

Integration of Automated Neural Processing into an Army-Relevant Multitasking Simulation Environment

Jon Touryan¹, Anthony J. Ries¹, Paul Weber², and Laurie Gibson²

¹ Human Research and Engineering Directorate
U.S. Army Research Laboratory, Aberdeen Proving Ground, MD 21005, USA
{jonathan.o.touryan.ctr, anthony.j.ries2.civ}@mail.mil

² Science Applications International Corporation, Louisville, CO, USA
{paul.r.weber, laurie.d.gibson}@saic.com

Abstract. Brain-computer interface technology has experienced a rapid evolution over recent years. Recent studies have demonstrated the feasibility of detecting the presence or absence of targets in visual imagery from the neural response alone. Classification accuracy persists even when the imagery is presented rapidly. While this capability offers significant promise for applications that require humans to process large volumes of imagery, it remains unclear how well this approach will translate to more real-world scenarios. To explore the viability of automated neural processing in an Army-relevant operational context, we designed and built a simulation environment based on a ground vehicle crewstation. Here, we describe the process of integrating and testing the automated neural processing capability within this simulation environment. Our results indicate the potential for significant benefits to be realized by incorporating brain-computer interface technology into future Army systems.

Keywords: Simulator, Brain-Computer Interface (BCI), Visual Search.

1 Introduction

Over the past decade, there has been a substantial improvement in the accuracy of neural signal classification algorithms. One notable area is in the classification of the neural response, as measured via electroencephalography (EEG), elicited by an image. It has been well established that there is a significant difference in the evoked response between images containing a task-relevant target and images without a target [1, 2]. This difference persists even when the image is visible for only a few milliseconds [3]. However, only recently have signal processing algorithms and techniques been sophisticated enough to accurately classify the EEG signal after a single presentation of the image. Initial applications of this technology were for a rapid review or triage of imagery without requiring a manual response to each image [4]. Using a paradigm of rapid image presentation, called rapid serial visual presentation (RSVP), various groups have shown that this approach can identify targets in large ensembles of images an order of magnitude faster than a manual search [5, 6]. However, these

initial applications typically used obvious target objects and the only task required of the operator was to focus on the image presentation.

The main goal of this Army-industry collaboration (funded through the Institute of Collaborative Biotechnologies 6.2 Translational Research Program) was to develop a simulation environment to test the performance of state-of-the-art neural classification techniques in a more operational context. The first stage in developing this simulation environment focused on determining the optimal parameters for classification of the neural response in an RSVP paradigm. The parameters investigated included target presentation properties (e.g., size, eccentricity and rate [7]), the effect of changes in attentional state on classification accuracy [8], and the effect of operator multitasking on system performance [9]. The second stage of development, described here, focused on the specific application of the automated neural processing to an Army relevant system. Our intent was to replace the manual visual search task currently utilized to both identify targets and maintain situational awareness in Manned-Ground Vehicles (MGV). Specifically, the RSVP paradigm, in combination with automated classification of the neural response, would replace the manual control of an imaging sensor on the vehicle. Therefore, instead of an operator manipulating the pan-tilt-zoom (PTZ) camera to scan the environment, images of the vehicle's surroundings, containing potential targets, would be rapidly presented and subsequently sorted based on the operator's neural response. The operator could then review the most relevant images for target confirmation.

This second stage of development consisted of two elements. First, we sought to quantify the potential tradeoff of replacing a manual search with RSVP. To accomplish this we conducted an experiment to compare the time-to-target and accuracy of manual search and RSVP paradigms. Second, we developed a simulation environment based on the MGV crewstation. This simulator was designed to switch between the two paradigms and was fully integrated with a real-time EEG processing system. In addition, the simulator incorporated multitasking aspects of the crewstation, including auditory and text communications. Together, these results demonstrate the feasibility and potential benefits of integrating automated neural processing technology into Army systems.

2 Visual Search and RSVP

Sixteen participants were recruited for this experiment, 10 from the general population and six from the project collaborators. They ranged in age from 23 to 55 (mean = 34.8) and included 14 males. Thirteen of the participants were right handed, two were left handed and one was ambidextrous. All individuals recruited from the general population received compensation of \$20 per hour. The voluntary, fully informed consent of the persons used in this research was obtained as required by Title 32, Part 219 of the Code of Federal Regulations and Army Regulation 70-25. The investigator has adhered to the policies for the protection of human subjects as prescribed in AR 70-25.

In this experiment, participants alternated between manual visual search and RSVP tasks. In the manual visual search blocks they were required to move a controllable portal (PTZ) over large scenes (1920 x 1080 pixel images) of a simulated urban environment. Images were screen captures from a popular video game (Call of Duty®, Activision Publishing Inc). Above this portal was a low resolution context display of the entire scene with an indicator as to the current location of the portal. The portal was a circular vignette (radius of 150 pixels) initially revealing approximately 3% of the large image; but this could be either increased or decreased based on the zoom factor. Participants were required to use the keyboard (arrow keys, “+”, “-”) to move the portal and scan the scene for a target (a soldier with a gun). Target identification was indicated by a key press (“t”). Likewise, if no target was found, participants terminated the search with an alternate key press (“n”). The initial placement of the portal was randomly distributed within a given window of 500 pixels around the target. In most cases, the target was not visible in the context display and required the search portal to be detected. The context display served primarily to influence the participant’s search path and provide information on likely target locations (doors, windows, cars, etc.).

In the RSVP blocks participants were presented with 100 x 100 pixel image chips representing a region of interest (ROI) from a high resolution image. These ROIs represented salient locations within the image. While the ROIs were manually selected for this experiment, in a real-world application they would be selected by pre-filtering computer vision algorithm [10, 11]. ROIs were displayed at 2 Hz (500 milliseconds), and participants were required to press a button (spacebar) when they saw a target. Since the purpose of the RSVP blocks was to estimate accuracy, only 10 ROIs were included in each RSVP block with a maximum of one target ROI per block. The likelihood of a target was 50 percent for both visual search and RSVP blocks. Participants completed 15 blocks of each task in alternating succession.

A summary of the results for 16 participants is shown in figure 1. As expected, the search time for target-present images was significantly shorter than for target-absent ($p < 0.001$; Wilcoxon rank sum test), 18 seconds versus 60 seconds. While the accuracy for the manual search component was high (mean total accuracy = 0.85), it was substantially lower than for the RSVP component (mean total accuracy = 0.99). There was a significant correlation across participants in accuracy between the search and RSVP components ($r = 0.62$, $p = 0.01$; Pearson’s correlation coefficient). For this experiment we decided to keep the RSVP length fixed at 10 ROIs (image chips) or 5 seconds. Under these conditions the RSVP length could be tripled (incorporating up to 30+ ROIs) and still outperform the manual search; requiring less than 15 seconds for completion.

One interesting observation from the manual visual search task was that the search time did not strongly correlate with the initial portal accuracy (i.e., distance from the portal center to the target). The correlation coefficient between search time and initial portal accuracy was 0.09 ($p = 0.07$). Unless the portal was placed within 100 pixels of the target, the initial placement of the portal did not influence the search time because participants followed a search path of potential target locations. This observation

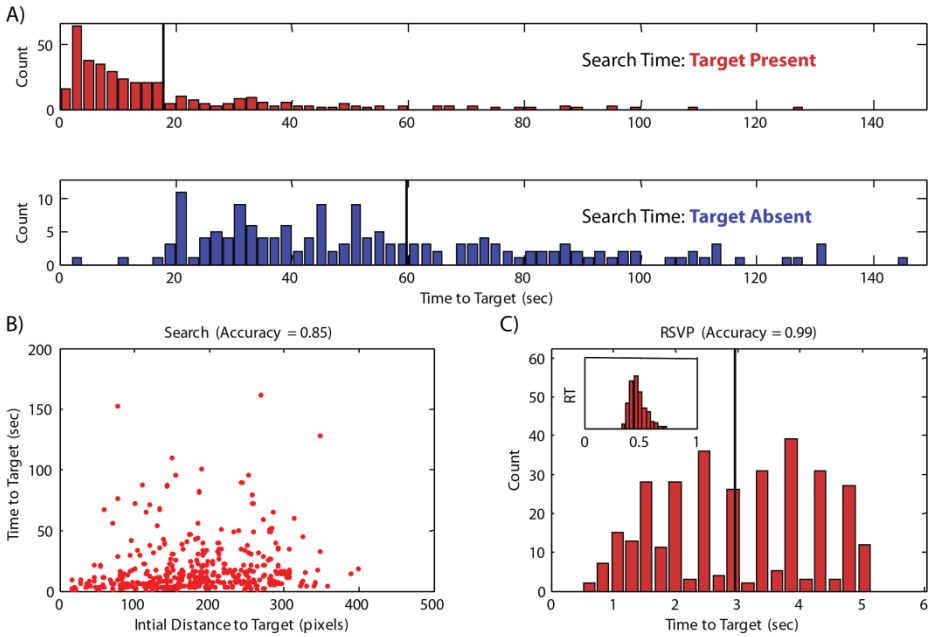


Fig. 1. Manual search and RSVP behavioral summary. A) The manual search time distributions for target-present (red) and target-absent (blue). Vertical lines indicate distribution mean. B) The relationship between initial portal placement and search time. C) The time-to-target in the RSVP condition with a 10 ROI sequence (9 non-target and 1 target image). Inset shows the reaction time (RT) distribution.

speaks directly to the importance of the slew-to-cue accuracy (orientation of the imaging sensor in response to external or environmental cues) in MGV. If the slew-to-cue accuracy can be well quantified, it will be imperative to instruct crewstation operators to stay within the area of initial placement and suppress their instinct to follow a contextual search path. In a similar fashion, the intelligent RSVP should be programmed to give priority to ROIs that fall within the cued area.

Another potential key parameter is portal speed. To test this directly we manipulated portal speed for a subset of participants ($N = 10$). These participants performed the search experiment in two sessions. In each session their portal speed (in PTZ) was set to a value of either baseline (1x condition) or twice baseline (2x condition). The order of the conditions alternated such that half of the participants had the 1x condition in the first session and half the 2x condition. Over the population we found that there was no significant difference in search time for the two conditions ($p > 0.05$; Wilcoxon rank sum test). However, we did find a significant reduction in search time between session one and two, indicating a significant practice effect ($p < 0.05$). These results suggest that training, rather than PTZ speed, is more important for system performance.

3 Simulator Design

The simulation environment (figure 2) was designed to test the translation of automated neural processing into a more real-world environment. Specifically, this environment was modeled after an Army MGCV crewstation. As in the visual search experiment described above, the operator's primary task was to search the environment for targets (soldiers with guns) while the vehicle navigated an urban landscape. The principal performance comparison for this simulator is the speed and accuracy of target detection between a manual search, via a gimballed camera described above, and an RSVP presentation of pre-filtered ROIs around the vehicle. During the RSVP presentation, each ROI is sorted based on an interest score derived from the evoked brain response [4]. The operator is then immediately presented with the images that generate the highest interest score to manually verify which of the top scoring images contained targets. In addition to this primary task, several other secondary tasks are required of the operator. These secondary tasks are designed to both increase difficulty and replicate real-world environments.

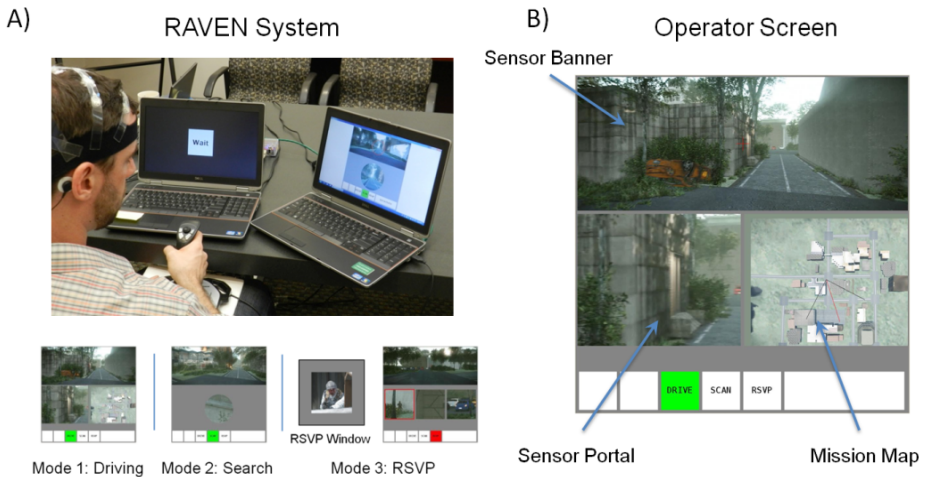


Fig. 2. RSVP-based Adaptive Virtual Environment with Neural-processing (RAVEN) system. A) RAVEN system with main operator screen and RSVP window. The primary task is identification of dismounts while the vehicle is navigating a simulated environment. Secondary tasks include identification of potential IED locations, monitoring and responding to communications (audio and text). The BCI component is engaged during RSVP (Mode 3) and top scoring images are presented on the main operator screen. B) Operator screen during driving (Mode 1) including sensor banner, sensor portal, and mission map windows.

After the optimal design parameters were identified, the simulation environment (called RSVP-based Adaptive Virtual Environment with Neural-processing or RAVEN) was developed to quantify the benefit of incorporating neural processing techniques into this Army relevant operational context. To measure the performance and potential benefits of the RSVP approach, we outlined two validation experiments.

The first experiment, described below, was designed to quantify the speed and accuracy of the two search paradigm with minimal interference from secondary tasks. The second experiment is currently being conducted at the University of California, Santa Barbara, and will focus on the effects of multitasking and task difficulty on overall system performance.

4 Simulator Validation

Fourteen participants were recruited for this experiment, eleven from the general population and three from the project collaborators. Six of these individuals also participated in the visual search experiment (described above). Participants ranged in age from 23 to 59 (mean = 36.9) and included 10 males. Eleven of the participants were right handed and three were left handed. All individuals recruited from the general population received compensation of \$20 per hour. The voluntary, fully informed consent of the persons used in this research was obtained as required by Title 32, Part 219 of the Code of Federal Regulations and Army Regulation 70-25. The investigator has adhered to the policies for the protection of human subjects as prescribed in AR 70-25.

In this experiment, participants again alternated between manual visual search and RSVP tasks, this time within the context of the RAVEN simulator (figure 2). Briefly, the task was a simulated patrol of an urban landscape. The vehicle was driven by the computer but the commander (experimental participant) was required to perform several tasks as the vehicle navigated through the environment. The primary task was visual target detection in order to identify threats. At each intersection (24 in all), the vehicle stopped and the participant searched for the target (soldier carrying a gun). At half of the intersections the search was via a controllable portal; in the other half, the search was performed through an RSVP sequence of pre-filtered image chips (ROIs). The majority of intersections contained a target (approximately 80 percent), which could appear at various locations within the scene. A set of potential locations was identified before the experiment but the final target location was randomly chosen by the computer at each intersection. The parameters of the primary task were similar to the visual search experiment described above. However, in this case we used a presentation rate of 5 Hz (200 milliseconds) for the RSVP component. In addition to the intersection search, the participant was required to perform two other tasks while the vehicle was navigating the environment: 1) identify potential IEDs near the roadside (e.g., trash bags, boxes, tires) while the vehicle is moving, and 2) respond to specific radio communications.

Electrophysiological recordings were digitally sampled at 256 Hz from 20 scalp electrodes, located on the standard 10-20 coordinate grid, using an ABM x24 system configured with the single-trial ERP sensor strip (Advanced Brain Monitoring, Carlsbad, CA). EEG was acquired during the entire simulation but real-time analysis, via single trial classification of the evoked response using custom software, was only engaged during RSVP. Individual neural classification models were constructed for each participant from a separate RSVP session prior to the simulator experiment. The

single-trial classification models were linear discriminate functions applied to the neural response elicited by each image. The models were constructed using a machine learning algorithm, described elsewhere [12], and typically achieved area under the ROC curve values between 0.85 and 0.95.

The principal comparison in this validation experiment was speed and accuracy between the manual search and RSVP conditions. At the manual search intersections, participants scanned the environment via the controllable portal until they identified the target or decided that no target was present. At the RSVP intersections, the neural response to each image was scored and the top three ROIs from each intersection search were shown to the participant for final target selection (or confirmation that no target was present). We used metrics of accuracy and time-to-target to quantify the performance in both conditions. Figure 3 shows the results from the 14 participants. At this faster RSVP presentation rate (5 Hz) the difference in mean time-to-target (manual search $\mu = 0.64$ minutes, $\sigma = 0.49$; RSVP $\mu = 0.23$ minutes, $\sigma = 0.11$) was significant ($p < .001$, $t = 7.8$) while the difference in mean accuracy (manual search $\mu = 0.80$, $\sigma = 0.4$; RSVP $\mu = 0.85$, $\sigma = 0.35$) was not statistically significant ($p = 0.259$, $t = -1.13$). These results indicate that even under more realistic conditions, the integration of automated neural processing can enhance overall operational performance.

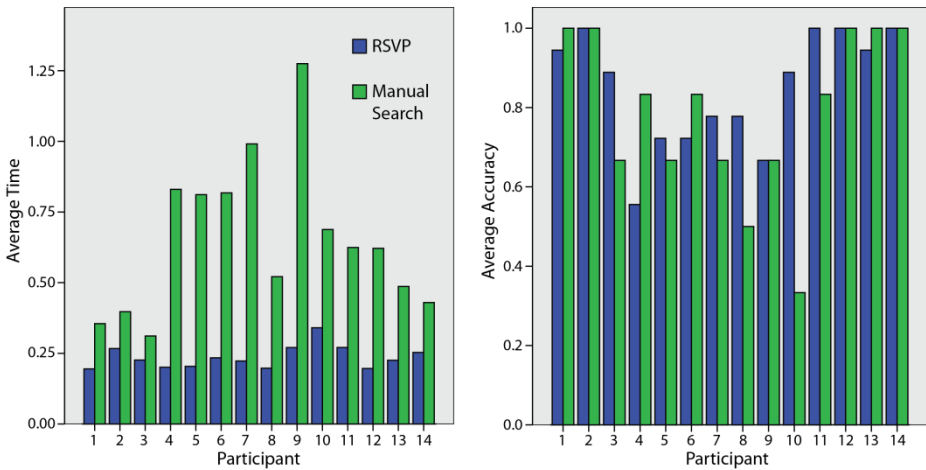


Fig. 3. Performance summary for fourteen participants with 5 Hz RSVP presentation rate. Graphs compare time (in minutes) to find the target and accuracy for RSVP and manual search.

5 Summary

For this study, we defined and implemented a simulation environment that offers a platform to test automated neural processing applications within a real-world context. We identified a common task for MGVS operators, manual search of the vehicle environment, which could potentially be replaced with a brain-computer interface. Single-trial classification of the neural response to ROIs of the vehicle surroundings can identify targets more rapidly than the manual search. This performance enhancement

persisted even when the task was embedded within an operation scenario. One important component underling the success of this study is that the accuracy of the single-trial classifier was sufficient to find the target at each intersection, even with a rapid presentation rate (5 Hz). Further tests will be conducted to explore how performance is modulated by increased difficulty and the imposition of multitasking. This simulation environment demonstrates the potential for brain-computer interface technology to meet the challenges of the ever increasing complexity of soldier-systems.

There are several important factors that will influence the success of this brain-computer interface application. One critical factor is how well EEG can be acquired and processed in the real-world vehicle environments. Noise from electrode movement and muscle activation can dramatically affect the EEG signal. Various approaches are being explored to process EEG in real-time within the context of a high noise environment. Another key factor is the performance of the automated systems in the ground vehicle. First, the slew-to-cue accuracy of the external vehicle sensors will directly impact how much benefit can be gained by this neural processing approach. If the automated systems can reliably and accurately locate targets, this triage approach may be unnecessary. Second, the efficacy of the computer vision pre-filtering algorithm will significantly influence the RSVP speed. The length of the RSVP is directly related to how many false alarm ROIs are generated by the pre-filtering algorithm. Given that targets may often be camouflaged or embedded within a dense context, it is likely that numerous false alarms will be detected for every correct target identified. Likewise, the RSVP application described here relies on the pre-filtering algorithm to include the target in the ROI ensemble. Thus, integration of this type of brain-computer application will be a dynamic process that depends on the capabilities of the complementary automated systems.

References

1. Squires, N.K., Squires, K.C., Hillyard, S.A.: Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology* 38, 387–401 (1975)
2. Snyder, E., Hillyard, S.A.: Long-latency evoked potentials to irrelevant, deviant stimuli. *Behavioral Biology* 16, 319–331 (1976)
3. Thorpe, S., Fize, D., Marlot, C.: Speed of processing in the human visual system. *Nature* 381, 520–522 (1996)
4. Sajda, P., Pohlmeier, E., Wang, J., Parra, L.C., Christoforou, C., Dmochowski, J., Hanna, B., Bahlmann, C., Singh, M.K., Chang, S.F.: In a Blink of an Eye and a Switch of a transistor: Cortically Coupled Computer Vision. *Proceedings of the IEEE* 98, 462–478 (2010)
5. Sajda, P., Gerson, A., Parra, L.: High-throughput image search via single-trial event detection in a rapid serial visual presentation task. In: *Proceedings of the First International IEEE EMBS Conference on Neural Engineering*, pp. 7–10 (2003)
6. Touryan, J., Gibson, L., Horne, J.H., Weber, P.: Real-time classification of neural signals corresponding to the detection of targets in video imagery. Presented at the International Conference on Applied Human Factors and Ergonomics (July 17, 2010)

7. Gibson, L., Touryan, J., Ries, A., McDowell, K., Cecotti, H., Giesbrecht, B.: Adaptive Integration and Optimization of Automated and Neural Processing Systems-Establishing Neural and Behavioral Benchmarks of Optimized Performance. DTIC Document (2012)
8. Giesbrecht, B., Eckstein, M.P., Abbey, C.K.: Neural decoding of semantic processing during the attentional blink. *Journal of Vision. Abstract* (2009)
9. Cecotti, H., Kasper, R.W., Elliott, J.C., Eckstein, M.P., Giesbrecht, B.: Multimodal target detection using single trial evoked EEG responses in single and dual-tasks. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, pp. 6311–6314 (2011)
10. Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* 40, 1489–1506 (2000)
11. Torralba, A., Oliva, A., Castelano, M.S., Henderson, J.M.: Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological review* 113, 766 (2006)
12. Touryan, J., Gibson, L., Horne, J.H., Weber, P.: Real-time measurement of face recognition in rapid serial visual presentation. *Front Psychol.* 2, 42 (2011)