2011 NDIA GROUND VEHICLE SYSTEMS ENGINEERING AND TECHNOLOGY SYMPOSIUM
MODELING & SIMULATION, TESTING AND VALIDATION (MSTV) MINI-SYMPOSIUM
AUGUST 9-11 DEARBORN, MICHIGAN

LEVERAGING BRAIN COMPUTER INTERACTION TECHNOLOGIES FOR MILITARY APPLICATIONS

Anthony J. Ries & Brent Lance
U.S. Army Research Laboratory
Human Research and Engineering Directorate
Aberdeen Proving Ground, MD

Paul Sajda
Department of Biomedical Engineering
Columbia University
New York, NY

ABSTRACT

Recent advances in neuroscience, signal processing, machine learning, and related technologies have made it possible to reliably detect brain signatures specific to visual target recognition in real time. Utilizing these technologies together has shown an increase in the speed and accuracy of visual target identification over traditional visual scanning techniques. Images containing a target of interest elicit a unique neural signature in the brain (e.g. P300 event-related potential) when detected by the human observer. Computer vision exploits the P300-based signal to identify specific features in the target image that are different from other non-target images. Coupling the brain and computer in this way along with using rapid serial visual presentation (RSVP) of the images enables large image datasets to be accurately interrogated in a short amount of time. Together this technology allows for potential military applications ranging from image triaging for the image analyst to target geo-tagging for ground troops.

INTRODUCTION

One area of research that may potentially lead to disruptive innovations is that of Brain-Computer Interaction Technologies (BCIT). BCITs use non-invasively measured neural data, such as electroencephalography (EEG), in combination with other physiological and behavioral measures to enhance joint human-system performance for healthy individuals [1]. Recent advances in neuroscience, signal processing and machine learning have made it possible to exploit and integrate visual, motor, and cognitive processing of the brain with advanced computational algorithms of the computer [2]. There are many potential Army-relevant tasks that may benefit from these types of BCITs, including processing large quantities of image data.

Advancements in sensor technology and digital storage continue to worsen the mismatch between our ability to collect and store data on the one hand and process and analyze it on the other. Large amounts of data present high computational demands on the limited processing capacities of both humans and computers, resulting in information overload and slow and often inaccurate performance. As the quantity of information increases, so does the challenge of sifting through large amounts of data in order to quickly and accurately detect potential targets of interest.

Throughput for image analysis is confined by processing limits of both the human brain and computer algorithms. Manually searching through large images or datasets is time consuming and impractical, especially when decisions about potential targets must be made quickly, whether searching for targets in satellite imagery or looking for enemy vehicles in video taken from an unmanned aerial vehicle (UAV). Given the critical nature of identifying targets in military operations, automated capabilities for processing image data are still limited.

The human and computer each have their limitations in processing visual imagery; however, their strengths may be combined to overcome these shortcomings. Humans are capable of rapidly making semantic distinctions based on visual input, while computers are capable of quickly searching large image databases. In this paper we describe a method for developing mutual human-machine systems that utilize the capabilities of both the human and computer systems to improve capabilities for analyzing image data by using brain-based signals related to target recognition such as the P300 event-related potential (ERP) and techniques such as displaying image data in rapid serial visual presentation.

NEURAL SIGNATURES OF TARGET DETECTION

A neural signal commonly used in EEG-based BCIT applications to identify when an observer detects a target of interest is the P300, or P3 event-related potential (ERP); e.g. [3-5]. An ERP reflects averaged EEG epochs time-locked to a specific stimulus and is used to improve the signal to noise ratio of the EEG signal [6]. The P3 is a large positive deflection in the ERP waveform occurring roughly 300ms after stimulus onset and is maximal over frontal/central and central/parietal electrodes.

Typically both P3 amplitude and latency are measured, with amplitude assessed by comparing the pre-stimulus baseline amplitude to the largest positive peak of the ERP within a variable time window post-stimulus, and latency measured as the point of maximum positive peak amplitude within a time window beginning at the onset of stimulus presentation [7]. While the P3 can be clearly seen in an averaged ERP, it can also be measured in single trials, (as shown by Makeig et al. [8]; and for simultaneous EEG/fMRI by Goldman et al. [9]).

The P3 has most commonly been evaluated using the 'oddball' paradigm where an infrequent target stimulus is presented within a series of frequently occurring non-target distractor stimuli [10]. In a two-stimulus oddball paradigm an infrequent "oddball" target stimulus is presented with many frequent distractor stimuli where only the infrequent oddball stimulus requires a response. The time between stimulus presentation in a series ranges between 1-3 seconds and the probability of an oddball stimulus appearing is generally between 5-20%. The infrequent target evokes a P3 component, specifically the P3b, which reflects the detection of a task-relevant stimulus [7], [11]. The P3 is also obtained when an observer views a highly familiar image or one that is currently held in memory [12], [13]. Overall, P3-related activity reflects neural process associated with attentional orienting toward a task-relevant stimulus in the sensory information stream.

RAPID SERIAL VISUAL PRESENTATION (RSVP)

Given the speed and precision of the human visual system to enable subtle semantic distinctions, and distinguish between relevant and non-relevant images, information throughput can be expanded by presenting stimuli at a much faster rate than is encountered in traditional self-paced searches. This is accomplished by using a rapid serial visual presentation (RSVP) paradigm in which a series of images are rapidly (e.g. 10Hz) and sequentially presented in the same spatial location to an observer who is searching for a predefined target image or class of images appearing in a stream of non-target distractor images (Figure 1). Essentially, an RSVP task containing one predefined target is a speeded version of the two-stimulus oddball task such that a predefined low probability target is presented among many high probability distractors in a rapid sequence.

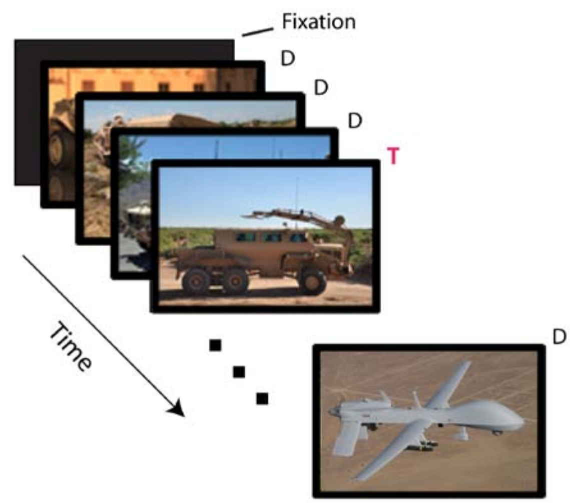

Figure 1: Example RSVP sequence. Sequence begins with a fixation screen followed by rapid (e.g. 100ms per image) and sequential image presentation. D = Distractor, T = Target. Photos courtesy of U.S. Army

By using the RSVP paradigm in conjunction with the P3-based ERP brain signal, it is possible for a human to quickly identify target stimuli, and for that identification to be detected by an external system (e.g., a BCIT).

RSVP-BASED BCITs

The human visual system can quickly categorize a predefined, semantically distinct target or class of targets from other non-target distractors in as little as 100ms [14]. This rapidity of image processing can be leveraged by using an RSVP or alternate image display paradigm. Detecting the resulting neural response induced by target detection makes it possible to develop BCIT systems for analyzing large quantities of image data. Several systems have been developed based on this principle [15], [16].

The system we will focus on in this paper is the Cortically-Coupled Computer Vision system, or C3V system [17-19]. The C3V system integrates the RSVP paradigm with P3 ERP detection and a computer vision algorithm in three different methods. The three methods are "computer first," where the computer vision algorithm is used to triage images in order to improve human performance at target detection, "human first," where target detection data collected from a human performing the P3 RSVP task are used to improve the performance of the computer vision algorithm, and a "tightly-coupled" method, where the P3 RSVP task and the computer vision run in an iterative fashion to improve joint performance at target detection.

An appealing feature of this approach is that the algorithms are trained as general interest or relevance detectors. There is evidence to suggest that even though an algorithm may be configured when an operator is searching for one class of targets, the same algorithm may be as efficient when searching for an entirely different class of targets [13]. This enables operators to process different classes of image data without having to retrain the C3V algorithms.

C3V: Computer First Method

One method for applying C3V technology uses computer vision followed by EEG and RSVP [18]. This occurs when the viewer has prior information about target features. The target features are then used to train a computer vision system as an image pre-processor. The image pre-processor then is then used to extract likely target images from the overall data set, and places potential regions of interest (ROI) in the center of the image. The images are then presented using RSVP while EEG is recorded. The EEG signals for each image are recorded and used to create an index of target relevance. A classifier is applied to discriminate between the target and distractor images, and those images that produce the highest target relevance score are placed at the front of an image triage for in-depth manual review.

C3V: Human First Method

The second method for applying C3 Vision is to first perform the RSVP task and use the resulting information to train the computer vision system [20]. One benefit of this method is that it does not require prior knowledge of specific target characteristics. To use this method, EEG signals are recorded during RSVP and images are ranked based on the viewer's perceived target relevance as derived from the P3-related neural activity. These image ranking metrics are then used to apply labels to the images. The computer vision algorithms then use the image ranking metrics obtained during RSVP and EEG acquisition to identify other images in the database with similar low-level visual features and propagate the labels to those images.

C3V: Tightly-Coupled Method

The final method for applying this technology is the "tightly-coupled" method. There are two potential implementations of this method, each with its own implications for interaction and application. In the first implementation, RSVP is performed and EEG collected while computer vision simultaneously analyzes the same images, leading to a combined target relevance measure for each image. Because this implementation executes in near-real time, it may be possible for it to run interactively, for example improving a Soldier's Local Situational Awareness by processing recent image data of nearby areas.

The tightly-coupled method can also be implemented as an iterative process, where the outputs from the RSVP and EEG analysis are used to provide training and labeling data to the computer vision system in order to improve its performance. Then, the computer vision system is used to triage the image database, providing more relevant images to the RSVP participant [19]. This iterative process, on the other hand, is more relevant for the offline improvement of target recognition algorithms such as image analysis for battlespace preparation before a mission. Either way, by using these tightly-coupled methods, the joint target detection capability of the human-machine system can be improved.

ARMY-RELEVANT APPLICATIONS OF RSVP-BASED BCITS

Searching Satellite Imagery Data for Intelligence Information

Figure 2: A. Aerial image to be searched. B. Highlighted – "mowing the lawn" visual scan-path. C. Highlighted – image chips showing the highest probability of target presence based on C3 Vision.

There are many potential Army-relevant applications of RSVP and EEG-based target detection systems such as the C3V system. One area where these technologies have been successfully applied is in the domain of intelligence analysis for satellite imagery, where the goal is to find specific targets in large satellite images, as shown in Figure 2a [2], [18]. The current method for analyzing this image data is to use manual search patterns, such as the "mowing the lawn" pattern seen in Figure 2b. However, the computer-first C3V method has been successfully applied to this problem. To do this, the large satellite image is first broken up into smaller image "chips," which are processed by the C3V computer vision algorithm. The goal of the computer vision is to locate potentially relevant chips, and change the focus of the chip so that potential ROI appear in the center of the image chip.

The image chips are shown to an intelligence analyst using an RSVP paradigm. The P3 signals extracted from the analyst's EEG data then provide an index of target relevance for each chip. Those image chips with high target relevance can then be highlighted in the original large image (See Figure 2c) for additional inspection. This process enables the entire large image to be interrogated quickly and accurately using foveal/parafoveal vision and few eye movements. The performance of the C3V system has been shown to enhance target detection performance. Parra and colleagues [2] compared the speed and accuracy over time of detecting helipads in large overhead imagery between baseline and C3 Vision approaches using trained image analysts (IAs).

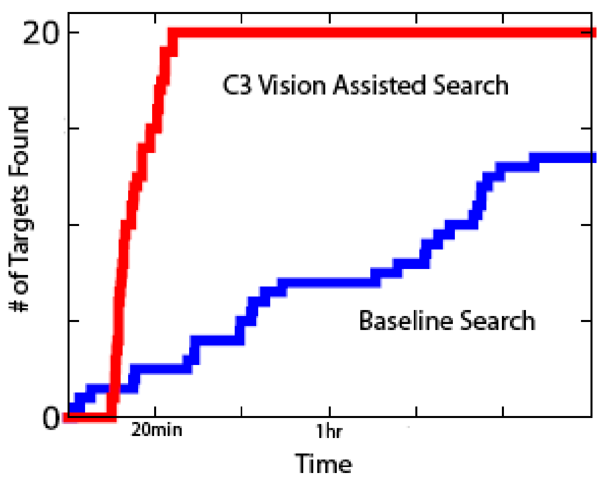

Figure 3: Number of targets identified over time in the baseline search (blue line) and C3 vision based search (red line). Figure adapted from [2].

The IAs searched the imagery at full resolution in the baseline condition zooming in and out using a keyboard and marking each target helipad with a mouse click. In the C3V condition, IAs viewed image chips using RSVP and each chip was assigned a probability of target presence based on EEG metrics. Those images with the highest probabilities were moved to the front of an image triage for additional inspection. Figure 3 shows the results from six subjects searching two different satellite images, demonstrating that targets are identified more quickly using the EEG-based C3V-prioritized search when compared to baseline self-paced search.

Ground Vehicle Map Propagation

Another potential translational application of C3 Vision technology to Army operations is to use its speed and adaptability to improve battlespace intelligence preparation by populating the common operating picture (COP) on a vehicle crewstation prior to a mission. Video and image data obtained from forward scouts, satellites, or UAVs of the to-be-entered battle space could be processed using C3 Vision virtually anywhere prior to mission execution. This would allow large areas of interest to be processed for potential targets at a rapid pace. The image data can be processed online in real-time or offline at a later point. Importantly those images that generate the highest target relevance score based on the C3 Vision metrics could be geo-tagged using latitude, longitude and elevation.

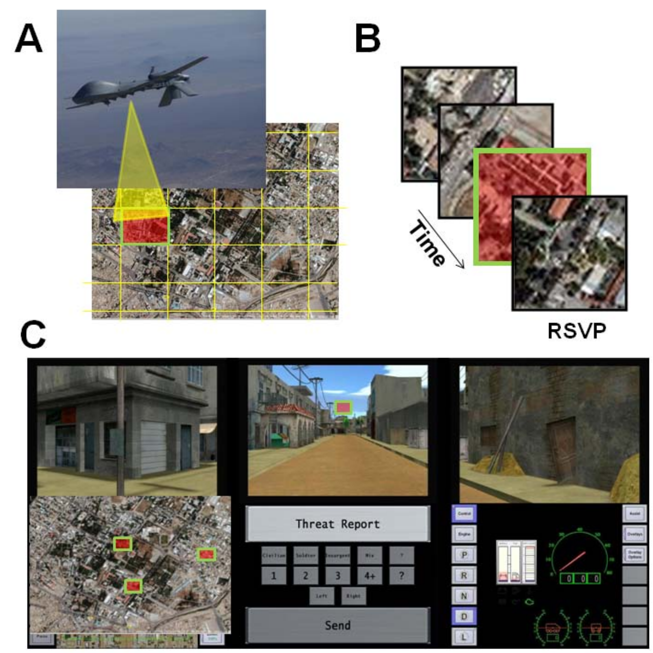

Figure 4: A. Image data obtained using UAV. B. Images presented to analyst using RSVP to identify potential targets. C. Target locations propagated to vehicle crewstation maps.

This information can then be propagated to distributed digital maps accessible through the crew station as potential 'hot spots' (see Figure 4). Using C3V technology in this way would provide a means for ground troops to exploit recent image intelligence by providing potential locations of probable targets prior to entering the battlespace.

We are currently performing work to demonstrate further applicability of this technique through the development of mutually-derived human-machine situational awareness. This mutual situational awareness (MSA) will be based on large databases of images collected from the local battlespace through the use of UAVs, unmanned ground vehicles (UGV), unmanned ground sensors (UGS), and 360º local situational awareness (LSA) systems installed on manned ground vehicles. The vision of mutually-derived situational awareness is that as P3 brain signals indicate that certain images are probable targets, simultaneous computer vision will search the databases of battlespace images and propagate labels to images with similar low-level visual features to the detected targets, and then further propagate those targets to distributed digital maps, showing potential targets similar to those identified by the human user. The initial phases of this work have shown promising preliminary results in experiments performed in simulated environments, with the system correctly identifying a high percentage of the targets placed in the simulated environment based on the EEG data.

Enhanced Aided Target Recognition (ATR)

Battlespace intelligence can change rapidly, and it is important for the technology that Soldiers use to be able to quickly adapt to these changes. The fact that C3V system functions as a general interest or relevance detector makes it possible to use this technology to rapidly reconfigure aided target recognition algorithms. For example, there are known examples of insurgents hiding improvised explosive devices (IED) in holes in the thick walls found in Iraq and Afghanistan, and covering the IEDs with posters [21]. Often the posters would be either pro-American or anti-American, leading Soldiers to approach the poster in order to read it or tear it down.

Developing a computer vision algorithm that can perform the semantic identification of posters on the wall, much less pro-US or anti-US posters, is not a simple task. However, the human brain is capable of making this semantic distinction quickly and accurately. By using the iterative version of the tightly-coupled C3V method described above, it should be possible to develop training sets for rapidly training computer vision algorithms to identify images with low-level visual similarities to targets in the training set, which can then be used for ATR. The success of this application will depend on the human ability to rapidly make the needed semantic distinction, the computer vision algorithm's capability to obtain low-level visual features from the images, and the relationship between the high-level semantic distinction and the commonality of low-level visual features in target images that represent the semantic distinction. While it is certain that this method will not generalize to every possible ATR scenario, it still provides promising capabilities to develop many useful systems.

The current focus has been on BCITs using visual target recognition signals largely related to the P3; however, many other neural and behavioral measures can be used in BCIT applications, ranging from other ERP type signals such as the error-related negativity, (ERN) to powers and phases in oscillatory activity such as in the theta (4-8hz), alpha (8-12Hz) and gamma (30-70Hz) bands. In addition, behavioral measures such as eye-tracking or other information streams investigated via other sensory modalities, such as auditory and tactile, could also potentially benefit BCITs, with an ultimate goal being BCITs that optimize across the rich multisensory experiences which are ubiquitous in military operations.

CONCLUSION

In this paper, we have described a neural measure, the P3 event-related potential, and a specific experimental paradigm, the rapid serial visual presentation paradigm and how they have been combined to develop a new type of brain-computer interaction technology, specifically the Cortically-Coupled Computer Vision system. The C3V system has been successfully applied to one militarily-relevant task, the intelligence analysis of satellite imagery data, and demonstrated improved performance over the current methodologies for performing this task. Additional Army-relevant applications are discussed that use C3V-based systems that should provide a means to further basic and translational research.

REFERENCES

[1] B. Lance, J. Capo, and K. McDowell, "Future Soldier-System Design Concepts: Brain-Computer Interaction Technologies," in *Designing Soldier Systems: Current Issues in Human Factors*, In Press.

[2] L. C. Parra et al., "Spatiotemporal linear decoding of brain state: Application to performance augmentation in high-throughput tasks," *IEEE Signal Processing*, vol. 25, no. 1, pp. 95–115, 2008.

[3] E. Donchin, K. M. Spencer, and R. Wijesinghe, "The mental prosthesis: assessing the speed of a P300-based brain-computer interface," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 174–179, 2002.

[4] M. Donnerer and A. Steed, "Using a P300 Brain-Computer Interface in an Immersive Virtual Environment," *PRESENCE: Teleoperators and Virtual Environments*, vol. 19, no. 1, pp. 12–24, 2010.

[5] D. S. Klobassa et al., "Toward a high-throughput auditory P300-based brain-computer interface,"

Clinical Neurophysiology, vol. 120, no. 7, pp. 1252–1261, 2009.

[6] T. C. Handy, *Event-related potentials: A methods handbook*. The MIT Press, 2005.

[7] J. Polich, “Updating P300: an integrative theory of P3a and P3b,” *Clinical Neurophysiology*, vol. 118, no. 10, pp. 2128–2148, 2007.

[8] S. Makeig, S. Debener, J. Onton, and A. Delorme, “Mining event-related brain dynamics,” *Trends in Cognitive Sciences*, vol. 8, no. 5, pp. 204–210, 2004.

[9] R. I. Goldman et al., “Single-trial discrimination for integrating simultaneous EEG and fMRI: Identifying cortical areas contributing to trial-to-trial variability in the auditory oddball task,” *Neuroimage*, vol. 47, no. 1, pp. 136–147, 2009.

[10] E. Donchin and E. F. Heffley, “Multivariate analysis of event-related potential data: A tutorial review,” *Multidisciplinary perspectives in event-related brain potential research*, pp. 555–572, 1978.

[11] J. Dien, K. M. Spencer, and E. Donchin, “Parsing the late positive complex: Mental chronometry and the ERP components that inhabit the neighborhood of the P300,” *Psychophysiology*, vol. 41, no. 5, pp. 665–678, 2004.

[12] M. D. Rugg and M. C. Doyle, “Event-related potentials and recognition memory for low-and high-frequency words,” *Journal of Cognitive Neuroscience*, vol. 4, no. 1, pp. 69–79, 1992.

[13] J. Touryan, L. Gibson, J. H. Horne, and P. Weber, “Real-Time Measurement of Face Recognition in Rapid Serial Visual Presentation,” *Frontiers in Psychology*, vol. 2, no. 42, pp. 1-8, 2011.

[14] S. Thorpe, D. Fize, and C. Marlot, “Speed of processing in the human visual system,” *nature*, vol. 381, no. 6582, pp. 520–522, 1996.

[15] P. Poolman, R. M. Frank, P. Luu, S. Pederson, and D. M. Tucker, “A single-trial analytic framework for EEG analysis and its application to target detection and classification,” *NeuroImage*, vol. 42, pp. 787-798, 2008.

[16] A. Kapoor, P. Shenoy, and D. Tan, “Combining brain computer interfaces with vision for object categorization,” in *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[17] 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*, vol. 14, no. 2, pp. 174-179, 2006.

[18] P. Sajda et al., “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.

[19] E. A. Pohlmeyer, J. Wang, D. C. Jangraw, B. Lou, S. F. Chang, and P. Sajda, “Closing the loop in cortically-coupled computer vision: a brain–computer interface for searching image databases,” *Journal of Neural Engineering*, vol. 8, p. 036025, 2011.

[20] P. Sajda, E. Pohlmeyer, J. Wang, B. Hanna, L. C. Parra, and S. F. Chang, “Cortically-Coupled Computer Vision,” in *Brain-Computer Interfaces*, Springer, 2010, pp. 133-148.

[21] TRADOC, “TRADOC DCSINT Handbook No. 4: Identification of Improvised Explosive Devices (IEDS, Non-Country Specific).” US Army Training and Doctrine Command, Jul-2006.