasive EEG-based brain computer interface. Proposed online system modulates the upcoming sound sources through gain adaptation which employs probabilistic decisions (soft decisions) from a classifier trained on offline calibration data. In this work, calibration EEG data were collected in sessions where the participants listened to two sound sources (one attended and one unattended). Cross-correlation coefficients between the EEG measurements and the attended and unattended sound source envelope (estimates) are used to show differences in sharpness and delays of neural responses for attended versus unattended sound source. Salient features to distinguish attended sources from the unattended ones in the correlation patterns have been identified, and later they have been used to train an auditory attention classifier. Compared to the existing results in the literature, in this paper we have two main contributions. First, using the auditory attention classifier, we have shown high offline detection performance with single channel EEG measurements of shorter duration compared to the existing approaches in the literature which employ large number of channels with longer EEG measurements. Second, using the classifier trained offline in the calibration session, we have shown the performance of the online sound source modulation system. We observe that online sound source modulation system is able to keep the level of attended sound source higher than the unattended source.

1. Introduction

Approximately 35 million Americans (11.3% of the population) suffer from hearing loss; this number is increasing and is projected to reach 40 million by 2025 [1]. Within this population only 30% prefer using current generations of hearing aids that are available on the market. One of the most common complaints associated with hearing-aid use is understating speech in the presence of noise and interferences. Effects of interfering sounds on masking the speech intelligibility and audibility have been widely studied [2,3]. Specifically, it has been shown that increase in SNR needed for the same level of speech understanding given a background noise for people with hearing loss can be as high as 30 dB more compared to people with normal hearing [3]. Therefore, amplifying the target signal versus unwanted noises and interferences to facilitate hearing and increase speech intelligibility and listening comfort is one of the basic concepts exploited by hearing aids [3]. Identifying the signal versus noise is a main step required for the design of a hearing aid. This identification step can be a difficult task in complicated auditory scenes like a cocktail party scenario in which both signal and interferences have acoustic features of speech and can instantly switch their roles based on the attention of the listener and can not be detected based on the predefined assumptions on signal and noise features. Our brain distinguishes the audio sources based on their spectral profile, harmonicity, spectral or spatial separation, temporal onsets and offsets, temporal modulation, and temporal separation [4,5] and focus on one sound to analyse the auditory scene [6] in the so called cocktail party effect [7]. Existence of each cue can reduce informational and energetic masking of competing sources and help focusing our attention on the target source. A general overview of the computational efforts in bottom-up or top-down modelling of auditory attention in a cocktail party setting is provided in [8].

* Corresponding author.
E-mail address: haghighi@ece.neu.edu (M. Haghighi).

http://dx.doi.org/10.1016/j.bspc.2017.08.008
1746-8094/

Brain/Body Computer Interface (BBCI) systems can be used to augment the current generations of hearing aids by discriminating among attended and unattended sound sources. They can be incorporated to provide external evidence based on top-down selective attention of listeners [9]. Attempts have been made to incorporate bottom-up attention evidences in the design of the hearing aids. Direction based hearing aids that detect attention direction from eye gaze and amplify sounds coming from that direction can be examples of bottom-up attention evidence incorporation [10]. Moreover, there are attempts to use electroencephalography (EEG)-based brain computer interfaces (BCIs) for the identification of attended sound sources. EEG has been extensively used in BCI designs due to its high temporal resolution, non-invasiveness, and portability. These characteristics, in addition to EEG devices being inexpensive and accessible, make EEG a practical choice for the design of a BCI that can be integrated into hearing aids to identify auditory attention. A crucial step in such an integration is to build an EEG-based BCI that employs auditory attention.

EEG-based auditory BCIs that rely on external auditory stimulation have recently attracted attention from the research community. For example, auditory-evoked P300 BCI spelling system for locked-in patients is widely studied [11–16]. It was shown that fundamental frequency, amplitude, pitch and direction of audio stimuli are distinctive features that can be processed and distinguished by the brain. Also, more recent studies using EEG measurements have shown that there is a cortical entrainment to the temporal envelope of the attended speech [17–19]. A study on the quality of cortical entrainment to auditory stimulus envelope by top-down cognitive attention has shown enhancement of obligatory auditory processing activity in top-down attention responses when competing auditory stimuli differ in spatial direction [20] and frequency [21].

Recently, EEG-based BCI has also been used in cocktail party problems for the classification of attended versus unattended sound sources [22–24]. In the identification of the attended sound source in a cocktail party problem, stimulus reconstruction to estimate the envelope of the input speech stream from high density EEG measurements is the state-of-the-art practice [25,22]. In the aforementioned model, envelope of the attended stimulus is reconstructed using spatio-temporal linear decoder applied on neural recordings. In one study that considered the identification of the attended sound source in a dichotic (different sounds playing in each ear) two speaker scenario, 60 s of high density EEG data recorded through 128 electrodes were used in the stimulus reconstruction. Two decoders using the attended and unattended speech were trained and it was shown that estimated sound source using the attended decoder has higher correlation with the attended speech envelope compared to the estimated stimuli using unattended decoder with unattended speech envelope [22]. For practical purposes, further studies have attempted to examine the stimulus reconstruction approach using smaller number of EEG electrodes [23] or even two bilaterally placed around the ear electrode arrays (cEEGrids) [26]. Furthermore, the robustness of the attended speech envelope reconstruction in noisy real world acoustic scenes has been demonstrated [27]. In contrast to the stimulus reconstruction methods, studies with system identification approaches to solve this problem, have tried to reconstruct the neural measurements using the linear forward map of sound sources [28–31]. In a recent related study, a single in-Ear-EEG electrode and an adjacent scalp-EEG electrode were used for auditory attention detection in a diotic two speaker scenario [30]. On the other hand, in a different class of target speaker detection approaches, studies have tried to extract informative and distinguishable features of EEG measurements with respect to the attended and unattended sound sources to train a classifier [24,32]. In a related study, authors have compared three types of features extracted from speech signal and EEG measurements to learn a linear classifier for the identification of the attended speaker using 20 s of data from high density 128 channels EEG recordings [24]. In our previous related work, we have investigated the role of spectral and spatial features of competing sound sources in an auditory BCI system with the purpose of detecting the attended auditory source in a cocktail party setting. We reported high single channel classification performance for attended sound source versus unattended one based on their spectral and spatial separation of the sources (diotic and dichotic paradigms) using 60 s of EEG and stimuli data [32].

In continuation of the above described literature and our previous work, this paper presents two contributions to the literature of EEG-based auditory attention detection in a cocktail party setting:

- First, we show successful identification of attended speaker source in a diotic (both sounds playing in both ears) two speaker scenario using 20 s of EEG data recorded from 16 channels. The presented classifier outperforms EEG-based auditory attention detectors previously presented in the literature in terms of accuracy, with smaller number of EEG channels (sparse 16 versus typically dense 96 or more), and using time-series of shorter durations (20 s versus typically 60 s). In fact, using 20 s of EEG data from only one of the 16 EEG channels, we demonstrate high classification performance for the auditory attention detection. This extends our results from [32], which showed high single-channel classification accuracy when the EEG duration was 60 s.
- Second, we introduce a novel online system that gives feedback on attention of the user in the form of attended to unattended source energy ratio amplification. The level of amplification of attended versus suppression of unattended source is assigned based on a probabilistic model defined over the classifier trained on the offline data including temporal dependency of the user's attention. The goal of the online system is using the probabilistic information of the user's attention to enhance the concentration of the user on the target source in multi-speaker scenarios. The introduced framework for online system is a proof of concept for design perspective of an EEG-augmented hearing aid system. Finally, we show the introduced online system in average is able to keep the level of attended source higher despite statistical changes happening in online data compared to the offline data used for training the classifier.

2. System overview

The diagram represented in Fig. 1 summarizes the steps of the proposed BCI system. The proposed system gets the mixture of sounds from the environment as the input and modifies the gain of each specific sound. The output of this system is the input to the ear channel.

The decision on gain modification of each sound is made by the BCI module which consists of three submodules of gain controller, auditory attention inference system and hearing aid DSP system. Hearing aid DSP system estimates independent sound sources from the mixture of sounds in the environment and outputs the information to the gain controller and attention inference module. In this work, we assume that we have the estimated sources which are the outputs of the DSP system based on blind source separation.

Auditory attention inference system estimates the probability of attention on each specific sound source using EEG measurements and estimated sound sources. Gain controller system takes the estimated probabilities from the attention inference system to modify gains of each specific sound. The details of the attention inference system and gain adjustments are provided in the following sections.