33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Impact of target probability on single-trial EEG target detection in a difficult rapid serial visual presentation task

Hubert Cecotti, Joyce Sato-Reinhold, Jocelyn L. Sy, James C. Elliott,
Miguel P. Eckstein, Barry Giesbrecht

Abstract— **In non-invasive brain-computer interface (BCI), the analysis of event-related potentials (ERP) has typically focused on averaged trials, a current trend is to analyze single-trial evoked response individually with new approaches in pattern recognition and signal processing. Such single trial detection requires a robust response that can be detected in a variety task conditions. Here, we investigated the influence of target probability, a key factor known to influence the amplitude of the evoked response, on single trial target classification in a difficult rapid serial visual presentation (RSVP) task. Our classification approach for detecting target vs. non target responses, considers spatial filters obtained through the maximization of the signal to signal-plus-noise ratio, and then uses the resulting information as inputs to a Bayesian discriminant analysis. The method is evaluated across eight healthy subjects, on four probability conditions (P=0.05, 0.10, 0.25, 0.50). We show that the target probability has a statistically significant effect on both the behavioral performance and the target detection. The best mean area under the ROC curve is achieved with P=0.10, AUC=0.82. These results suggest that optimal performance of ERP detection in RSVP tasks is critically dependent on target probability.**

I. Introduction

The research directions in non-invasive brain-computer interface (BCI) depends on the target user and therefore the application. The target user directly drives the BCI performance expectation, from a working system to a system that reaches or outperforms other interfaces. Although BCI are mainly designed for people suffering from severe disabilities, a current trend is to consider BCI for healthy users in applications like target detection. For these users and due to ethical issues, BCI remain non-invasive. Such BCI can be based on the analysis of electrical signals recorded at the surface of the scalp (electroencephalography (EEG)).

The detection of event-related potentials, *e.g.* the P300, can be used for creating a BCI, *e.g.* the P300 speller [1]. The P300 is a positive deflection in the EEG that occurs about 300ms after a task-relevant target has been detected. A long line of studies has demonstrated that the P300 amplitude is affected by a variety of task factors, including target probability [2]. While these studies have shown the effects of target probability on the amplitude of the P300 ERP component, it remains unclear whether single-trial pattern classification is affected in a corresponding manner.

This research was supported by the Institute for Collaborative Biotechnologies through contract W911NF-09-D-0001 from the U.S. Army Research Office.

Department of Psychological & Brain Sciences, and Institute for Collaborative Biotechnologies, University of California Santa Barbara, Santa Barbara, CA 93106-9660, USA email: {lastname}@psych.ucsb.edu

The purpose of the present work is to investigate the effect of target probability on single trial detection in a problem where targets are difficult to detect, *i.e.* the behavioral performance is expected lower than for easier tasks. Understanding the influence of target probability on single-trial detection in difficult tasks will provide key insights into BCI designs that will be robust in a variety of task contexts. While the search of a well-identifiable target is already a challenge, the search of noisy targets increases the difficulty of the detection, both within the EEG signal, but also for the behavioral performance. This parameter is also relevant for other real applications as the target probability may vary over time, *e.g.* weapons in luggage or tumors in mammograms.

We propose in this paper to evaluate a difficult visual target detection task using a rapid serial visual presentation (RSVP) task. An RSVP paradigm is a useful tool for researchers working on visual attention, allowing researchers studying the temporal characteristics of neural information processing [3]. In this paradigm, each image replaces the previous one at the same spatial location. In [4], an RSVP system has been efficiently used for face recognition. Single trial detection of ERP has been addressed for target detection by using RSVP, *e.g.* the search of targets in satellite images [5], [6], [7]. Target detection has also been the subject of a recent competition in an international machine learning workshop [8]. In this competition, the area under the ROC curve (AUC) was about 0.82 for the best participants. We focus here on the impact of target probability on the overall detection.

II. Experimental protocol

A. Paradigm

Images of faces and cars were presented as stimuli to the observers who performed the behavioral task of identifying the correct label of the image (face/car) and their neural signals were recorded via electroencephalography. Figure 1 shows examples of the visual stimuli with and without noise (participants only saw the versions with noise). These images depict the difficulty of the task. Each image was presented for 500 ms, with no blank interval between subsequent images, resulting in a presentation rate of 2 Hz. In each session, target probability, *i.e.* face probability, was either 0.05, 0.10, 0.25, or 0.50 and remained constant throughout the session. Target images were set by block of two minutes keeping a relative constant target probability over time. Each subject had to complete four sessions (a session at each probability condition). Each session contained 12,000 trials. The order

978-1-4244-4122-8/11/$26.00 ©2011 IEEE

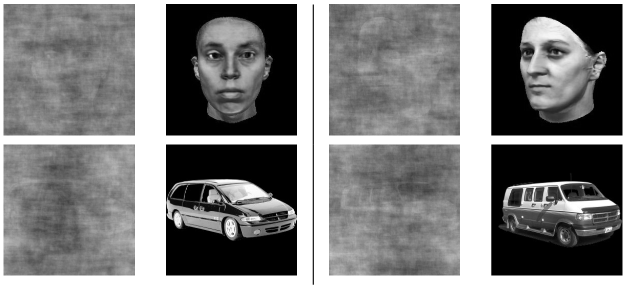


Fig. 1. Samples of visual targets (faces, **top**) and non-targets (cars, **bottom**) with their corresponding model.

of the session was counterbalanced across subjects. In the following sections, we denote by $P(0.05)$, $P(0.1)$, $P(0.25)$ and $P(0.5)$ the conditions with the target probability 0.05, 0.10, 0.25, and 0.50, respectively.

Eight healthy subjects participated to the experiment (mean= 23.5, sd=8.38, 3 females). They were instructed to respond to the presence of a face as quickly and accurately as possible by hitting the enter key on a standard keyboard.

B. Signal acquisition

The EEG signal was recorded from 32 Ag/ AgCl sintered electrodes mounted in an elastic cap and placed according to the International 10/20 System. The horizontal and vertical electro-oculograms (EOG) were recorded from electrodes placed 1 cm lateral to the external canthi (left and right) and above and below each eye, respectively. The data were sampled at 512 Hz and referenced offline to the signal recorded from the mastoids. Trials containing ocular artifacts (blinks and eye movements) detected by EOG amplitudes exceeding $\pm 75\mu V$ were excluded from both the ERP analysis and the behavioral performance evaluation.

C. Signal processing

Before the classification step, a set of features were determined for what best discriminates the ERP to targets from the ERP to non targets. The experimental protocol suggests the presence of a P300 in the ERP of each target. Therefore, we can estimate the location of an ERP between the beginning of the visual stimulus and less than 1s after its beginning. The sampling rate is 512Hz and the signal is first bandpassed filtered (Butterworth filter of order 4) with cutoff frequencies at 1 and 10.66Hz. Then the signal is downsampled to obtain a signal at a sampling rate equivalent to 32Hz. This new sampling rate corresponds to the sampling frequency used by the winning team of the MLSP competition 2010 [9]. For the following steps, we used the observed signal over 625ms after the start of a visual stimulus. As the sampling frequency is now 32Hz, it corresponds to 20 sampling points.

The next step consisted of enhancing the relevant signal by using spatial filters. Let us denote by $U \in \mathbb{R}^{N_s \times N_f}$, the spatial filters, where N_s is the total number of sensors and N_f is the number of spatial filters. The signal after spatial filtering is defined by $X_{filt} = XU$ where $X \in \mathbb{R}^{N_t \times N_s}$ is the recorded signal, N_t is the number of sampling points.

For spatial dimension reduction, we consider that the expected ERP is stable over time. Although, the latency and amplitude of the P300 may vary over time for a given task, we assume a spatially stationary waveform of the ERP. With this hypothesis we can consider a single set of spatial filters that can be applied all over the signal. We consider here the xDAWN algorithm [10], [11]. This method has been already successfully applied in BCI for P300 detection in the P300 speller paradigm. An algebraic model of the recorded signals X is composed of three terms: the responses on targets $(D_1 A_1)$, a response common to all stimuli, *i.e.* targets and non-targets confound $(D_2 A_2)$ and the residual noise (H)

$$X = D_1 A_1 + D_2 A_2 + H. \quad (1)$$

where D_1 and D_2 are two real Toeplitz matrices of size $N_t \times N_1$ and $N_t \times N_2$ respectively. D_1 has its first column elements set to zero except for those that correspond to a target onset, which are represented with a value equal to one. For D_2, its first column elements are set to zero except for those that correspond to stimuli onset. N_1 and N_2 are the number of sampling points representing the target (the P300 response) and superimposed evoked potentials, respectively. H is a real matrix of size $N_t \times N_s$.

The goal of applying spatial filters $U \in \mathbb{R}^{N_s \times N_f}$ is to enhance the signal to signal-plus-noise ratio (SSNR) of the enhanced P300 responses $(D_1 A_1 U)$, where N_f is the number of spatial filters

$$XU = D_1 A_1 U + D_2 A_2 U + HU. \quad (2)$$

We define the SSNR in relation to the spatial filters by:

$$\text{SSNR}(U) = \frac{Tr(U^T \hat{A}_1^T D_1^T D_1 \hat{A}_1 U)}{Tr(U^T X^T X U)} \quad (3)$$

where $\hat{A}_1$ corresponds to the least mean square estimation of A_1 :

$$\hat{A} = \begin{bmatrix} \hat{A}_1 \\ \hat{A}_2 \end{bmatrix} = ([D_1; D_2]^T [D_1; D_2])^{-1} [D_1; D_2]^T X \quad (4)$$

where $[D_1; D_2]$ is a matrix of size $N_t \times (N_1 + N_2)$ obtained by concatenation of D_1 and D_2.
Spatial filters are obtained through the Rayleigh quotient by maximizing the SSNR [11]:

$$\hat{U} = \text{argmax}_{\ U} \ \text{SSNR}(U). \quad (5)$$

For the classifier input, we used four spatial filters ($N_f = 4$). Therefore, the number of features that are given as input to the classifier is 80. For each input vector, the signals were normalized so that they had a zero mean and a standard deviation equal to one for each spatial component. Finally, the input vector is obtained by the concatenation of the N_f time-course signals across spatial filters. The Bayesian linear discriminant analysis (BLDA) classifier is considered for the detection of the P300 wave [12]. This classifier is relatively robust to noise in the training data as regularization is used during learning [13]. As the number of samples between targets and non-targets are not equal for conditions $P(0.05)$, $P(0.1)$ and $P(0.25)$, we rebalance the data set artificially. For each classifier, we upsample the data set by replicating cases from the minority class in order to have an equal number of samples for each class.

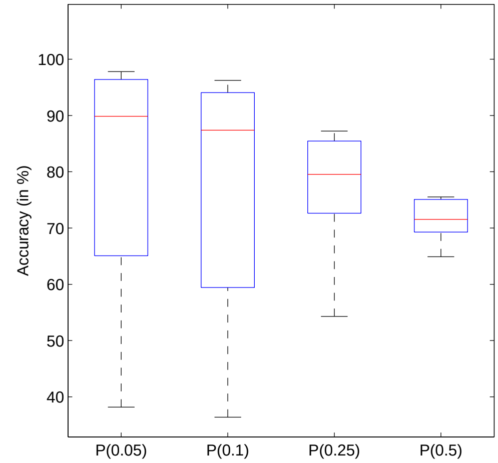
III. Results

The performance of every subject is evaluated through the behavioral performance and the EEG signal classification results. For the evaluation of the classifier, we provide the results obtained after a K-fold cross-validation where $K = 5$. The classifier evaluation is based on the AUC as ROC curves are insensitive to changes in class distribution [14].

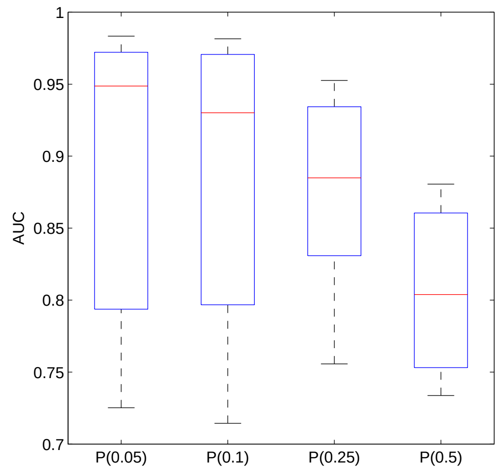
A. Behavioral performance

A response was considered as correct if the target was presented to the user less than 800ms before the user pressed a button. If the subject did not press a button within 800ms after the presentation of a target, it was considered as a mistake. It was also a mistake if the subject pressed a button when there was not a target on the screen.

The behavioral performance of each subject and for each condition is presented in Figure 2. These figures highlight first the difficulty of the task as no subject was able to complete the task with perfect accuracy. The mean accuracy was 78.7, 76.4, 77.0 and 71.5% for $P(0.05)$, $P(0.1)$, $P(0.25)$ and $P(0.5)$, respectively. The best accuracy was achieved by Subject 1 at $P(0.05)$ with an accuracy of 97.8%. A pairwise two-tailed t-test comparison indicates that there is no statistically significant difference across conditions for the accuracy. In Figure 2, the AUC of the behavioral performance is estimated as the normal cumulative distribution function of $d'/\sqrt{2}$ where d' is the sensitivity index $d' = Z(hit\ rate) - Z(false\ alarm\ rate)$ and $Z(p)$, $p \in [0, 1]$, is the inverse of the cumulative Gaussian distribution.



(a) Accuracy



(b) Estimated AUC

Fig. 2. Behavioral performance for each subject and each condition. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.

B. Single trial detection

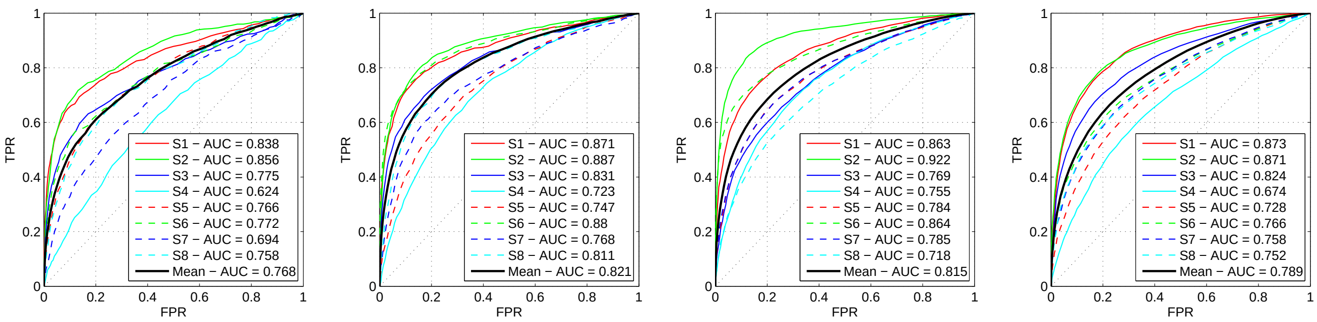
Figure 3 presents the ROC curves for every condition and every subject. In these curves the classifier and the spatial filters were evaluated each time on the same condition. The mean AUC across subjects is 0.768, 0.821, 0.815 and 0.789 for conditions $P(0.05)$, $P(0.1)$, $P(0.25)$ and $P(0.5)$, respectively. The standard deviation of the AUC is 0.074, 0.063, 0.068, and 0.070 for conditions $P(0.05)$, $P(0.1)$, $P(0.25)$ and $P(0.5)$, respectively. The best performance is achieved by Subject 2 with the condition $P(0.1)$, where the AUC reaches 0.922.

A pairwise two tailed t-test reveals that $P(0.1)$ provides better performance than $P(0.05)$ ($t_{(7)} = -3.77$, $p < 0.01$). $P(0.1)$ also provides better performance than $P(0.5)$ ($t_{(7)} = 2.50$, $p < 0.05$). A pairwise two-tailed t-test indicates no difference between $P(0.05)$ and $P(0.25)$, $P(0.05)$ and $P(0.5)$, $P(0.1)$ and $P(0.25)$ or between $P(0.25)$ and $P(0.5)$. The best probabilities for target detection are therefore 0.10 and 0.25.

The spatial distribution of the ERP present in 625ms after a visual stimulus corresponding to a target is depicted for every condition in Figure 4. The spatial distribution varies across conditions, with a higher activity in the occipital area than the parietal area for high target probabilities.

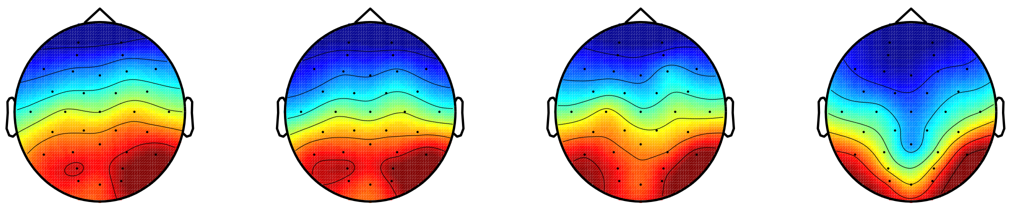
IV. Discussion and Conclusion

Target probability in an RSVP task has an influence on both behavioral and single trial detection performance. In ERP measurements, it is well known that the target probability has a high importance for P300 measures. Indeed, it is assumed that the longer the time between consecutive target occurrences within the typical oddball task, the larger is P300 amplitude and the shorter is its peak latency [15]. Single trial ERP detection confirms these findings as the average ERP detection is lower with P=0.5 than for P=0.25. However, with a low probability P=0.05, the ERP detection is not higher than with P=0.10 or P=0.25. It shows that there may exist an optimal target probability, which would allow the best target detection. It is therefore useless to decrease



(a) $P(0.05)$ (b) $P(0.1)$ (c) $P(0.25)$ (d) $P(0.5)$

Fig. 3. ROC curves for each condition.



(a) $P(0.05)$ (b) $P(0.1)$ (c) $P(0.25)$ (d) $P(0.5)$

Fig. 4. Average spatial distribution across subjects for each condition obtained through spatial filters analysis based on the xDAWN algorithm [11]. Blue/red colors denote positive/negative values.

the target probability to expect better performance as the attention of the subject may be impaired by a too low target probability. The different neural processes in relation to the target probability are confirmed by the analysis of the spatial distribution of the ERP. It suggests that the ERP to detect varies across target probabilities.

For BCIs that are used for target detection, the target probability is a critical parameter for determining the performance of an RSVP based application where the distribution of the target can be unpredictable. The differences observed at the spatial distribution of the ERP, at the behavioral performance and the classification level indicates that the ideal system should incorporate this information over time. Identifying the instantaneous distribution of the target probability may help finding or selecting the best spatial filters and the best classifiers to use anytime. Other parameters in the RSVP task could be dynamically changed in relation to the current target probability. With a high/low probability, the inter-stimulus interval could be reduced/increased. Further works will deal with the automatic selection of classifiers and RSVP parameters by determining the current target probability of the targets.

References

[1] L. Farwell and E. Donchin, “Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 70, pp. 510–523, 1988.
[2] J. Polich, “Updating P300: An integrative theory of P3a and P3b,” *Clinical Neurophysiology*, vol. 118, pp. 2128–2148, 2007.
[3] M. M. Chun and C. M. Potter, “A tow-stage model for multiple target detection in rapid serial visual presentation,” *J. Exp. Psy. Human Perception and Performance*, vol. 21, no. 1, pp. 109–127, 1995.
[4] J. Touryan, L. Gibson, H. J. Horne, and P. Weber, “Real-time measurement of face recognition in rapid serial visual representation,” *Frontiers in psychology*, vol. 2, no. 42, 2011.
[5] N. Bigdely-Shamlo, A. Vankov, R. R. Ramirez, and S. Makeig, “Brain activity-based image classification from rapid serial visual presentation,” *IEEE Trans. on Neural Systems and Rehab. Eng.*, vol. 16, no. 5, 2008.
[6] A. Gerson, L. Parra, and P. Sajda, “Cortically-coupled computer vision for rapid image search,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 174–179, 2006.
[7] L. C. Parra, C. Christoforou, A. D. Gerson, M. Dyrholm, A. Luo, M. Wagner, M. G. Philiastides, and P. Sajda, “Spatio-temporal linear decoding of brain state: Application to performance augmentation in high-throughput tasks,” *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 95–115, 2008.
[8] K. E. Hild, M. Kurimo, and V. D. Calhoun, “The sixth annual MLSP competition, 2010,” *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 107–111, 2010.
[9] J. M. Leiva and S. M. M. Martens, “MLSP competition, 2010: Description of the first place method,” *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 112–113, 2010.
[10] H. Cecotti, B. Rivet, M. Congedo, C. Jutten, O. Bertrand, E. Maby, and J. Mattout, “A robust sensor selection method for p300 brain-computer interfaces,” *Journal of Neural Engineering*, vol. 8, 2011.
[11] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, “xDAWN algorithm to enhance evoked potentials: application to brain-computer interface,” *IEEE Trans Biomed Eng.*, vol. 56, no. 8, pp. 2035–43, 2009.
[12] D. J. C. MacKay, “Bayesian interpolation,” *Neural Comput.*, vol. 4, no. 3, pp. 415–447, 1992.
[13] U. Hoffmann, J. Vesin, K. Diserens, and T. Ebrahimi, “An efficient P300-based brain-computer interface for disabled subjects,” *Journal of Neuroscience Methods*, vol. 167, no. 1, pp. 115–125, 2008.
[14] T. Fawcett, “An introduction to roc analysis,” *Pattern Recognition Letters*, vol. 27, pp. 861–874, 2006.
[15] C. J. Gonsalvez and J. Polich, “P300 amplitude is determined by target-to-target interval,” *Psychophysiology*, vol. 39, pp. 388–396, 2002.