

Optimising the Number of Channels in EEG-Augmented Image Search

Graham Healy

CLARITY: Centre for Sensor Web Technologies
School of Computing
Dublin City University, Ireland
ghealy@computing.dcu.ie

Alan F. Smeaton

CLARITY: Centre for Sensor Web Technologies
School of Computing
Dublin City University, Ireland
Alan.Smeaton@dcu.ie

Recent proof-of-concept research has appeared showing the applicability of Brain Computer Interface (BCI) technology in combination with the human visual system, to classify images. The basic premise here is that images that arouse a participant's attention generate a detectable response in their brainwaves, measurable using an electroencephalograph (EEG). When a participant is given a target class of images to search for, each image belonging to that target class presented within a stream of images should elicit a distinctly detectable neural response. Previous work in this domain has primarily focused on validating the technique on proof of concept image sets that demonstrate desired properties and on examining the capabilities of the technique at various image presentation speeds. In this paper we expand on this by examining the capability of the technique when using a reduced number of channels in the EEG, and its impact on the detection accuracy.

Brain computer interface (BCI), image search, electroencephalograph (EEG), P300

1. INTRODUCTION

Traditionally the problems addressed by BCI (Brain Computer Interfaces) focused on the restoration of functionality and/or communication with people suffering from a variety of disorders such as ALS, stroke, and brain damage to name a few. There are many signals detectable from the brain, and many techniques for capturing these signals which can then be used to drive these systems. The systems tend to be classified based on their invasiveness from the less invasive being EEG/nIR/fMRI/MEG through to scenarios where recording sensors are placed directly on the surface of the brain to measure electrical signals (ECoG). These modalities of sensing provide different levels of functionality and benefit, whilst also bringing varying costs and dangers. Recently, however, it has come to light that some of the same fundamental principles employed to allow brain-computer communication can be used in a different application scenario to allow us to detect a users level of arousal or attentional orientation in response to viewing an image (Gerson, et al. (2006); Bigdely-Shamlo, et al. (2008); Huang, et al. (2008)). By detecting neural signals related to attentional orientation in response to particular images within a presentation stream, we can build a system to label or rank images directly,

driven by neural signals elicited from a participant by each of the images. Of primary interest to the BCI community is electroencephalograph (EEG) due to its relative low cost, availability, safety, and demonstrated applicability in regard to BCI applications. Due to its high level of temporal accuracy it allows for finer-grained analysis of a type of signal known as an ERP (Event Related Potential), which occurs in response to the presentation of an image to a participant. In this paper we present a brief overview of previous work in this area to equip the reader with an understanding of the fundamental techniques employed. Following this we provide a description and results of an experiment carried out utilizing EEG signals to drive an image search system. The primary contribution of our work in this paper is in demonstrating that similar or even better accuracy may be achieved using fewer EEG channels or nodes which makes for a less computationally demanding process, and which is ultimately more comforting for the participant involved.

2. BACKGROUND

Our work addresses the well-known problem of information overload, which is a fundamental challenge to search. In the case of searching image

data, the field of computer vision has developed feature sets which can be extracted from images and used to support content-based access to large image datasets. However, for high-level image interpretation, a human is required in the loop to assist the process of image interpretation, or perhaps is needed to entirely guide the process.

The idea of driving an image search system by a user's neural signals is relatively new, but the fundamental physiological phenomenon that this process can be based on has been known for a long time (Sutton, et al. (1965)). The seminal work in this area (Gerson, et al. (2006)) highlighted the applicability of the technique in triaging a sequence of images where a proportion of these contained a figure of a person in a forest vs. just a picture of trees. These target images were inserted into blocks of 100 non-target images where the block was then presented at a speed of 10 images per second (10Hz) to a participant connected to an EEG machine. Presenting images in this fashion is commonly known as RSVP (Rapid Serial Visual Presentation). What the authors found was that in response to target image presentations a temporally defined signal perturbation presented itself that was not present for the non-target images, i.e. the users involuntary attention was orientated toward the target image, and this could be seen from their brainwaves. Others have explored this technique on a variety of other datasets including satellite imagery with experienced intelligence analysts (Bigdely-Shamlo, et al. (2008); Huang, et al. (2008)). Other work (Shenoy, et al. (2008)) has examined the role of neural signals related to implicit processing where the user can be unaware of the task yet their brainwaves can still guide the process. One such example of these signals are those in response to faces. Signals of this type tend to be category-specific meaning that such an implicit analysis does not always extend. In this paper we are concerned with analysis of EEG signals captured at precisely the same time as a participant is involved in the explicit processing and detection of target images where the user is aware of the target be searched for.

3. EEG AND ERP

The EEG signal that previous authors have detected for explicit processing is more commonly known as a P300 (or an oddball response) and is one of the most studied EEG signals in regard to novelty and target detection within streams of images presented to participants. The signal occurs at or after 300 ms upon exposure to the (visual) stimulus, with its latency and amplitude modulated by factors such as task difficulty, saliency of the target, and probability

of the target. Being able to detect this signal in response to a specific image allows us to label or rank that image as having somehow stood out in comparison to the other images presented in the RSVP stream. This signal of interest is however often masked by the noise of other ongoing activity within the brain that also characterizes its presence by generating electrical activity detected on the surface of the scalp like the P300. The P300 signal comes in many forms (Polich (2007)) depending on the task, and attentional strategy of the participant. These subcomponents of the P300 manifest themselves with different temporal and scalp topographies (Makeig, et al. (2004)). Since these signals are inertly noisy and often partly concealed by ongoing unrelated neural activity, different techniques can be used to study them. One such technique is EEG signal averaging where a number of epochs (time regions following the presentation of a visual stimulus) are averaged together to produce a single waveform where activity unrelated to the stimulus should cancel out revealing a response related to the stimulus. By doing this we can study the neural responses and how they are differentiated by comparing EEG signal averages for target and non-target cases. An example of such an average is in Figure 1 where P300 activity can be seen most prominently at around 400 ms.

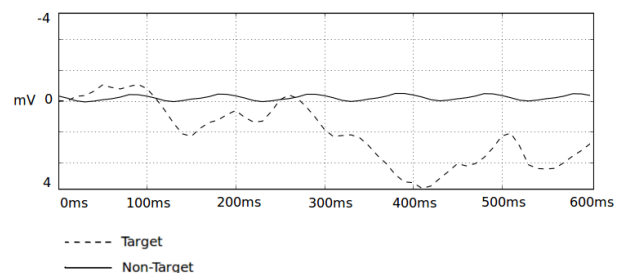


Figure 1: A typical ERP average showing a P3 component peaking at around 400ms

Recording EEG signals requires the placement of electrically conductive nodes directly upon the scalp of a participant using a conductive gel or paste. The placement of these nodes on different areas of the scalp corresponds to functionally distinct regions of the brain. Examining the spectral power of frequency for the channels for each of these nodes we can see distinct spectral characteristics that have been shown to correspond to various levels of attentional engagement in tasks, and factors such as level of arousal. These are examples of features that are extracted and examined without regard to temporal onset of sensory events such as images in an RSVP stream. Within these streams of EEG signals are perturbations related to specific cognitive and sensory events such as seeing images. Of particular interest to us are those related to sensory events,

whose timing and content can be controlled (a presented sequence of images where we know the presentation time of each image). These responses related to a sensory event are known as ERPs (Event Related Potentials). The P300 is one such class of ERP. Since numerous brain regions are involved in the production of this oddball P300 signal we aim to gain maximum coverage of the scalp hence we used the 10-20 system for node placement shown in Figure 2.

What is the focus of our work and what is addressed later in this paper is whether a similar level of accuracy can be achieved using a fewer number of nodes. In the experiments introduced in the next section of this paper we set out to see how such a reduction in number of channels affects accuracy.

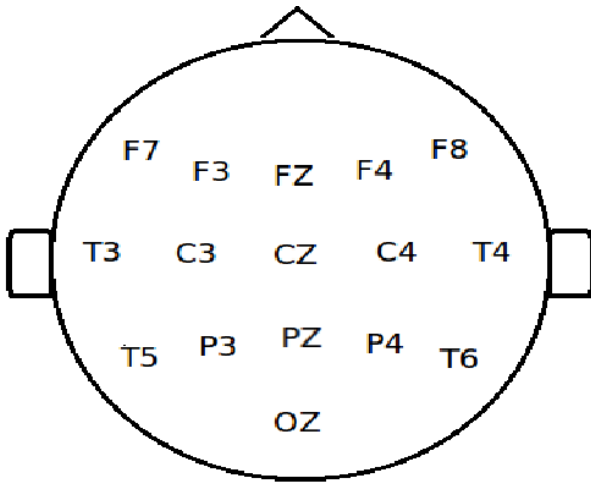


Figure 2: 10-20 Placement system for EEG nodes on the scalp

4. EXPERIMENT

4.1. Dataset

Images from the ALOI (Amsterdam Library of Object Images) were used (Geusebroek, et al. (2005)) in our experiments because they are well-known and have been used previously by others. This image set is comprised of 1,000 objects, each photographed from a number of camera angles and under a number of different lighting conditions. This image set was also chosen because it allowed use of a wide variety of non-target images which display visually salient and attentional arousing properties whilst allowing for a large number of different camera angles/lighting conditions for each object (i.e. our target object was represented by a large number of different images). Examples of some of these images are shown in Figure 3.



Figure 3: Example images from the ALOI dataset

4.2. Setup and Recording Procedure

For recording of EEG signals we used a KT88-1016 EEG system with a left 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 .1Hz to 20Hz. Stimulus presentation of images and recording of EEG data were carried out on the same computer to ensure time stamps could be matched between EEG data and the presentation times. The Curiosity Cloning Image viewer from the European Space Agency was used to present the images. A press button was also used for the user to signal target detection. This was placed on a table on the side of their dominant hand so that they could rest their arm and employ minimum physical effort in pressing the button. Button presses were recorded on the KT88 apparatus to allow for time-stamping of behavioural responses to the EEG data. 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. With ethical approval granted to carry out these experiments from the university ethics board we recruited a total of 8 participants from the postgraduate and staff population on campus. 5 males and 3 females were recruited with an average age of 27.5 years with standard deviation of 4.5 years.

5. EXPERIMENTAL PARAMETERS AND DESCRIPTION

Participants were shown a number of images of a target object that they were to search for prior to starting the experiment. Upon the appearance of this object the participant was instructed to press the button. In total 4800 images were shown to the user at a rate of 10Hz. Amongst these images 60 target images were randomly distributed accounting for 1.25%. The total duration of the task was 8 minutes. Four different targets were randomly selected from the ALOI dataset, with each target searched for by 2 users. Participants 1 & 5, 2 & 6, 3 & 7, 4 & 8 searched for ALOI targets 161, 455, 18 and 373 respectively. Each block sequence was constructed by randomly

sampling the pool of available target and non-targets. The images of the target object could be from any of a number of perspectives or lighting conditions, thus ensuring the actual target image would always be different. The start of the presentation of each block was signalled by a countdown.

6. RESULTS AND ANALYSIS

The purpose of EEG-augmented image search is to enhance the detection capabilities of a user searching for a target image within a large database. In this regard we evaluate in this section the increased accuracy achieved by using EEG in combination with behavioural responses (button press), and where trade-offs exist between the number of channels used. To examine the EEG signals and derive a set of measures of their detectability we used a SVM (Support vector machine) linear kernel (Chang, et al. (2001)). For each image in the stream we extracted the EEG signals from 16 channels for the 1 second following its presentation, sampled at 100 Hz. We also extracted an additional channel which recorded the button presses. We set out to examine the effects of a reduced number of channels on classification accuracy of the EEG signals and behavioural metrics. To achieve this we used a SFFS (Sequential Forward Feature Selection) scheme (Somol, et al. (1999)). This scheme finds subsets of features which offer optimal discriminative capacity between two classes, by starting with an empty set and adding the feature (or set of features) that provide the greatest increase in accuracy on each iteration. This algorithm for each forward iteration also evaluates back-steps by seeing if removing a feature (or set) offers an increase in accuracy. In this way local minima are avoided and optimal subsets are found by this floating search method.

Using this algorithm in combination with a linear kernel SVM we were able to find subsets of channels which offered optimal solutions. We did examine use of a SVM-RBF kernel with wide gridsearching for cost and gamma parameters, but this seemingly provided little gain at the cost of much increased running times of the SFFS algorithm. Using the SFFS algorithm with a linear svm we employed a cross validation approach where on each iteration a test set of 10/790 and 50/50 non overlapping target/non-targets were randomly selected from the available pool of samples. The training partition of 50/50 targets/non-target were fed into the SFFS algorithm that then evaluated subset combinations of channels. The SFFS algorithm evaluated channel subset combinations by further partitioning its training set into a test and training set of sizes 10/10 and 40/40 respectively. On each

iteration the SFFS algorithm produced a set of the channels for subset sizes 1 to 15 which represent the best found channel combination for that subset size. These subsets were evaluated on the initially removed test set of 10/790. The feature vector corresponding to a channel subset being evaluated was created by concatenating the EEG signal for those channels. Additionally a second feature vector was created using only the button press signal channel. SVM models using these two feature sets were trained on the training set of size 50/50, where an additional SVM model to fuse their outputs was created by using their predictions in a 10-fold cross validation on this set. These two models were then used on the originally removed testing set of size 10/790 to produce prediction values for EEG and button presses, where the third model was used to combine the predictions. These predictions were then evaluated using an accuracy measurement function ($P@n$) for each of the 15 channel subsets.

Table 1: Detailed results per participant

Participant	Max $P@n$ (#n of channels)	Button	% Increase
1	.4 (3)	.31	28.6%
2	.67(3)	.511	30.1%
3	.6(3)	.375	62.6%
4	.76 (5)	.466	63.1%
5	.62 (6)	.315	96.8%
6	.41(5)	.328	25%
7	.4(3)	.287	41.4%
8	.44 (5)	.247	76.2%

We repeated this 20 times, and averaged the $P@n$ accuracies as identified by their channel subset size (i.e. 20 accuracy values for channel subsets of size 4 were averaged to give an accuracy value for 4 channels). This scheme of keeping independent testing sets was necessary to ensure that subset solutions found by the algorithm were not simply biased by random relationships in the training data which did not generalise to the rest of the data. By keeping a test set of size 10/790 separate from the beginning on each iteration, we can ensure the models applied and evaluated are in no way biased. Precision@n is the fraction of true positives within the first n elements of an ordered list. We set n at 10 since our test set contained 10 targets, as this reflects the target to non-target ratio of the pool data collected (10/790 to 60/4740). Shown in Figure 4 are the results of this for all 8 Participants. As we can see the inclusion of the EEG signals improved accuracy over just using behavioural metrics (x-axis value=0) for all users. We can also see that the inclusion of additional EEG channels in some cases can actually reduce detection performance (i.e. participant 3) albeit not very much. This may be due to that

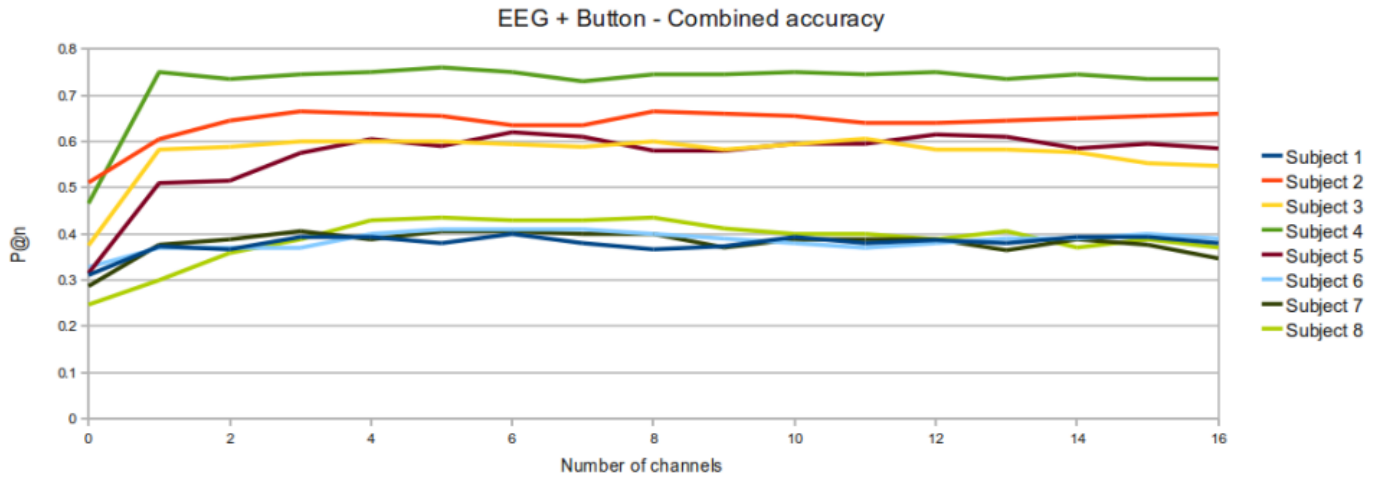


Figure 4: Combined $P@n$ accuracy of EEG + Button

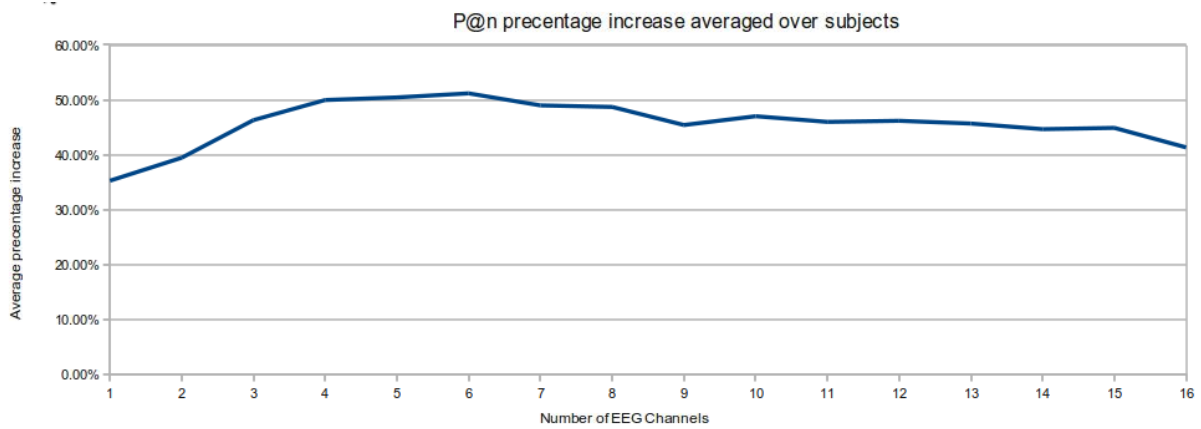


Figure 5: Average percentage increase of $P@n$ across users against number of channels

fact additional channels do not provide any further discriminate information, and only serve to introduce noise. Examining the button press channel following target presentations it was found that some users failed on occasion to respond within one second (i.e missed the target). Participants 7 and 3 missed 9 and 3 targets respectively, with participants 3 and 6 missing 2. This may explain the lowered accuracy in some cases.

Table 1 summaries some of this detail from Figure 4 for each participant. In column 1 (c1) we show the maximum $P@n$ achieved along with the associated number of channels. In column 2 (c2) we show the $P@n$ achieved using only button presses (x-axis value = 0). In column 3 we show the percentage increase calculated by $(c1 - c2) / c2 * 100$. The average increase by including EEG data was 52.8% that of using only the button press.

Of interest to us in this paper is examining the effects that a greater/fewer number of EEG channels has on performance of signal detection. In Figure 5 we

show the average increase across the set of 8 users achieved by adding an additional channel. We can see that by using 4 channels of EEG we achieve near 50% of an increase compared to only using button press responses. The optimum seems to be indicated at 6 channels with a 51.17% increase, but this negligible gain if statistically significant hardly seems worth introducing 2 more channels for.

For each iteration of the SFFS algorithm we kept a score for how many times each channel was selected to be included when the channel set being evaluated was of size 4. A collapsed list of these channel counts across all participants revealed that the most frequently chosen channels for inclusion resided on the posterior points of the scalp, which is consistent with the discriminating activity typically produced by a P3b ERP (Polich (2007)) in response to target detection. The channel counts across users were ranked as follows: Oz,P3,Pz,Cz,C3,T6,T5,Fz,C4,T4,F8

These results show that EEG does provide an increase in accuracy when combined with the button press, and this increase can be realised using a subset of the available channels.

7. CONCLUSIONS AND FUTURE WORK

In this paper we have shown that EEG signals can be used to augment the image search process by fusing them with behavioural responses and using a reduced number of EEG channels. This is of significance as there is a growing availability of cheap consumer-grade EEG hardware. With an activity as pervasive as image search, there is much scope to evaluate the types of image search tasks which may benefit most from including EEG data.

ACKNOWLEDGEMENT

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

8. REFERENCES

- Bigdely-Shamlo, N. and Vankov, A. and Ramirez, R.R. and Makeig, S. (2008) Brain Activity-Based Image Classification From Rapid Serial Visual Presentation. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 16, 432-441.
- Gerson, A.D., Parra, L.C. and Sajda, P. (2006) Brain Activity-Based Image Classification From Rapid Serial Visual Presentation. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 14, 174-179.
- Geusebroek, J. M. and Burghouts, G. J. and Smeulders, A. W. M. (2005) The Amsterdam Library of Object Images. *International Journal of Computer Vision*, 61, 103-112.
- Huang, Y. and Erdogmus, D. and Mathan, S. and Pavel, M. (2008) Large-scale image database triage via EEG evoked responses. *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, Nevada - Las Vegas, 4th April 2008, 429-432.
- Makeig, S. and Delorme, A. and Westerfield, M. and Jung, T. and Townsend, J. and Courchesne, E. and Sejnowski, T. J. (2004) Electroencephalographic Brain Dynamics Following Manually Responded Visual Targets. *PLoS Biol*, 2.
- Polich, J. (2007) Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118, 2128-2148.
- Shenoy, P. and Tan, D. S. (2008) Human-aided computing: utilizing implicit human processing to classify images *SIGCHI conference on Human factors in computing systems*, Florence, Italy, 5th April 2008, 845-854. ACM, New York, NY, USA.
- Sutton, S. and Braren, M. and Zubin, J. and John, E. R. (1965) Evoked-Potential Correlates of Stimulus Uncertainty. *Science*, 150, 1187-1188.
- Somol, P. and Pudil, P. and Novovičová, J. and Paclík, P. (1999) Adaptive floating search methods in feature selection. *Pattern Recogn. Lett.*, 20, 1157-1163.
- Chang C and Lin C. (2001) LIBSVM: a library for support vector machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>