Journal of Neuroscience Methods 298 (2018) 1–15

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

Improved artefact removal from EEG using Canonical Correlation Analysis and spectral slope

Azin S. Janani [a,b,*], Tyler S. Grummett [a,b,c], Trent W. Lewis [a,b], Sean P. Fitzgibbon [d], Emma M. Whitham [e], Dylan DelosAngeles [a,c], Hanieh Bakhshayesh [a,b], John O. Willoughby [c,f], Kenneth J. Pope [a,b]

[a] College of Science and Engineering, Flinders University, Adelaide, Australia
[b] Medical Device Research Institute, Flinders University, Adelaide, Australia
[c] Centre for Neuroscience, Flinders University, Adelaide, Australia
[d] Wellcome Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford, United Kingdom
[e] Department of Neurology, Flinders Medical Centre, Adelaide, Australia
[f] College of Medicine and Public Health, Flinders University, Adelaide, Australia

HIGHLIGHTS

- We investigated limitations of traditional Canonical Correlation Analysis (CCA), and considered other artefacts.
- We introduced a new automatic muscle-removal approach based on traditional CCA and the spectral slope of components.
- We validated the effectiveness of this approach using EMG-free data from subjects that were given a neuromuscular blockade.

ARTICLE INFO

Article history:
Received 1 September 2017
Received in revised form 19 January 2018
Accepted 19 January 2018

Keywords:
Electroencephalogram
Electromyogram
Canonical Correlation Analysis
Spectral slope
Automatic muscle removal
Neurophysiological response

ABSTRACT

Background: Contamination of scalp measurement by tonic muscle artefacts, even in resting positions, is an unavoidable issue in EEG recording. These artefacts add significant energy to the recorded signals, particularly at high frequencies. To enable reliable interpretation of subcortical brain activity, it is necessary to detect and discard this contamination.
New method: We introduce a new automatic muscle-removal approach based on the traditional Blind Source Separation-Canonical Correlation Analysis (BSS-CCA) method and the spectral slope of its components. We show that CCA-based muscle-removal methods can discriminate between signals with high correlation coefficients (brain, mains artefact) and signals with low correlation coefficients (white noise, muscle). We also show that typical BSS-CCA components are not purely from one source, but are mixtures from multiple sources, limiting the performance of BSS-CCA in artefact removal. We demonstrate, using our paralysis dataset, improved performance using BSS-CCA followed by spectral-slope rejection.
Result: This muscle removal approach can reduce high-frequency muscle contamination of EEG, especially at peripheral channels, while preserving steady-state brain responses in cognitive tasks.
Comparison with existing methods: This approach is automatic and can be applied on any sample of data easily. The results show its performance is comparable with the ICA method in removing muscle contamination and has significantly lower computational complexity.
Conclusion: We identify limitations of the traditional BSS-CCA approach to artefact removal in EEG, propose and test an extension based on spectral slope that makes it automatic and improves its performance, and results in performance comparable to competitors such as ICA-based artefact removal.

1. Introduction

Electroencephalogram (EEG) is an important biological signal reflecting electrical changes in networks of cortical and subcortical neurons. It has an important role in brain disease diagnosis, brain computer interfaces (BCI), and brain research (Vinhas et al., 2008;

* Corresponding author at: College of Science and Engineering, Flinders University, GPO Box 2100, Adelaide 5001, South Australia, Australia.
E-mail address: azinsadat.janani@flinders.edu.au (A.S. Janani).

https://doi.org/10.1016/j.jneumeth.2018.01.004
0165-0270/

Zhang et al., 2015). Scalp measurements are contaminated by different non-neural biological artefacts, especially electromyogram (EMG) (Zhang et al., 2015). Removing this muscle contamination is one of the main issues in EEG research.

Phasic contractions of cranial muscles are large enough in amplitude that they can be easily detected by visual inspection or mathematical algorithms and then, either, the data are ignored or the signal imperfectly "cleaned" (Freeman et al., 2003; Goncharova et al., 2003; Fitzgibbon et al., 2015). In contrast, tonic muscle activity, especially the steady contractions that maintain the posture of head and neck, are typically too low in amplitude to be easily detected in scalp measurements but significantly affect their power spectra (Whitham et al., 2008; Pope et al., 2009). Recent research on paralysed, conscious subjects indicates that tonic muscle contamination increases spectral power of EEG from frequencies as low as 10–20 Hz, with increases up to 200 times higher than EEG especially at peripheral channels (Whitham et al., 2007; Whitham et al., 2008; Pope et al., 2009).

Different methods have been implemented which aim to provide EMG-free EEG using various referencing methods (Fitzgibbon et al., 2013; Fitzgibbon et al., 2015) or blind source separation (BSS) (Pope and Bogner, 1996), but none of them have been completely successful (Amari et al., 1996; McMenamin et al., 2010; Fitzgibbon et al., 2013; Gabsteiger et al., 2014). Independent components analysis (ICA) (Shackman et al., 2009; McMenamin et al., 2011; Fitzgibbon et al., 2016) and canonical correlation analysis (CCA) (De Clercq et al., 2006; Karhunen et al., 2012) are the most common BSS approaches in removing muscle contamination.

ICA separates a set of signals (EEG channels) into a set of sources or components, via a linear transformation. Like principal component analysis (PCA), ICA algorithms produce components that are uncorrelated, but ICA additionally maximises their independence (Prakash and Roy, 2016). Different characteristics of components such as spectral slope, topographic maps and their temporal properties can be evaluated to detect and reject muscle components (Shackman et al., 2009; Fitzgibbon et al., 2016). The disadvantages of these techniques are their computational complexity (Hyvärinen and Oja, 2000), the need for many samples of data, particularly when there are many EEG channels (Korats et al., 2012), and their inability to separate sources with Gaussian distributions or white spectra, depending on the ICA algorithm (Prakash and Roy, 2016).

CCA is a statistical method using a linear transformation to find the correlation structure between two multivariate datasets (Hotelling, 1936). In 2006, Clercq and colleagues proposed CCA as a BSS approach to remove muscle contamination of EEG. BSS-CCA finds components uncorrelated with each other and maximally autocorrelated at a lag of one sample by considering the scalp recording data as the first dataset and a delayed version as the second dataset (De Clercq et al., 2006). The broad band spectrum of EMG, resembling white noise, and its concomitant low autocorrelation are exploited to identify muscle components (Chen et al., 2014). Since then, several researchers have reported the superiority of BSS-CCA to the traditional ICA method of extracting phasic muscle components (De Clercq et al., 2006; Gao et al., 2010; Karhunen et al., 2012).

The standard approach to muscle reduction using CCA has some limitations. First, muscle activity, both phasic and tonic, does not have a flat spectrum like white noise (Engel et al., 1992; Bertrand and Tallon-Baudry, 2000; Goncharova et al., 2003). For example, Fitzgibbon et al. (2016) and Goncharova et al. (Goncharova et al., 2003) have shown that spectral power of muscle components increases with frequency in the range 7–75 Hz. Second, there is little discussion about how to choose which components are to be discarded. Perhaps the clearest advice is to discard components with low correlation coefficient one by one until enough muscle contamination is removed (Górriz et al., 2011). Third, the effect of environmental noises in the recorded mixtures are ignored. Two significant sources are mains power noise and white noise, which have autocorrelation coefficients in the range of brain and muscle respectively. Fourth, while the effectiveness of the approach in removing phasic muscle contamination has been tested, its effect on tonic muscle has not been addressed. It has been shown that constant tonic muscle contamination of EEG is significant even in resting positions (Whitham et al., 2007).

In this paper, we address these limitations by examining the correlation coefficients of brain, muscle, mains power and white noise signals, and use datasets that have no muscle (paralysis), and no muscle and no brain (background noise) to identify improvements in the approach. We then propose a new algorithm that uses the spectral slope of components in addition to their correlation coefficient to improve muscle removal, especially tonic muscle. Finally, we compare our proposed algorithm with a standard ICA muscle removal approach, based on Infomax (Bell and Sejnowski, 1995), and demonstrate improved performance can be achieved.

2. Method

2.1. Datasets

To evaluate different muscle removal approaches, we applied them to three different existing datasets recorded from healthy participants. All participants signed a consent form and these experiments were approved by the Clinical Research Ethics Committee of Flinders Medical Centre and Flinders University.

The first dataset contains scalp recordings from 6 participants as described in Table 1. The participants were asked to complete a series of tasks including: baseline eyes closed, baseline eyes open, the auditory verbal learning task, serial subtraction, and exposure to a strobe light. The tasks were performed twice, once before and once during pharmacologically-induced paralysis. So, the first set of these data contained muscle artefact (pre-paralysis or EMG-contaminated) while the second has no muscle artefact (post-paralysis or EMG-free). More information about this dataset can be found in (Whitham et al., 2007; Whitham et al., 2008).

The second dataset consists of 13 participants as described in Table 1. An auditory stimulus with a 1500 Hz carrier amplitude modulated by a 40 Hz message was presented for all participants under three conditions as previously described (DeLosAngeles, 2010). Under all conditions, the brain should exhibit an auditory steady state response to this stimulus. The total recording time was about 6 min.

The third dataset includes 93 participants (large sample) as detailed in Table 1. They were asked to complete a series of tasks: baseline eyes closed, baseline eyes open, the auditory verbal learning task, serial subtraction, auditory discrimination, visual discrimination, visual rotation, reading, traversing a maze, and fin-

Table 1
EEG parameters and subject demographics.

Dataset number	females	males	Age	Number of EEG channels	Reference channel	Sampling frequency (Hz)	Length of data (min)
1	1	5	28–73	115	Left ear	1000	12
2	7	6	7–80	112	Left ear	1000	6
3	48	45	29–62	115	Linked-ears	1000	22