International Journal of Psychophysiology 111 (2017) 156–169

Contents lists available at ScienceDirect

International Journal of Psychophysiology

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

Best practice for single-trial detection of event-related potentials: Application to brain-computer interfaces

Hubert Cecotti [a,*], Anthony J. Ries [b]

[a] Faculty of Computing and Engineering, Ulster University, Magee campus, Londonderry BT48 7JL, Northern Ireland, UK
[b] Human Research and Engineering Directorate, US Army Research Laboratory, Aberdeen, Proving Ground, MD 21005, United States

ARTICLE INFO

Article history:
Received 15 December 2015
Received in revised form 2 July 2016
Accepted 16 July 2016
Available online 22 July 2016

Keywords:
Event-related potentials
Brain-computer Interface
Biomedical engineering
Spatial filtering
Multivariate pattern analysis
Classification

ABSTRACT

The detection of event-related potentials (ERPs) in the electroencephalogram (EEG) signal is a fundamental component in non-invasive brain-computer interface (BCI) research, and in modern cognitive neuroscience studies. Whereas the grand average response across trials provides an estimation of essential characteristics of a brain-evoked response, an estimation of the differences between trials for a particular type of stimulus can provide key insight about the brain dynamics and possible origins of the brain response. The research in ERP single-trial detection has been mainly driven by applications in biomedical engineering, with an interest from machine learning and signal processing groups that test novel methods on noisy signals. Efficient single-trial detection techniques require processing steps that include temporal filtering, spatial filtering, and classification. In this paper, we review the current state-of-the-art methods for single-trial detection of event-related potentials with applications in BCI. Efficient single-trial detection techniques should embed simple yet efficient functions requiring as few hyper-parameters as possible. The focus of this paper is on methods that do not include a large number of hyper-parameters and can be easily implemented with datasets containing a limited number of trials. A benchmark of different classification methods is proposed on a database recorded from sixteen healthy subjects during a rapid serial visual presentation task. The results support the conclusion that single-trial detection can be achieved with an area under the ROC curve superior to 0.9 with less than ten sensors and 20 trials corresponding to the presentation of a target. Whereas the number of sensors is not a key element for efficient single-trial detection, the number of trials must be carefully chosen for creating a robust classifier.

1. Introduction

An event-related potential (ERP) is the measured brain response evoked by specific sensory, cognitive, or motor event. More generally, it is any stereotyped electrophysiological response to a stimulus. The ERP technique provides a powerful non-invasive tool for exploring the human brain, particularly for research related to the temporal measurement of cognitive mechanisms (Luck, 2005). An ERP component corresponds to the scalp-recorded neural activity generated from cortical sources and can be reliably measured using electroencephalography (EEG), a procedure that measures electrical activity of the brain over time using electrodes placed on the scalp (Luck, 2004). Since EEG measurements reflect thousands of simultaneous post-synaptic neural activations, the brain response to a single stimulus or event of interest is not usually visible in the EEG recording of a single trial (Childers et al., 1987). Therefore, experimenters typically average many trials together in order to see the brain's response to a stimulus. This causes random brain activity to be averaged out, and the relevant waveform to remain, i.e. the ERP. One of the most widely studied ERP components, first reported over 50 years ago, is the P3 ERP component (also referred to as the P300), a large positive wave that peaks around 300 ms after the stimulus onset (Sutton et al., 1965). The P3 is often used to index processing related to stimulus categorization and as an input signal in many brain computer interface (BCI) systems for both patients and healthy individuals (Donchin et al., 2000; Donnerer and Steed, 2010). The P3 has been traditionally studied using a two-stimulus oddball task where an infrequent "oddball" target stimulus is presented among a series of frequent non-target distractor stimuli where only the infrequent target stimulus requires a response by the observer (Polich, 2007). The P3 amplitude is affected by target probability such that infrequent targets produce larger amplitudes compared with frequent targets, however, P3 amplitude is also affected by stimulus status since larger P3 amplitudes have been obtained for stimuli defined as targets when having the same probability of occurrence as non-target stimuli (Duncan-Jhonson and Donchin, 1977; Johnson, 1988). The time between target presentations also affects P3 amplitude showing increased amplitude as target to target interval increases (Croft et al., 2003). The latency of the P3 is correlated with the reaction time to a

* Corresponding author.
E-mail address: h.cecotti@ulster.ac.uk (H. Cecotti).

http://dx.doi.org/10.1016/j.ijpsycho.2016.07.500
0167-8760/

target stimulus making the latency at which this component peaks a useful metric for estimating the time it takes to evaluate and categorize a target stimulus (Dien et al, 2004; Folstein and Van Petten, 2011). Targets that are easier to categorize produce faster reaction times and earlier P3 peak latencies than more difficult targets. For reviews on the P3 ERP (see Luck and Kappenman, 2011; Polich, 2007). Other ERPs components such as the N2 (a negative wave that peaks 150-350 ms post-stimulus), which usually precedes the P3, have been used with single-trial detection as a single component, or with the P3. This ERP component has been extensively studied to find out its relationship with selective attention to specify stimulus location in certain area of the visual field (Kiss et al., 2008). In the subsequent sections of this paper, the ERP definition includes the brain response relative to any sensory, cognitive, or motor event stimulus. Hence, it includes both visual evoked potentials (VEPs), and auditory evoked potentials (AEPs) (King et al., 2013). Finally, to limit the scope of this paper dedicated to ERP single-trial detection, this paper focuses on multivariate pattern analysis (MVPA) methods that allow the extraction of meaningful and reliable information about the presence of a particular ERP at the single-trial level (e.g. P3; target vs. non-target trial). This differs from methods that extract relevant information from a population of trials by using features from each individual trial (e.g. amplitude or latency measurements across trials). In this later case, while the analysis is performed at the single-trial level, only the group analysis can provide information about neural processes.

In typical ERP analysis in cognitive neuroscience, the grand average response across trials is used to analyze and compare differences between ERP characteristics (amplitude, latency) across subjects. However, it is important to be cautious about the interpretation of ERP waveforms. In most of the ERP experiments, the different ERP waveforms are isolated by using the grand average response. Yet, the variations across trials may not be captured by the grand average response, which may provide a biased view of the single-trial waveforms. This effect is enhanced when ERP component latencies vary significantly across trials. Hence, it is ideally better to not assume that an averaged ERP waveform represents the single-trial waveform prototype. Only a few studies have taken this information into account (Marathe et al, 2014), by extracting shift-invariant features from the EEG signal (Cecotti, 2015b).

Modeling trial-to-trial variability in EEG signal has become a major focus in single-trial classification. Latencies at the single-trial level are typically by peak-picking and template-matching (Smulders et al., 1994). Already in Kutas et al. (1977), it was shown at the single-trial level that the latency of P3 corresponds to stimulus evaluation time and is independent of response selection. In Delorme et al. (2015), they have developed a method based on an ERP-image visualization tool in characteristic such as potential and spectral power are represented as colour coded horizontal lines that are then stacked to form a 2-D colored image. Moving-window smoothing across trial epochs can make otherwise hidden ERP features in the data more perceptible. Stacking trials in different orders, for example ordered by subject reaction time, by context-related information such as inter-stimulus interval, or some other characteristic of the data (e.g., latency-window mean power or phase of some EEG source) can reveal aspects of the multifold complexities of trial-to-trial EEG data variability.

In this paper, we define a complete ERP signal-trial detection pipeline that includes temporal filtering, spatial filtering, and classification. For each component, we review the best methods and the best parameters that are currently used. We compare the performance of several state-of-the art techniques from the BCI literature in order to assess the performance of these different approaches to highlight the most efficient pre-processing steps and classification procedure. To compare the different techniques, we consider a database of 16 healthy subjects performing a visual target detection task, and where it can be expected to find a major N2 and P3 component with a high amplitude. Moreover, this task illustrates how oddball paradigms have evolved to complex and more applied problems that can be applied in novel BCI systems for potential threat detection applications in real-world scenarios. The remainder of this paper is organized as follows: First, we present how ERP detection is used in BCI in Sections 2 and 3. After the description of the experimental protocol related to the data used in this paper, in Section 4, we describe the system architecture in Section 5. The pre-processing and classification techniques are detailed in Sections 6 and 7. Finally, the results are presented in Section 8 and discussed in Section 9.

2. Brain-computer interface

Brain-Computer Interface (BCI) or Brain-Machine Interface (BMI) systems have been introduced as a new means of communication for severely disabled people who are unable to communicate with conventional devices (e.g. mouse, keyboard, switch), for rehabilitation purposes (Wolpaw et al., 2002; Millán et al, 2010), and to enhance performance of healthy individuals (Lance et al., 2012). BCIs based on ERP detection require subjects to pay attention to a specific sequence of stimuli (typically visual or auditory) in order to produce a robust and detectable ERP. The stability of the spatial distribution, the amplitude, and the latency of a brain evoked responses are key features that allow robust single-trial detection. It is now possible to reliably detect brain evoked responses using efficient signal processing methods that denoise the signal and enhance its main discriminant characteristics. This principle has been used in BCI to detect specific event-related potentials (Wolpaw et al., 2002). Virtual keyboards based on the detection of ERPs have been used in BCI, the most famous variation is the P3 speller and (Farwell and Donchin, 1988), and new variations based on other ERP components have been proposed (Hong et al., 2009). A large number of studies have been dedicated directly to the P3 speller, in order to understand the impact of its parameters, and how this system can be efficiently used (Guger et al., 2009). While the P3 ERP is relatively stable in P300 speller paradigms, accurate and reliable detection of the specific neural responses often requires averaging multiple responses. For instance, it is common that about ten trials are averaged in BCI virtual keyboards to optimize the accuracy (Cecotti et al., 2011). The requirement of several trials is mainly due to the noise in the signal coming from eye movements, muscular contractions, and ongoing brain activity that is unrelated to the experimental task. Although averaging the signal from multiple brain responses can increase the efficiency of detection, it also decreases the information transfer rate of the BCI due to the increase of time to acquire additional trials that are needed to reach a robust decision (Cecotti, 2011). Moreover, there exist tasks where it is not possible to repeat the visual stimuli: they appear only one time (Cecotti, 2015b). It happens in paradigms where the repetition may have an effect on the brain evoked response (use of memory), or when the application does not allow the repetition of the stimuli, e.g. when a subject watches a video; each frame of the video is presented only one time. In situations where novel incoming stimuli are presented in real-time, it may not be possible to repeat the presentation of visual stimuli in order to combine the decision scores from their corresponding brain responses. For this reason, single-trial detection has to be used for target detection where it is not possible to determine if an image belongs to a target or a non-target class by considering multiple presentations of the same image. Yet, if images can be presented several times, it is possible to combine the decision outputs from the different presentations like in the P3 speller (Farwell and Donchin, 1988). Thus, the real time constraint justifies the necessity to find new strategies for increasing the performance of single-trial detection. Finally, the N2pc (posterior-contralateral) has been used in some recent BCI studies (Awni et al., 2013; Matran–Fernandez and Poli, 2016; Sirvent Blasco et al., 2012), to extract information about the spatial location of attentional allocation to targets in images: a stronger deflection amplitude is expected in the area of the visual cortex which is opposite to the location of the target stimulus.