Collaborative Brain-Computer Interfaces for the Automatic Classification of Images

Ana Matran-Fernandez, Riccardo Poli and Caterina Cinel

***Abstract*— In this paper, we propose a collaborative brain-computer interface for the automatic discrimination of images containing specific targets. When a user looks at a stream of images that are displayed using the rapid serial visual presentation protocol, images containing targets elicit a P300 event-related potential that can be detected. This allows the images to be automatically labelled as targets and non-targets. While this is relatively unreliable with single users, by combining the evidence from groups of users, we obtained relative median improvements in the accuracy (assessed by the area under the receiver operating characteristic curve) of 8.2% and 14.3% for groups of 2 and 3 users, respectively.**

I. INTRODUCTION

A Brain-Computer Interface (BCI) [1], [2] converts signals generated by the brain into commands that serve as input to another device, e.g., a computer or a prosthesis. In the most widely used type of noninvasive BCIs, electric activity from the brain of the user is recorded by means of an electroencephalograph (EEG) which converts it into digital signals. The BCI processes these in order to infer the desired command before sending it to the external device.

One of the most common ways of controlling a noninvasive BCI is based on the oddball paradigm [3], where stimuli are sequentially presented to a user. Most of these stimuli are not of interest for the observer (non-targets or distractors). However, there is a second type of stimuli (targets) that appear much less frequently in the sequence and present some distinctive feature. Their occurrence elicits Event-Related Potentials (ERP). One of these ERPs is the P300, a positive peak that appears about 300 ms after the occurrence of the target and can be detected in the EEG recordings.

BCIs have been used for the automatic detection of targets in images by means of the EEG with reasonably good results [4], [5], [6]. Combining the oddball paradigm with Rapid Serial Visual Presentation (RSVP) techniques, a P300 wave appears in the presence of targets that allows for the labelling of the images of interest.

To date, computer vision systems have not been able to outperform humans in visual search and recognition tasks. Such systems do not perform well in situations in which there is no specific description of what type of object has to be found [7]. For example, in broad-area search, intelligence analysts look for things that fall in the "threats" category. However, the same object can mean very different things depending on its context, and computers do not have the necessary flexibility to reproduce human behaviour in this respect. All these factors limit greatly the extent to which automatization of the search task is feasible. Hence, this area of research is relevant given the large amounts of images that need to be processed, which may also be time-sensitive, depending on the field of application.

A. Collaborative Brain-Computer Interfaces

Collaborative BCIs work by merging the EEG signals (or the corresponding control commands) from several users with the aim of controlling a single device [8]. Unlike traditional BCIs, these are conceived as devices that augment the capabilities of able-bodied users [8], [9].

Amongst other applications, this type of BCIs can be used for group decision making [10]. P300-based BCIs typically record the neural response over multiple trials and then average the EEG epochs in order to decide the most probable target. However, in a context where this is not possible (a person cannot make the same decision several times), aggregating the signals from a number of users has proven to be useful. More specifically, in [10] a collaborative BCI integrated the activity of up to 7 subjects for decision making based on a visual perception task. Errors were reduced from approximately 11.5% of single-user decisions to around 4% for groups of 6 or 7 users.

In [11], we reported on preliminary work on the collaborative classification of broad-area aerial images. By using pairs of observers (from a pool of five), we were able to speed up the process of revising the images and we obtained noteworthy higher accuracies than with single observers. This paper extends that work to 10 participants and also to a larger number of collaborating users. It also augments it with a statistical analysis.

B. Rapid Serial Visual Presentation

RSVP is a presentation technique in which items are sequentially displayed at a very fast rate over the same spatial location. Originally, this technique was designed for the study of reading and language processing by means of serial presentation of words or groups of words forming sentences [12]. Images were introduced in RSVP in order to study memory and retention [13].

A phenomenon widely reported and studied in the RSVP literature is the attentional blink (AB), where the second of two targets fails to be reported by the subject if it is

*This work partially supported by the UK's Engineering and Physical Sciences Research Council (grant EP/K004638/1).

The authors are with the Brain Computer Interfaces Laboratory, School of Computer Science and Electronic Engineering, University of Essex, {amatra,rpoli,ccinel}@essex.ac.uk

The authors would like to thank Dr Adrian Stoica of NASA JPL for his input on preliminary versions of this work.

presented shortly after the first one [14]. The attentional blink (i.e., the masking of the second target) varies in latency and duration depending on the type of stimuli that are being presented [15].

When considering the use of RSVP in test experiments and real world applications, it is important to take into account the effect of the AB. In broad-area search, the occurrence of targets may be very low, but consecutive target images could appear if targets are present in neighbouring regions (in this case shuffling the images could solve the problem). To avoid the issue, in [4] a maximum of 1 target in every 51-picture stream was allowed, whereas Healy *et al* [5] fixed the proportion of targets vs. distractors at 10%.

C. Collaborative Classification of Images

Organisations such as NASA and ESA have an interest in using BCIs for the automatic classification of images by intelligence analysts [4], [5]. The potential advantage of a BCI approach to this problem is that users could sift through images at high speed and would not need to stop the stream to indicate the presence of a potential target image. However, the results published so far, although promising, still leave much room for improvement.

Here we propose to combine the RSVP technique with a *collaborative* BCI approach to improve the accuracy of the system by averaging classifiers' outputs from several users without reducing its throughput.

In [11], we implemented this idea in a collaborative BCI based on RSVP and the oddball paradigm with the aim of classifying aerial images without any manual input from the users. We used different presentation rates [5] and asked our five participants to look for different targets in different experiments. The results were analyzed for pairs of users, and we determined that by pairing subjects we could increase the accuracy of the method.

Yuan *et al* [9] also performed a simple visual task on groups of 3 people using visual evoked potentials. Their experiment consisted on a fixation cross presented for a random time and followed by a synthetic visual pattern. They reported that visual pattern detection was "signaled" faster through a BCI than through manual input from the users. In addition, detection was faster with collaborative BCI than with single-user BCI. They also reported an average increase of 11% in the accuracy of groups vs. individuals.

There have also been attempts to classify medical images, in particular mammograms [16].

As we indicated above, in this paper, we have applied the concept of collaborative BCIs to the classification of images, extending on our previous work [11] by increasing the number of participants and group members and including a statistical analysis of the results.

II. METHODOLOGY

A. Data Acquisition

We gathered data from 10 volunteers (mean age 27.3; 3 females) with normal or corrected-to-normal vision. They all signed the consent form approved by the Ethics Committee of the University of Essex and reported no family history of epilepsy. One of the subjects was discarded for the analysis due to extremely poor performance during the training part of the experiment.

Participants were seated at approximately 80 cm from an LCD screen where the stimuli were presented. EEG data were acquired by using a BioSemi ActiveTwo system with 64 electrodes mounted in a standard electrode cap following the international 10-20 system, including electrodes on the earlobes of the subjects (impedance <20 kΩ). The EEG was referenced to the mean of the electrodes placed on the earlobes. The initial sampling rate was 2048 Hz. Data were low-pass filtered with a cutoff frequency of 16 Hz before down-sampling to 32 Hz.

B. Stimuli

The images shown to the volunteers consisted of 1,200 aerial pictures of London, divided up into 100-image bursts. These were converted to grayscale and their histograms were equalised. These pictures were used to create training and test datasets for the system, keeping the ratio of target vs. non-target images to 10%. Target pictures contained the image of an artificially inserted plane that could be rotated in any angle and placed anywhere in the picture. Target pictures were randomly placed within a burst with the only restriction that there had to be at least one non-target image between two targets. Fig. 1 shows examples of target and non-target images.

We generated two pairs of training and test sets: one that always contained the same plane in target pictures and another in which the planes were randomly chosen (with replacement) from a set of three with different shapes and sizes.

Participants were presented with bursts of 7 levels of difficulty. The parameters for the levels of difficulty are given in Table I. Each level consisted of a training and a testing phase, both of which required the presentation of the 1,200 images. Thus, within each phase (training or testing), within each level, we presented a total of 12 bursts to each participant.

Subjects were instructed to try to reduce eye blinks and movements in order to obtain EEG signals with as few artifacts as possible. They were asked to mentally count the number of planes they saw within each burst to ensure

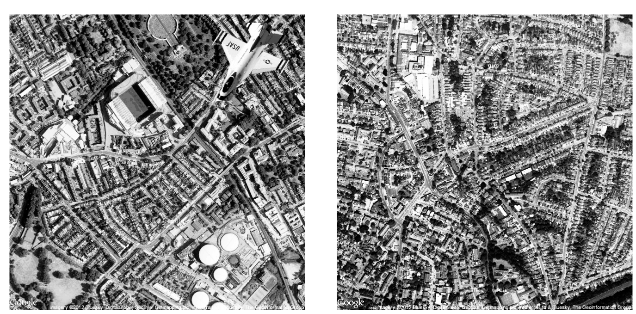

Fig. 1. Examples of target (a) and non-target (b) images used in our experiments.

they were focused on the task. After each burst, subjects reported how many planes they had identified. They could rest between bursts and decide when to move on with the experiment. Eeach participant completed the tasks in 90 minutes or less.

C. Signal Processing

Following the onset of each picture on the screen, an 800 ms epoch of the signal in each EEG channel was extracted (26 samples per channel) and concatenated to form a feature vector of $64 \times 26 = 1664$ elements. Naturally, with the RSVP technique there is significant overlap between the epochs (and ERPs) for different images, especially at high presentation rates, so one epoch contains data from more than one target image. However, we chose to keep epochs this long because late components of the P300 can give useful information for automatic classification [5]. After this pre-processing stage, we performed the training and testing of the classifiers as described below.

1) Individual classification: For each level of difficulty, we used the 1,200 epochs collected in the training phase to train an ensemble of two hard-margin linear Support Vector Machines (SVMs) for every participant. We labelled the correct outputs as $+1$ for the epochs corresponding to a target image and -1 for non-targets. Rather than using binary classification, the output of the SVMs was used as a score for the labelling of an independent set of 1,200 test images: the higher the output, the more likely a picture contained a target. This method is labelled as SBCI (Single BCI) in figures.

2) Collaborative classification: We used two methods to merge the signals from multiple participants. In an approach called Single Classifier Collaborative BCI (SC-cBCI), we averaged the raw signals across users (thus creating 1,200 "compound" EEG epochs) and trained a single classifier for every group. In the Multiple Classifier Collaborative BCI (MC-cBCI), instead, we averaged the outputs of the individually tailored classifiers (see Sec. II-C.1) across images for groups of 2 and 3 subjects. Due to the high number of permutations of size 3 that can be done with 10 subjects, we used a representative random sample of groups for the analysis of the performance (labelled as 3MC-cBCI in the figures).

III. RESULTS

We used the Area Under the Curve (AUC) of the receiver operating characteristic as a measure of the performance of different BCI systems. This is a well-established measure of the performance of a classifier in machine learning. The AUC for a perfect classifier (no false alarms, no false negatives) is 1, whereas its value for a random classifier is 0.5.

A. Average Performances

To see if there is an advantage in a collaborative approach to classification, we first need to assess the performance of a single-user BCI. So, we calculated the median values of the AUC across all individuals for every difficulty level.

In the case of the groups, we analysed all the possible combinations of pairs of subjects with the SC-cBCI and the MC-cBCI approaches. For the groups of 3 we analysed the performance of 50 random groups with the 3MC-cBCI method. Fig. 2 shows a summary of the performance for each difficulty level and for the different methods used.

B. Statistical Significance

Fig. 3 shows the p-values returned by the one-tailed Kolmogorov-Smirnov test when comparing the different methods based on the AUC scores. Here, we compared every collaborative approach vs. individuals, as well as the 3MC-cBCI vs the best of the two methods used for pairs of subjects (MC-cBCI).

IV. DISCUSSION

A. Single and Pairs of Users

The average performance of pairs of users is higher than that of individuals for the two methods of creating the collaborative BCI. The relative improvements in the median AUC values in the range 3.7%–8.3% in the SC-cBCI method and 3.5%–8.2% for the MC-cBCI.

Moreover, we found that performance of the collaborative BCI does not improve over the individual BCI in two cases: (1) when the users score very differently, in which case the pairs obtain a performance that is nearer the better AUC-scorer of the two, but below this value; and (2) when subjects have a similar but low score individually, in which case the overall performance when using the SC-cBCI method may actually be worse than the worse of the individuals. Most probably this happens because low scorers report to have seen very few targets, and it is unlikely that the same target pictures will elicit a P300 in both. So, there is no advantage in merging the signals together.

As can be seen in Fig. 3, even though pairing has a positive effect on the performance, this is not at a statistically significant level when compared to single-user BCIs, except for level 7.

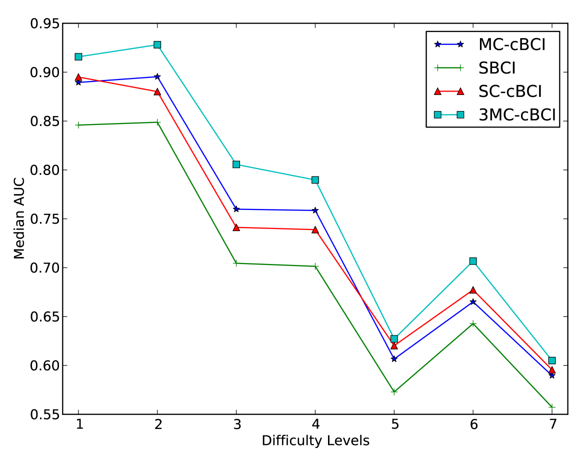

Fig. 2. Median values of the AUC across subjects and groups for all methods over the different difficulty levels.

TABLE I

PARAMETERS OF THE DIFFERENT LEVELS OF THE EXPERIMENT.

Feature	*Level 1*	*Level 2*	*Level 3*	*Level 4*	*Level 5*	*Level 6*	*Level 7*
Presentation rate (pictures/sec)	5	6	6	10	10	12	15
Number of different targets	1	1	3	1	3	1	1
Burst duration (sec)	20	16.67	16.67	10	10	8.33	6.67

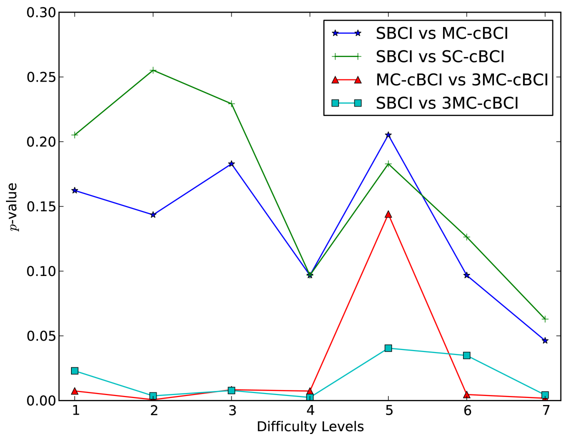

Fig. 3. Kolmogorov-Smirnov p-values for comparisons based on AUC values.

B. Groups of Three Users

With grouping of 3 users we registered relative improvements in the range 8.3%–14.3% in the median AUC over individual performance. Hence, increasing the size of the groups had a positive effect in the accuracy of the classification of images. It can be seen from Fig. 2 that the median value of the AUC for the 3MC-cBCI method is higher than any other.

As can be seen in Fig. 3, the 3MC-cBCI method is significantly superior to the SBCI case at a 5% confidence level in every case. Also, except for level 5, it is superior to the better of the two methods based on pairs (MC-cBCI).

V. CONCLUSIONS

We proposed a collaborative BCI for the high-speed hands-free discrimination of images containing specific targets.

We tested two alternative ways of combining brain signals from multiple users: directly merging their EEG into a single signal and then classifying it or using separate classifiers for different users, but then integrating the outputs of such classifiers. We found that the latter works better than the former in most cases.

We applied the approach to groups of 2 and 3 subjects, finding that in all cases median performance of our collaborative BCIs is markedly superior to single-user one (although we did not reach statistical significance in all cases).

In the future we intend to explore alternative ways of combining the evidence from multiple users, and we will test the applicability of our approach to video surveillance.

REFERENCES

[1] B. Z. Allison, E. W. Wolpaw, and J. R. Wolpaw, "Brain-computer interface systems: progress and prospects." *Expert Review of Medical Devices*, vol. 4, no. 4, pp. 463–474, Jul 2007.

[2] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–91, June 2002.

[3] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, Dec 1988.

[4] A. Kruse and S. Makeig, "Phase I Analysis Report for UCSD / SoCal NIA Team Acknowledgments," Institute for Neural Computation, University of California San Diego, La Jolla, Tech. Rep. January, 2007.

[5] G. Healy, P. Wilkins, A. F. Smeaton, D. Izzo, M. Rucinski, C. Ampatzis, and E. M. Moraud, "Curiosity Cloning: Neural Modelling for Image Analysis, Technical Report," Dublin City University; European Space and Technology Research Center (ESTEC), Dublin, Ireland; The Netherlands, Tech. Rep. 0, 2010.

[6] A. D. Gerson, L. C. Parra, and P. Sajda, "Cortically coupled computer vision for rapid image search." *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 14, no. 2, pp. 174–9, June 2006.

[7] D. J. Huber, D. Khosla, K. Martin, and Y. Chen, "A low-bandwidth graphical user interface for high-speed triage of potential items of interest in video imagery," pp. 873 608–873 608–8, 2013.

[8] Y. Wang and T.-P. Jung, "A Collaborative Brain-Computer Interface for Improving Human Performance," *PLoS ONE*, vol. 6, no. 5, May 2011.

[9] P. Yuan, Y. Wang, W. Wu, H. Xu, X. Gao, S. Gao, *et al.*, "Study on an online collaborative bci to accelerate response to visual targets." in *Conference proceedings:... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, vol. 2012, 2012, pp. 1736–1739.

[10] R. Poli, C. Cinel, F. Sepulveda, and A. Stoica, "Improving decision-making based on visual perception via a collaborative brain-computer interface," in *IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*. San Diego (CA): IEEE, February 2013.

[11] A. Stoica, A. Matran-Fernandez, D. Andreou, R. Poli, C. Cinel, Y. Iwashita, and C. Padgett, "Multi-brain fusion and applications to intelligence analysis," pp. 87 560N–87 560N–8, 2013.

[12] K. I. Forster, "Visual perception of rapidly presented word sequences of varying complexity," *Perception & Psychophysics*, vol. 8, no. 4, pp. 215–221, 1970.

[13] M. Potter and E. Levy, "Recognition memory for a rapid sequence of pictures," *Journal of experimental psychology*, vol. 81, no. 1, pp. 10–15, 1969.

[14] M. Chun and M. Potter, "A two-stage model for multiple target detection in rapid serial visual presentation," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 21, no. 1, pp. 109–127, 1995.

[15] W. Einhäuser, C. Koch, and S. Makeig, "The duration of the attentional blink in natural scenes depends on stimulus category," *Vision research*, vol. 47, no. 5, p. 597, 2007.

[16] C. Hope, A. Sterr, P. Elangovan, N. Geades, D. Windridge, K. Young, and K. Wells, "High throughput screening for mammography using a human-computer interface with rapid serial visual presentation (rsvp)," in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2013, pp. 867 303–867 303.