Eye Fixation Related Potentials in a target search task

Graham Healy, Alan F. Smeaton

***Abstract*— Typically BCI (Brain Computer Interfaces) are found in rehabilitative or restorative applications, often allowing users a medium of communication that is otherwise unavailable through conventional means. Recently, however, there is growing interest in using BCI to assist users in searching for images. A class of neural signals often leveraged in common BCI paradigms are ERPs (Event Related Potentials), which are present in the EEG (Electroencephalograph) signals from users in response to various sensory events. One such ERP is the P300, and is typically elicited in an oddball experiment where a subject's attention is orientated towards a deviant stimulus among a stream of presented images. It has been shown that these types of neural responses can be used to drive an image search or labeling task, where we can rank images by examining the presence of such ERP signals in response to the display of images. To date, systems like these have been demonstrated when presenting sequences of images containing targets at up to 10Hz, however, the target images in these tasks do not necessitate any kind of eye movement for their detection because the targets in the images are quite salient. In this paper we analyse the presence of discriminating signals when they are offset to the time of eye fixations in a visual search task where detection of target images does require eye fixations.**

I. INTRODUCTION

Recently there is growing interest in using EEG (Electroencephalograph) signals to label images [1], [2]. By examining neural signals from users in response to presenting images to them, we can determine information about the images through the interpretation of them by the subject. The EEG signals that are generated in response to a stimulus such as image presentation are more commonly referred to as ERPs (Event Related Potentials) and are known to have idiosyncratic components reflecting attention-orientating events, such as a subject noticing a particular target within a stream of images. This phenomenon is more commonly known as P300 and has a well-established history of study [3]. The oddball paradigm is commonly used to elicit P300, where a subject is asked to count or respond when a particular stimulus appears on-screen. The idea is that the subject is unaware when the target stimuli will appear, thus the appearance generating this ERP component reflective of orientation of attention.

Using EEG signals to label or rank images is of practical interest as many types of images cannot be automatically labelled by a computer, and thus still require a human-in-the-loop. Examples of this include radiologists examining medical images or intelligence analysts viewing geo-spatial data. Aside from the demonstrated image sets used in previous studies that leverage this EEG annotation technique, a large open question remains as to what other application domains could this technique be used in. In previous work, the properties defining the target images to be detected did not necessitate any eye movement because the target is dominant within the image and so a RSVP (Rapid Serial Visual Presentation) paradigm can be used with speeds as high as 10Hz.

From vision science we know that some targets within image streams are not always salient, and can often require search involving eye movements [4]. Obvious examples of this include an airport security screener searching for a broad range of targets fixating on numerous locations within an x-ray image, or a radiologist similarly fixating on regions of a medical scan to search for abnormalities. These types of search activities require a number of eye fixations where each fixation within the image reveals information as to whether that region or the image is a target. Past work has assessed parameters of EEG-annotation techniques on image sets which have been displayed in fast RSVP paradigms. In our work we seek to examine if the general technique could be extended to determine whether individual fixations on an image could reveal target information. Thus, by performing such an action we would not only identify targets, but also their potential locations within images.

EEG signals extracted with regard to a fixation time are known as EFRPs (eye-fixation related potentials). Sajda [5] has explored a technique for detecting faces or people as targets and has confirmed that in such cases, pre-fixation differentiating EEG activity is present, showing that the user could see the target before fixation, and thus fixated to confirm. While visual search is often guided by cues, and target objects to be detected can be salient, we sought to determine the case where fixations do not occur as a confirmation of a target but are necessary in order to detect the target. In this regard we show that a different pattern of neural activation occurs when a subject searches for a target without prior knowledge pre-fixation as to whether it is a target or not.

In this paper we determine the capability to detect these ERP signals when they are offset from the time of a fixation. The idea is that fixations related to detecting targets should display a differentiated EEG signal. Fixations were detected by means of EOG (Electrooculogram) by examining the VEOG and HEOG channels. In section 2 we outline our experiments, the reasoning behind these, its parameters, and

This work is supported by Science Foundation Ireland under grant 07/CE/I1147.

Graham Healy and Alan Smeaton are with CLARITY: Centre for Sensor Web Technologies, School of Computing, Dublin City University, Glasnevin, Dublin 9, Ireland. ghealy@computing.dcu.ie

our experimental setup. In section 3 we present our results and in section 4 we present a conclusion and future work.

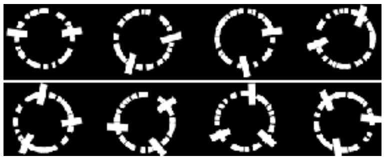

Fig. 1. The top row shows examples of the stimulus target objects used, and the bottom the non-target objects.

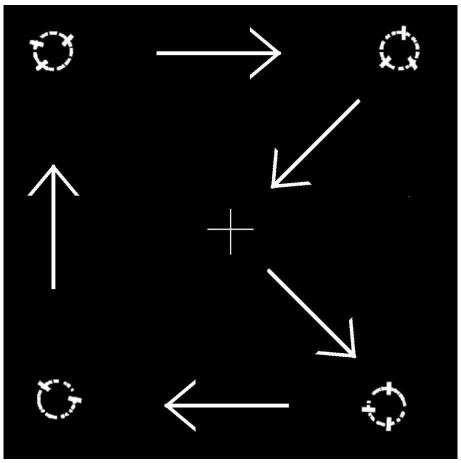

Fig. 2. An example of a frame. The bottom left corner contains a target stimulus object example. The object stimuli are increased 200% in size so that they can be seen more easily here.

II. Experiment

To address whether we can extract differentiated EEG activity related to target detection offset from the time of eye fixations we constructed an experiment whereby each subject was required to perform a search task on an LCD screen. In each of the 4 corners of the 24 inch (1680x1050) screen, a small stimulus was present which was either a target or a non-target object. The experiment was designed so that when the subject's gaze remains fixed on the central fixation cross (Fig. 2), they would remain unaware as to whether any of the objects are a target until the time of fixation. These same target and non-target object stimuli were confirmed not to be pre-attentively salient in a later experiment using the same subjects. By evaluating the reaction times for detecting the same objects in an array serial search task we found that an increase in distractor objects both increased the detection times and hindered the detection performance of a subject in an RSVP paradigm, thus confirming that these targets did not "stand out" and thus required serial search [4]. The target object to be detected and counted was a broken circle with 2 lines, while the non-targets were a broken circle with 3 lines. Examples of these are given in Fig. 1. By using such stimuli we were able to contain detection of the target item to the time of fixation. Subjects also confirmed whilst staring at the central fixation cross that they were unaware as to whether any of the corner objects were indeed targets.

The experiment was broken into 16 blocks, with each block containing 16 frames. Preceding each block, a search pattern was presented on-screen for 10 seconds to indicate the route to be followed to examine the objects for that block (shown by the arrows in Fig. 2). A white circle then appeared in the centre of the screen to indicate that a fixation cross would appear in 500 milliseconds after which the subject is expected to follow the given search pattern. At the end of a block, a subject then reports the total number of targets observed. Each frame was displayed for 2,500 milliseconds. Within that time, the subject was expected to view all 4 corner objects following the outlined pattern, and to then return their focus to the central fixation cross. This central fixation cross would then be replaced by the warning white circle where after 500 milliseconds the fixation cross would reappear, indicating the next frame was about to appear.

The search pattern within each block was kept consistent, but changed from block to block, hence displaying the search pattern at the beginning of the block. The arrows used to indicate the search pattern were superimposed over all frames for that block so that the subject would not forget the pattern. With A,B,C,D referencing each corner (see Fig. 2) on the screen (with E as the central fixation) we permuted this sequence to create 8 distinct search sequences, each consisting of 5 movements. 32 frames, each containing 4 corner stimuli, were then generated for that sequence. The probability of any one object stimulus being a target was kept to 0.125. Each of these 8 populated sequences were then cut in half to create the 16 blocks. In this way the target count per block would not be predicted. The order of these blocks for each subject was randomised.

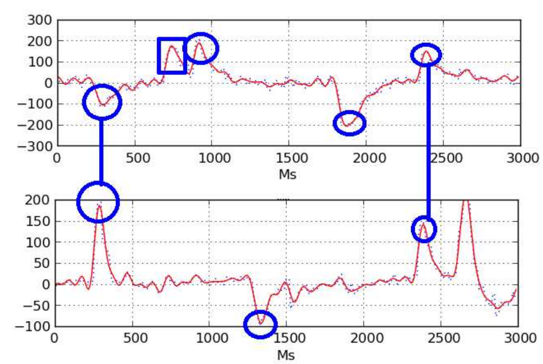

Fig. 3. EOG Channels: HEOG on top and VEOG on the bottom.

A. Data collection

For data recording, we used a KT88-1016 EEG system with a linked mastoid reference and the chin as ground. Ag/AgCl electrodes were used with a 10-20 placement cap at locations F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, Oz. Signals were digitized at 100Hz and subsequently band-passed from 0.1Hz to 20Hz. A 2 channel pendant EEG device was used to record EOG (VEOG and HEOG). Subjects were seated 1.2m away from the screen.

This meant each object stimulus was perceivable within .72 degrees.

An Intel Quad Core PC 2.4GHz with 3.2 gigabytes RAM and an Nvidia 8600GT graphics card was used for stimulus presentation and recording. All time stamping was carried out on this machine. With ethical approval granted to carry out these experiments from the university ethics board we recruited a total of 7 subjects from the postgraduate and staff population on campus. 4 males and 3 females were recruited with an average age of 27.3, and a standard deviation of 4.7. One of these was left handed.

B. EFRP Extraction

By using the EOG channels (VEOG and HEOG) we were able to find the time indexes of fixations on the object stimuli. Eye movements along one plane (i.e. horizontal) generate signals more prominently on one channel than the other, and the voltage deflections are sensitive to the direction of eye movement. Eye movements in any direction are typically characterised by either positive or negative voltage deflections on both channels. Since search patterns were consistent within blocks, the EOG patterns remained fairly consistent in that they always displayed a stereotyped sequence of deflections, other noisy EOG components were often present though. An example of a subject's eye movement search pattern for one such frame is shown in Fig. 3. With 8 basic eye movements used across the blocks, we could detect the fixations in the EOG signals using a simple scheme of matching these deflection patterns to the movement most likely to have generated them. Deflections present in the EOG signals not conforming to the stereotyped sequence for that block were discarded. In the case of two consecutive eye movements occurring in the same direction, the second peak was taken as the fixation upon the object (the first assumed to be upon the arrow). The time at which the EOG signal(s) peaked were taken as the index time from which to extract EEG activity. The peak times were detected by finding zero-crossings of the first derivative of the signal. To mitigate noise in the EOG signals, we disregarded eye movements where the combined absolute value of the peak height(s) fell below 2 standard deviations for that movement.

In an ideal circumstance we should have been able to extract 128 target fixations, and 896 non-target fixations in total for each subject. In practice, for each subject (1 to 7) respectively we extracted the following target/non-targets counts: 111/778, 107/768, 117/825, 109/772, 113/761, 118/838, 114/794.

Using these labeled time indexes of fixations, we extracted windows of the EEG signal starting post-fixation 0ms to 1000ms for each of the 16 channels. These were then concatenated to form a feature vector of length 640 which was then normalised to the range [-1,1]. No distinction was made to the eye movement associated with each target and non-target, only that that feature vector represented a target or non-target fixation.

III. Results

A. ERP Analysis

Both early visual EFRP and later discriminating components are visible in the grand average scalp plots shown in Fig. 4. The first notable component is the fixation lambda potential [6] (related to the visual P100) which peaks at occipital sites at 80ms (visible on the grand average of channel Oz in Fig. 5). At this time a negative component was also present at frontal sites which subsequently peaked around 120ms, where it then followed a wide spatial and temporal spread continuing to 200ms. Early frontal negativities have been shown to occur in combination with the lambda potential following this time-course [7], while the latter activity is consistent with the visual N1. A posterior negative component was seen across subjects typically peaking between 250ms and 350ms, and occurring later and more generally enhanced in amplitude for target objects across subjects. This activity is consistent with a posterior visual N2 in a feature discrimination task [8]. A positivity was seen far frontally between 280ms and 400ms peaking typically at 320ms for both object classes, and was diminished across users for targets. This dimished activity may be due to the an ealier counterpart anterior negativity related to posterior N2 activity observed for targets.

Differentiating activity between the detection of the target and non-target objects could be seen emerging at 250ms for most subjects, but prominent differences appear on the grand average scalp maps at 500ms with the presence of a widely distributed positive component present over occipital and parietal regions, which is consistent with P3b activity expected to occur with an oddball task such as this [3]. This posterior positivity began for most subjects at 460ms and continued on to 600ms. A frontal negativity emerged for subjects for the target objects at typically 600ms (starting as the p3b activity diminished) and continued for up to 1000ms typically diminishing with a parietal distribution. Previous work examining target detection in search tasks have shown a similar late occurring component with target detection [1]. This component may be reflective of a self-monitoring process.

B. Machine Learning Analysis

To examine and to derive a set of measures of the detectability of the EEG signal (P300) associated with the target fixations, we used a support vector machine (SVM) with radial basis function. Using 20-fold cross-validation for each subject, we randomly sampled a training set of 80 target and 80 non-target examples, and then used these to train an SVM model. An independent testing set of 27 target and 27 non-target examples were randomly sampled from the remaining feature vectors. The SVM models' gamma and cost parameters were found by using a gridsearch approach on the training data only. The test sets used to generate final results were always kept seperated from the training set. For each iteration of the cross-validation, an ROC (Receiver Operating Characteristic) curve was generated and its AUC

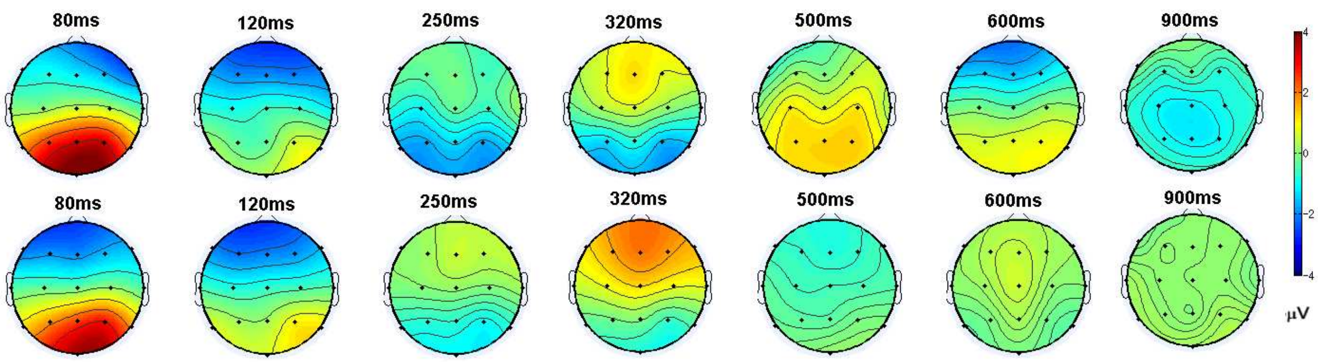

Fig. 4. Grand average scalp plots - target plots shown on top, non-target plots shown on bottom

(Area Under Curve) calculated. These AUC values were then averaged and are displayed in Table 1 for each subject. The AUC measure provides a ratio independent measure of the general discriminative capability of the constructed classifier. We also formed another 3 separate feature vectors, the first using only signals from anterior nodes (F7, F3, Fz, F4, F8, VEOG, HEOG), the second using posterior nodes only (T5, P3, Pz, Oz, P4, P6) and the third using signals from all 16 channels but only extracting 600ms pre-fixation. We wanted to confirm that the discriminative information learned by the classifier was not largely derived from the EOG activity alone (anterior sites), and that this activity only appeared after fixation. The AUC averages for these are displayed in Table I.

Using the full features from all channels we obtained an average AUC of .76 across subjects. Using only signals from the frontal nodes we still obtained an above-chance classification rate, however, this lowered rate confirms that a majority of the discriminative information learned by the classifier came from posterior nodes. This behavior fits with the typical scalp topography of the P3b. No discriminative information was learned in the EEG signals pre-fixation further confirming object detection was offset to the time of fixation.

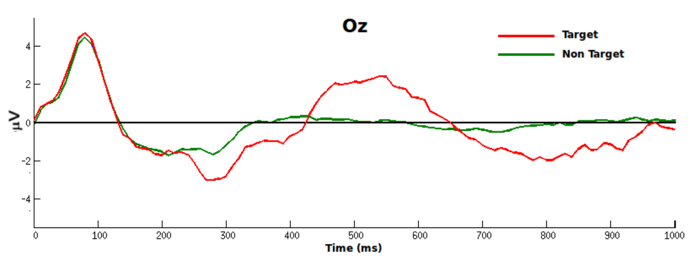

Fig. 5. Grand Average for site Oz

IV. Conclusions and Future Work

In this paper we have presented evidence that EFRPs can be extracted from an EEG signal (using EOG) that show differentiating activity related to target object detection. This contrasts and improves upon previous work highlighting that not all visual search tasks allow for a subject to be aware pre-fixation of whether an object/area is or contains a target [5]. Eye movements related to target search in real world tasks are often known to be guided by bottom-up features, global image properties, and factors such as expertise. Our future work will focus on evaluating the application of these EFRP signals in such real world search scenarios focusing on natural datasets.

TABLE I
AUC results from classifiers

Subject	AUC-All	AUC-Posterior	AUC-Anterior	AUC-PreFixation
1	.74	.67	.58	.49
2	.81	.73	.56	.51
3	.79	.73	.66	.51
4	.85	.75	.66	.50
5	.74	.66	.58	.51
6	.68	.68	.55	.48
7	.68	.61	.48	.52
Average	.76	.69	.58	.5

References

[1] P. Sajda, E. Pohlmeyer, J. Wang, L. Parra, C. Christoforou, J. Dmochowski, B. Hanna, C. Bahlmann, M. Singh, and S.-F. Chang, “In a blink of an eye and a switch of a transistor: Cortically coupled computer vision,” *Proceedings of the IEEE*, vol. 98, no. 3, pp. 462 –478, 2010.

[2] A. Gerson, L. Parra, and P. Sajda, “Cortically coupled computer vision for rapid image search,” *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 14, no. 2, pp. 174 –179, 2006.

[3] J. Polich, “Updating P300: an integrative theory of P3a and P3b.,” *Clinical neurophysiology*, vol. 118, pp. 2128–2148, Oct. 2007.

[4] J. M. Wolfe, “Guided Search 4.0: Current Progress with a model of visual search,” in *Integrated Models of Cognitive Systems* (W. Gray, ed.), pp. 99–119, New York: Oxford, 2007.

[5] A. Lou and P. Sajda, “Do we see before we look?,” in *Neural Engineering, 2009. NER ’09. 4th International IEEE/EMBS Conference on*, 29 2009.

[6] K. Kazai and A. Yagi, “Comparison between the lambda response of eye-fixation-related potentials and the p100 component of pattern-reversal visual evoked potentials,” *Cognitive, Affective,; Behavioral Neuroscience*, vol. 3, pp. 46–56, 2003. 10.3758/CABN.3.1.46.

[7] P. RM and T. BACCINO, “Eye fixation related potentials (efrps) during object identification,” *Visual Neuroscience*, vol. 27, no. 5-6, pp. 187–192, 2010.

[8] B. F. O’Donnell, J. M. Swearer, L. T. Smith, H. Hokama, and R. W. McCarley, “A topographic study of erps elicited by visual feature discrimination,” *Brain Topography*, vol. 10, pp. 133–143, 1997. 10.1023/A:1022203811678.