Multi-brain fusion and applications to intelligence analysis

A. Stoica[a], A. Matran-Fernandez[b], D. Andreou[b], R. Poli[b], C. Cinel[b], Y. Iwashita[a] and C. Padgett[a]

[a]NASA Jet Propulsion Laboratory, Pasadena, CA, USA;
[b]BCI Lab, School of Computer Science and Electronic Engineering, University of Essex, UK

ABSTRACT

In a rapid serial visual presentation (RSVP) images are shown at an extremely rapid pace. Yet, the images can still be parsed by the visual system to some extent. In fact, the detection of specific targets in a stream of pictures triggers a characteristic electroencephalography (EEG) response that can be recognized by a brain-computer interface (BCI) and exploited for automatic target detection. Research funded by DARPA's Neurotechnology for Intelligence Analysts program has achieved speed-ups in sifting through satellite images when adopting this approach. This paper extends the use of BCI technology from individual analysts to collaborative BCIs. We show that the integration of information in EEGs collected from multiple operators results in performance improvements compared to the single-operator case.

Keywords: Brain-Computer Interfaces, collaborative BCI, Rapid Serial Visual Presentation, visual search

1. INTRODUCTION

1.1 Background

When performing broad-area search, intelligence analysts look for targets that, although well-defined for a human, are too generic to be recognized by a computer vision system. The task can thus be automatised to a very limited extent. Yet, manually analysing thousands of aerial or satellite images is extremely tiring and time-consuming. So, there is a need for software to help intelligence analysts perform this task.

Research funded by the DARPA's Neurotechnology for Intelligence Analysts program[1] showed that by using brain signals a good classification rate could be achieved. Moreover, they were able to speed up this task by using Rapid Serial Visual Presentation (RSVP) techniques. This is possible because the detection of specific targets in a stream of pictures triggers a characteristic EEG response which can be recognized by a brain-computer interface (BCI) and exploited for automatic target detection (i.e., without the operator having to stop the presentation to explicitly signal the presence of a target). In the system presented in[1] , a series of image bursts were rapidly classified to highlight regions of interest within a broader area, so that the analysts could later focus on those regions and pay less attention to areas where no interesting events had been detected.

The work in[1] uses the so-called "oddball" paradigm. In this paradigm, sequences of stimuli are presented to an observer where most stimuli, called non targets, are similar to each other and present no interest for the observer. However, other stimuli with some distinctive feature appear in the sequence. These are called targets and occur much less frequently. Their presentation triggers a specific neural response (oddball response). Similar research in automatic classification of images based on the oddball paradigm has been performed by other research groups as well[2–4] .

As we will show later, our system exploits the oddball paradigm as well, but it does so in a collaborative BCI.

Further author information: (Send correspondence to A. Matran-Fernandez)
A. Stoica: E-mail: adrian.stoica@jpl.nasa.gov, Telephone: +1 818 354 2190
A. Matran-Fernandez: E-mail: amatra@essex.ac.uk
D. Andreou: E-mail: dandreb@essex.ac.uk
R. Poli: E-mail: rpoli@essex.ac.uk, Telephone: +44 (0) 1206 872338
C. Cinel: E-mail: ccinel@essex.ac.uk, Telephone: +44 (0)1206 873291
Y. Iwashita: E-mail: Yumi.Iwashita@jpl.nasa.gov

Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2013,
Proc. of SPIE Vol. 8756, 87560N · © 2013 SPIE · CCC code: 0277-786X/13/$18
doi: 10.1117/12.2016456

1.2 Brain-Computer Interfaces

A Brain-Computer Interface (BCI) converts signals generated by the brain into commands that serve as input to another device, e.g. a computer or a prosthesis. In a typical BCI an electroencephalography (EEG) system records the electric activity from the scalp of the subject and, after some type of signal processing and classification, the command is sent to the external device.

One of the most common ways of controlling a BCI is based on the use of the oddball paradigm. When an interesting target event is detected by the user, an Event-Related Potential (ERP) is elicited and can be seen in his/her brainwaves. One of these ERPs is the P300, a positive peak that appears about 300 ms after the occurrence of the target. One of the advantages of using P300-based BCIs is that P300s occur spontaneously in response to specific events and, therefore, users do not need long training sessions to learn how to control them. This is in contrast to other BCIs which rely on the ability of the user to modulate and control particular brain features (for example slow brain potentials). Learning to do this can require extensive training.

One of the problems with P300-based BCIs is that, because of the typical EEG signal noise, the detection by the system of a P300 is based on the average of many signal repetitions of the same target events. Therefore, to avoid low information transfer rates, target/non-target sequences have to be presented at a very fast pace. As we will show later, the RSVP technique can take advantage of this fact.

BCIs have been typically developed as a form of assistive technology for people with severe disabilities. However, new advances in EEG technology have given rise to BCI systems to be used by able-bodied users as an extension of their capabilities, e.g. as a new input channel. A new field that is benefitting from these changes is that of collaborative BCIs, in which several users aim to control the same device at the same time. In order to achieve this, the EEG signals from two or more users are merged or decoded together, so the final command is direved from their collective intentions, rather than from a single user.

The aggregation of the signals in collaborative BCIs can be done in different ways. The simplest method would be to just average the raw EEG signals across several users prior to their classification. In this way, a unique classifier is used for all the subjects and there is a reduction in the noise level. However, this method may not be the most accurate, since the EEG signals from different subjects may have different latencies with respect to the particular event that they are tied to. A second possible option is classifying the signals individually (training one classifier per subject) and implementing the merging step afterwards. In this case, each classifier will be taylored to an individual user (as is typical in single-user BCIs), so the overall performance can be increased if an appropriate merging system (e.g. a voting system) is implemented.

Even though the field of collaborative BCIs is a new area of research, there are some promising results in the literature. Wang and Jung[5] showed that multiple users are able to generate a movement command to a prosthetic limb faster than single users. Poli et al.[6] performed an offline analysis merging signals from pairs of users to control a mouse on a screen by means of a BCI, reporting straighter trajectories than those achieved individually.

Collaborative BCIs are also found in the context of group decision making. Due to the noise present in EEG signals, in a context where it is not possible to average them over multiple trials (a person cannot be asked to make the same decision multiple times), aggregating the signals from several users in order to achieve a better outcome whilst reducing the level of noise has proven useful[7] .

1.3 Rapid Serial Visual Presentation

As discussed above, in P300-based paradigms the target/non-target sequences need to be presented at a fast rate in order to achieve a satisfactory information transfer rate. This is the reason why it can be an advantage combining oddball paradigms with RSVP paradigms.

In RSVP items are sequentially shown at a very fast rate and same spatial location on a screen. 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[8] .

Some tasks within such experiments consisted of matching a target named in advance (e.g. deciding whether a given letter or word had been shown in a stream of letters or words)[9] . This could be viewed as the first form of target detection within the RSVP paradigm. In some cases, words inside a sentence were replaced by images[10] with the purpose of studying semantic processing.

Potter and Levy[11] also studied recognition memory for pictures at presentation rates ranging between 2 seconds/picture and 0.125 seconds/picture. After a short film showing the target pictures, a new film in which distracters had been inserted was shown, and subjects had to verbally state whether a given picture was a target or not. They also studied eye movements (i.e. saccades) during the presentation of the images. For films lasting between 2 and 4 seconds, eye movements were almost completely suppressed at rates greater than 4 pictures/second, whereas saccades where more common for lower presentation rates. This is relevant in the field of BCIs, because EEG systems are very sensitive to noise produced by eye movements.

A phenomenon reported in the RSVP literature is that of the attentional blink, where a target is missed if resented within a certain time interval after the previous target in the same stream of stimuli[12] . The attentional blink (i.e. the masking of the second target) varies in its latency and duration depending on the type of stimuli that are being presented. This effect has to be taken into account when designing an RSVP experiment. For example, in[1] , in order to avoid attentional blinks, a maximum of 1 target every 51-picture stream was allowed. Furthermore, the targets were placed within the central part of the image array to avoid cross over effects (they reported that targets could not be detected in the last two images due to the "shock" produced by the end of sequence). On the other hand, Healy et al.[2] constructed a more generic environment in which the proportion of targets vs. non-targets was fixed at 10% and their results were not bound by these limitations, which we consider to be more appropriate. Furthermore, this 10% is in accordance with the oddball paradigm.

1.4 RSVP and BCIs

As it was pointed before, the idea of using the oddball paradigm in conjunction with the RSVP technique is not new. By inserting a few target images amongst a large amount of distracters, or non targets, one can expect a P300 wave to be produced when the events of interest are presented within a stream of pictures. It has been shown that images can therefore be classified as targets or distracters using only the user's EEG signals corresponding to each type of event[1,2,4] .

The results published so far, although promising, still leave room for improvement. By showing the same pictures to several analysts, for example, the level of uncertainty, when classifying images, can be reduced, and higher accuracies can be obtained. This suggests that we can implement a collaborative BCI system to try and improve the performance of the existing systems.

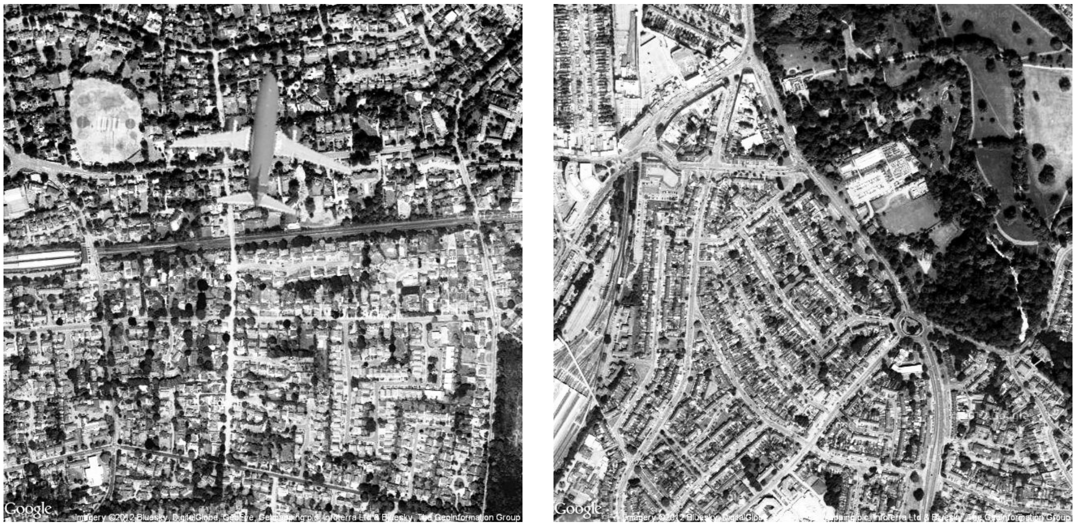
In this paper, we implemented a collaborative BCI based on RSVP/oddball paradigm with the aim of classifying images without any input from the users other than their EEG signals. We used different presentation rates, since it has been reported that every person has a different performance for different speeds[2] . The results were analysed for pairs of users.

2. METHODS

2.1 Data Acquisition

We gathered data from 5 participants (1 female, mean age 29.6 years). All had normal or corrected-to-normal vision and reported no history of epilepsy. Volunteers were seated at approximately 80 cm from an LCD screen.

EEG data were acquired by using a BioSemi ActiveTwo system using 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). EEG was referenced to the mean of the electrodes placed on the earlobes. The initial sampling rate was 2048 Hz. Data were then low-pass filtered with a cutoff frequency of 16 Hz before downsampling to 32 Hz.



(a) (b)

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

2.2 Stimuli

In order to construct the stimuli, we first captured a total of 1,200 620×620 pictures of London taken directly from Google Maps. The pictures were converted to grayscale and then their histogram was equalised to reduce the possibility of producing spurious “oddballs” (e.g., due to high luminance or contrast effects).

These pictures were used to create training and test datasets for the system. These were build in such a way that the proportion of target vs non-target images was 10%. Targets pictures were those containing a plane which we artificially inserted in them by using image processing tools. The plane could be rotated in any angle and placed anywhere within a picture (with the condition that the whole plane would be contained in it). Figure 1 shows examples of target and non-target images. Sequences of target and non-target pictures were generated; we will call them “bursts” hereafter. 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.

We generated two pairs of training and test sets. In the first pair, we always used the same plane in target pictures. However, to make the task more realistic, we also created a second pair in which the planes were randomly chosen (with replacement) from a set of three. These had different shapes and sizes.

With these two setups, we created an experiment in which participants were presented with bursts of 7 levels of difficulty. Difficulty was varied by varying both the speed of presentation (5–15 pictures per second) and the number of different planes that could be seen in a burst (either 1 or 3). We arranged the levels in increasing difficulty order (approximately) and presented them in the same order to every participant. The parameters for each level are given in Table 1.

Each level of difficulty consisted of a training and a testing phase. Each required the presentation of 1,200 images, divided up into 100-image bursts. Thus, within each phase (training or testing), within each level, we presented a total of 12 bursts to a participant.

Subjects were instructed to try to reduce eye blinks and movements during the bursts 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 they were indeed focused on the task. After each burst, subjects were asked to verbally report how many planes they had identified. Subjects could rest between burst and they themselves decided when to move on with the experiment. After preparation, each participant completed the experiment in 90 minutes or less.

Table 1. 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

2.3 Signal Processing for Image Classification

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). The data from each electrode were concatenated to form a feature vector of $64 \times 26 = 1,664$ elements. Naturally, with RSVP there is significant overlap between the epochs for different images. However, late components of the P300 can give useful information for automatic classification,[2] which is why we use 800 ms epochs.

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 individual user. We labelled the correct outputs as $+1$ for the epochs corresponding to a target image and -1 for non-target ones. After training, the output of the SVMs was used as a score for the classification of the test images: the higher the output, the more likely a picture contained a target.

2.4 Collaborative BCI

Since all the subjects had been exposed to the same pictures in the same sequence for all the training and test data, after the experiments we could simulate (offline) the conditions of pairs of subjects performing the task simultaneously together with a collaborative BCI. As in our previous work,[6] we used two methods to merge the EEG signals of pairs of users: (1) averaging the raw signals and training a single classifier per pair of users, and (2) averaging the output of two classifiers, each trained (and used) on the data from one user. In both cases, the pre-processing techniques were the same as for the single-user case.

In BCIs that operate based on the oddball paradigm, such as Donchin's matrix speller,[13] every stimulus is typically flashed several times before making a decision on what the user's intention is. By averaging the epochs relative to each stimulus, a reduction of the noise level is achieved and the ERP appears clearer, thus achieving higher system accuracy. By adopting the first approach to collaborative BCI, a similar reduction of noise is obtained by averaging the signals across multiple users instead of across multiple trials from the same user. However, since users tend to have different ERP waveforms, this method could also reduce the amplitude of the P300s and make matters worse.

The second approach to collaborative BCI consists of creating specific classifiers for each user (one ensemble of 2 SVMs per user) and average the outputs of such classifiers. As it was said before, these outputs are not binary, but proportional to the certainty that the system has of the presence of a P300. Therefore, the resulting system can be seen as a kind of voting system that gives more weight to the more certain user.

3. RESULTS

3.1 Single User Performance

Let us begin with the performance of our system when the signals from a single user are used for the classification of images into targets and non-targets. We used the Area Under the Receiver Operating Characteristic Curve (AUC) as a measure of the performance of the machine learning component. The value of the AUC of a perfect classifier is 1, whereas its value for a random classifier is 0.5.

Table 2 reports the values of the AUC for every subject and level of difficulty. These are consistent with previous published results: the user performance decreases with the presentation rate and in almost every case the best classifiers are associated with the first (easiest) difficulty level. Another expected result are the big drops in AUC when moving from only one type of plane in target images to three types of planes. Less expected was the fact that even at the very high rates of presentation used in the last two levels of difficulty (12 and 15 Hz, respectively), the average AUC value across all subjects is nowhere near 0.5 (the performance of a random

Table 2. AUC values across subjects for every difficulty level in our single-user experiments.

Subject	*Level 1*	*Level 2*	*Level 3*	*Level 4*	*Level 5*	*Level 6*	*Level 7*
1	0.937	0.912	0.809	0.734	0.605	0.673	0.595
2	0.865	0.863	0.746	0.705	0.633	0.620	0.54
3	0.939	0.942	0.835	0.815	0.708	0.789	0.592
4	0.827	0.786	0.578	0.65	0.509	0.53	0.546
5	0.672	0.634	0.567	0.612	0.522	0.526	0.512

Table 3. AUC values for coupled users (averaging the raw EEG signals) for every level.

Subject	*Level 1*	*Level 2*	*Level 3*	*Level 4*	*Level 5*	*Level 6*	*Level 7*
1 & 2	0.913	0.923	0.739	0.781	0.673	0.718	0.598
1 & 3	0.955	0.960	0.892	0.807	0.685	0.812	0.628
1 & 4	0.934	0.908	0.741	0.759	0.576	0.719	0.608
1 & 5	0.910	0.880	0.797	0.728	0.594	0.669	0.597
2 & 3	0.924	0.927	0.720	0.831	0.696	0.774	0.655
2 & 4	0.869	0.869	0.646	0.736	0.595	0.609	0.574
2 & 5	0.806	0.788	0.646	0.739	0.620	0.610	0.543
3 & 4	0.933	0.921	0.766	0.796	0.657	0.765	0.648
3 & 5	0.906	0.894	0.776	0.795	0.663	0.748	0.600
4 & 5	0.848	0.754	0.603	0.645	0.560	0.560	0.540

classifier), indicating that the subject's visual system was still able to discriminate, at least to some degree, targets from non-targets.

Subjects 4 and 5 show worse performances than average. In particular, subject 4 was over-zealous: at the end of the experiment the subject reported that he had blinked to signal each target (even though he had explicitly been told to try to avoid blinks). Due to the artifact rejection that we have in place in the system, the classifiers for this subject could not be properly trained. Subject 5, instead, was drowsy: the subject regularly reported a number of targets seen in each burst much smaller than the real value. Also, when we plotted the grand averages of the signals for targets and non-targets for this subject, we found that there were almost no differences between them.

In normal conditions, BCI researchers would exclude subjects such as 4 and 5 as they didn't behave as was required by the experiment. However, we will not do this here, as this actually servers the purpose of illustrating the benefits of a collaborative BCI.

3.2 Collaborative Classification

As mentioned before, we tested two methods of aggregation in our collaborative BCI system. Table 3 reports the results for the first method (averaging the raw EEG signals and using a joint SVM classifier) while the second method's results (a voting system based on the certainty of the individual classifiers of the presence of a target within an image) are reported on Table 4.

Figure 2 compares these methods against single-user performance by plotting the average value of the AUC for each method and level. The average performance of individuals is worse than the collaborative case. It can also be seen from this figure that, for the easier levels, the 2-SVM collaborative BCI outperforms the 1-SVM collaborative method, whereas they cross over for higher difficulty levels.

4. DISCUSSION AND CONCLUSIONS

In this paper, we have studied the possibility of using a BCI for the automatic classification of images and the improvements when the observations are averaged across several users.

One of the difficulties present in this work is the environment in which the signals were collected. Whereas most research groups that work with EEG signals record them in rooms that are insolated from electromagnetic noise, we carried out our experiments in a normal room where the volunteer could be subject to disruptions and general noise at any point. In addition, there was no isolation from electromagnetic noise. Still, for low speeds

Table 4. AUC values for coupled users (averaging the outputs of the individual classifiers) for every level.

Subject	*Level 1*	*Level 2*	*Level 3*	*Level 4*	*Level 5*	*Level 6*	*Level 7*
1 & 2	0.934	0.944	0.835	0.780	0.655	0.690	0.590
1 & 3	0.965	0.971	0.900	0.830	0.706	0.806	0.617
1 & 4	0.935	0.915	0.764	0.751	0.585	0.640	0.590
1 & 5	0.903	0.885	0.749	0.728	0.599	0.632	0.568
2 & 3	0.935	0.953	0.845	0.823	0.712	0.776	0.598
2 & 4	0.895	0.895	0.714	0.734	0.605	0.608	0.560
2 & 5	0.866	0.845	0.711	0.738	0.597	0.601	0.548
3 & 4	0.933	0.924	0.783	0.804	0.643	0.745	0.606
3 & 5	0.903	0.900	0.794	0.792	0.653	0.739	0.581
4 & 5	0.836	0.784	0.596	0.674	0.533	0.544	0.555

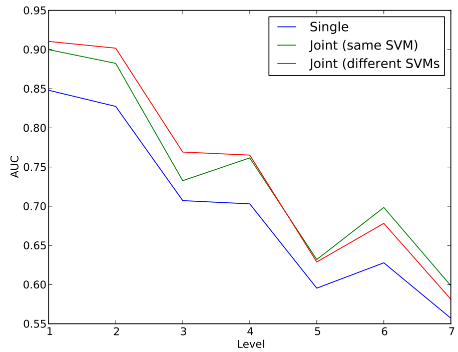


Figure 2. Average AUC values plotted for every level and method.

of presentation, we achieved an acceptable classification rate (for a presentation rate of 10 images/second, we obtained an average AUC of 70%).

In the collaborative case, we merged signals from pairs of users using two different methods. In both cases, the performance of the classifiers was better than in the individual case. We believe that these results can be further improved if we take into consideration the level at which every subject performs best and use those signals to create the collaborative BCI.

It has been reported that the bigger the group, the higher the accuracy of the BCI. In this study we only averaged over pairs of subjects. Given the reduced number of subjects, we thought that bigger groups would not be representative or easy to generalize, since we would not even be able to form two separate groups of 3 people.

Hence, one of the future challenges is to form bigger groups to classify images.

There is certainly an interest in the use of the RSVP technique to classify images without the user having to manually select the interesting ones. We have applied it to the field of target detection in broad area search. However, there are other fields where this method could be appreciated. For instance, there is a vast amount of medical images that have to be seen by skilled clinicians on a daily basis for clinical purposes. In this sense, Hope and collaborators[14] applied the concept to the screening of mammographies by experts. In medical imaging, the collaborative BCI would be of great help, since a diagnosis can be subjective to a professional. By averaging across multiple experts, part of this bias could potentially be eliminated.

ACKNOWLEDGMENTS

The research was jointly carried out in the School of Computer Science and Electronic Engineering of the University of Essex and at Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. The authors would like to thank the UK's Engineering and Physical Sciences Research Council (EPSRC) for financially supporting this research (grant EP/K004638/1).

REFERENCES

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

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

[3] Bernat, E., Shevrin, H., and Snodgrass, M., "Subliminal visual oddball stimuli evoke a P300 component.," *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology* **112**, 159–71 (Jan. 2001).

[4] Gerson, A. D., Parra, L. C., and Sajda, P., "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* **14**, 174–9 (June 2006).

[5] Wang, Y. and Jung, T.-P., "A Collaborative Brain-Computer Interface for Improving Human Performance," *PLoS ONE* **6** (May 2011).

[6] Poli, R., Cinel, C., Matran-Fernandez, A., Sepulveda, F., and Stoica, A., "Towards cooperative brain-computer interfaces for space navigation," in [*Proceedings of the International Conference on Intelligent User Interfaces (IUI)*], (19–22 March 2013).

[7] Poli, R., Cinel, C., Sepulveda, F., and Stoica, A., "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)*], IEEE, San Diego (CA) (February 2013).

[8] Forster, K. I., "Visual perception of rapidly presented word sequences of varying complexity," *Perception & Psychophysics* **8**(4), 215–221 (1970).

[9] Petrick, M. S., *Acoustic and semantic encoding during rapid reading*, PhD thesis, Massachusetts Institute of Technology, Department of Psychology (1981).

[10] Potter, M., "Rapid serial visual presentation (RSVP): A method for studying language processing," in [*New methods in reading comprehension Research*], ch. 5, 91–118 (1984).

[11] Potter, M. and Levy, E., "Recognition memory for a rapid sequence of pictures," *Journal of experimental psychology* **81**(1), 10–15 (1969).

[12] Chun, M. and Potter, M., "A two-stage model for multiple target detection in rapid serial visual presentation," *Journal of Experimental Psychology: Human Perception and Performance* **21**(1), 109–127 (1995).

[13] Farwell, L. A. and Donchin, E., "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and Clinical Neurophysiology* **70**, 510–523 (Dec 1988).

[14] Hope, C., Sterr, A., Elangovan, P., Geades, N., Windridge, D., Young, K., and Wells, K., "High throughput screening for mammography using a human-computer interface with rapid serial visual presentation (rsvp)," 867303–867303–8 (2013).