Journal of Neuroscience Methods 275 (2017) 80–92

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

The extraction of motion-onset VEP BCI features based on deep learning and compressed sensing

Teng Ma[a], Hui Li[a], Hao Yang[a], Xulin Lv[a], Peiyang Li[a], Tiejun Liu[a,b], Dezhong Yao[a,b,*], Peng Xu[a,b,*]

[a] Key Laboratory for NeuroInformation of Ministry of Education, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, 610054, China

[b] Center for Information in BioMedicine, University of Electronic Science and Technology of China, Chengdu, 610054, China

HIGHLIGHTS

- Deep learning is combined with compressed sensing to mine discriminative mVEP information.
- The deep learning and compressed sensing approach can generate multi-modal features.
- The proposed multi-modal feature can effectively improve the BCI performance with approximately 3.5% accuracy incensement.
- The deep learning and compressed sensing approach is more effective for subjects with relatively poor performance.

ARTICLE INFO

Article history:
Received 22 April 2016
Received in revised form 28 September 2016
Accepted 8 November 2016
Available online 11 November 2016

Keywords:
Motion-onset VEP
Brain computer interface
Compressed sensing
Deep learning
Multi-modality feature

ABSTRACT

Background: Motion-onset visual evoked potentials (mVEP) can provide a softer stimulus with reduced fatigue, and it has potential applications for brain computer interface(BCI)systems. However, the mVEP waveform is seriously masked in the strong background EEG activities, and an effective approach is needed to extract the corresponding mVEP features to perform task recognition for BCI control.

New method: In the current study, we combine deep learning with compressed sensing to mine discriminative mVEP information to improve the mVEP BCI performance.

Results: The deep learning and compressed sensing approach can generate the multi-modality features which can effectively improve the BCI performance with approximately 3.5% accuracy incensement over all 11 subjects and is more effective for those subjects with relatively poor performance when using the conventional features.

Comparison with existing methods: Compared with the conventional amplitude-based mVEP feature extraction approach, the deep learning and compressed sensing approach has a higher classification accuracy and is more effective for subjects with relatively poor performance.

Conclusions: According to the results, the deep learning and compressed sensing approach is more effective for extracting the mVEP feature to construct the corresponding BCI system, and the proposed feature extraction framework is easy to extend to other types of BCIs, such as motor imagery (MI), steady-state visual evoked potential (SSVEP)and P300.

1. Introduction

A Brain Computer Interface (BCI) is used to establish communication between humans and output devices, such as a computer application or a neuroprosthesis via noninvasive or invasive approaches. For noninvasive BCI, a scalp-recorded electroencephalogram (EEG) from a participant conveys intentions according to some well-defined paradigms (e.g., motor imagery or some external visual stimuli), and it is the mostly widely used signal for BCI control due to its low cost and portability. There are various BCI types, based on the EEG signals used to perform BCI control; of these, the VEP-based BCI is one of the important branches in an EEG-based BCI system (Liu et al., 2004; Page et al., 2005; Wolpaw et al., 2000; Wolpaw and McFarland, 2004). The main merit of VEP-based BCI is that it can evoke the relative stable control signal and

* Correspondence to: #4, Section 2 of North Jianshe Road, Chengdu, Sichuan, 610054, China.

E-mail addresses: dyao@uestc.edu.cn (D. Yao), xupeng@uestc.edu.cn (P. Xu).

http://dx.doi.org/10.1016/j.jneumeth.2016.11.002
0165-0270/

does not need a comprehensive training procedure for the subject, so that it can provide the effective control (Bin et al., 2009a; Lee et al., 2006; Momose, 2007). However, most of the VEP-based BCIs use a light flash or pattern reversal (i.e., a high contrast and a bright luminance of a visual object) to evoke the related brain signals for control (Bin et al., 2009b; Müller-Putz and Pfurtscheller, 2008; Y. Zhang et al., 2012). In practical situations, these approaches may cause visual fatigue in the BCI user, especially after long-term use (Beveridge et al., 2015; Gao and Gao, 2014). Recently, a new VEP stimulus, i.e., motion-onset VEP (mVEP), was introduced that can provide a relatively softer stimulus and has been used to build the online-BCI system (Guo et al., 2008; Marshall et al., 2015). The mVEP represents a visual motion response from the middle temporal area (MT) and the medial superior temporal area (MST). Compared to the traditional VEP-based BCI using pattern-onset or pattern-reversal stimulus, the mVEP-based BCI requires no sudden change in luminance or high contrast of visual objects; thus, subjects have less visual fatigue and feel more comfortable. Motion-onset evoked potentials (mVEP) are mainly characterized by three components: P1, N2, and P2. P1 usually occurs at approximately 130 ms. N2 is produced in the temporal-occipital regions at 160–200 ms and mainly reflects neural activity (Kuba et al., 2007) and processed motion vision (Hollants-Gilhuijs et al., 2000). P2 commonly occurs at approximately 240 ms and is mainly distributed in the parietal and central areas of the brain (Kuba and Kubová, 1992).

mVEP is deeply masked in strong background noise, and several repeated stimuli are usually required to perform the averaging for a reliable mVEP waveform. Generally, the more repeated trials that are involved in averaging, the higher quality the mVEP is. However, for an online BCI system, it is necessary to extract the EEG features and perform the recognition as quickly as possible. Therefore, the repeated stimulus number is usually sufficiently small that the mVEP waveform has a low signal to noise ratio (SNR) and the three components of mVEP are, consequently, difficult to extract reliably.

Currently, the most frequently used feature extraction method for mVEP is to directly utilize the amplitude information of N2 and P2 to perform the recognition (Kuba et al., 2007). The low SNR of the mVEP encountered in an online BCI system will weaken the BCI performance. Therefore, the effective feature extraction is crucial to improving the mVEP BCI performance.

Recently, compressed sensing and deep learning have been widely utilized in feature extraction. Compressed sensing has mainly been applied in the fields of image feature extraction (Li et al., 2011; K. Zhang et al., 2012), speech recognition (Gemmeke et al., 2010), and speech compression (Gunawan et al., 2011). The most important role of compressed sensing is to extract the primary information from the redundant original signal effectively in real time, reconstruct signals and satisfy the requirements of both prevention against overfitting and dimension reduction and the need of fast real-time at the same time especially for online-BCI system. Deep learning, which can extract the distributed features of a finite sample, is mainly used in fields such as speech recognition (Bengio et al., 2013; Sainath, 2014) and computer vision (Cui et al., 2014; Sohn et al., 2011).

Derived from the potential of compressing sensing and deep learning for feature extraction, we developed a novel approach combined with two known approaches to extract the mVEP features for the BCI system in this study.

2. Methodology

2.1. EEG recording

Overall, 11 subjects (3 females, 8 males, aged 23.6 ± 1.2 years) participated in the experiment. They all had normal or corrected to normal vision. The experimental protocol was approved by the Institution Research Ethics Board of the University of Electronic Science and Technology of China. All participants were asked to read and sign an informed consent form before participating in the study. After the experiment, all participants received monetary compensation for their time and effort.

Ten Ag/AgCl electrodes (CP1, CP2, CP3, CP4, P3, P4, Pz, O3, O4, Oz) from an extended 10–20 system were placed for EEG recordings with a Symtop amplifier (Symtop Instrument, Beijing, China). All electrode impedances were kept below 5 kΩ, and the AFz electrode was used as a reference. The EEG signals were sampled at 1000 Hz with a 50 Hz notch filter.

The visual stimuli were presented on a 14-in. LCD monitor with a 60 Hz refresh rate and 1280×1024 resolution and were viewed from a distance of 50 cm. Fig. 1(a) shows the graphical user interface (GUI) for the subjects in the experiment, with a visual field of $30^{\circ} \times 19^{\circ}$ on the screen. Six virtual buttons representing 1, 2, 3, 4, 5, and 6 were embedded in the rectangle interface, and each had a visual field of $4^{\circ} \times 2^{\circ}$. In each virtual button, a red vertical line appeared on the right side of the button and moved leftward until it reached the left side of the button to form a brief motion-onset stimulus. The entire move took 140 ms, with a 60-ms interval between consecutive moves. The motion-onset stimulus in each of the six buttons appeared in a random order, and a trial was defined as the complete and successive appearance of a motion-onset stimulus in all six virtual buttons. The interval between two trials was 300 ms, and each trial lasted 1.5 s (see Fig. 1(b)). Five trials comprised a block, which took 7.5 s.

During the experiment, the subject needed to focus on the button that was indicated in the center of the graphical user interface, and the instructed number was randomly generated. To increase their attention, the subject was asked to silently count the time between moving stimuli appearing in the target buttons. A total of 72 blocks, including 360 trials, was collected for each subject in two separate equal sessions, with a 2-min rest period between sessions. The first session was used as a training set, and the second session was used as a test set for the analyses. For details on the EEG data set, please refer to our previous work (Zhang et al., 2015). In this study, we used the collected data set to evaluate the feasibility of using combined sensing and deep features to improve the performance of an mVEP BCI system.

2.2. EEG processing

In the EEG recordings, each flashing button had 5 trials in each block. Before feature extraction, processing including ocular artifact removal, bandpass filtering, and averaging were applied to the trials corresponding to each button. This EEG analysis procedure is shown in Fig. 2.

In the current study, ± 75 μv was used as the threshold for ocular artifact removal, and trials with an EEG amplitude exceeding this threshold were discarded from the analysis. Then, artifact-free trials were bandpass filtered within 0.5–10 Hz for each subject. Finally, the filtered trials were averaged according to the flashing box. After the processing procedure in Fig. 2, 6 averaged ERPs with each corresponding to one flashing box were estimated in each block. Among the 6 ERPs, the only ERP corresponding to the indicated number cue was defined as the target, which theoretically had the strongest mVEP. The remaining 5 trials were treated as the non-target. The aim of the following feature extraction was to extract discriminative information such that only one target would be recognized from the six ERPs. Because the characterized components of mVEP were N2 and P2, a time window of 131–322 ms, which covered these two components, was used as our time window for mVEP-related feature extraction at electrodes P3 and Pz in the occipital region.