

Using Behavioral Information to Contextualize BCI Performance

Stephen M. Gordon^(✉), Jonathan R. McDaniel,
Jason S. Metcalfe, and Antony D. Passaro

DCS Corporation, Alexandria, VA 22310, USA
{sgordon, jmedaniel, jmetcalfe, apassaro}@dcscorp.com

Abstract. Brain-computer interface (BCI) systems often require millisecond-level timing precision in order to function reliably. However, as BCI research expands to an ever-widening array of applications, including operation in real-world environments, such timing requirements will need to be relaxed. In addition, overall BCI system design must be improved in order to better disambiguate the numerous, seemingly similar, neural responses that may arise in such environments. We argue that this new area of operational BCI will require the integration of neural data with non-neural contextual variables in order to function reliably. We propose a framework in which non-neural contextual information can be used to better scope the operational BCI problem by indicating windows of time for specific analyses as well as defining probability distributions over these windows. We demonstrate the utility of our framework on a sample data set and provide discussion on many of the factors influencing performance.

Keywords: Brain-computer interface (BCI) · Electroencephalography (EEG) · Asynchronous event detection

1 Introduction

Previous work to monitor individuals in operational environments using electroencephalography (EEG) has focused on continuous state monitoring. Systems have been developed to estimate states such as a user's alertness or fatigue [1, 2]. In order to perform reliably in operational settings many of these systems focus on, relatively, slowly varying signals that often exhibit broad scalp topologies. These systems do not attempt to monitor the moment-to-moment dynamics associated with event-related, phase-locked, processing unless precise event timing information is available. Within the field of brain-computer interfaces (BCI) there has been little work in developing tools capable of detecting such events in real-time without precise timing.

Most BCI classifiers are built using some form of time-locked signal representation, which in turn is based on the onset of some known event. However, in real-world applications the onset of events is often not known. Without knowing the time periods to which the classifier should be applied it is very difficult to control the number of false positives produced. This problem is exacerbated when one considers that in real-world environments there are numerous tasks and side-tasks that an individual may perform.

Therefore one of the major problems associated with asynchronous, online, EEG-based event detection is one of scope. Furthermore, we argue that it is unlikely that advances in EEG technology or signal processing methods for EEG analysis alone will be able to clearly disassociate the myriad brain states and phase-locked events that a system could encounter in operational settings.

In this regard, researchers and system developers must find alternative methods to analyze and interpret real-time EEG data. In order to facilitate application of current BCI tools to operational settings we feel that well-designed systems need the following three critical components:

1. *Trigger mechanism*: This mechanism indicates a period of high likelihood for an event's occurrence and, thus, scopes the classification problem.
2. *Likelihood function*: Even with an appropriate trigger mechanism, the temporal ambiguity of the neural event must be quantified or estimated in order to optimize classifiability.
3. *Decision criterion*: This enables a go/no-go decision to be made using the current data. In the case of a no-go response, the classifier may wait for more data, or output that there is insufficient information to make a decision.

Our approach is based on the notion that in operational settings there is an inextricable link between behavior and brain dynamics [3, 4] that must be taken advantage of in order to build functioning BCI systems.

2 Operational BCI Model

2.1 Trigger Mechanism

Most of the prior work in the field of asynchronous BCI [5, 6] has focused on strategies to reason over the output of multiple classifiers without utilizing additional information, such as that provided by other physiological and behavioral sources, to contextualize or scope the problem. Scoping the problem, however, is a critical need that must be met in order to perform BCI analysis in operational environments. In its most basic form, scoping the problem defines an analysis window over which the BCI system must reason and may be used to select the appropriate set of classifiers to apply to that analysis window. We propose the use of a trigger mechanism to initially scope the problem.

In operational environments there may be a number of potential trigger mechanisms. Triggers may occur before, during, or after an event and, in some cases, may only occur with a particular probability. However, when present, the trigger mechanism determines an analysis window for the underlying BCI components. In addition, the trigger itself may form one layer of event detection with the presence or absence of a neural signature forming a subsequent layer for event detection. Trigger mechanisms may be unique to each data set and to each cognitive event.

Specific trigger mechanisms could come from behavior, such as saccadic eye movements or changes in pupil dilation in response to the onset of visual stimuli or gross movement patterns in response to actual or perceived errors. Alternatively,

triggers could come from more general state changes such as changes in attentional focus or current task. In the prior case, the analysis window may be a relatively short, well defined period of time, whereas in the latter case the trigger may indicate a shift from one set of asynchronous BCI tools to another.

2.2 Likelihood Function

The second need for operational BCI is that, even given an approximate analysis window, there will still be a significant amount of temporal variability associated with the precise timing of the cognitive event. This is due, in large part, to the ambiguity of the natural world and the difficulty associated with determining the precise timing, both at the physiological level and at the hardware level, of events. It is important to establish, to the extent possible, a likelihood function describing the probability of the event within the analysis window. There has been recent work to account for temporal ambiguity in the event response, e.g. Marathe et al. [7], but such approaches have not incorporated probability distributions derived from known behavior and event dynamics.

There are many ways in which likelihood functions may be obtained, but the key is that these functions link measurable properties of the situation, e.g. the trigger mechanism, to the probability of the cognitive event. These functions allow the system to interpret the output of the BCI classifier, or an ensemble set of classifiers, at each time point in the analysis window. In addition, if the likelihood function can be linked to other measurable signals, either physiological or behavioral in origin, then this allows the integration of the BCI output into a more structured belief network describing the current state of the operator. Integration strategies that could be used include approaches such as generalized linear models, multivariate regression, Bayesian belief networks, Markov chains, or fuzzy methods. In addition, the likelihood function could be used to integrate the output of multiple classifiers in a manner similar to the approach of [7].

2.3 Decision Criteria

Given the ambiguity associated with defining both the analysis window and the likelihood function, operational BCI systems will also need some form of confidence metric or decision criteria to determine whether the current output should be accepted or rejected. If the output is rejected, the system may continue to analyze additional data as it arrives or assert that insufficient information was available to make the decision. Incorporating a likelihood function allows a natural extension to develop confidence values in the BCI output. An appropriate decision criterion must be tuned for each scenario to balance (1) the speed with which a decision is reached, (2) the number of misclassifications produced by the system, and (3) the number of missed cognitive events. There has been much prior work in machine learning to develop confidence metrics for both the input data and the output of the classifier [8]. We simply suggest using metrics that take advantage of the known dynamics of the situation, e.g. the likelihood function, the analysis window, and the signal to be classified.

3 Methods

We illustrate our proposed framework using an experiment in which participants were required to detect and report targets that appeared in the environment. Target reports were made by pushing a button with either the left or right index finger. We focus on classifying the target reports (i.e. distinguishing which hand the participants used). We use saccadic eye movements as the trigger mechanism and use empirically derived estimates of the distribution of response times given a saccadic eye movement as the likelihood function. We compare our approach with an asynchronous method in which no such information is available.

3.1 Participants

Participants were 13 right-handed males age 20–40 (mean, 31.3 std. 2.6). All subjects reported normal, or corrected to normal, vision and reported no known neurological issues. All participants provided written, informed consent in accordance with procedures approved by the Institutional Review Board of the US Army Research Laboratory, and all testing conformed to the guidelines set forth by the 1964 Declaration of Helsinki.

3.2 Stimuli and Procedure

Participants completed a simulation task in which they were driven, as passengers in a vehicle, through a simulated environment and were responsible for detecting and classifying targets and reporting the type of target, by pressing a button with either the left or right index finger. The simulated environment was an urban landscape and targets appeared at random, but in logically congruent, locations in that environment. There were two basic types of targets: threats and nonthreats. Participants were trained to recognize each type of target during an initial training phase. Training continued until each participant reached a minimum performance level of 80 % classification accuracy. Targets were presented roughly every three seconds and participants were asked to respond as quickly and accurately as possible. Participants completed two 15 min blocks, each comprising approximately 180 targets.

3.3 Physiological Recording

Electrophysiological recordings were sampled at 1024 Hz from 64 scalp electrodes arranged in a 10–10 montage using a BioSemi Active Two system (Amsterdam, Netherlands). External leads were placed on the outer canthus and below the orbital fossa of the right eye to record monopolar electrooculography (EOG). External leads were also placed on both mastoids to provide electrical reference as well as on the forehead and along the ridge of the masseter to record electromyographic (EMG) signals. The continuous data were referenced offline before being digitally filtered 0.3–50 Hz. To reduce muscle and ocular artifacts in the EEG signal we removed EOG and EMG components

using independent components analysis (ICA) by finding the ICA components that maximally correlated with the horizontal and vertical EOG channels [9].

Eye tracking was recorded at 60 Hz using the Facelab (www.seeingmachines.com) two-camera, video-based eye tracking system. Participants were calibrated for the eye tracker prior to the start of the experiment. In addition, participants performed a baseline eye tracking task involving a quick saccade task in which the participant had to saccade to different points on the screen and click on colored targets using the mouse pointer. This task was completed before and after the experiment.

3.4 Contextualized BCI System Construction

Baseline Classifier. We built a classifier for each participant to detect the finger movements associated with target responses. We used ICA to derive a set of components that captured the phase-locked, event-related potentials (i.e. ERPs) and non-phase-locked (i.e. spectral changes in the alpha and beta frequency bands) but event-related features associated with motor control for each subject. We used these components as features and a forward feature selection algorithm to select the best components for each subject. Classification was performed using a support vector machine with a radial basis function kernel implemented with the LibSVM toolbox for Matlab [10]. For each subject we trained on the first 15 min block of data and tested on the second.

Integrating Context. To improve detection and disassociation of cognitive events associated with target onset, target discrimination, and reporting, we developed the model shown in Fig. 1. For the results described in the following section, we focused on the model components highlighted by the dashed line.

We analyzed the relationships between target onset and the saccadic eye movement and between the saccadic eye movement and the motor response. We first smoothed the

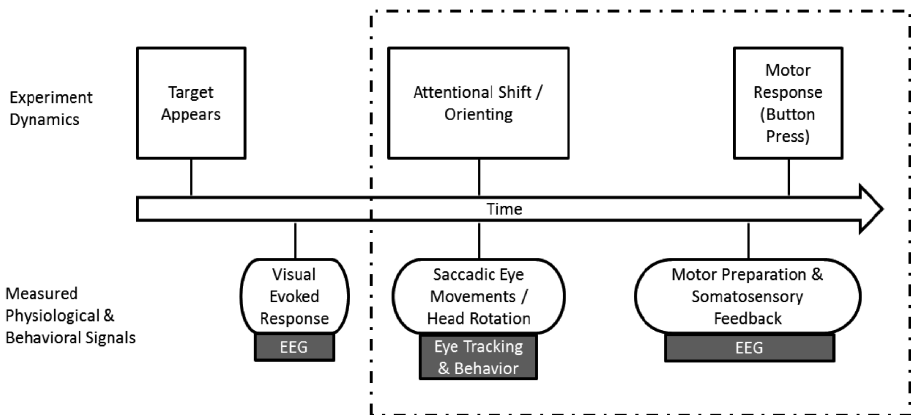


Fig. 1. Time course of experiment dynamics and measurable physiological and behavioral changes associated with target onset, detection, and reporting.

eye tracking data and then performed threshold-based saccade detection. For our current analysis we used the saccade as the trigger mechanism. We then used the probability of response given a saccade as the likelihood function, which was empirically estimated from the available training data.

We analyzed subsets of this data by stepping the BCI system over the analysis window. The analysis window extended from the onset of the saccade for 1.5 s. We limited our analysis to this window because no motor responses occurred outside of this time range for any subject. For each participant a set of classifiers were trained using different subsets of the available training data to predict the motor response. The classifiers were each trained using one second of data. We then stepped these classifiers over the analysis window in 100 ms increments and for each of these analysis points we summed the output from this ensemble of classifiers to obtain a single value. We used the distribution of reaction times given a saccadic eye movement to combine the outputs within a single analysis window.

Eye Tracking Quality. Eye tracking quality was determined using the baseline saccade task. As previously stated, the task required participants to saccade to different points on the screen. For each trial, the minimum distance between the gaze coordinates measured by the eye tracker and the known screen coordinates for the current point was used to create an error for that trial. These errors were z-scored and then averaged over all trials for each participant and used to compute a total quality for each subject. We standardized these values before averaging because we found that the errors produced at the end of the experiment were significantly larger than those produced before the experiment.

4 Results

4.1 Epoch BCI

The top row in Table 1 shows the classification results for all subjects when BCI training and testing was performed using knowledge of the exact timing for each finger movement, i.e. precisely timed, epoched data. These results are presented as a baseline since there was no need to incorporate trigger events or the likelihood function. The bottom row in Table 1 shows the standardized eye tracking error as computed using the baseline saccade tasks.

Table 1. BCI classification accuracy and eye tracker error for each subject. Top row: classification accuracy for each subject when the BCI system was trained and tested on precisely timed, epoched data. Bottom row: standardized mean-square error for each subject's eye tracking data.

	<i>S1</i>	<i>S2</i>	<i>S3</i>	<i>S4</i>	<i>S5</i>	<i>S6</i>	<i>S7</i>	<i>S8</i>	<i>S9</i>	<i>S10</i>	<i>S11</i>	<i>S12</i>	<i>S13</i>
BCI	0.75	0.80	0.83	0.64	0.84	0.77	0.74	0.76	0.85	0.83	0.84	0.71	0.73
Eye	-0.39	0.56	-0.17	1.35	0.47	-0.39	-0.73	-0.77	-0.25	0.30	-0.41	-0.58	1.22

4.2 Contextualized BCI

Next, we performed asynchronous BCI detection using the same data that was used to generate the results in Table 1. As described previously, we focused on analysis windows starting with the onset of the saccade, but we removed all experimental information describing the timing of those events within the analysis window. We performed two types of asynchronous detection. First we used the continuous fire approach where we obtained final BCI output by integrating the individual outputs computed over the entire analysis window. Second, we used the likelihood function that related the probability of a motor response to the onset of the saccade. We applied these likelihood values to the individual outputs produced at each analysis point in the analysis window. This likelihood function was determined using the same training data that was used to develop the BCI classifiers. Finally, we compared the performance of both approaches by analyzing how the accuracy varied as a function of confidence level.

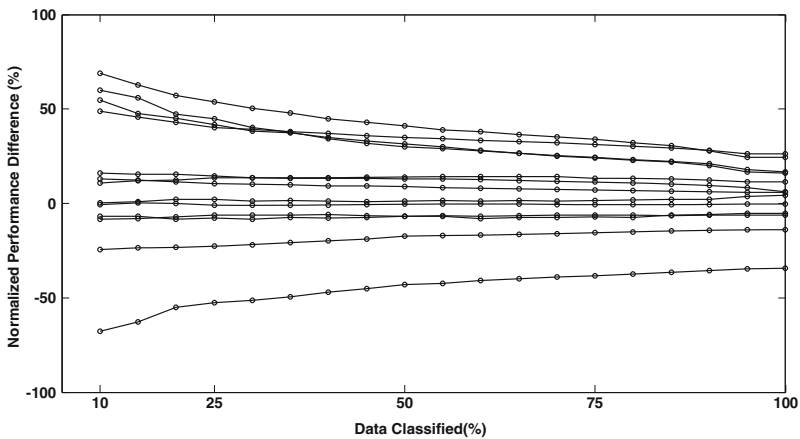


Fig. 2. Percent improvement in classification for the contextualized BCI over the non-contextualized approach. For each subject the differences have been normalized by the baseline classification level achieved with the epoch-based classifiers from Sect. 4.1.

Figure 2 shows the performance differences between the contextualized approach and the non-contextualized approach for each subject as we iteratively relaxed the decision criteria to capture more of the motor events. The x-axis shows the percent of data classified based on selecting the data with the strongest confidence values. The y-axis shows performance differences between the two approaches, normalized for each subject by the baseline classification levels achieved in Table 1. Positive values indicate that the contextualized approach performed better.

As can be seen from Fig. 2, the results are mixed. Four subjects showed substantial improvement and had, generally, monotonically decreasing curves indicating that the confidence metrics were functioning properly. One subject performed substantially worse and the remaining eight subjects were distributed within the approximately $\pm 15\%$ range, indicating a net zero effect on performance across these subjects.

To better understand the performance differences we next analyzed the extent to which these differences were a function of eye tracker quality as well as baseline performance on the epoch-based BCI. These results are presented in Fig. 3. The x-axis shows the standardized eye tracker error (from Table 1), while the y-axis shows the standardized averaged performance gains (from Fig. 2). Blue o's indicate the top six performers on the baseline epoch-based BCI (from Table 1). Black x's indicate the bottom seven performers on the baseline epoch-based BCI. While the conclusions that can be drawn from such a small data set are limited, there appears to be general trend among the best performers (blue o's) that the performance gains are a negative function of eye tracker error. In other words, as the quality of the eye tracking data decreases the utility of our approach also decreases. For the worst performers (black x's) the trend is less clear. It is unclear the extent to which the performance differences are a function of eye tracking quality, poor BCI classifiability (as measured using the epoch-based BCI), other factors not discussed here, such as EEG quality, or some combination of multiple factors.

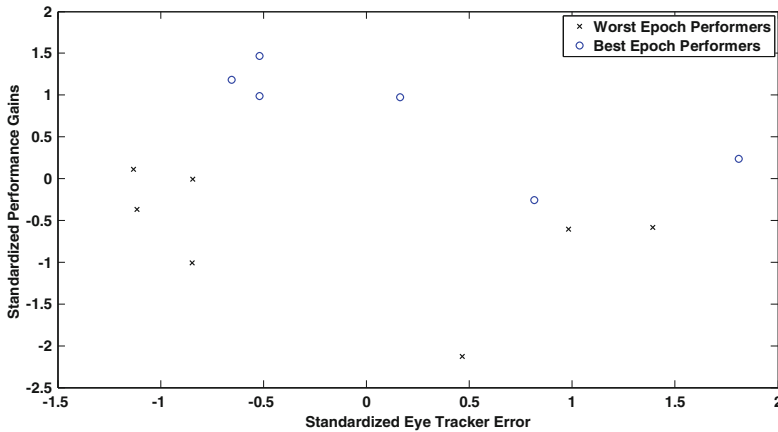


Fig. 3. Performance differences between the contextualized BCI and the non-contextualized BCI as a function of eye tracker error. Both axes have been standardized. Higher values indicate the contextualized BCI performed better. Data points marked by (o) indicate the six best performers on the epoch-based BCI. Data points marked by (x) indicate the seven worst performers on the epoch-based BCI (Color figure online).

5 Discussion

The results presented here suggest that, under the right circumstances, physiological and behavioral signals can be used to improve online BCI performance. While the results comparing the performance of the contextualized approach to the non-contextualized approach for all subjects were mixed, a more detailed analysis showed a possible trend between the quality of the contextualizing information and the performance gains for those subjects with better baseline BCI performance. We believe that this can be

explained, in part, by the fact that these subjects had more robust BCI classifiers allowing us to see the trend in eye tracking quality. However, further data is needed to substantiate this claim.

In addition, we did not focus on comparing different types of BCI classifiers. We attempted to optimize an SVM-based classifier for each subject, but future work should compare performance gains with our approach across a range of different classifier methods. In addition, future work should also attempt to model, to the extent possible, the moment-to-moment variations in signal quality for both the EEG and contextualize resources. As these sources of variability become better understood, this can pave the way for more comprehensive systems that incorporate multiple forms of trigger events as well as multivariate probability distributions.

As BCI research and development extends into more complex domains, we believe that approaches, such as the one presented here in which contextual information is incorporated into the BCI problem, will provide the best opportunities to build functioning, reliable systems. This represents a new challenge for traditional BCI but it is also an opportunity to expand towards a more comprehensive view of brain-body-behavior modeling.

Acknowledgements. The authors would like to thank Dr. Kelvin Oie of the Army Research Laboratory for his help in designing the research study. This research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0022. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

1. Lin, C.-T., Wu, R.-C., Liang, S.-F., Chao, W.-H., Chen, Y.-J., Jung, T.-P.: EEG-based drowsiness estimation for safety driving using independent component analysis. *IEEE Trans. Circuits Syst. I Regul. Pap.* **52**(12), 2726–2738 (2005)
2. Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P.D., Craven, P.L.: EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviat. Space Environ. Med.* **78** (Supplement 1), B231–B244 (2007)
3. McDowell, K., Lin, C.-T., Oie, K.S., Jung, T.-P., Gordon, S., Whitaker, K.W., Li, S.-Y., Lu, S.-W., Hairston, W.D.: Real-world neuroimaging technologies. *IEEE Access* **1**, 131–149 (2013)
4. Makeig, S., Gramann, K., Jung, T.-P., Sejnowski, T.J., Poizner, H.: Linking brain, mind and behavior. *Int. J. Psychophysiol.* **73**(2), 95–100 (2009)
5. Leeb, R., Friedmann, D., Müller-Putz, G.R., Scherer, R., Slater, M., Pfurtscheller, G.: Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: a case study with a tetraplegic. *Comput. Intell. Neurosci.* **2007**, 79642 (2007)

6. Townsend, G., Graimann, B., Pfurtscheller, G.: Continuous EEG classification during motor imagery-simulation of an asynchronous BCI. *IEEE Trans. Neural Syst. Rehabil. Eng.* **12**(2), 258–265 (2004)
7. Marathe, A.R., Anthony, J.R., McDowell, K.: Sliding HDCA: single-trial EEG classification to overcome and quantify temporal variability. *IEEE Trans. Neural Syst. Rehabil. Eng.* **22**, 201–211 (2014)
8. Schapire, R.E.: The boosting approach to machine learning: an overview. In: Denison, D.E., Hansen, M.H., Holmes, C.C., Mallick, B., Yu, B. (eds.) *Nonlinear Estimation and Classification. Lecture Notes in Statistics*, pp. 149–171. Springer, New York (2003)
9. Makeig, S., Bell, A.J., Jung, T.-P., Sejnowski, T.J.: Independent component analysis of electroencephalographic data. In: *Advances in Neural Information Processing Systems*, pp. 145–151 (1996)
10. Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol. (TIST)* **2**(3), 27 (2011)