



***Defense Advanced Research Projects Agency
Neurotechnology for Intelligence Analysts (NIA) Program***

Phase I Analysis Report for UCSD/SoCal NIA Team

31 January 2007

Prepared for:

Dr. Amy Kruse

Program Manager

DARPA Defense Sciences Office

3701 North Fairfax Drive

Arlington, VA 22203-1714

Prepared under contract: HM1582-05-C-0044

Submitted by:

Scott Makeig

Director, Swartz Center for Computational Neuroscience

Institute for Neural Computation

University of California San Diego

9500 Gilman Drive

La Jolla CA 92093-0961

Acknowledgments

We thank the NIA program staff, Drs. Amy Kruse and Josh Schulman, and Grace Rigdon for their help, and also thank the three volunteer NGA intelligence analysts who participated in our metric experiments. The PI also thanks his collaborators, Nima Bigdely-Shamlo, Andrey Vankov, Rey Ramirez, Nuno Vasconcelos, Hamed Masnadi, Jason Palmer, Wolfgang Einhaeuser-Treyer, Laurent Itti, Christof Koch, and Pierre Baldi for their cooperative spirit and enthusiastic participation in the project, as well as Lori Koller and the UCSD grants management team, who saw to the j's and t's. Last, but not least, thanks to system administrator Robert Buffington, who keeps us computing.

Abstract

In Phase 1 of the DARPA “Neurotechnology for Intelligence Analysts” (NIA) project, our team composed of faculty, students, postdoctoral fellows, and technical staff from UCSD, USC, UCI, and Caltech, developed a prototype of a novel networked EEG response classification (NERC) system. The system delivers bursts of overhead imagery clips at high presentation rates (rapid serial visual presentation, RSVP) while recording up to 256 channels of electroencephalographic (EEG) information from the human scalp, plus subject button presses following each image burst. The goal of the system is to detect, in near real time, ‘flickers of recognition’ in the subject’s brain wave patterns whenever the subject detects a target object (such as a plane, helipad, or similar) in one of the briefly displayed images. Further, the system paints the areas that have been searched on another monitor screen colored according to the estimated target presence probability based on the EEG response. An image analyst can use screen controls on this viewer screen to pan and/or zoom into the map, change the color mapping, etc., to evaluate the results of the RSVP image triage. The NERC system is implemented as a set of interlocking processes that can operate on any number of CPUs or computers linked by a TCPIP and UDP network. Unique features of the system include: (1) A novel method for constructing, selecting, and sequencing small image clips, given a broad-area image to be searched. (2) A novel method of response classification based on separation by independent component analysis (ICA) of the scalp data into independent brain sources, followed by separation of the time courses of activity of the independent source processes, both in time-domain and time/frequency-domain, again by ICA, followed by linear discriminant analysis for each component. Finally, the estimated probabilities for the independent sources are multiplied to give a single target probability estimate for each presented image. In its current version, the system requires less than 50 ms per image to compute the probability estimates (or less than 6 ms per image omitting frequency-domain measures). This report describes results of a first metric experiment on three volunteer government image analysts and discusses fruitful directions for follow-on research. These directions include filtering of image clips and clip order to minimize visual distraction and attentional blink phenomena, further improvements in classification algorithms, online ICA training, incorporation of subject alertness and attention monitoring, incorporating subject eye movement information in the classifier, and delivering sensory feedback to the analyst subject to better maintain and improve image classification accuracy.

Executive Summary

Overview of the Team Effort. In Phase 1 of the DARPA “Neurotechnology for Intelligence Analysts” (NIA) project, our team composed of faculty, students, postdoctoral fellows, and technical staff from UCSD, UCSD, UCI, and Caltech, developed a prototype of a novel networked EEG response classification (NERC) system that delivers bursts of overhead imagery clips at high presentation rates (rapid serial visual presentation, RSVP) while recording up to 256 channels of electroencephalographic (EEG) information from the human scalp, plus subject button presses, with a goal of detecting in near real time ‘flickers of recognition’ in the subject’s brain wave patterns whenever the subject detected a given target feature (such as a plane, helipad, or similar feature) in one of the briefly flashed images.

Further, the system paints the areas that have been searched on another monitor screen colored by the probability estimated from the EEG responses that each area searched contains a target feature. The subject, an image analyst, can use screen controls on this results viewer screen to pan and/or zoom into the map, change the color mapping, etc., to evaluate the results of the RSVP screening. The NERC system is implemented as a set of interlocking processes that can operate on any number of CPUs or computers linked by a TCPIP and UDP network.

Team collaborations during the Phase I research included work by team members Nuno Vasconcelos and Hamed Masnadi at UCSD to develop a method for classifying EEG responses to image clip presentations using a fast cascade architecture, work by Pierre Baldi and students at UCI to test applicability of conventional neural network classifiers of EEG-based classification, work by Einhauser-Treyer and Christof Koch of Caltech to better understand the nature of the attentional blink in RSVP paradigms, and work by Laurent Itti and students at USC to understand the relation of RSVP target detection to low-level visual surprise. This report describes results of a first metric experiment on three volunteer government image analysts and discusses fruitful directions for follow-on research.

Overall Significance of Findings. Results of EEG experiments on 15 pilot subjects, including commercial GIS analysts, and on three volunteer government image analysts demonstrate conclusively that multi-dimensional information is available in the EEG allowing accurate discrimination of subject response to single presentations of images containing or not containing specified target features. This is true even at relatively high rates of image presentation (e.g., 12/s). Novel methods are introduced for maximizing the efficiency of performing broad-area search while minimizing subject discomforts associated with RSVP viewing, and for distributing the real-time computing load among computers linked in a standard local area network.

Adequacy / Maturity of Measurements. EEG data collection technology is fairly mature. Our approach incorporates active chip sensors (Biosemi, Inc). Further advances in EEG recording methods will come in the form of high-density dry electrode caps that do not involve the need to prepare the scalp or apply water-based gel. A good deal of the effort in the project went toward identifying informative features in the EEG signals and maturing the classifier algorithms used in the system. We believe that further improvement is possible.

Adequacy of Development Approach Used (What Would You Do Differently?) The approach we pursued, combining high-density EEG data collection, spatial and temporal filtering by independent component analysis (ICA), and multiple linear discriminant analysis applied to features of the independent component processes, proved to be effective for near real-time classification of brain responses. Our software approach, involving a modular network of cooperating processes, also proved to be successful, and may be flexibly extended in several directions. Give more time, funding, and trained personnel, we would prefer to collect and analyze data on a larger number of subjects in a fixed paradigm; the highly compressed time frame of the project did not allow for this, since each experiment needed to test some aspects of our development system.

Operational Relevance and Impact. From our contacts with government overhead imagery analysts, both in our metric experiments and at the NIA project meetings in the Washington DC area, we learned that there is a need for software that aids analysts in performing careful broad-area search for specific but varying classes of targets. We believe the heptunx path search technology we developed during this project could be a basis for improved operational software for this purpose. While our EEG results represent a leap in accuracy of classification of responses based on single events, we expect that the eventual best uses of EEG neurotechnology in the analyst workplace may be some yet untested combination of monitoring, classification and feedback, for analyst training, performance maintenance, and performance efficiency.

Lessons Learned. We learned many things during the course of the project. The major lesson was in the value of research teamwork – without cooperative efforts from highly skilled participants, we would not have made the amount of progress we did make. Our work to optimize EEG-based target response classification taught us to respect the large amount of information about subject state, response, and intentions that is available in near real-time in scalp EEG data. This bodes well for the future of neurotechnology research and the useful systems that will eventually emerge from the research.

Existing Questions Requiring Further Basic or Applied Research. Clearly, one most fruitful direction for further research is to better understand the information about subject responses to new visual information acquired not only through RSVP, but through normal 3-4/s saccades during visual search. Combining single-trial EEG classifiers with active eye gaze monitoring appears to us to be a neurotechnology research frontier. One area of applied research in this direction could be to assist analysts in reviewing overhead or other video intelligence. Another application would be toward a system that substitutes assisted visual (saccadic) search for RSVP search. Another important area of basic research is on the effect of ‘closing the loop’ for EEG classification by feeding back to the analyst subject the probability of a target response estimated by the EEG-based system following known targets introduced into a target search task. The question here is whether the analyst subject can use this feedback to learn to produce more distinct EEG responses that further improve the performance of the classifier in detecting target responses to previously unknown targets, e.g., those acquired by the analyst during broad-area search. The ability of subjects to make use of other forms of EEG-based feedback concerning their current levels of alertness and attention to the presented images also needs careful investigation. Might a similar feedback approach, applied to EEG responses to search saccades in normal visual search, teach analysts to perform more efficient visual search? The answer to that question may have practical importance, since training of imagery analysts is currently a lengthy and difficult process.

Table of Contents

Acknowledgments.....i

Abstract ii

Executive Summary..... iii

Introduction 1

Research Methods..... 1

Problem Statement..... 1

 1.1 *Objectives Statement*..... 2

 1.2 *Technical Approach Statement*..... 2

 A Networked EEG Response Classifier System 5

 1.3 *Goals of the Research*..... 8

 1.4 *Participants* 8

 1.4.1 *Participant characteristics* 8

 1.4.1.1 *Primary focus: IA subjects (Metric experiments)*..... 8

 1.4.1.1.1 *demographics, years of experience in the field, subject-matter expertise* 8

 1.4.1.2 *Secondary Focus: Pilot/novice subjects* 8

 1.4.2 *Facilities*..... 8

 1.5 *Experimental Design* 8

 1.5.1 *Task and experimental design for Training, Triage Mode, Baseline Search* 9

 1.5.1.1 *Workstation / Interface*..... 9

 1.5.1.2 *Imagery* 9

 1.5.1.2.1 *Presentation parameters*..... 24

 1.5.1.3 *Sensor Setup*..... 24

 1.5.1.4 *Recording Parameters* 24

 1.5.1.5 *Data Handling Procedures*..... 25

Bad-channel detection and compensation. 25

Analysis Methods..... 27

 1.6 *Signal properties* 27

 1.6.1 *Time, Frequency, Space*..... 27

 1.7 *Classification strategy*..... 31

 1.7.1 *Rationale*..... 31

 1.7.2 *Implementation*..... 31

 1.7.3 *Robustness across sessions, subjects, and sensor numbers*..... 37

 1.7.4 *Processing speed*..... 43

Results..... 43

 1.8 *Image Analyst Recordings*..... 43

 1.8.1 *Detection (True Positives), Failures (False Positives, False Negatives)*..... 43

 1.8.2 *Individual Results Viewer screens and their average across subjects*..... 46

 1.9 *Non-IA/pilot data*..... 48

 1.9.1 *Detection (True Positives), Failures (False Positives, False Negatives)*..... 48

 1.9.2 *Individual subjects and average across subjects* 49

 1.9.3 *Comparison with IA-Subject results* 50

 1.10 *Statistical Power*..... 50

Lessons Learned 50

 1.10.1 *Unforeseen Challenges Encountered During the Conduct of the Experiments*..... 50

 1.10.2 *Findings Outside the Scope of the Program*..... 53

Future Development Plans..... 54

Networked EEG Response Classification system enhancements 54

Further improvements in response classification. 54

Neurobiology and brain source localization...... 55

Neural dynamics and training of visual search. 56

Neural dynamics of video search. 56

Feedback to maintain alertness and attention. 56

Other possible intelligence applications...... 56

References **57**

 List of Acronyms 62

 Program Effort – Publications To Date 63

DARPA Neurotechnology for Intelligence Analysts Program

Phase I Analysis Report for UCSD/SoCal NIA Team

Introduction

The ever-increasing amount of imagery produced by today's intelligence satellites, and the practical problems for intelligence agencies in training enough experienced image analysts to analyze the images they produce, have led to an interest in more efficient ways of reviewing satellite and other intelligence imagery. RSVP (Rapid Serial Visual Presentation) is a potentially promising approach to speeding the image review process. In particular, it may be useful for the broad area search task in which a low number of targets or target features are sought among large collections of images, or over a very wide image area that does not contain the target features of interest. Research has shown that human observers can recognize the "gist" of a scene very rapidly [2], achieving above-chance recognition of fairly abstract target categories (e.g., animal versus non-animal subjects) at presentation durations as short as 20 ms [3], or when multiple images are presented in rapid sequence, at presentation rates up to 40 Hz [1]. Similarly, in our experiments, subjects were able to detect with more than 80 percent accuracy whether or not an image containing a target airplane was present in a 4-s burst of satellite image views presented at a rate of 12/s.

Any system that uses RSVP to expedite image review requires a way to obtain subject responses to detected targets quickly and unobtrusively enough that it can accurately identify images containing targets in rapid RSVP sequences without distracting the operator viewing the image sequences. At high presentation rates, simple button presses produced by the image analyst (IA) upon target detections may not allow accurate enough time resolution to identify which images were perceived by the analyst as containing a target feature. Button pressing might also interfere with analyst appraisal of later images, or would interrupt the presentation of the image sequence. The goal of our NIA Phase I project was to build, test, and demonstrate a real-time system detecting the brain's 'flicker of recognition' of target features in scalp electroencephalographic (EEG) signals recorded from the IA. Our goal was to build a Network EEG Response Classification (NERC) system to more accurately and less obtrusively classify presented RSVP clips as containing targets or not, without active intervention of the analyst. Such a system should moreover provide useful visual results to the IA in near real time. Here, we report on our program of Phase I research culminating in demonstration of such a system, and suggest the most important directions for further research and development.

Research Methods

Problem Statement

Given the paradigm of a very large satellite image to be inspected in detail by an IA subject conducting a broad area search for a given target feature, and adequate EEG signal recording to determine the nature of their brain response to each image, the would-be builder of a real-time classification system is faced with these design questions:

1. How to extract small image 'clips' from the large image?
2. What image preprocessing steps are useful?

3. What image clip size, shape, and presentation rate to use to maximize system performance and user acceptability?
4. What sub-image ordering and grouping strategy will feel most convenient and natural to the user while maintaining image classification performance?
5. What is the nature of the neural signature(s) (the ‘flicker[s] of recognition’) associated with target detection in the recorded EEG signals? Which brain areas, response latencies, and dynamic properties are involved?
6. How best to detect this neural signature?
7. How best to use the EEG signature to accurately classify EEG signals in near real time?
8. How best to present the classification results for use by the analyst subject?

Our previous research on individualized models for EEG-based alertness monitoring (Makeig & Inlow, 1993; Makeig & Jung, 1996; Jung, Makeig, et al., 1997) strongly suggest that best target classification performance is possible only using individualized EEG models, additional questions may be answered once a classification system has been developed:

9. What detection/classification factors are in common, and which differ across individuals?
10. What neural signature differences are there between trained and untrained subjects?
11. Can the analyst subject learn to produce more easily classifiable brain responses, and can the classification system adapt at the same time to changes in the analyst’ brain signals?

1.1 Objectives Statement

Our goal in Phase I has been to build and test a real-time image classification system that goes beyond simple ‘proof of concept’ toward an integrated end product. To achieve this, we had to tackle several engineering challenges concerning real-time image presentation, processing of the EEG data, image response classification, and interactive viewing of results of the classification.

1.2 Technical Approach Statement

Early EEG signs of visual target recognition. Human intelligence is highly adapted to visual search for objects or aspects of a scene whose identity or statistics match those of a sought object or category. The visual system is adept at fast visual association and pattern recognition through successive analysis in a series of visual cortical regions. Thorpe et al. (1996) have suggested that divergence of posterior brain evoked responses to target and non-target category images, beginning 130 ms after image presentation, reflects activity within feed-forward visual pathways. Note, though, that this result was only obtained for the average of hundreds of trials, not from the single-trials themselves. However, recently it has become clear that the visual system is more ‘temporally compact’ than previously realized, with feedback from upper to lower visual areas (Hupe et al, 2001), as well as between primary auditory and visual areas beginning as little as 30 ms after abrupt presentation of an image (Schroeder and Foxe., 2002). By 130 ms after visual stimulus onset, extensive phase reorganization of local field activity is already occurring in many brain areas, as shown in a face image task recorded during invasive human presurgical monitoring by Klopp et al. (2002). These spatially complex field potential effects may, however, be highly variable from trial to trial, further complicating the signal processing necessary to automatically recognize them.

Dependence on image statistics. A further difficulty of using EEG to determine whether the intelligence analyst has detected a target arises from the fact that the images in which this target is embedded may have drastic variability of appearance. For example, if the analyst were shown an intermixed sequence of images with two different types of texture (city and forest, desert and farmland, etc.), each containing an occasional target feature, it is highly likely that significant variations of the EEG visual response to target features will be masked by lower-level visual detection areas of the brain that will respond differently to the different textures. These variations are likely to make it more difficult to detect the signatures associated with the detection of target relevance by the analyst. A similar problem has recently been investigated by Johnson and Olshausen (2003) in a study which clearly showed that, in the context of generic object recognition, EEG signatures associated with the detection of targets can be easily confused with EEG signatures arising from low level feature variability that has no information relevant for recognition.

Early frontolimbic responses. Another cause of image response variability is subtle changes in internal state of attention of the subject. The classical visual information pathway branches into brain arousal, evaluation, and motor systems very early. For example, changes in the firing rate of brainstem dopamine neurons in monkeys may begin as early as 50 ms after presentation of a stimulus with unexpected positive or negative reward value (Schultz, 2000). Dopaminergic innervation of amygdala and frontal inferior and medial cortex is believed to play an essential role in learning from experience, as well as in quick decision-making and ‘top-down’ responding to the anticipated consequences of events. This (and other neurotransmitter-labeled sensory arousal systems) may act prior to ‘conscious’ perception of the exact identity or category of the perceived object or scene. Both of the earliest aspects of the human brain response (posterior and frontal) to target category images appear in the EEG before the fastest manual responses. For example, fast finger press responses to an unexpected simple visual target stimulus in a spatial selective attention experiment (Makeig et al., 1999, 2004) are initiated at the mean peak latency of the earlier (P3f) aspect of the late positive complex, which originates in inferior frontal cortex. This early far-frontal response also appears in experiments not involving motor responses, in which it also appears to index subject detection of task relevance of visual stimuli (Potts et al., 2004), suggesting that rather than participating directly in the production of motor responses, it may index limbic system activity that can override the inhibitory effect of inferior frontal cortex on impulsive motor responding.

Our research has shown that this far-frontal P3f response preceding fast manual responses to visual targets is largely composed of a 1.5-cycle theta-band oscillation (Delorme and Makeig, in preparation). This oscillatory event in inferior frontal cortex partially precedes an ensuing, partially phase-coherent oscillation in or near dorsal anterior cingulate cortex, motor cingulate or adjacent supplementary motor cortex, and hand somatomotor cortex (Makeig et al., 2004). Such distributed theta synchronization events (TSEs) may organize early limbic system ‘top-down’ responding to the perceived affective implications of sensory events, including adjusting the distribution of sensory and mnemonic attention across cortical regions and reshaping motor planning.

Like other EEG phenomena, event-related theta activities may also vary widely from trial to trial, even when the single trials share superficial attributes. For example, in a working memory experiment, the small mean increase in frontomedial theta activity with increased

memory load, replicating previously reported results, was tiny relative to the nearly 30-dB range of trial-to-trial variations in theta activity (Onton, Delorme and Makeig, 2005). Efficient single-trial signal processing of both posterior and anterior EEG activity for real-time response classification must therefore concentrate on separating out this variability into meaningful processes or components, and intelligently combining the information they contribute to form an overall target probability estimate for each successive image in the RSVP sequence.

Independent Component Analysis. Electrodes recording electrical potentials from the scalp are electrically far from the brain itself. This has advantages and disadvantages. The disadvantages are that the signals reaching the scalp electrodes are very weak (uV) after passage through the skull, and each represent a complex mixture of underlying brain sources, since each such source (one or two compact regions of cortex), by basic biophysics, projects by volume conduction to nearly all the scalp electrodes. The advantages of EEG recording are that together, the electrodes ‘see’ the whole cortex, and that the mixtures of source activities are linear, allowing linear decomposition of the channel mixtures into underlying brain (and non-brain artifact) sources. The brute-force approach to solving this inverse problem (finding source activities given sensor activities) is technically difficult, since it requires a detailed electrical model of the head and skull. A more elegant and possibly more accurate approach, given the technical difficulty of making such models, is to separate the recorded signals into their distinct information sources. Since coupling between cortical areas is estimated to be, on average, a million time weaker than within cortical activities, islands of locally synchronous field activity can emerge and operate nearly independently, controlled for the most part by various neuromodulatory systems that do not routinely enforce phase synchrony between the activities of the various locally-synchronous EEG source domains.

Independent Component Analysis (ICA), developed in the 1990’s by the applied mathematics / signal processing community, is an effective method for separating recorded mixtures into maximally independent source activities, as first shown by us for the case of EEG data (Makeig et al., 1996). ICA learns spatial filters for independent sources of information – temporally distinct or independent waveforms – in the recorded data. Many of these have the signature of activity arising in a single (or sometimes dual-symmetric) cortical patch, i.e. a near-dipolar scalp map (Makeig et al., 2004).

An important finding of our Phase 1 project (illustrated in a later section) is that the independent components of high-density EEG signals that contribute the most information about image response classification are the same ICA components (ICs) with dipolar scalp maps, plus some ICs accounting for subject eye movement artifacts – apparently, subjects’ response to target appearance could also be estimated, to some extent, from their pattern of eye movements as well as from their cortical brain activity.

Our results (reported in a later section) confirm our initial hypothesis that extracting information about the nature of the subject response to RSVP images from independent signal components, instead of from the single scalp channels themselves, can give higher response classification performance. Therefore, the real-time response classifier uses ICA sources rather than raw EEG scalp channels to achieve higher performance.

A Networked EEG Response Classifier System

The Networked EEG Response Classifier (NERC) system we have developed during the Phase I project uses both time-domain and frequency-domain data collected from a dense array of EEG electrodes during analyst-initiated presentations of 4-s RSVP image bursts. The system software, written in Borland C++ under Windows XP, operates through a local network (intranet) that may include any number of computers connected by Ethernet TCPIP. The system classifies EEG signals for each presented image, estimating the probability that subject has responded to the image as including a search target feature. A movie of its operation is available by ftp at http://scn.ucsd.edu/pub/nia/nis_team6_ucsd_demo.mov

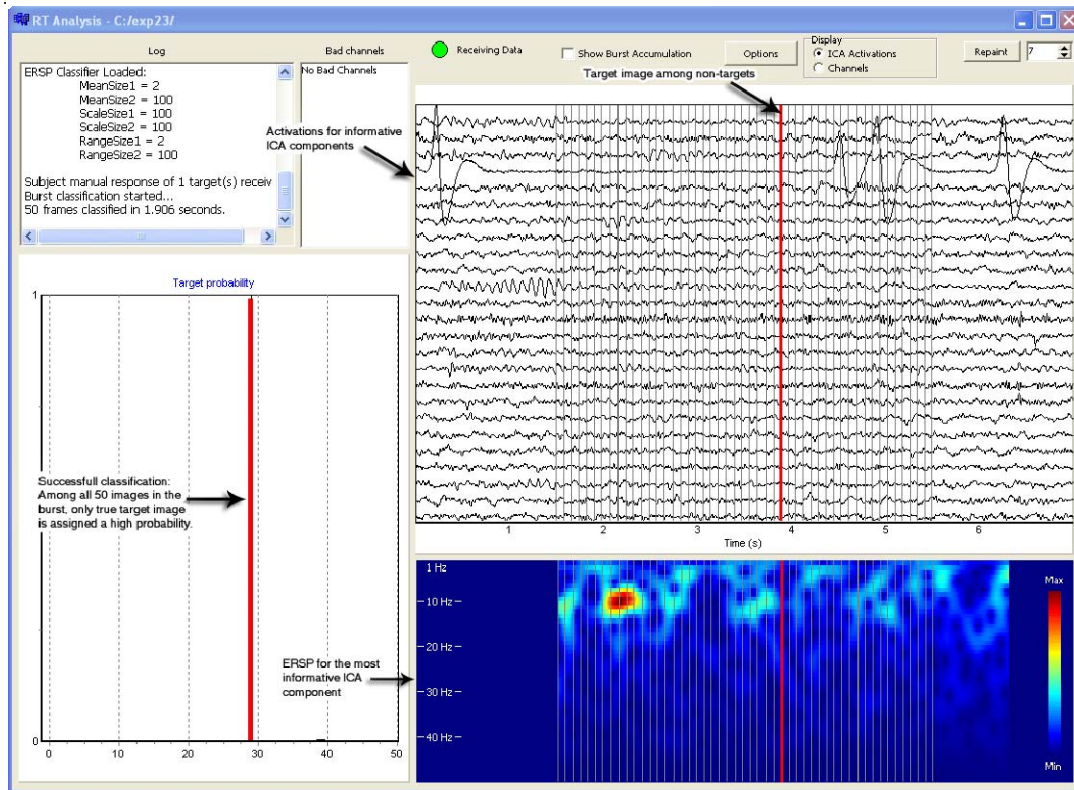


Figure 1. A snapshot from the Networked EEG Response Classifier program.

Figure 1 shows a screen shot of the NERC image classification screen. On the top right, independent components of the recorded EEG data are scrolled, while in the lower right panel the normalized spectrogram or event-related spectral perturbation (ERSP; Makeig, 1994) of the most informative component appears. Note the large eye blink potentials captured by one of the independent components (upper right), and the alpha burst confined to another single component before the burst onset. For demonstration purposes, the red vertical lines on these panels show the actual locations of a target image in the 51-image RSVP test sequence (thin black vertical lines).

The lower left panel reports the probability that each of the first 49 images in the burst contains a target. (The last two images in each burst are dummy images added to mitigate

confounds produced by the burst offset responses, allowing successful classification of the final actual burst images). In the successful burst classification shown here, only the actual target (lower left, vertical red line) is identified as being a probable target based on the EEG dynamic signature built for this subject based on earlier pilot data and used by the classifier to estimate target probability for each image in the burst. The upper left windows give information to the operator about program operation, including the number and indices of any channels whose signals have been lost during the session.

The Response Classifier program (Figure 1) passes the computed target probability estimates (lower left) to a real-time Results Viewer (Figure 2). The Viewer visualizes the locations of the presented RSVP image ‘clips,’ producing a target-like brain response in near-real time by translucently coloring-in their locations on the large area map from which they are drawn.

Software Modules. Since the implemented NERC system uses TCP/IP and UDP standards for data transfer over intranet, each module may run on a separate computer, or on a different CPU node of a multi-core processor. This leads to greater scalability and minimizes performance bottle-necks. The version of the NERC system demonstrated during Phase I consists of four modules running on three networked PC workstations:

- **Image Presentation:** One machine presents the image clips (4-second bursts of 50 images) to the subject and handles subject interactions. To present the visual stimuli, we used a Stim2006 program developed in our laboratory, based on the previously developed Variete© scripting language¹. Each presented image is assigned a unique ID number that travels through the system along with EEG data to assure correct assignment of calculated probabilities to each image. This makes the system immune to disruptions of image order based on fluctuating network and operating system task load.
- **Data Acquisition:** EEG data are collected via a 256-channel BIOSEMI Active2 system using custom data acquisition software. The EEG data are digitized at 256 Hz and streamed along with the image event synchronization codes from the stimulus presentation system to an inter-process communication buffer, part of an ADAPT© real-time network². The buffered data is then broadcast to the real-time Classification program. Simultaneously, it is recorded to a file for off-line analysis and archiving, to a real-time Results Viewer program, via broadcast across the local TCPIP network.
- **Response Classification:** A listening thread in the Classification program reads data from the data broadcast buffer as soon as it becomes available and passes it to the classifier routine. (**Error! Reference source not found.** shows a snapshot from this program). After calculating the probability of being a target for each image, the information is passed to the Viewer program, which may be running on one or more remote computers.
- **Results Viewer:** After receiving unique image ID numbers along with calculated probabilities from the Classification program through the NERC TCP/IP daemon, the Results Viewer program matches these ID numbers to a previously read image database and finds corresponding locations on the wide-area image from which the image clips

¹ Variete© , originally developed by EEG Solutions, LLC, Andrey Vankov, principal.

² ADAPT© , originally developed EEG Solutions, LLC

have been extracted. The Results Viewer then generates a probability map using image clip probabilities and locations. This map is translated into a colored image and superimposed on the large area map using user-selected color-map and transparency options. The operator can move around it and its superimposed target probability map and has the option of zooming in or out of this image. The Results Viewer program can handle large images (its comfortable interactivity has been tested for images up to 4K pixels in each dimension), and produces high-quality anti-aliased pictures even at high zoom levels. Because of the overall speed of the system, subject is able to view likely target positions in the large area image a few seconds after the end of each presented RSVP burst. (Figure 2 shows a screen snapshot of the Viewer program).

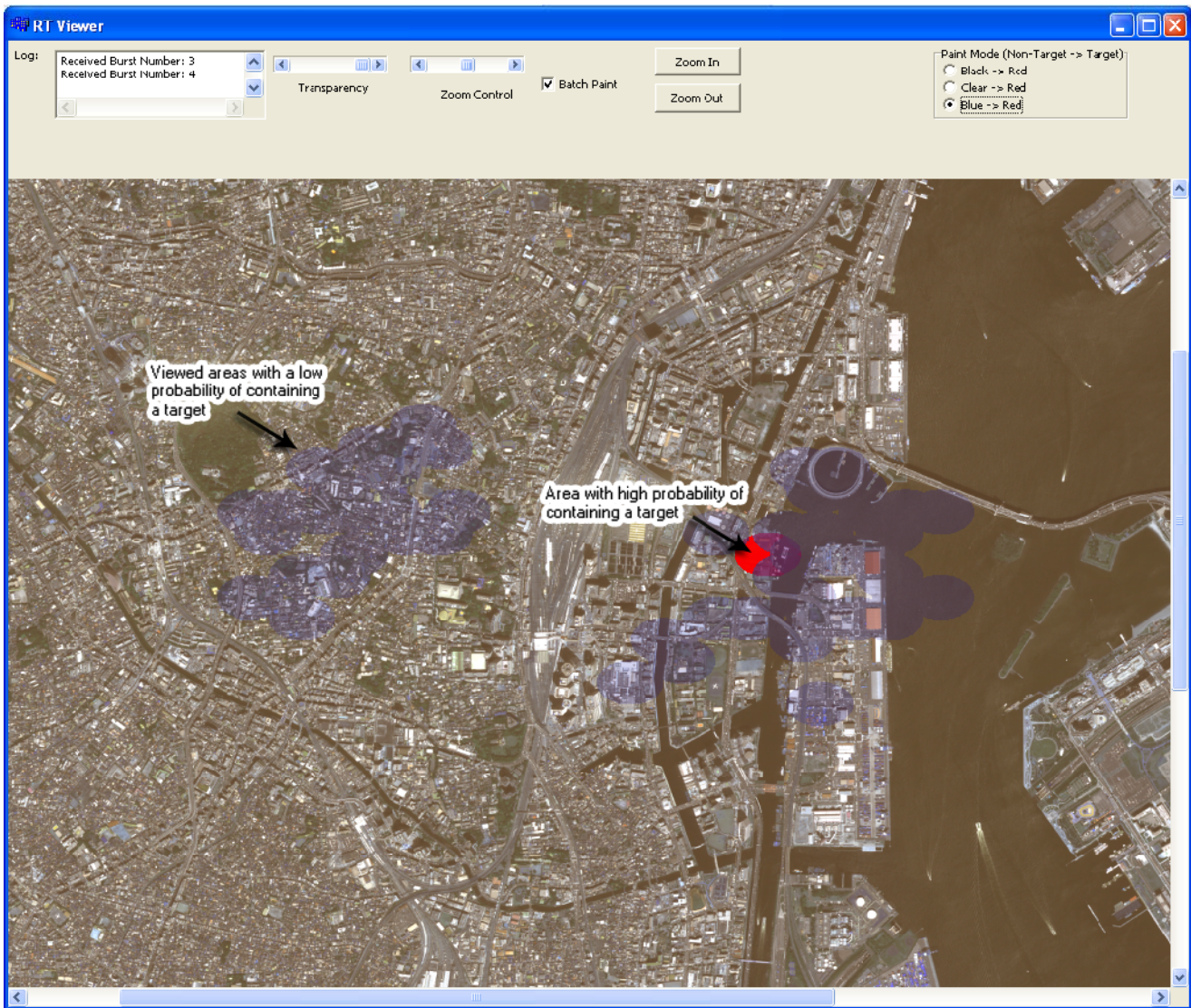


Figure 2. A snapshot of the Results Viewer program in operation.

A video demonstrating our NERC system at work in a NIA demo experiment involving an IA subject is located at: http://sccn.ucsd.edu/pub/nia/nia_team6_ucsd_demo.mov

1.3 Goals of the Research

The research goals were to build, refine, and test a networked EEG response classifier (NERC) system that allows efficient ‘triage’ of a large portion of a broad-area satellite image as unlikely to contain targets, allowing the IA to concentrate their time for performing broad-area search only in searching the portions of the broad-area image most likely to contain a target, based on a rapid initial RSVP search.

1.4 Participants

1.4.1 Participant characteristics

1.4.1.1 Primary focus: IA subjects (Metric experiments)

1.4.1.1.1 demographics, years of experience in the field, subject-matter expertise

Subject	Age	Gender	Years of experience	subject-matter expertise
1	30	male	NA	NA
2	33	male	NA	NA
3	44	male	2 years	homeland security

Table 1. The three IA subjects.

1.4.1.2 Secondary Focus: Pilot/novice subjects

Our pilot experiment subjects were mostly graduate students at UCSD. Later experiment mostly used professional GIS image analysts recruited from the community, plus one female graduate student who studied microscopic images in her research and exhibited very high levels of behavioral and EEG-based image classification in our RSVP testing.

1.4.2 Facilities

Our experiments were conducted in our Swartz Center EEG research laboratory located in a medical building La Jolla CA adjacent to the UCSD campus (<http://sccn.ucsd.edu>). Subjects sat in a standard office chair and viewed a CRT computer screen, giving a behavioral response at the end of each burst presentation and then prompting the program to continue to the next burst using a hand-held finger button. The experiments used low-level illumination in the experiment room, though this is unlikely to be crucial to system performance

1.5 Experimental Design

All our experiments consisted of two or more sessions collected on different days, the first a model Training session, and the second a real-time system Testing session. Our metric experiments on the three IA volunteers also included a Baseline task, conducted just before the Training session. In the Baseline task, the IA subjects were asked to go through a broad-area image similar to but not the same as the broad-area imagery covered in the Training and Test sessions. They were asked simply to report the number of target helipads located in the Baseline task image. The subjects seemed to perform rather quickly and inaccurately in this task, likely

the function of our explicit or implicit task instructions (we had misunderstood that we did not need to run a Baseline condition until just before the IAs arrived). Thus, our Baseline task speed results are likely unrepresentative. Details of the Training and Testing session protocols are given in the descriptions below.

1.5.1 Task and experimental design for Training, Triage Mode, Baseline Search

1.5.1.1 Workstation / Interface

All components of our NERC system run on the Windows XP operating system and are written in C++ (Borland builder environment) to maximize execution speed. During Testing sessions, classifier data (ICA matrix, feature vectors, LDA classifier weights, channel locations...) are read from files created from EEG and behavioral data collected in a previous Training session by an automated Matlab script that reads the pilot session data, applies ICA, selects informative components, and stores all the model information used in subsequent testing sessions by the NERC system, with no need for manual optimization.

1.5.1.2 Imagery

We used a portion of one of the large images provides by NGA in Triage mode in experiment on IAs (file: /data/nia2/NIA_Experiments_nobmp/Helipad/06JAN05024037-P1BS-005540078010_01_P002/ 06JAN05024037-P1BS-005540078010_01_P002_nup.tif). For non-IAs, we mostly used (/Large Unchipped Tiff/000000059158_01_P001/02MAY02012703-M2AS_R1C1-000000059158_01_P001.tif). We used airplane targets added to non-target image clips in Training sessions for IA Subjects 2 and 3. During the Training session of IA Subject 1, we presented actual helipad target images.

Image Preprocessing: The most important consideration for RSVP image preprocessing is ‘chipping’ or extracting from the broad-area search image the smaller image clips to be shown in each burst at a high presentation rate. There are three desirable criteria for selecting and ordering these image clips:

1. **Context:** One of the main differences between novice and expert analysts is in their use of broader image context. Experts tend to better incorporate contextual information into their search strategies. For example, a nuclear power plant might not look very different from any other structure in a satellite image, but by following power lines and searching for places close to water an expert analyst can easily pinpoint its location. RSVP sequences that present image clips at random do not allow the operator to maintain their trusted sense of relative image location – giving the IA a rapid tour of random trees, but without allowing the analyst to understand which parts of the forest they come from.
2. **Sequence locality:** Target classifiers may not be perfect. In particular, there can be an error in matching EEG signature of target detection to the exact position in the presented RSVP sequence (especially at rates as high as 12 Hz in which the interval between presenting two adjacent images is only 83 ms). Usually, this leads to misclassifying an image that is presented shortly before or after the actual target as a perceived target. If subsequent image clips come from uncorrelated and random locations in the larger image, a small error in calculating the latency of target detection event produces a large error in estimating the

position of the target. But if successive clips are sampled from nearby locations in the large image, a small error in calculated latency leads to a relatively small error in estimating the location of the target. For this reason, retaining locality is a desirable aspect of the image ‘chipping’ process.

3. **Subject acceptability:** A simple way to provide locality in the RSVP sequence is by scanning through the image using stripes in some fixed direction (‘mowing the lawn,’ in IA parlance). Although this approach is simple and efficiently preserves sequence locality, it has the unpleasant and unacceptable side effect of inducing motion sickness when used for high-rate RSVP. If an RSVP-based image triage system is going to be used on a daily basis, this would raise insurmountable user acceptability issues and would reduce analyst behavioral performance as well!

We have addressed these design concerns by adopting the following design principles:

Principle I: Retain image context through scene warping. To include contextual information in our RSVP-based classification system, we developed a scene warping scheme for image chipping that allows continuous display of different spatial scales of a high-resolution image around a small ‘foveal’ search area in each image clip.

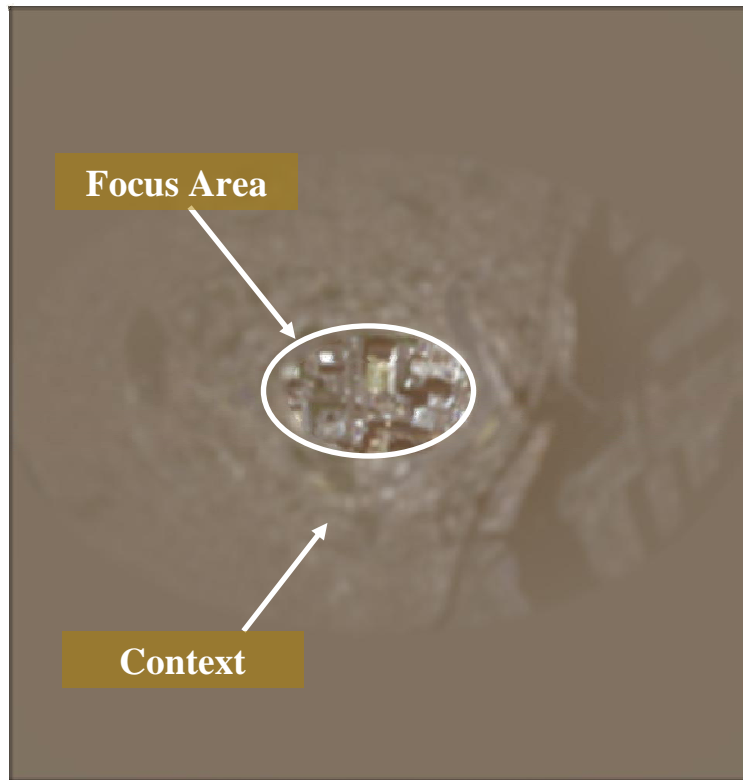


Figure 3. An image clip from an RSVP burst showing scene warping.

In our system, each RSVP image clip consists of two functionally different regions (Figure 3):

- a. Fovea:** This area presents, in high definition, the relatively small portion of the broad-area image to be examined for potential targets by subject.
- b. Contextual Surround:** The area surrounding the foveal area (even encompassing the entire broad-area search image) is presented in ‘warped’ fashion at very low definition and contrast (in practice, even lower-contrast than shown in Figure 3). The subject does not and cannot examine this area for high-definition target features. Instead, it provides peripheral information about the position of the foveal image clip relative to major features of the broad-area image (rivers, cities, etc.).

In an eventual working classification system (not yet implemented in our NERC system, but straightforward to implement in future), each RSVP burst would begin with the same broad-area image, dynamically zooming in (like current Mac and future Microsoft OS graphics) to the scene-warped view like that shown in Figure 3, and then zooming out to the broad-area image again at the end of the RSVP burst.

By this means, the IA will retain knowledge of and visual contact with the wider image context, even during the RSVP bursts. In an advanced system, he or she might also partially direct the locale of each RSVP burst sequence using hand or eye movement control. This would have several benefits. First, the analyst might use their higher-order knowledge and experience to avoid searching unpromising areas. Second, the IA would retain a feeling of control over the search process – a feeling vital to ready acceptance of neurotechnology in the workplace.

A further refinement would place the foveal area not at the center of the warped image clip, but at a location corresponding to the location of the foveal region in the broad-area image. This would allow the analyst to focus on the area of upcoming RSVP search while still viewing the native broad-scale image, without needing to make a saccade to the screen center. The locale of the search could even be selected to be the point-of-gaze of the analyst at the moment he or she presses a ‘RSVP start’ key.

To avoid unintentional search area overlap or gaps in the areas searched, the actual search path produced by such point-of-gaze selection could be based on the underlying, and perfectly tiling heptunx grid.

The scene-warped RSVP image clips in our system are created in three steps:

1. **Warp the broad-area image** around a smaller image focus (for example, an area of 600x600 pixels) centered on, or at least warped around the foveal area of interest as shown in Figure 4.

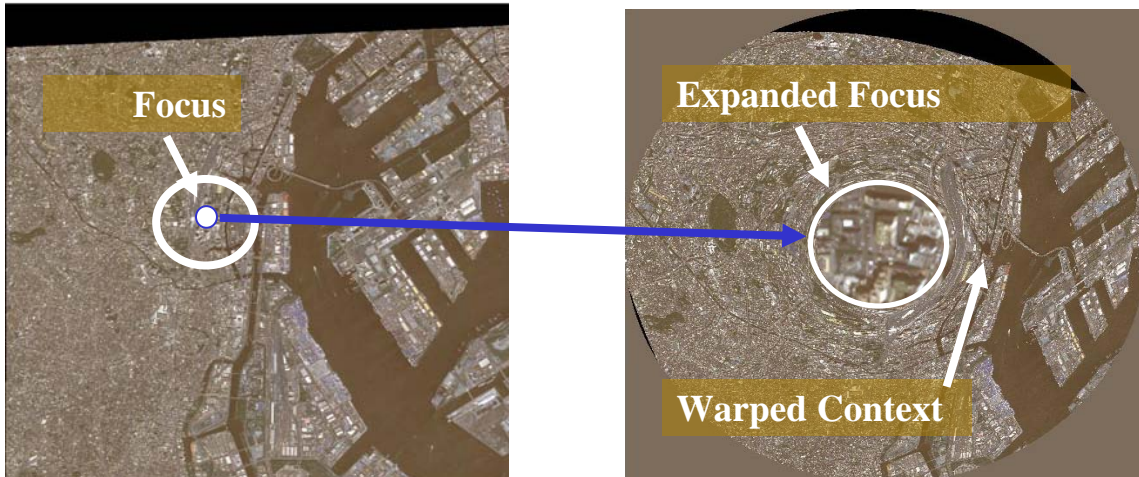


Figure 4. Warping the broad-area image around a selected focus.

2. **Reduce peripheral detail** by blurring and reducing contrast of the warped image sufficient to retain large-scale features such as roads, water, forest, etc., while reducing peripheral RSVP flicker and apparent peripheral movement to a minimum (see Figure 5).

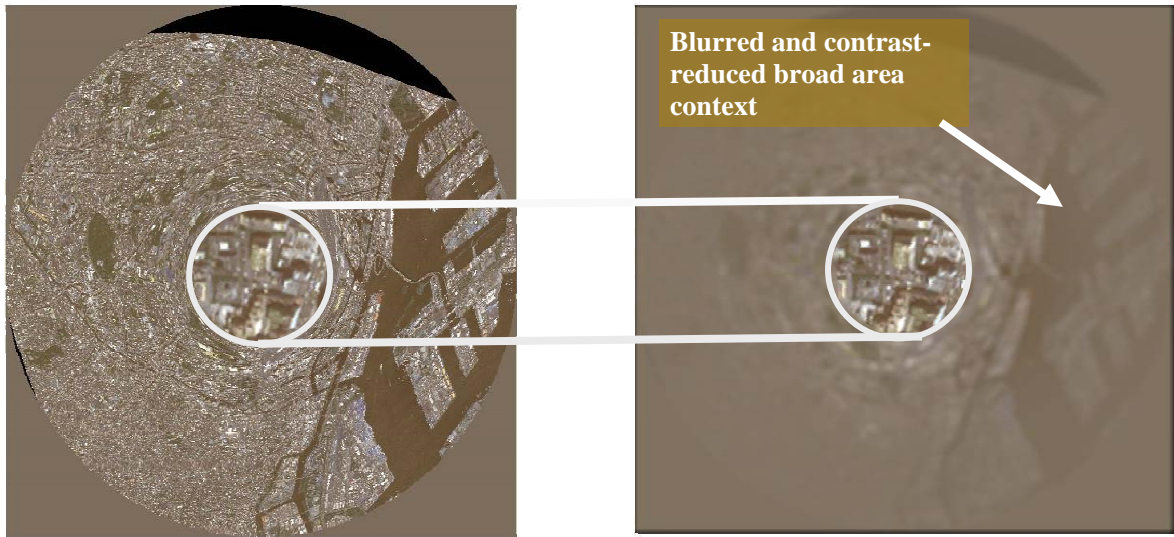


Figure 5. Blurring and contrast-reducing the image clip surround.

Principle II: Adopt the natural geometry of human vision.

The human visual field, retinal receptor distribution and oculomotor range are anisotropic, extending further in the horizontal than in the vertical direction. While saccades under natural viewing conditions preferentially follow cardinal directions, there is also a strong preference for horizontal eye-movements as compared to vertical ones. As a consequence, fixated locations are elliptically distributed, with a constant aspect ratio of about 1.6 (horizontal/vertical) for a variety of naturalistic tasks and display sizes (Osterberg, 1935; Guitton, 1984; Einhauser-Treyer et al., unpublished). Hence, we hypothesize, an elliptical

field of view may feel more natural and may be more effective for RSVP target feature search by human observers.

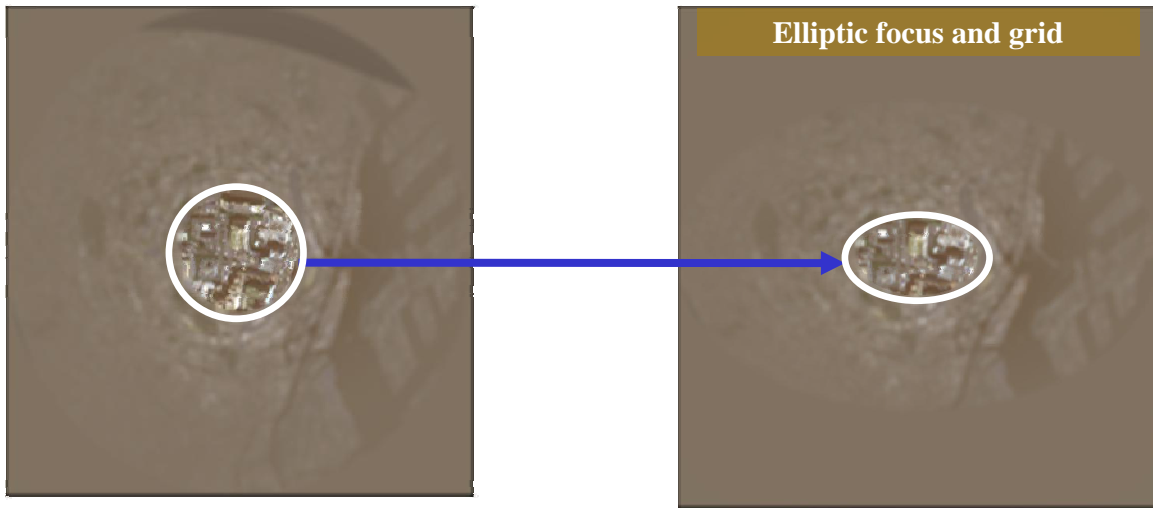


Figure 6. Elliptic focus area.

Based on this assumption, we use an elliptical foveal area with eccentricity of 1.6 (see **Error! Reference source not found.**). The grid of points on which the image clips are centred is also horizontally stretched and/or vertically squeezed to equalize overlap between adjacent elliptical focus areas.

Principle III: Use a randomized, self-similar search path on a hexagonal grid. To maximize both image sequence locality and subject acceptability, we devised an innovative method for creating an RSVP image clip sequence from a broad-area image. Heptunx path search is a path-generation algorithm implemented on a 2-D hexagonal (rather than rectangular) grid. It consists of hierarchical self-similar patterns that tile the grid in a non-overlapping manner. The basic idea behind the heptunx-based search strategy is to exhaustively search each general location while preventing motion sickness by minimizing movements in any one direction for more than one image transition.

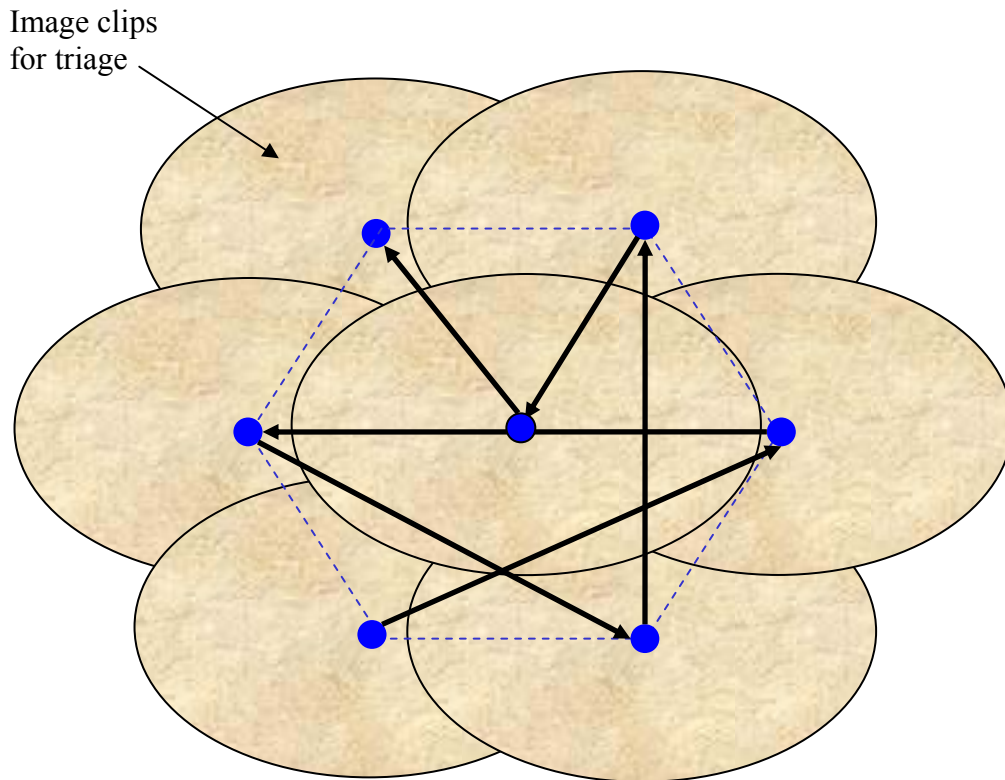


Figure 7. Level-1 heptunx pattern

The basic heptunx shape (level-1 heptunx) is a hexagon with an additional node in the center (see Figure 7) – thus the (new) term ‘heptunx’ (after ‘quincunx’, the 5-dot pattern used on dice, from the Latin word for 5/12). Image clips are created by focusing on small portions of the broad-area image centered on the heptunx nodes. In the heptunx RSVP search path, the 7 clips from each level-1 heptunx are presented consecutively in random order. Historical note: A self-similar path order on the multi-level heptunx grid, noted a generation ago by H. William Gosper, been called ‘Gosper’s flowsnake’ and the multi-scale heptunx pattern it traces, ‘Gosper’s snowflake.’ Gosper’s concept was to start at the largest image scale and burrow down into smaller scales with a self-similar path, as in Peano’s earlier space-filling path built on a Cartesian grid. To minimize the chance of RSVP subject vertigo, we propose, instead, a random RSVP search path at each heptunx scale (see below).

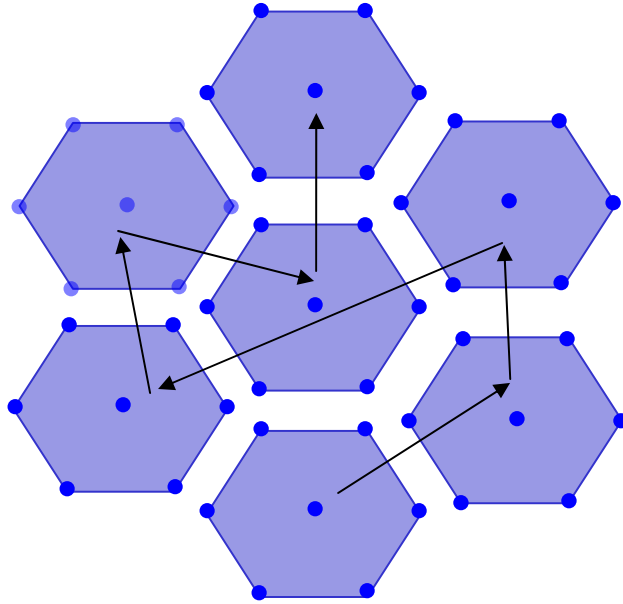


Figure 8. Level-2 heptunx pattern incorporating seven level-1 heptunx patterns

After displaying image clips obtained for a level-1 heptunx, the search path algorithm moves to the super ordinate level-2 heptunx. This scale consists of 7 level-1 heptunx patterns arranged in a similar hexagonal pattern with, again, a seventh level-1 node in the center (see Figure 8). Following the same strategy as at level-1, the algorithm chooses the order of search through its level-1 heptunxes at random, randomly selecting a new search path through each level-1 heptunx image cluster. (Optionally, these orders can be constrained to avoid continued circular motions, though these will be rare under true pseudo-random ordering).

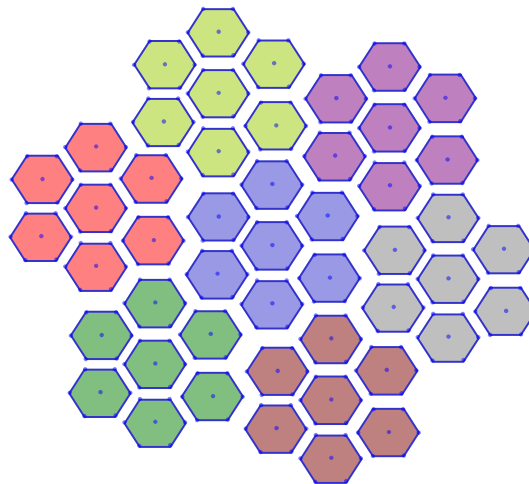


Figure 9. Level-3 heptunx pattern incorporating seven level-2 heptunx.

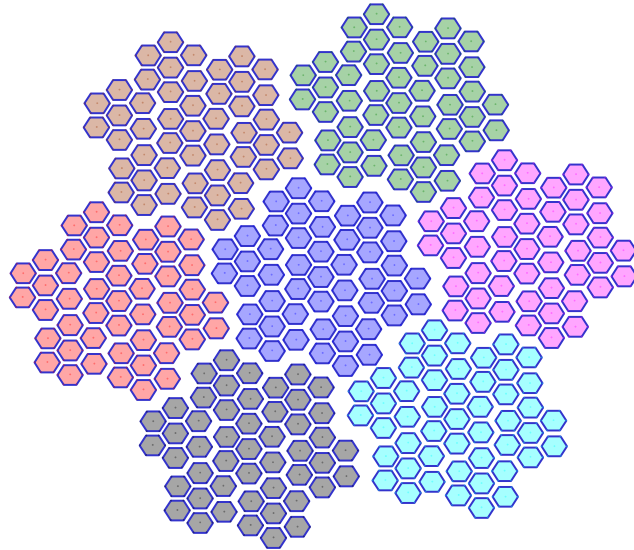


Figure 10. Level-4 heptunx incorporating seven level-3 heptunx. Note: The appearance of gaps between the level-3 heptunx patterns here is artificial; the level-3 heptunx again evenly tile the hexagonal grid.

Higher-level heptunx patterns are constructed in a similar manner (see **Figure 9** and Figure 10). The multi-scale heptunx patterning can grow to any desired size to tile an arbitrarily large broad-area image. Since the algorithm exhausts each scale before moving to the next-higher one, consecutive images presented during RSVP tend to come from the same neighborhood in the large image.

The resulting heptunx search path resembles fractal trajectories observed in animal food foraging and mimics qualities of gaze paths used by subjects in 2-D visual search tasks, while maximizing the stability of REVSP image context without inducing motion sickness. The hierarchical heptunx structure also minimizes the number of large jumps in the search path that may introduce radical changes in texture (for example jumping from desert to urban areas). These sudden alterations can induce EEG signature of surprise. Since this signature shares similar features with signature of target detection, it may confuse the EEG classifier and reduce response classification performance.

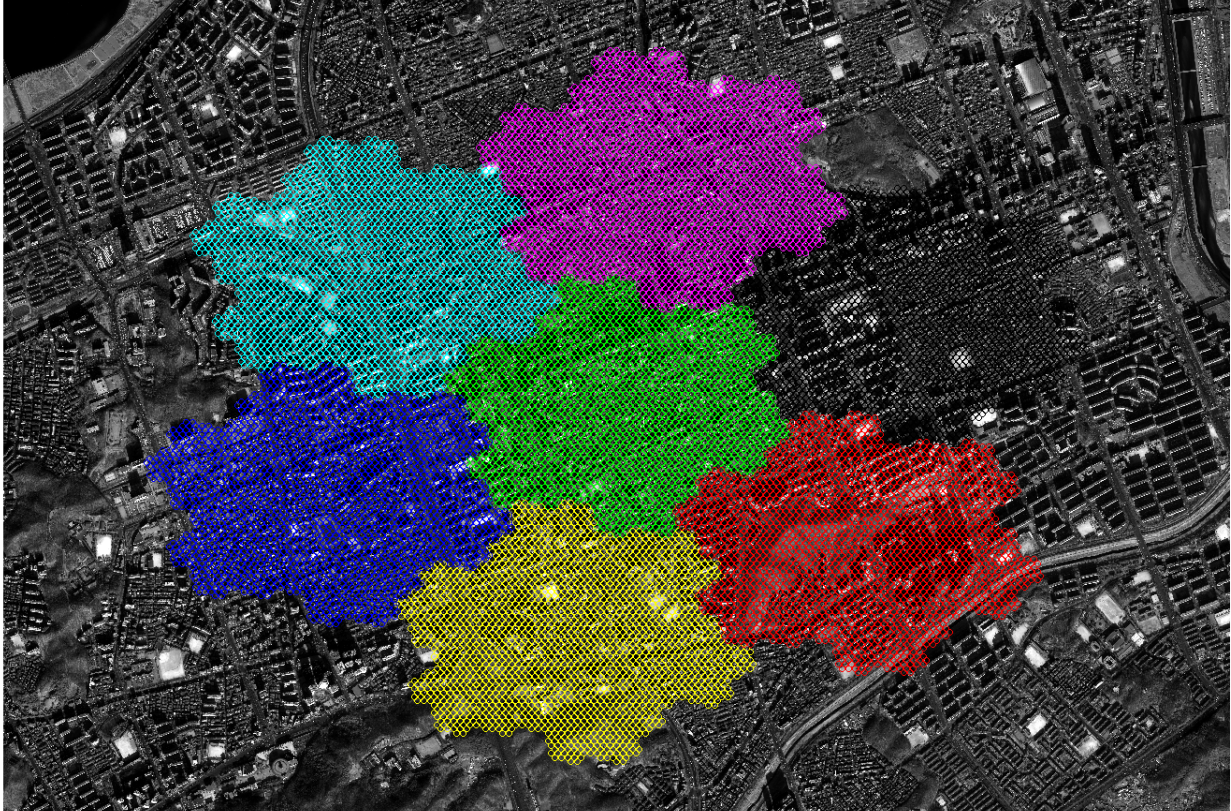


Figure 11. Level-5 heptunx (incorporating seven level-4 heptunx patterns) used to create image clips from the background broad-area image in the IA metric experiments.

In our metric experiment on IA subjects, we used a level-5 heptunx pattern to construct the image clip path through a large portion of the broad-area image (see Figure 11). This consisted of $7^5 = 16,807$ foveal image clips. Each RSVP burst in these experiments represented a search through one 49-clip level-2 heptunx. Thus, $7^3 = 343$ RSVP bursts were required to search through the level-5 search area.

More complex heptunx search paths can also be generated. For example, one can define a similarity measure between different image clips (for example, based of likeness between their brightness or texture) and solve the ‘traveling salesman’ problem to find a hierarchic heptunx search path that maximizes the sum of this measure at a desired heptunx level. In certain circumstances, this could produce search paths that are ‘easier on the eye’ (i.e., having fewer sudden brightness or texture changes).



Figure 12. Combining heptunx path and image segmentation.

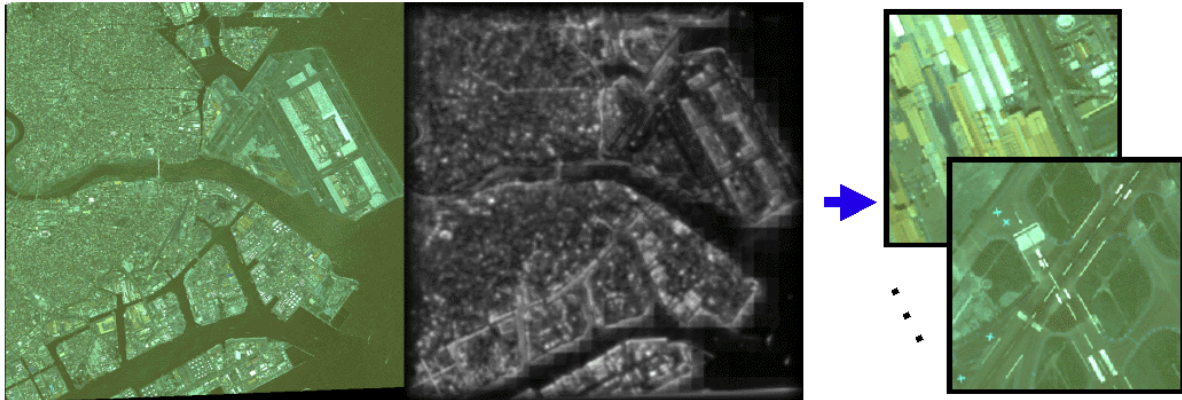
Another improvement we have experimented with is to combine building of the heptunx-based search path with image segmentation. There may be areas within the broad-area image that are unlikely to contain the specified target. For example, an active military tank should not appear in a lake. These areas can be segmented out using relatively simple image processing operations. Figure 12 shows such an example. We did not apply image segmentation to images used in our metric experiments because of the short amount of time available for image preprocessing. However, we recommend that this option be more fully explored in future research, as detailed below.

Image surprise. Identifying un-surprising (i.e. ‘boringly similar’) regions of a large image can be used to prune the number of locations to be searched. To this end, we (Itti and colleagues, USC) developed a tool that allows one to identify the more interesting or surprising regions in a large image, and to generate a path through these regions that optimizes the traveling salesman problem (i.e., to minimize the total distance along the path while visiting each surprising location once). The rationale for attempting to find the shortest route is to minimize any feeling of spatial disorientation or vertigo the analyst may experience while viewing the resulting RSVP sequence.

In practice, however, the image clips useful for high-rate RSVP proved to be small and disjoint enough that this technique did not exhibit a definitive advantage in preliminary testing.

In addition to looking for visually surprising scene elements, therefore, we have also developed a technique by which the system can be “biased top-down” to more predominantly attend to locations that resemble visual features of interest (Figure 13)

Find anything surprising:



Emphasize round buildings, suppress water:

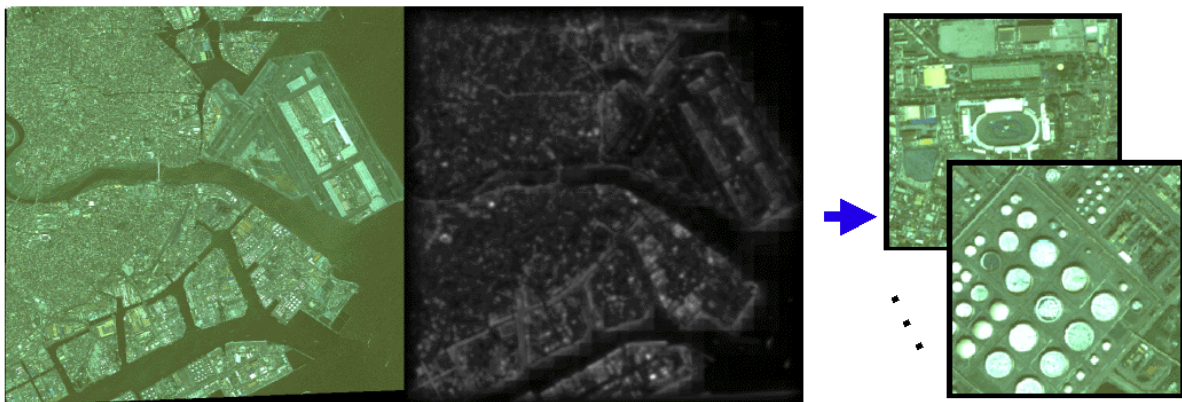


Figure 13. Computing a surprise map (grayscale images in middle column) from a satellite image (green, left) and using it to automatically extract ‘interesting’ image clips (right). The system can either operate in a default mode (Extract all clips containing anything visually surprising, top) or in a biased mode (Find surprising clips emphasizing round buildings, bottom).

Mitigating misses by regulating surprise detection. We hypothesized that surprising events occurring during the presentation of an RSVP sequence, such as for example abrupt changes in overall luminance across successive clips, or the appearance of a rare color, may parasitically capture the attention of the viewer, thereby effectively reducing the attentional resources available to process this and subsequent frames. In effect, hence, low-level visually surprising events may mask subsequent targets, an effect similar to that known in the psychophysics literature as the “attentional blink”. RSVP sequences of overhead imagery clips, however, are very different from computer-generated laboratory stimuli typically used for psychophysical testing. Hence, it was unclear whether an attentional blink effect would be observed with our more realistic and specific stimuli.

We therefore designed an experiment to address the question of which stages in the visual processing hierarchy limit performance in an RSVP target detection task – early attentional selection, later object recognition, the decision process itself, random fluctuations in alertness, or other factors. To investigate the relative contributions of attention, recognition, and other processes, eight human observers performed an animal detection task in sequences of natural images presented at 20 Hz (see Figure 14). Observers were highly consistent, with the number of the target images that were detected by all the observers being 15 times higher than expected if observer errors had been independent. This is compatible with operator performance in any trial depending primarily on visual processing, including attentional selection and recognition, rather than on other, more idiosyncratic, factors. Two statistical properties were found to significantly predict performance on individual sequences: contrast-variation within the target image and an information-theoretic measure of “surprise” in the adjacent images in the sequence.

In a second experiment, we re-ordered the sequences of images that all subjects had correctly recognized so as to elicit increased “surprise” before and/or after the target. Whereas a new set of observers still readily detected the targets in the original sequences, surprise enhancement causally impaired target detection significantly across all observers (see Figure 15). Hence, *visually “surprising” events efficiently mask adjacent items in the RSVP sequence, even when the target images themselves remain unchanged.* Consequently, and contrary to several previously published findings, our results demonstrate that attentional limitations, rather than target recognition alone, affect the detection of targets in rapidly presented visual streams, a result we believe is fundamental to the use and design of RSVP sequences for target response recognition.

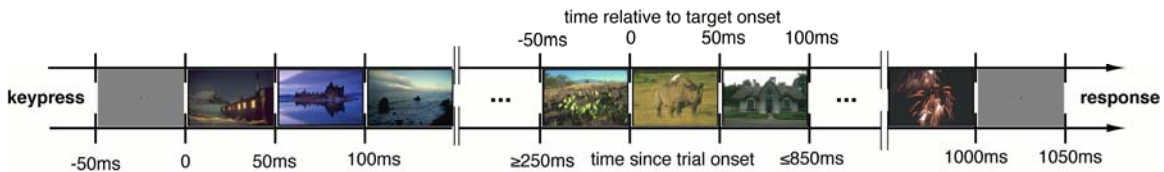


Figure 14. Example sequence. Observer starts a trial by a key press. In target-present sequences, the target (animal) can occur between the 6th and 15th frame, i.e., between 250 ms and 850 ms after trial onset.

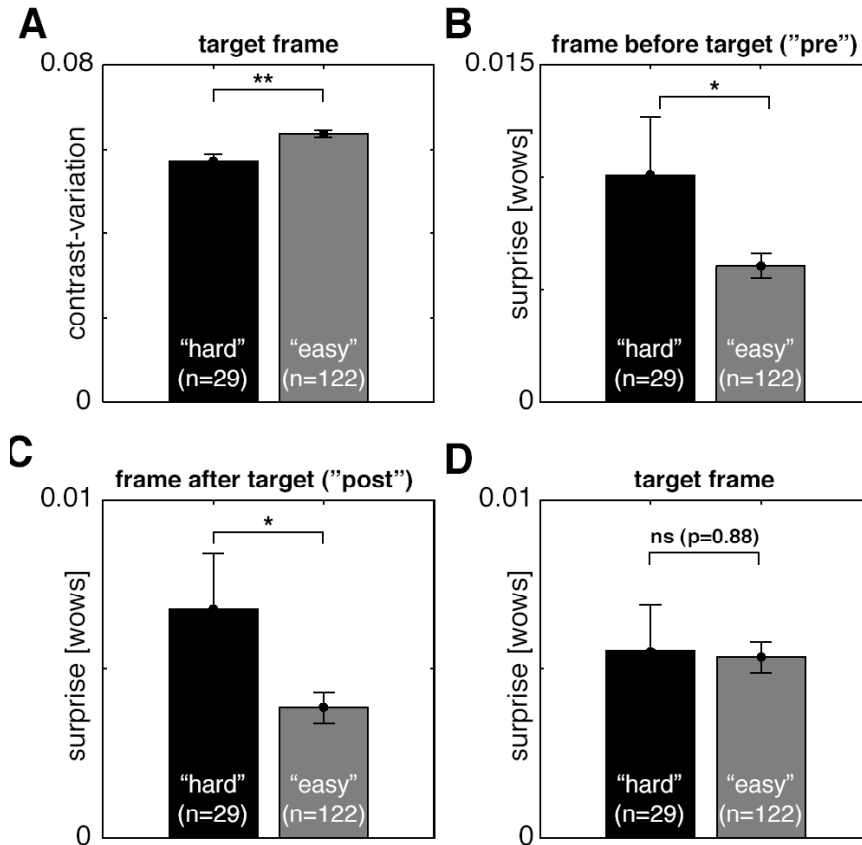


Figure 15. “Easy” sequences (in which all 8 observers correctly detected the target) differ from “hard” sequences (in which none of the observers detected the target): (A) in the contrast-variation in the target-frame, (B) in the degree of visual surprise in the frame preceding or (C) succeeding the target frame, but not (D) in the degree of visual surprise of the target frame itself. All panels depict mean \pm s.e.m. across sequences for the 29 hard (left, black bar) and the 122 easy (right, gray bar) sequences. Significance markers refer to results of t-tests (* $p < 0.05$, ** $p < 0.01$, * $p < 0.001$).**

This is a very important finding for projects using RSVP detection, as it shows that observers are essentially “blind” for a brief period of time following a visually surprising event. There are several things that can be done to mitigate this problem: (1) re-order the RSVP sequences so as to minimize visual surprise events arising from abrupt changes in luminance, color, contrast, etc.; (2) recycle images that appear just after a visual surprise, showing them again in another RSVP sequence; (3) attempt to alter the images so as to preserve their semantic content as much as possible and still allow target recognition, while controlling their visual surprise. Prospects for controlling the amount of visual surprise produced by a dynamic visual display is further discussed below.

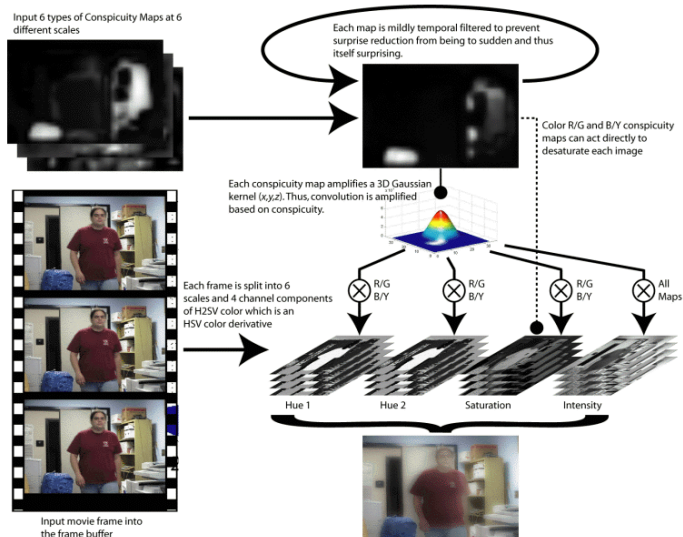
Principle IV: Filter the appearance of image clips using surprise modulation. Modifying image sequences to reduce the number and strength of parasitic visual surprise events is a complicated endeavor, but we have made significant progress in approaching it. We have developed a system that can detect, in a given RSVP or movie sequence, when a low-level

surprise event will occur, find the main reason for this surprise (e.g., an abrupt color change or a sudden upwards movement), and apply various filtering to attempt to reduce the degree of visual surprise (see

Figure 16 and Figure 17) This work is ongoing. It is not possible to invert our surprise model in closed form, so one has to use a number of heuristics and numerical techniques. Yet, it shows great promise, at least in visual inspection that this stage, where movie sequences which may previously have felt flashy, choppy, and overall annoying or tiring to watch tend to flow more smoothly once controlled for degree of visual surprise. A paper is currently under preparation. Further research is necessary in this part of the effort, though.

Reducing low-level surprise in RSVP sequences

Motivation: Avoid that analysts be distracted by low-level surprise events in RSVP sequences (e.g., abrupt changes in image luminance) so that they can focus more on semantics and high-level surprise.



Mundhenk & Itti,
in prep

Figure 16. Architecture of the surprise control system. Surprising events are detected in an RSVP or video sequence (here demonstrated by a movie clip rather than an RSVP stream), and the surprise model is “inverted” to yield image filters that reduce visual surprise while preserving overall intelligibility.

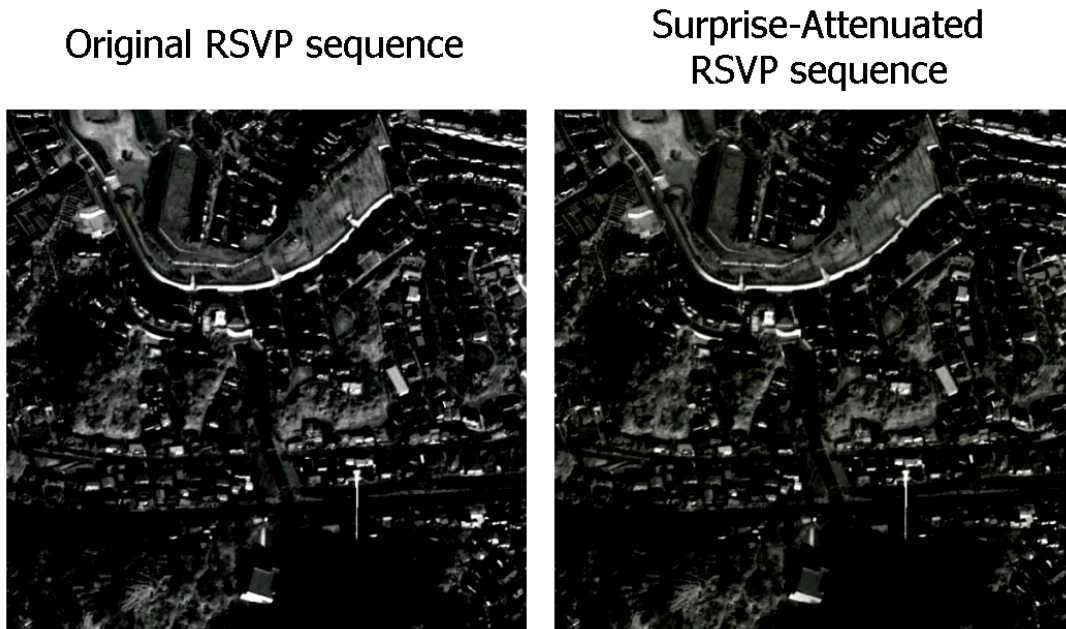


Figure 17. Comparison between an original image clip from an RSVP sequence (left) and its surprise-controlled version (right). Note how some contrasts and long contours or bright buildings have been significantly attenuated so as to provide a better “blending in” of this image with the previous and subsequent ones in the RSVP sequence. Yet, intelligibility (i.e., ability to recognize objects in the scene) is little affected.

The Attentional Blink. Several attentional limitations can limit RSVP detection performance – Repetition blindness (Kanwisher, 1987), the inability to detect a second of two identical items in direct succession, and the Attentional Blink (AB; Raymond, Shapiro, & Arnell, 1992), the impaired ability to recognize a second target image following closely, but not instantly, another target image. Unlike repetition blindness, the AB targets do not need to be identical; they merely need to be defined as targets distinct from the distractors (for details on dissociating repetition blindness from the AB, see Chun, 1997). For NIA RSVP application, repetition blindness does not present a major issue, as long as the presentation strategy preserves some topology of the search array, as it is the case in the heptunx path scheme: targets potentially missed in RSVP will be those that appear closely following a detected target in the RSVP sequence.

Experimental study. As most studies on the AB had either used artificial stimuli or had focused on superficial detection rather than detailed (exemplar) recognition (Evans & Treisman, 2005), we performed a pilot experiment using complex target items (watches, faces) embedded in visually similarly complex distractors. In this setting, we found a consistent AB, whose length

and duration depended on stimulus category (Einhauser, Koch & Makeig, 2007, see attached). Significant impairment of target recognition can last over half a second after a first target.

If target density is high (or targets are clustered in few regions) and the RSVP presentation scheme does not take the potential effect of the AB into account, a substantial number of targets could be missed. As the AB is category dependent, there is only limited possibility to predict which frames could be subject to the blink. Only by using an integrated system of hierarchically organized search patterns (as in heptunx search), combined with manipulation of visual surprise (such as described above) can the effects of the AB on real-time target classification and interactive viewing be minimized.

1.5.1.2.1 Presentation parameters

After trying different RSVP rates in a behavioral paradigm (Einhauser-Treyer et al., 2007), we decided to use 12 Hz as the RSVP presentation rate for the metric experiments. Each 4-s burst of 51 images contained 49 images forming a level-2 heptunx pattern, plus 2 ‘dummy’ images and the end of the burst, ignored by the classification system, to prevent trailing burst edge effects.

Since the first 49 images from every burst formed a level-2 heptunx (located in a compact section of the broad-area image), all large jumps between distant areas of the broad-area image occur between bursts. This greatly reduces sudden changes in presented RSVP image sequences, increasing subject convenience and preventing the potential negative effects of such scene alterations.

Principle V: Minimize appearance of successive targets. In our pilot experiments each burst could contain up to only one target. Targets were also restricted from being in the first or last 500 ms of each burst. This was to prevent burst edge EEG signature to influence the classifier. As our classifier improved, we relaxed these restrictions and allowed for multiple targets to appear in each burst at any latency.

1.5.1.3 Sensor Setup

Our demonstration NERC system uses a BIOSEMI ActiView-2 system with 256 channels, mounted in a whole-head elastic electrode cap (E-Cap, Inc) with a near-uniform custom montage across the scalp, neck, and bony parts of the upper face. As explained below, our real-time classifier system does not use data from more than 128 of the electrodes. The remainder of the data are stored for later review and numerical experiment. The Biosemi recording system has the advantage of high density, high resolution (synchronous 24-bit A/Ds), and active-electrode chip technology that reduces artifacts from moving electrode cables. Computer data acquisition is performed via USB using a customized acquisition driver.

1.5.1.4 Recording Parameters

Data is collected and recorded to file from all 256 electrodes at 256 Hz sampling rate. Currently we use approximately half of these electrode sites (about 128 channels) for real-time analysis and classification. This is, in part, because classification is performed on ICA component activations and the length of recorded Training session data (~1 hour) places a limit on the number of these components that can be learned accurately from the data using present algorithms.

Are so many channels necessary for good NERC system performance? We would ‘guestimate’ that an ICA-based NERC system using as few as 32 channels could be effective, as well as possibly being cheaper to build, wear, and maintain – at the cost of spatiotemporal resolution and, hence, some degree of classification accuracy. Such trade-offs could be studied in future using offline re-analysis of existing high-density data sets.

1.5.1.5 Data Handling Procedures

In our NERC system, EEG preprocessing consists of the following steps:

Channel re-referencing. We re-reference the active-reference Biosemi EEG data to one electrode for noise reduction (a claimed SNR increase of 40 dB for the Biosemi system).

Frequency-band filtering. We perform three types of frequency-domain filtering on the EEG data: (1) IIR High-pass filtering (using a Butterworth filter of order 2) at 2 Hz to remove low frequency trends. (2) IIR Low-pass (Butterworth of order 3) at 50 Hz to remove high frequency noise. (3) IIR Band-reject (Butterworth of order 3) between 40 Hz and 80 Hz to minimize 60-Hz line noise. Infinite impulse-response (IIR) filters with some desired frequency response are usually much shorter than finite impulse-response (FIR) filters with the same frequency response. In addition to reducing computation time, this eliminates the need to place incoming data in a long buffer (0.5 to 1 seconds) before it can be filtered by FIR filter, minimizing the delay in classification caused by filtering alone. However, the use of IIR filters has two potential drawbacks: nonlinearity in phase response and possible stability issues. Using low IIR filter orders (2 or 3) we can avoid the stability problem, even in the presence of occasional high-amplitude muscle or movement artifacts. To avoid performance loss in the classifier from phase distortions, the same IIR filters are implemented in our offline Matlab routines for individual subject classifier training from Training session data.

Channel selection. Next, the system selects and retains only those channels selected for ICA decomposition in the signal model learned from the Training data for this subject. These typically amount to about half of the recorded 256 channels.

Bad-channel detection and compensation.

An operational real-time system must offer some degree of error tolerance to guarantee its sustained operation. For an EEG-based classification system, the chief cause of problems may be inconsistency of the EEG acquisition. The 256-channel BIOSEMI system we are using is, in our experience, quite reliable compared to standard EEG recording systems. However, there are regularly some electrodes, among this large number, that lose connectivity during the course of the experiment, possibly from subject movements or insecure contact with the scalp. An important practical problem for an EEG-based classifier is that a target-response signature learned from an earlier training session cannot be directly applied in the testing session that records data from a somewhat different subset of ‘good’ electrode channels.

To deal with this problem, the NERC system detects electrodes whose signals become ‘bad’ (i.e., unreliable), and replaces their activities with interpolated signals based on the previously-learned ICA sources for the subject. Thorough offline testing has indicated that the system can maintain accurate response classification even when the ten most informative electrode signals become completely disconnected during the testing session.

When a channel becomes disconnected, its activity shows little correlation with signals from nearby electrodes. The real-time bad-channel detection routine therefore calculates the correlation between each channel and nearby channels every 0.5 seconds. If the maximum of absolute correlation between each channel signal and the signals of the other channels within a 5-cm radius drops below a threshold, it is flagged as a ‘bad’ channel. Once a channel is flagged as ‘bad’, the interpolation mechanism replaces its activity with interpolated activity based on the continuously computed ICA component activations. (Absolute correlation is used here as a computationally practicable estimate of mutual information between channels).

Some channel signals may continuously alternate between ‘good’ and ‘bad’ states. As the bad-channel detection routine checks channels only every 0.5 seconds (to conserve CPU usage), this alternating behavior could result in inclusion of noisy channel activity between bad-channel checks. To solve this problem we introduced 7-s hysteresis delay in clearing ‘bad channel’ flags. This means that a ‘bad channel’ state is only revoked for a data channel when the detection routine finds that the electrode activity has remained acceptable for at least 7 seconds. Once a channel signal is flagged as ‘bad’, the interpolation mechanism replaces its activity with interpolated activity based on the stored ICA decomposition for this subject. The algorithm is summarized in the following steps:

1. Eliminate the rows of the mixing matrix associated with noisy channels. This mixing matrix is computed from a previous recording (the training session). The new mixing matrix is referred to as the truncated mixing matrix.

$$\mathbf{M}_{trunc} = \mathbf{M}_{train}; \quad \mathbf{M}_{trunc}(noisy,:) = []$$

2. Compute the noise regularized pseudo-inverse of the truncated mixing matrix to obtain a truncated unmixing matrix.

$$\mathbf{U}_{trunc} = \mathbf{M}_{trunc}^+$$

3. Obtain a new set of activation functions using the truncated unmixing matrix and the new data (testing session).

$$\mathbf{A}_{trunc} = \mathbf{U}_{trunc} \mathbf{D}_{test}$$

4. Compute a virtual data matrix by multiplying the original mixing matrix and the new set of component activation functions.

$$\mathbf{D}_{virtual} = \mathbf{M}_{train} \mathbf{A}_{trunc}$$

5. Replace the noisy channels with the corresponding rows of the virtual data matrix.

$$\mathbf{D}_{denoised}(noisy,:) = \mathbf{D}_{virtual}(noisy,:)$$

In our testing, we found this algorithm to give better classification performance than channel replacement using interpolation based on nearest-neighbors channels or spherical splines. Importantly, our method can be successfully applied in cases even when a local group of channels covering a large area become noisy, since the projections of independent components to the scalp surface via volume conduction are typically quite broad.

Independent component decomposition. Next, our NERC system performs extended infomax ICA unmixing of the selected channel subset using the stored ICA unmixing matrix learned from the Training session data for this subject by an automated subject model training Matlab script.

The results are time series of activity for each of the retained independent component sources to be used in the classification algorithm. Essentially, multiplication of the data with the ICA unmixing matrix performs spatial filtering on the data, filtering for the activities of cortical or other EEG sources that are mixed together in the scalp channel data.

Analysis Methods

1.6 *Signal properties*

1.6.1 Time, Frequency, Space

The neurophysiological relevance of our ICA approach. When building an EEG-based classifier, one can adopt a ‘black box’ approach, directly using the electric potential signals measured on the scalp, e.g., to try to predict whether an analyst has seen a target or not. Indeed, as demonstrated by many groups, the EEG features obtained at the scalp level have discriminative information usable for classification. However, this black box approach ignores the fact that the signals measured by the EEG electrode array are mixtures of many underlying source signals located in many different brain regions. Thus, although measurements have information from many brain regions, the local dynamics cannot be used separately for classification since the signals are linearly mixed. By contrast, we first use blind source separation in the form of ICA to extract the underlying source signals, and treat these separately to extract their source-specific informative features. Each independent source signal reflects the coherent activity of a local area of neuropile oscillating collectively. Thus, our ICA source activity approach (first discussed in Delorme and Makeig, 2000; see also Delorme and Makeig, 2003) represents a clear departure from most approaches used for brain computer interfaces and other neurotechnologies that ignore the underlying neurophysiology and work directly with the sensor-level signals.

To extract information from the independent component (IC) time courses, after decomposing the data with infomax ICA, we cut each IC activity time-course into overlapping 800 ms time windows time-locked to each image and perform a ‘second-level’ ICA on the event-related epochs for each IC to learn a basis of features for its dynamics both in the time domain, or applied to the normalized spectrograms of the event-related epochs, in the time-frequency domain. A key finding here is that *the sensor maps of the independent components with the most informative features are highly dipolar.*

There is no *a priori* reason why ICA should learn such dipolar maps, which are clearly compatible with their neurophysiologically plausible generation by neural activity within a focal brain region, or within bilaterally symmetric focal regions. Moreover, there is no *a priori* reason why the components with the most informative features for target response classification should be dipolar. Thus, the fact that this happens strongly suggests that our classifier is using signals that directly reflect the dynamics of *functionally distinct cortical networks*. In Figure 18 (below) the scalp maps of the most informative components of a recording of one of our 3 image analysts are shown. Notice how they have dipolar (e.g., simple, smoothly projecting) scalp maps. Many

of them are consistent with the forward fields that would be expected from focal sources in frontal, parietal, occipital and temporal regions.

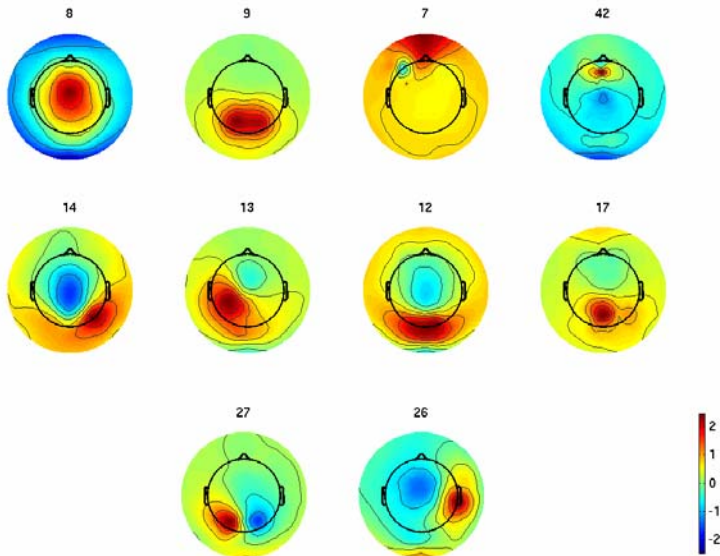
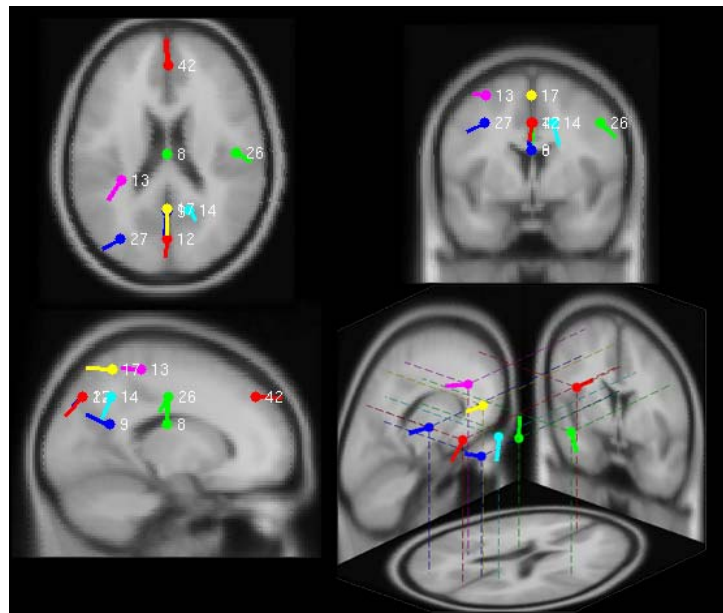


Figure 18. Scalp map of most informative ICA components for IA Subject 3.

In Figure 19 (below), we localize these components using a single dipole source model and a standardized BEM forward model. Although, more accurate and specific localization results could be obtained using more sophisticated inverse and forward head models, these preliminary results clearly suggest that the informative independent components are

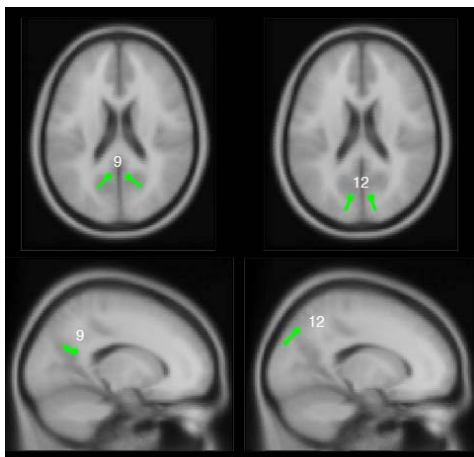
associated with sources located in brain regions that have been implicated in target detection.

Figure 19. Locations of equivalent dipoles for the most informative ICs for Subject 3 (Figure 18) indicating the regional origins of the component generators.



The scalp maps of the most informative ICs for IA Subject 2 share the same ‘dipolarity’ property (Figure 19 and Figure 20).

Figure 20. Posterior bilaterally-symmetric equivalent dipole models for IC9 and IC9, also among the most



informative ICs for IA Subject 3 (see Figure 18).

Information concerning image classification appears at a wide range of latencies in the component activities. Figure 21 (below) shows the time course of mutual information

(MI) between the activity value of the informative ICs for Subject 3 (Figure 18) and target classification. IC14 (magenta) shows a small peak in MI after 150 ms, while major increases in MI appear centered near 350 ms after target onset and then at 550 ms.

The time courses of time-domain MI between the activities of the most informative ICs for IA Subject 2 and target identity (Figure 23) show the same trends. Note the uniformly low amount of MI in the first 100 ms following target presentation, the early peak (after 190 ms) in two lateral posterior components (IC47, IC107) and the two peaks near 250 and 550 ms after target presentation.

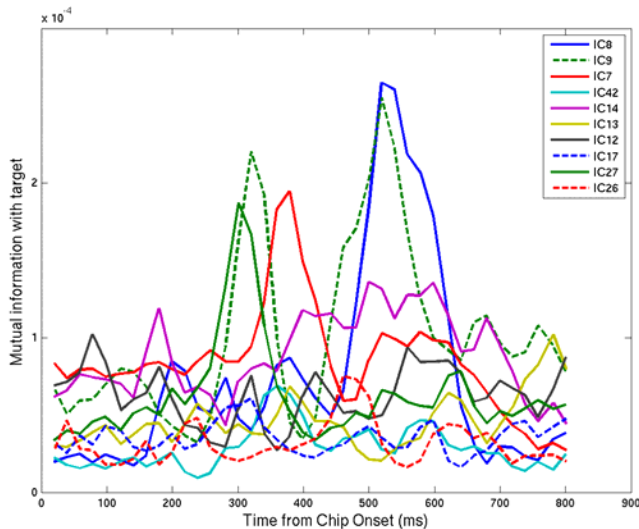


Figure 21. Time course of mutual information between the activity time courses of the most informative ICs for IA Subject 3 (see Figure 18) and target identity.

Note also the apparent homologies in scalp map and most-informative MI time courses between, respectively, IC14, IC8, and IC9 for IA Subject S3, and IC107, IC39, and IC16 for IA Subject 2. To statistically test the generality of this and other component homologies, it will be necessary to run

training and testing sessions on a much larger group of subjects (20 or more). Such an experiment could determine the set of brain areas whose different EEG signatures reliably contribute to target response classification.

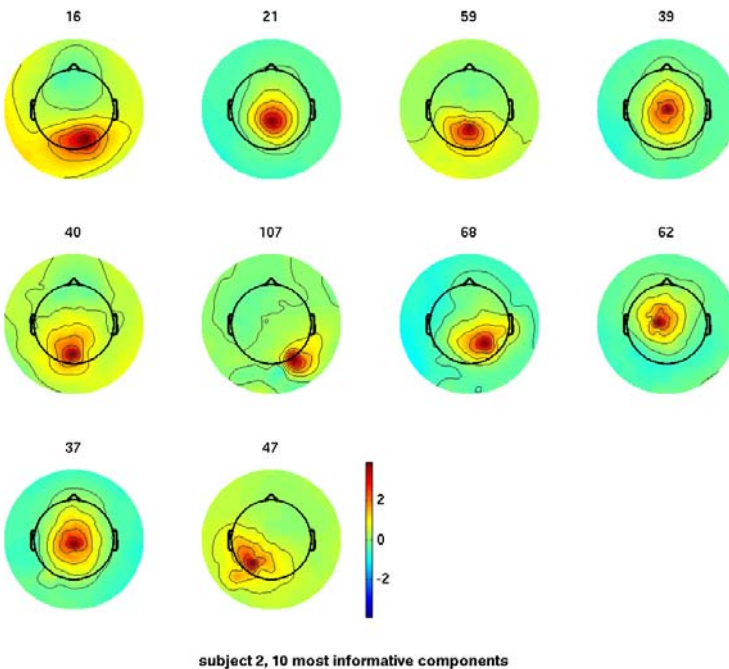
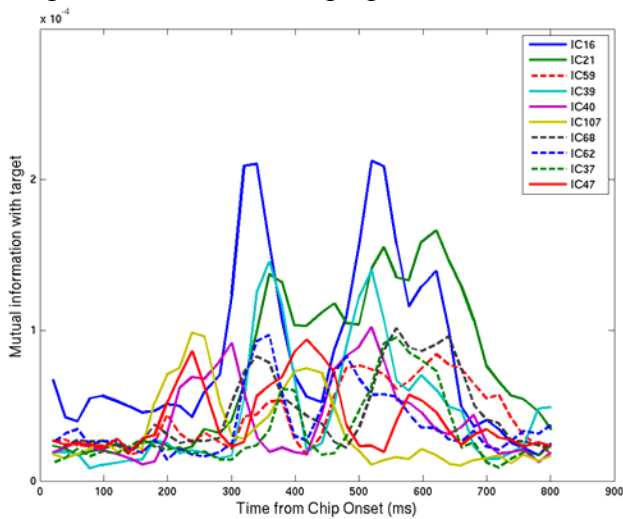


Figure 22. Scalp maps of the most informative ICs for IA Subject 2.

In addition to the ‘dipolarity’ of the most informative independent component scalp maps, the time and time-frequency dynamics of these components also suggest that the ICA-based classifier is using physiologically meaningful features. For example, the activities of the most informative components have time-courses and spectral characteristics consistent with brain rhythms in the theta, alpha, and beta frequency bands well known to be modulated by sensory inputs. The mean event-related time-courses (or component ERPs) of the most

informative components show a clear residual oscillation at alpha for both target and non-target

responses, but show a superposed theta wave only in response to target stimuli (Figure 24). Mean



theta band (4-8 Hz) power is also higher following targets, as shown by the difference of amplitude time-frequency image (not shown). Both of these results are consistent with the EEG and MEG literature implicating P300-like waveforms and transient theta responses as the neural correlate of target detection (Makeig et al., 2004a). Moreover, the learned time and time-frequency templates that are most informative also have P300-like waveforms and transient theta increases in power near 300 ms (see Figure 24).

Figure 23. Time course of mutual information between the activity time courses of the most informative ICs for IA Subject 3 (see Figure 18) and target identity.

The P300 peak and theta-like waveforms of the ERP are usually obtained by averaging the single trials time-locked to an anticipated but infrequent stimulus (Sutton et. al., 1965; Soltani and Knight, 2001). However, the P300 peak has variable latency. For example, for simple stimuli this peak may be around 300 ms post-stimulus, but for more complex stimuli the peak latency is longer. Although peaks in the ERP have been treated traditionally as reflecting the activity of a single “P300” source, there is a evidence that multiple source dynamics located in different brain regions throughout the frontal, parietal, occipital and temporal lobes contribute to these large amplitude theta waveforms, which in single trials may take the form of one- to two-cycle bursts (Makeig et. al., 2004b). In the context of RSVP target-specific responses, these waveforms are expected since targets are infrequent events with direct relevance to the analyst’s task. However, theta waves have also been implicated in many other related processes, such as attention, working memory, and error recognition (e.g., underlying the well-known Error-Related Negativity), (Ramirez et. al., 2005; Onton et. al., 2005; Luu et. al., 2004). These processes all share the common property of top-down influences in relevant stimulus appraisal, which is usually associated with activity in the limbic system and associated cortical areas.

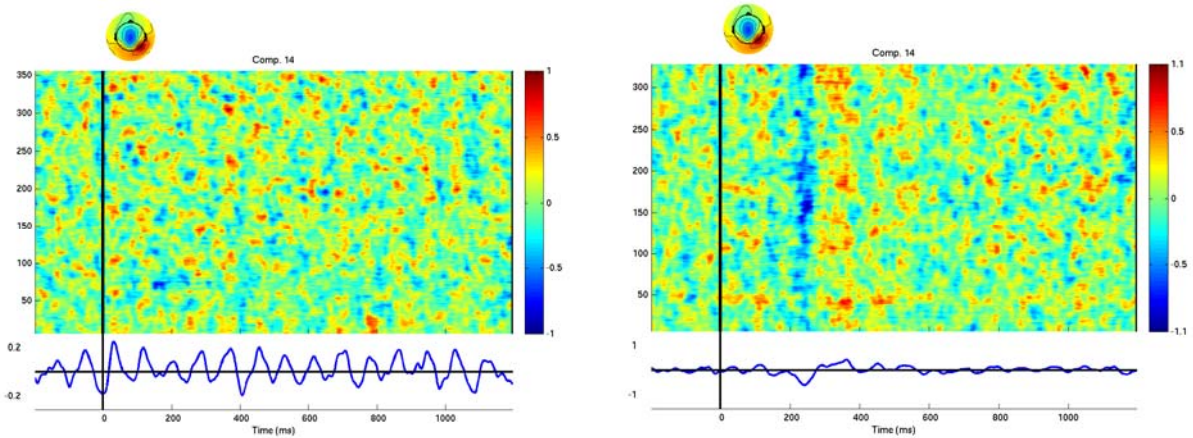


Figure 24. The time courses of activity of IC14 for IA Subject 3 following non-target images (left) and target images (right). The event-related averages (component ERPs) are shown below the color-coded single trials, smoothed vertically. Note the large difference in the vertical ERP scales, the absence of a non-target response in the time domain for this component process, and its clear target response (right) resembling (in the stimulus-locked average) a one-cycle 5-Hz theta oscillation.

1.7 Classification strategy

1.7.1 Rationale

As explained earlier, our classification approach is based on the well-founded assumption that there are numerous cortical areas that exhibit differential dynamics in response to perceived target versus non-target images, and that the information these areas contribute to the EEG scalp signals are largely independent of one another. Systematically applying this assumption has produced steady improvement in our EEG-based classification results across the course of our Phase 1 work. Thus, there is no single ‘flicker of recognition’ in the EEG indexing the recognition by the subject that an image contains a target feature. Rather, there are a large number of quasi-independent features of the activities of a number of cortical EEG sources that contribute to the response difference. As nothing contributing to the classification occurs on the scalp itself, we use an information-based spatial filtering approach (ICA) to separately monitor the major sources of the scalp EEG, combining time-domain and frequency-domain features of their activities into an optimal target probability estimate.

1.7.2 Implementation

To classify targets vs. non-target image clips, we used two Fisher linear classifiers: one acts on time domain features of ICA components and the other on their frequency domain features. Posterior probabilities from these two are combined using Bayesian fusion (multiplied together) to form the final calculated probability.

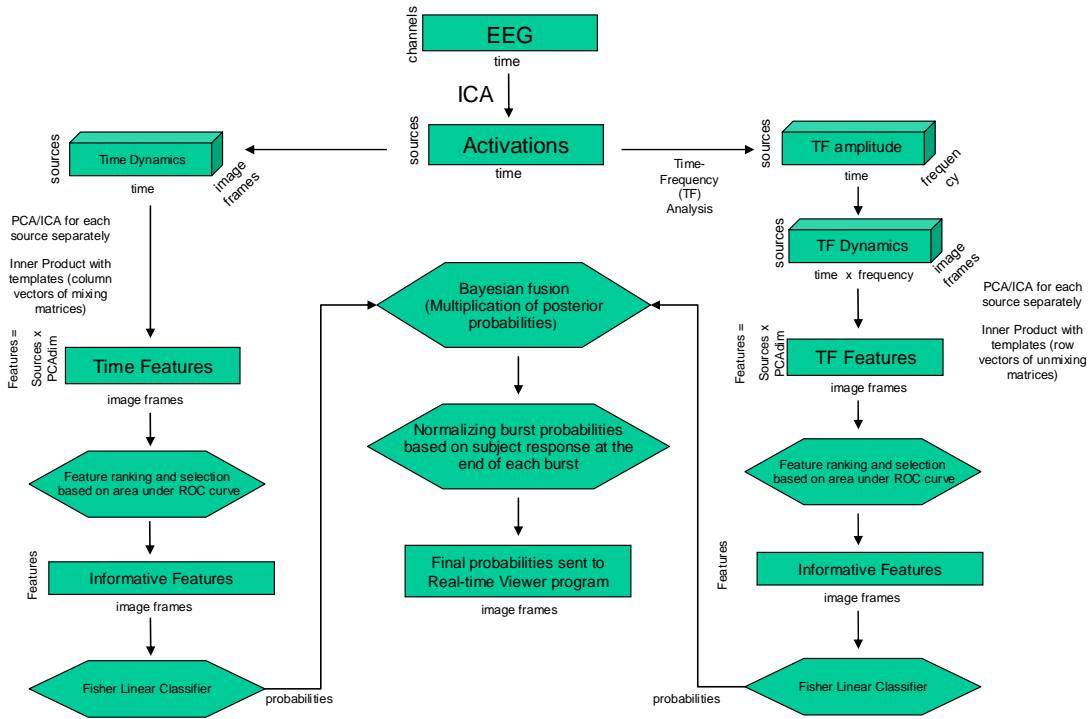


Figure 25. Diagram of the Networked EEG Response Classification (NERC) system.

After each 4 s burst, subject is asked to report an estimate of number of detected targets in the burst by choosing by button press between three options:

- No target
- One target
- More than one target

To incorporate this information into final calculated probabilities, we normalize the probabilities of each burst obtained from Bayesian fusion to sum to the reported number of detected targets. In particular, targets tended to appear in larger clusters in the heptunx presentation order in the IA subject metric experiments. This additional normalizing step appears to increase classification performance by compensating for some source of error likely to include response changes linked to the attentional blink (discussed earlier). If the subject reports detecting more than one target, in pilot experiments with randomized presentation order, we have normalized the summed probability as assuming two targets, while in experiments using the heptunx path search, we normalized to 4 targets per burst. An advantage of this approach is that it reduces the false positive rate.

In bursts in which subject reported detecting no target, the system reports zeros for final probabilities – based on the assumption that we are not able to detect ‘unconscious’ target detection events. This would occur when a number of brain regions correctly respond to a target as such, but for some reason the subject fails to respond. Our pilot data explorations indicated we could not detect such events in single trials with high confidence, and estimating target probabilities for bursts with a subject ‘no-targets’ response lead to unacceptably large number of false positives.

Our classifier gives lower classification performance on the two images presented at the end of the RSVP burst. This is mainly due to the similarity of EEG target-response signature and the novelty-response introduced by the visual surprise of changing from the steady RSVP display sequence to a blank screen. To avoid this problem, we pad the image sequence by repeating two random images from the beginning and middle of the sequence at the end of each burst. For example in a sequence of 49 images, we may repeat images number 10 and 20 after image 49 and display total of 51 images. These two dummy images are then ignored during classification to achieve a near-uniform performance across the 49 burst images.

Response feature computation. Target detection is by definition a rare event. This results in small numbers of positive examples (compared to many non-target events) for training an EEG-based classifier. To prevent over-training and improve cross-session performance, the dimensionality of feature space must be massively reduced. We have devised a new method for feature construction that prevents over-training by completely avoiding the use of labels in feature construction. Labels are only used in feature selection and training the fisher discriminant classifier.

Our approach is based on performing two levels of ICA decomposition. At first, ICA separates EEG sensor data into several independent components (brain and artifact sources). Then the second level ICA is performed on the activity (time-domain or frequency domain) of each component, generating 50 templates for each brain source corresponding to different 'modes' of its activity. These features form a natural spatiotemporal decomposition of the EEG signal with low interdependency. This makes them ideal candidates for features used in an EEG classifier. After observing that a number of these dimensions contained robust information about target responding, we decided to use these informative features as input to our Fisher discriminant linear classifier. For time-domain features we used columns of transpose of the mixing matrix associated with second level ICA as match filters (Figure 27). For frequency domain features, we used frequency-domain independent features (IFs) learned by second-level ICA in the classifier (Figure 28).

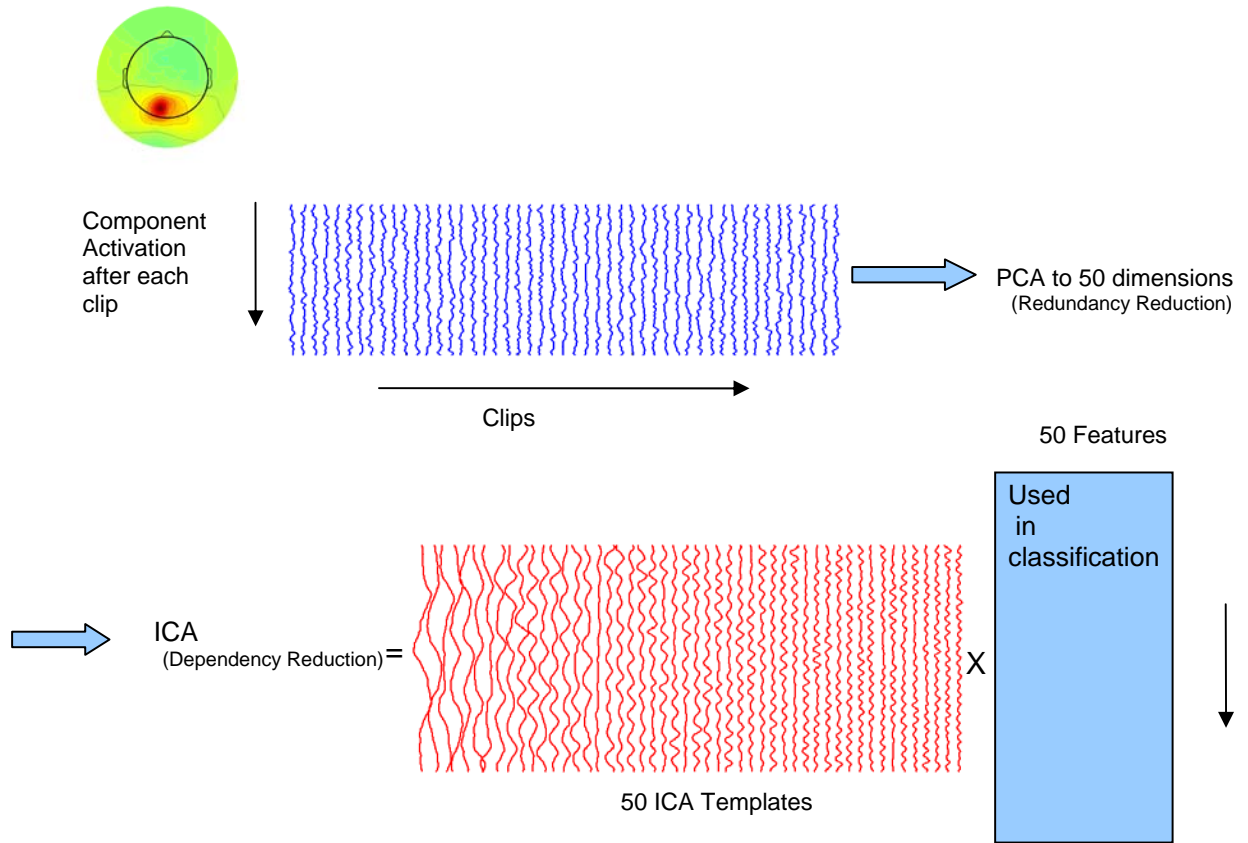


Figure 26. Constructing time-domain features from the activities (or ‘activations’) of independent component processes.

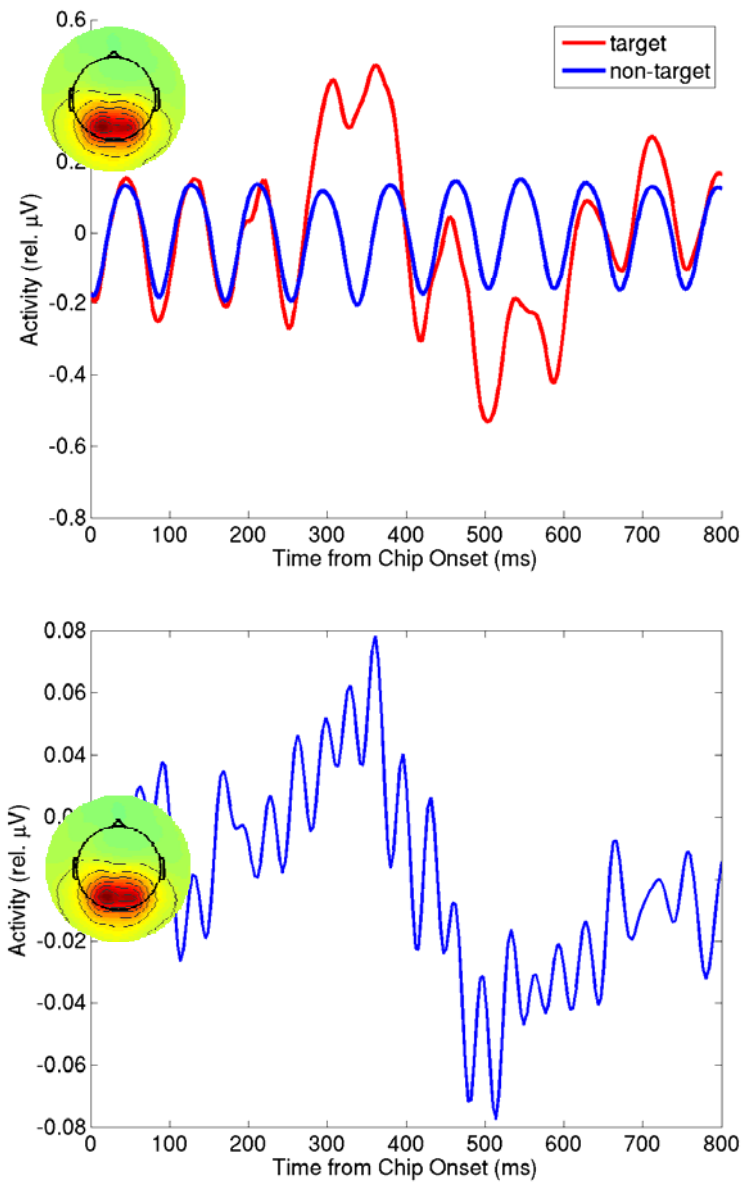


Figure 27. Time-domain dynamics of IC9 from IA Subject 3: (Top) Average component activation following targets and non-targets. (Bottom) Time-domain matched filter for this IC created by a learned linear combination of time-domain informative features.

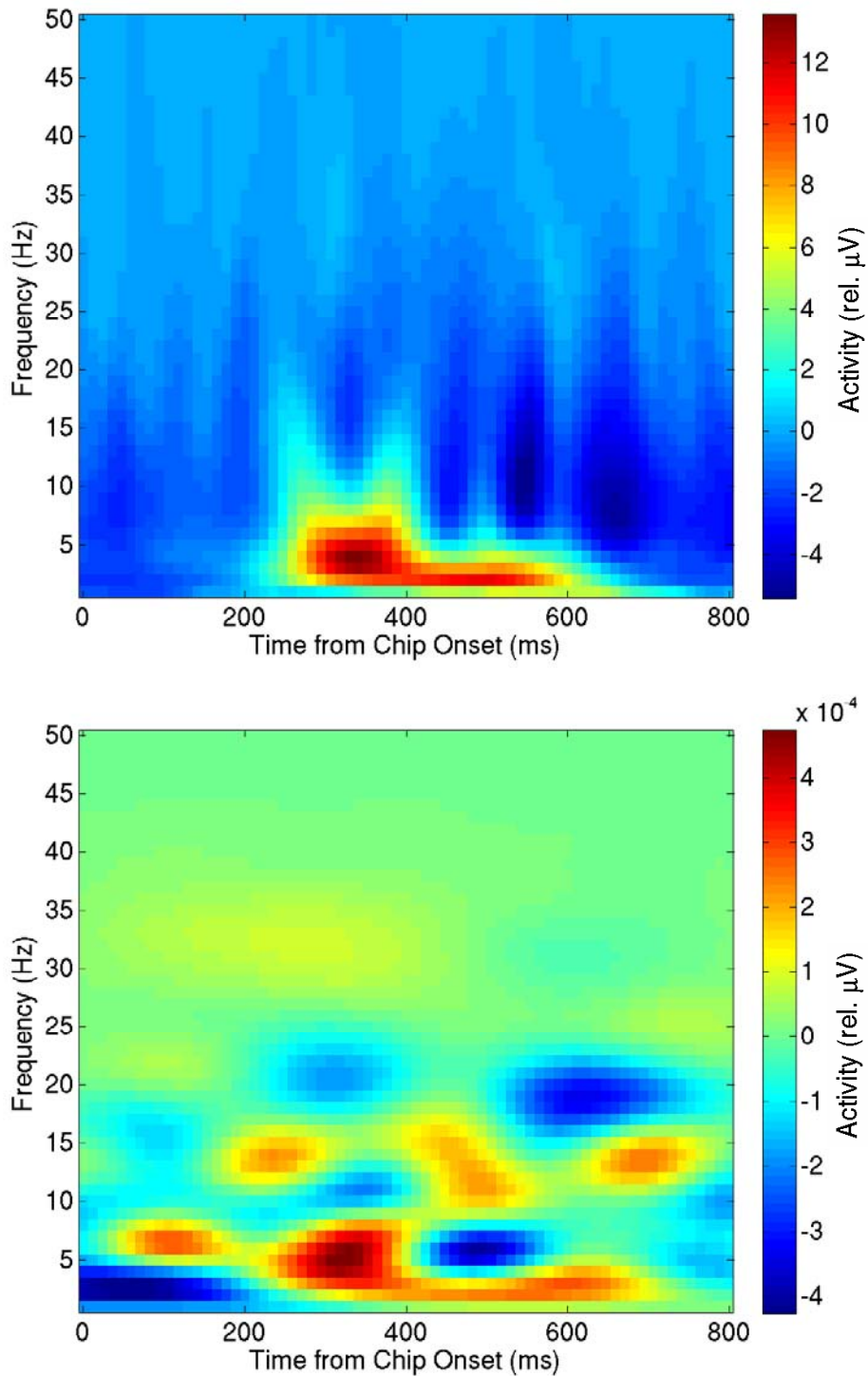


Figure 28. Time-frequency dynamics of IC9 from IA Subject 3: (Top) amplitude difference between targets and non-targets. (Bottom) Frequency-domain matched filter created by a learned linear combination of informative frequency-domain features. Note the prominent increase at 5 Hz, 300 ms (compare Figure 27).

1.7.3 Robustness across sessions, subjects, and sensor numbers

From the beginning of our classification efforts, we decided to separate subject behavioral performance and machine classification. To do so, the classifier was only trained on RSVP bursts in which subject had correctly identified the presence or number of targets. To test the trained classifier in ‘offline’ mode (*post hoc* in Matlab and not in the real-time system), we also used bursts containing detected targets. Only in ‘on-line’ or real-time mode did we use all bursts. Our results show that including bursts containing no targets does not meaningfully change the classifier performance. This is mainly because our experiments usually have presented a large number of images in target-bearing bursts (more than 10,000 of the image clips per experiment). Therefore, adding more non-targets would only lengthen the execution time for the training and testing procedures. Excluding bursts followed by an incorrect behavioral response kept behavioral errors arising from operator inattention, drowsiness, or distraction from adversely affecting classifier training.

We started by testing a non-linear RVM (Relevance Vector Machine) classifier on frequency-domain features (2nd-level ICA activations). A cross-session area under the ROC (Receiver Operating Characteristic) curve of 0.77 (by two-fold validation) was achieved on data with presentation rate of 60 Hz (using behaviorally correct trials only). To improve behavioral performance and image classification, we reduced presentation rate to 20 Hz. Hundred-fold validation using an RVM classifier applied to frequency-domain features gave a 0.76 area under the ROC value, still not satisfactory. By lowering the RSVP rate to 5 Hz and modifying some classifier parameters, we achieved an AUROC (area under the ROC curve) of 0.93 by 100-fold validation. Table 2 shows result for a GIS-analyst subject classifier by RVM and by a linear classifier on data from a still later experiment using a 12-Hz RSVP rate.

Area under the ROC curve using RVM			Area under the ROC curve using a linear classifier		
Features	EEG Channels	ICA Components	Features	EEG Channels	ICA Components
Time-window (10-ms average)	0.89	0.92	Time window (10-ms average)	0.92±0.05	0.93±0.03
Frequency domain	0.77	0.85	Frequency domain	0.76±0.04	0.86±0.04
Time-domain + Frequency-domain	0.90	0.94	Time + Frequency domain	0.92±0.04	0.95±0.02

Table 2 . Ten-fold validation results for a GIS analyst subject using a 12-Hz presentation rate.

The second classification method we tried was a linear classifier on time-window mean features. To construct these features, the algorithm examined all 10-ms time windows from 0 ms to 800 ms (205 time points at the 256-Hz EEG sampling rate) following image presentations. The average ICA activation within each 10-ms window was then calculated and the window that was most informative (e.g., had the highest AUROC value) was chosen for each component.

Although a relatively good in-session classification could be achieved by this method (AUROC by 10-fold validation, ~ 0.92) the cross-session classification performance dropped considerably (to ~ 0.87).

To improve classifier performance using time domain features, we performed a 2nd-level ICA on the activations of each (1st-level) ICA component. In this method we concatenated the component activations from 0 ms to 800 ms following image presentation, then performed a PCA decomposition to reduce the dimensions of the activation matrix to its largest 50 principal components. Finally we performed an ICA decomposition (by extended infomax) on this principal subspace, returning 50 maximally independent components. The inner product between columns of the mixing matrix from this (2nd-level) ICA decomposition and component activation were used as our classifier features (see Figure 2).

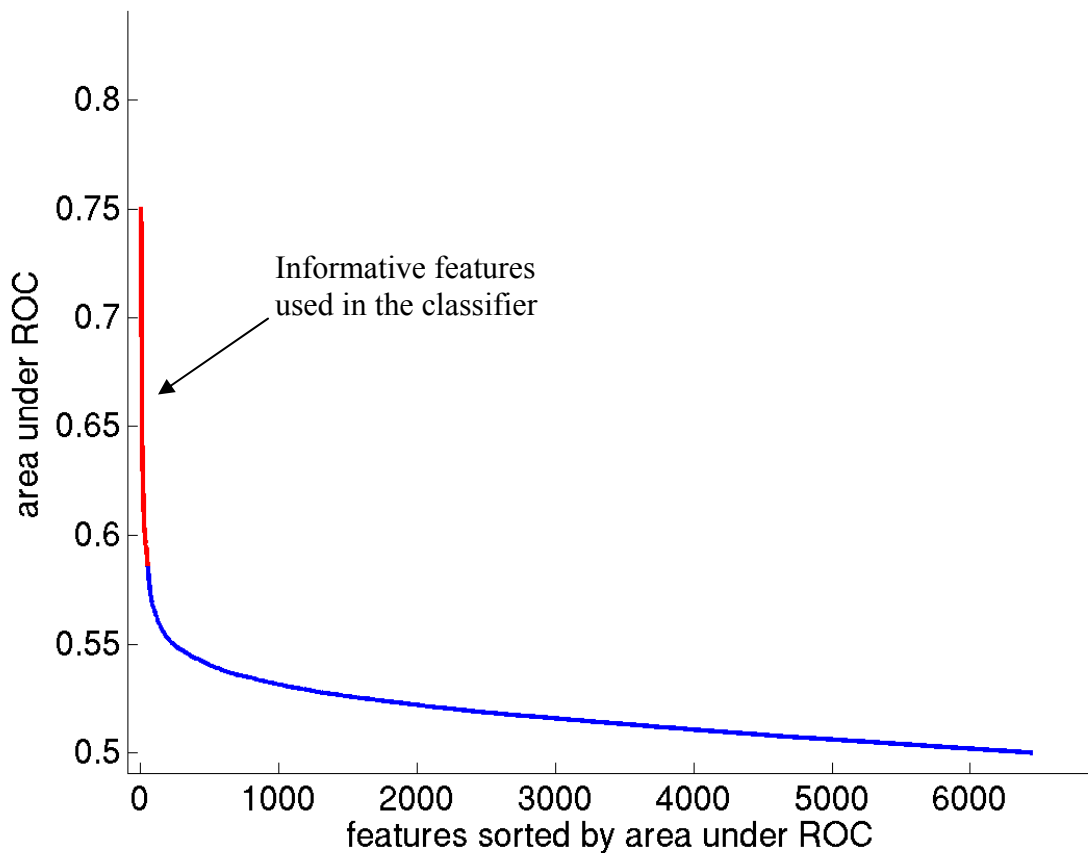


Figure 29. Most informative features are chosen for use in the linear classifier.

This method creates 50 features per (1st-level) ICA component. To choose the most informative subset of these features, we calculated the AUROC for each of these features separately and selected the most informative (~ 100) features from all components as input to a Fisher discriminant classifier (see Figure 29).

Classification performance was further improved by combining these time-domain features with the learned frequency-domain features. Comparing RVM and linear classifiers in Table 3, we see that the linear classifier performed better than RVM in this case. Training the linear classifier is also much faster than training the non-linear RVM. From this point on, we therefore decided to use a linear model for classification instead of RVM.

Subject	Train experiment	Test experiment	Feature type	Area under ROC (%)	Notes
10	19	23	time domain	94	-
10	23	19	time domain	96	Using ICA and time-domain features of exp19
10	23	24	time domain	91	After ignoring the last burst image and normalization with subject response
10	23	24	frequency domain	84	After ignoring the last burst image and normalization with subject response
10	23	24	time + freq. domain	94	After ignoring the last burst image and normalization with subject response
9	18	21	time domain	93	-
9	21	18	time domain	92	Using ICA and time-domain features of exp18

Table 3. Linear classifier results, excluding behavioral errors.

As shown in Table 3, our linear classifier proved to give robust cross-session classification.

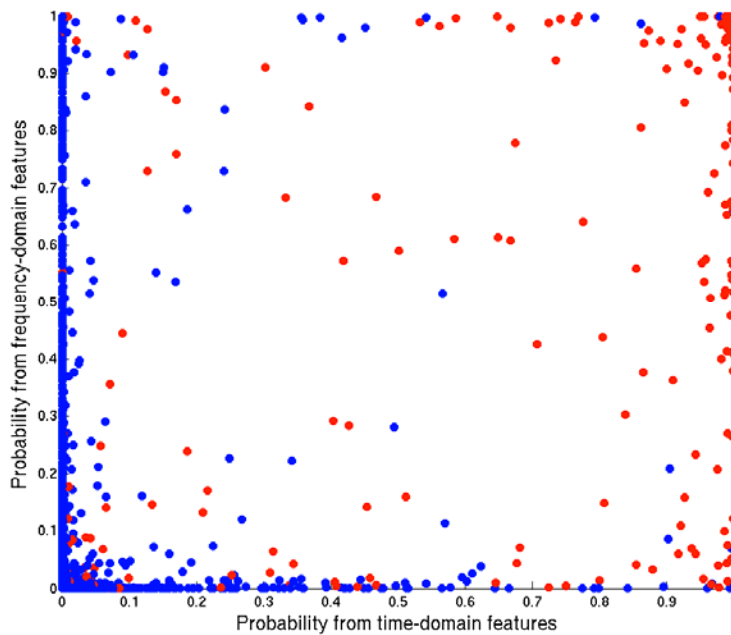


Figure 30. Scatter plot of probabilities calculated from time domain and frequency domain features (Red = targets, Blue = non-targets).

Time-domain features are sensitive to phase-locked EEG responses, while frequency-domain (spectral amplitude) features capture event-related modulations in amplitude of various portions

of the signal spectrum. As these are separate and even complimentary aspects of event-related signal dynamics, we decided to use these two feature types in two separate linear classifiers. Posterior probabilities calculated from each of these classifiers were then combined by Bayesian fusion. The assumption that these features contribute independent evidence to the classification leads to a simple form of Bayesian fusion, ‘naive Bayes’, which is easily implemented by multiplying the posterior probability estimates from both classifiers.

Figure 30 shows a scatter plot of probabilities calculated from time and frequency domain features of a training-session experiment. The star shape of non-target (blue dot) distribution around origin indicates high degree of independence among these values for non-targets. In fact the correlation coefficient for non-targets is ~ 0.13 which is quite low and justifies our use of naïve Bayes fusion.

Other classifier approaches. We also explored other classification methods, most notably asymmetric boosting. Although we chose not to implement this classifier in the real-time Phase I NERC system, we believe that significant further improvement in EEG-based classification may be achieved by combining classification results from different methods in real-time. We propose to explore the use of such *classifier committees* in Phase II of NIA project.

Asymmetric boosting for real-Time EEG classification. A new boosting algorithm called asymmetric boosting (Vasconcelos et al.) can be used to automatically construct a real-time cascade classifier for EEG signals. Asymmetric boosting can be thought of as a feature selection algorithm in which the most discriminating features are chosen from a feature pool such that they can later be placed in a cascade architecture. The cascade architecture uses the features to reject the more numerous negative examples with the least number of feature evaluations, thus improving robust generalization across data sets as well as extremely fast real-time performance when implemented optimally. A variety of numerical experiments using asymmetric boosting were conducted on EEG signal data from the NIA pilot experiments. Depending on the type of EEG data used, these numerical experiments can be divided into three categories:

- **1-D (time domain) EEG data:** In these experiments one-dimensional EEG signals were used either taken directly from the single EEG scalp-channel signals (channel data) or from the ICA component activations (component data). 1-D Haar-like wavelet features were trained using asymmetric boosting and are shown below superimposed on the average event-related potential for one scalp channel (see Figure 31). 1-D EEG signals allowed for the classification of phase-locked signals and generally produced better classification results for these data.
- **2-D (frequency domain) EEG data:** Two-dimensional EEG signals were produced from scalp channels or component activations by performing time-frequency analysis on the (1-D) EEG time series, producing moving spectrograms for each scalp channel or maximally independent signal component. A cascade of quickly applied 2-D Haar-like wavelet features were trained; they are shown superimposed on the ERSP images in Figure 32.

- 1-D and 2-D EEG data:** Asymmetric boosting was also used to learn real-time cascades in which the best features were chosen from both time-domain and frequency-domain data. Not surprisingly such classifiers gave the least amount of classification error. About three quarters of the features were taken from the time-domain feature pool confirming that time-domain features were generally more useful in boosting. However, the variable resolution of frequency-domain transformations leave open the possibility that we may not have optimized the frequency-domain measures used in these experiments.

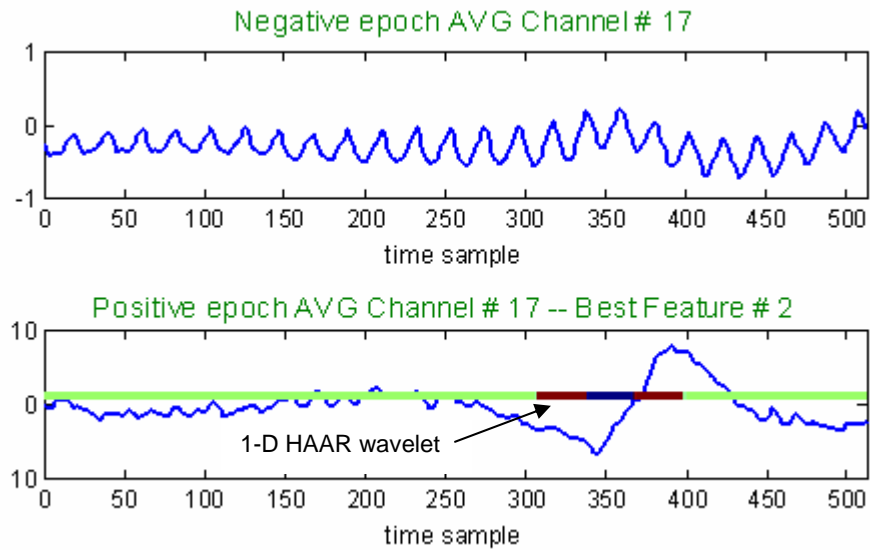


Figure 31. (Top) Average activation following non-target image presentation at time sample 0. (Bottom) Average channel-17 activity (in uV) following target presentation, with the most informative 1-D feature superimposed.

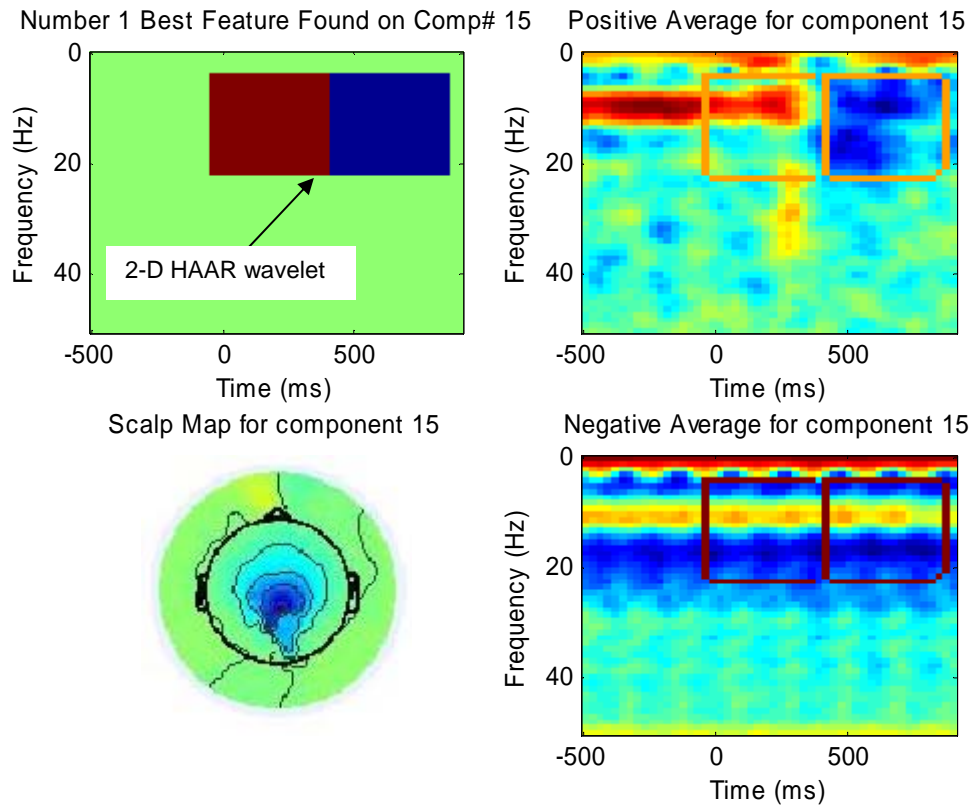


Figure 32. EEG signals in the time-frequency domain, and outlines of learned 2-D Haar wavelets superimposed on the average non-target (upper right) and target (lower right) frequency-domain responses of an ICA component whose scalp map is shown on the lower left.

Experimental results: Table 4 has a summary of some of our experimental results. It can be seen that 1-D channel data has the best classification with 95.6% area under ROC using only 200 features. 1-D component data using a level one ICA has an area under ROC of 90.89% with 200 features and using 2-D features alone with 100 features cannot do better than 85% area under ROC even on same-day same person data where test data is taken from the same person on the same experiment day.

Method	Setup	# Features	Area Under ROC
time-domain channel data	same-person different-day	200	95.6 %
time-domain component data	same-person different-day	200	90.89 %
frequency-domain component data	same-person same-day	100	85 %

Table 4. Summary of asymmetrical Adaboost results.

1.7.4 Processing speed

The demonstrated NERC system requires less than 0.3 seconds to classify the whole burst of 49 images using 100 time-domain features. Total time for time-domain plus frequency-domain feature-based burst classification (100 features each) is less than 2.5 seconds on a common workstation with a Pentium 4 CPU (3.2 GHz), with the majority of the extra time used for the time-frequency analysis. This relatively high classification speed is achieved by exploiting the SSE2 and SSE3 instruction sets offered by modern CPUs (using low-level CPU-specific optimization). In absence of this, classification time for frequency-domain would extend to more than 15 seconds, impeding the capability of the system to classify the next burst and making it impossible to use these relatively complex features in real-time. The time required for classification could be further reduced using readily-available dedicated signal processing chips.

Results

1.8 Image Analyst Recordings

1.8.1 Detection (True Positives), Failures (False Positives, False Negatives)

For our IA metric experiments, we attempted to simulate a real-life broad-area search scenario in which the location of the targets is not known *a priori*. In real-life, it is very difficult for subjects to detect small target features far from the search fovea. We reverse-engineered the DARPA image chipping process to address these issues. First, we selected the provided broad area image that contained the most (helipad) targets (*/NIA_Experiments_nobmp/Helipad/06JAN05024037-P1BS-005540078010_01_P002/ 06JAN05024037-P1BS-005540078010_01_P002_nup.tif*) and magnified it 2x to increase behavioral performance. Then we selected an area with a high concentration of targets (see Figure 34, below) and placed a large 5th-level heptunx there (Figure 11). We then generated 16,807 (=7⁵) scene-warped image clips centered on a hexagonal grid and ordered their presentation using a random heptunx search path (as explained above). Each clip contained an elliptical fovea and a scene-warped, blurred, and contrast-reduced broad-area surround as described in Section **Error! Reference source not found.**

For IA Subject 1, we designed a training session by placing a second 5th-level heptunx in another section of the same broad-area image (such that there was no overlap between the Training, Testing, or Baseline session images). After creating an image clip sequence following a heptunx path and adding the warped surround, in the Training task we removed bursts that contained more than one target. Finally the length of train session was increased by repeating selected bursts in a random order. Each burst contained 49 image clips drawn from a compact 2nd-level heptunx area.

To show that the EEG signature our classifier used signals target detection independent of the specific target shape, we decided **to train IA Subjects 2 and 3 on an airplane target task** and then **to test them on a helipad target task**. This is an important issue because the NERC system should be able to classify any kind of target, even when it has been trained only on one target category, thus eliminating the need to train the classifier on each possible type of search target.

Subject	Image ID	Target type	Total # of targets	# of true positives detected	# of true positives missed	# of false positives	Total image area (pixels)	Total Search Time (min)
1	Part of 06JAN05024034-P1BS-005540078010_01_P001	Helipad	113	36	77	NA	2.30E+08	36
2	Part of 06JAN05024034-P1BS-005540078010_01_P001	Helipad	113	37	76	NA	4.60E+08	14
3	Part of 06JAN05024034-P1BS-005540078010_01_P001	Helipad	113	11	102	NA	4.60E+08	51
Mean	-	-	113	28	85	NA	3.8E+08	34

Table 5. Baseline session results for IA subjects

Table 5 shows Baseline session results for the three IA subjects. Our subjects went through the Baseline task images rather quickly – possibly a function of the exact way in which we instructed them to perform the task – and they reported a correspondingly small number of the 113 targets that our later systematic (heptunx-based, but not RSVP) search revealed were in the search area – a mean of only 28% of the actual targets. Since we did not ask them to record the locations of the identified targets, we could not measure their false alarms.

IA Subject	AUROC by 10-fold validation using time-domain features on behaviorally correct bursts containing targets (%)
1	80
2	97
3	93
Mean	90

Table 6. AUROC for Training sessions

Table shows the off-line classifier results (AUROC, by 10-fold validation) using time-domain features for behaviorally-correct target-containing bursts from the Training session. The classifiers generated for IA Subjects 2 and 3 had high internal consistency, while EEG data from Subject 1, who had a low behavioral performance in the Training session because of a high number of false positives, did not allow consistent classification.



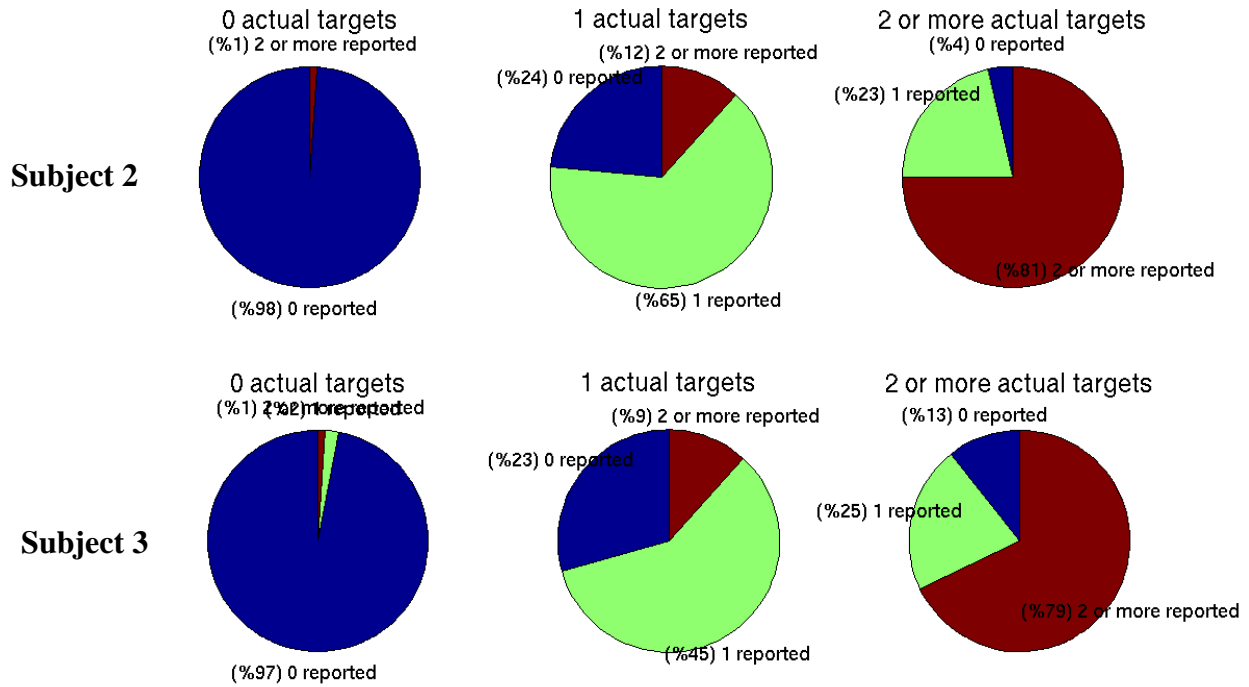


Figure 33. Behavioral performance of IA subjects in Test sessions

Figure 33 shows the behavioral performance of IA subjects in the Test sessions. Subject 1 again reported considerably more false-positive target detections than other two IA subjects, leading to a lower AUROC for this subject (cf. Table 6). The behavioral AUROC results for all bursts by the other two IA subjects were relatively high (0.93 and 0.90).

Subject	Base Image ID	Target Type	RSVP rate	# of clips	Total Triage time (Min)	Total # of targets	# of true positives detected	# of true positives missed	# of false positives
1	Part of 06JAN05024034-P1BS-005540078010_01_P001	Helipad	12 Hz	16,366	42	106	93	13	5,149
2	Part of 06JAN05024034-P1BS-005540078010_01_P001	Helipad	12 Hz	16,813	37	114	103	11	1,859
3	Part of 06JAN05024034-P1BS-005540078010_01_P001	Helipad	12 Hz	16,806	45	114	100	14	2,153
Mean	-	-	-	16,662	41	111	99	13	3,054

Table 7. Image classification results for the IA subjects in Test sessions

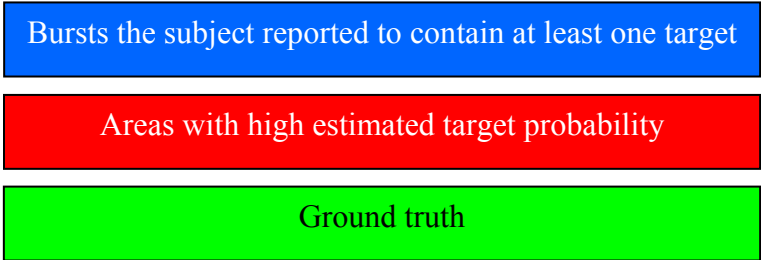
Subject	AUROC for all bursts (%)	AUROC for bursts with correct behavior (%)	AUROC for bursts followed by a correct behavioral response and containing at least one target (%)
1	81	97	65
2	93	97	66
3	90	97	71
Mean	88	97	67.3

Table 6. Area under the ROC curve: Results for the IA subjects in Test sessions

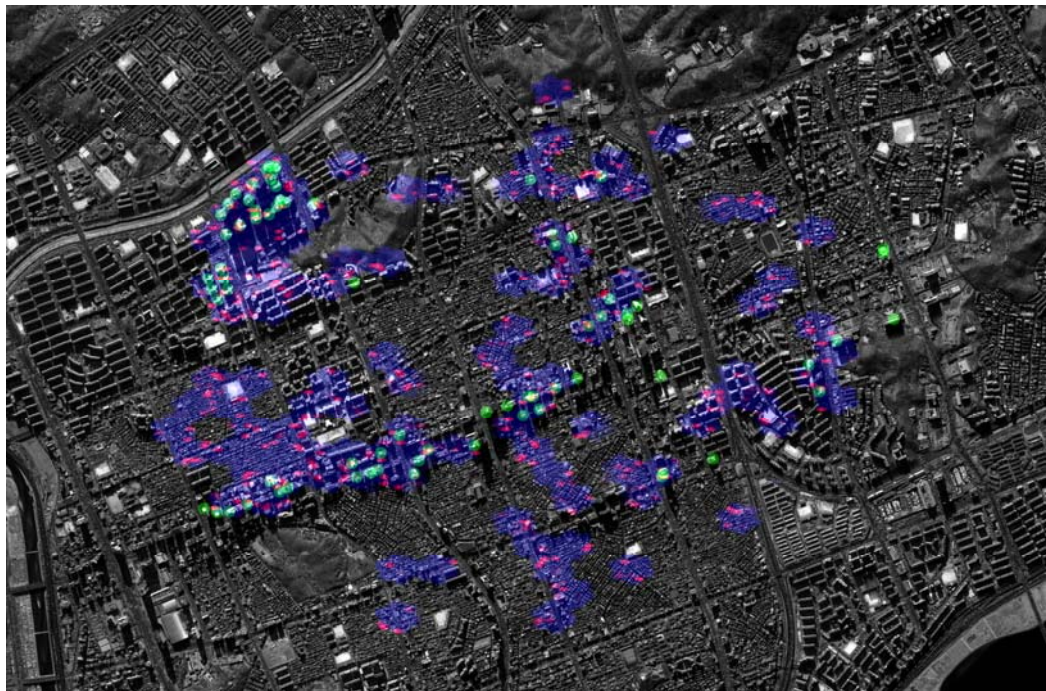
However, as Table 8 shows, the classification of image clips in bursts containing targets left considerable room for improvement. Below, we speculate the reasons for this relative drop in RSVP target classification performance from the levels achieved in the pilot experiments.

1.8.2 Individual Results Viewer screens and their average across subjects

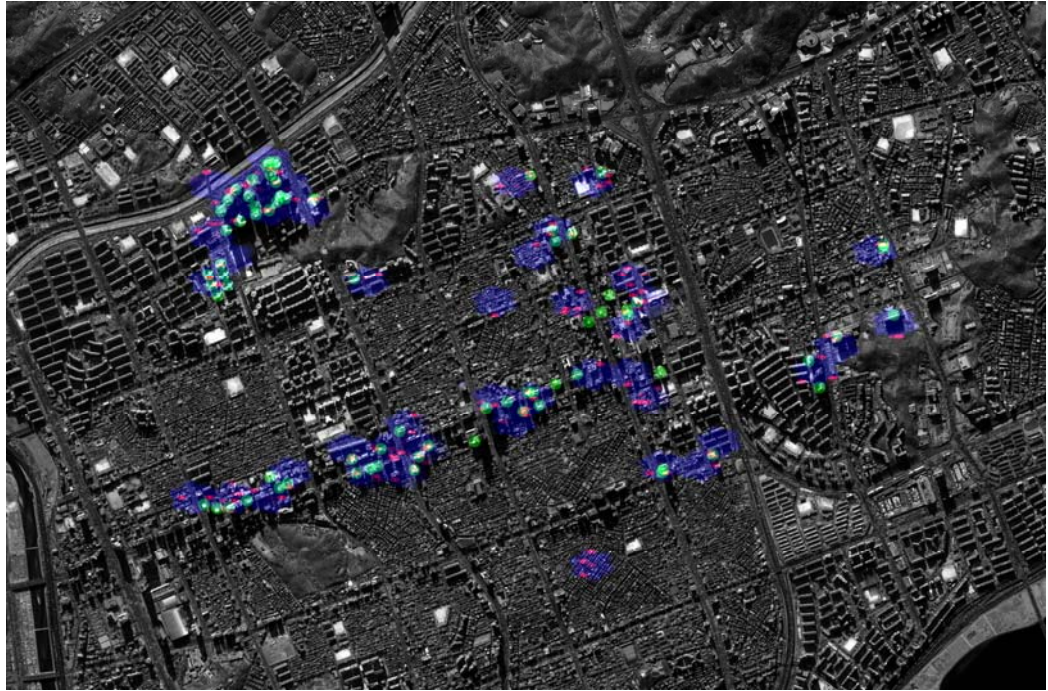
Legend



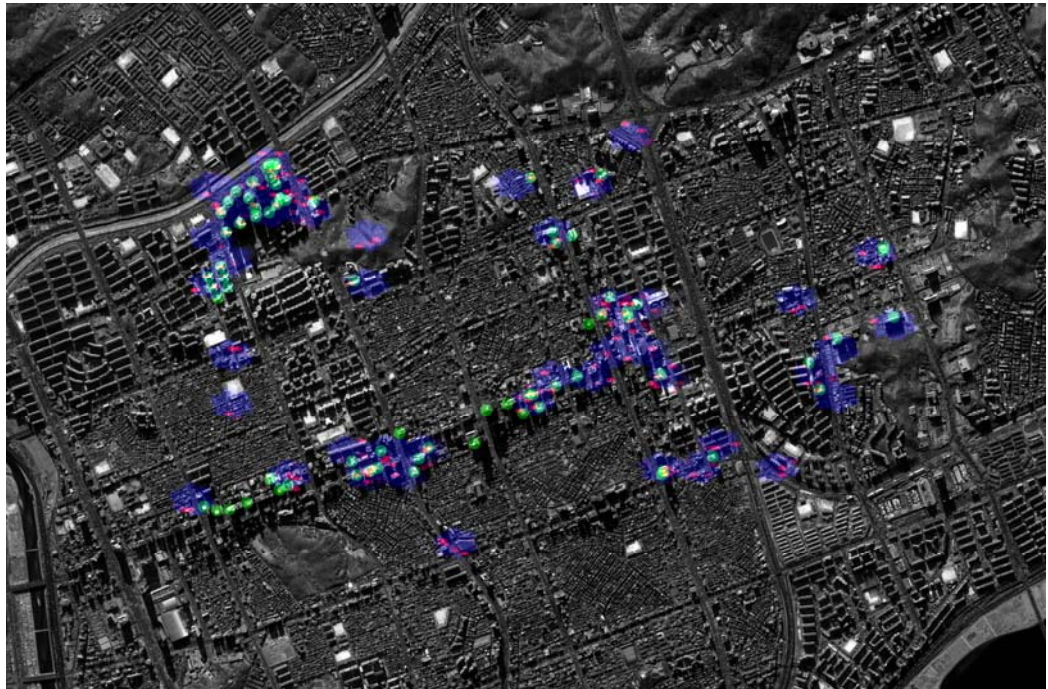
Subject 1



Subject 2



Subject 3



Average across
all 3 IA subjects

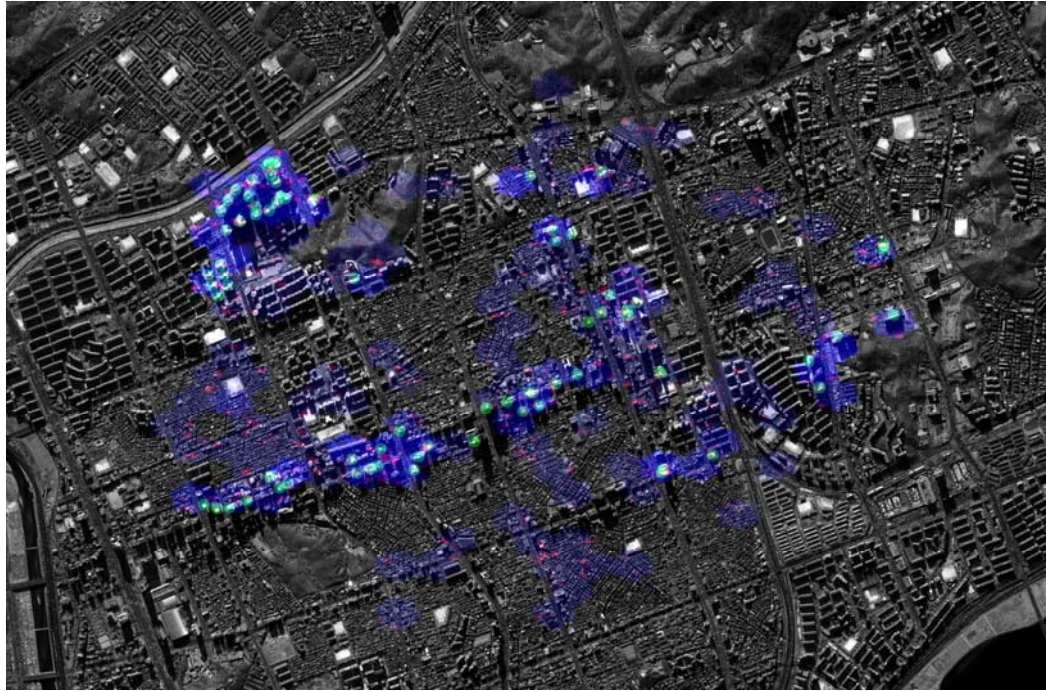


Figure 34. Individual and average Results Viewer screens for the three IA subjects

1.9 Non-IA/pilot data

1.9.1 Detection (True Positives), Failures (False Positives, False Negatives)

We conducted 41 experiments on 15 non-IA subjects, most experiments using airplane targets. After trying different experimental conditions, we decided to use an RSVP rate of 12 Hz.

Subject ID	Gender	Hand	Date of birth	Education level	Occupation
1	M	R	1975	masters	graduate student
2	M	R	1984	undergraduate	student
3	F	R	1981	BS	teacher
4	M	R	1984	undergraduate	student
5	M	R	1974	BA	teacher
6	M	L	1981	undergraduate	student
7	M	R	1984	undergraduate	student
8	F	R	1982	undergraduate	student
9	M	L	1973	B.A.	GIS remote sensing analyst
10	F	R	1978	graduate school	graduate student
11	M	R	1985	undergraduate	student
12	F	R	1983	BA	lab assistant
13	M	A	1974	BA	GIS Analyst
14	M	R	1976	BS	GIS Analyst
15	M	R	1976	PhD	post doctorate

Table 7. Non-IA subjects

Experiment	Targets detected	Targets missed	False positive	Total number of clips	False positive percentage	Area under ROC curve (%)	Time(sec)
Exp23-> exp19	265	66	1105	21456	5	93	4462
exp18->exp21	265	66	2007	21456	9	87	3827
exp18->exp25	177	76	2309	23406	10	84	4208
exp23->exp24	177	76	1108	23272	5	90	4691
Average	221	71	1632	22398	7	89	4297

Table 8. Non-IA RSVP triage results, 12-Hz presentation rate, airplane targets, including behavioral errors

1.9.2 Individual subjects and average across subjects

Exp. ID	Exp. Type	Rate (Hz)	Targets per burst	Target type(s)	Burst duration (s)	Target density (% of bursts containing at least one target)
1	RSVP	6 and 12	1	50 Planes with various difficulty levels, added manually	continuous	[0-3 targets per burst]
2	RSVP	5	1	highly salient planes	5	50
3	RSVP	60	1	highly salient planes	2	15
4	RSVP	60	1	highly salient planes	2	40
5	RSVP	60	1	highly salient planes	2	40
6	RSVP	20	2	salient planes and water-tanks	2.4	%25 no targets, %25 planes, %25 water-tanks
7	RSVP	20	2	salient planes and water-tanks	2.4	%25 no targets, %25 planes, %25 water-tanks
8	free search	0	1	planes with various saliencies	N/A	0-4 targets per clip
10	RSVP	N/A	N/A	faces	N/A	N/A
11	psychophysics	12	2	planes and water tanks	3.92 s	1
12	RSVP	12	1	planes	3.92	75
13	RSVP	12	1	planes	3.92	1
14	RSVP	12	1	planes	2	3

Table 9. Experimental conditions for non-IA subjects

Since we tried different conditions during our experiments on non-IA subjects (see Table 9), the average AUROC can only be calculated on a subset of experiments. As shown in Table 8, the mean area under the ROC curve for non-IAs was 0.89. Since plane targets were placed on individual clips, they did not have fixed locations across different experiment bouts (= 7 bursts) (the subjects were asked to consider them to be planes flying across the broad-area image); therefore there is no broad-area scene to show to summarize the pilot non-IA experiment results.

1.9.3 Comparison with IA-Subject results

As shown in Table 8, the mean area under the ROC curve for non-IAs was 0.89. This is roughly similar to average AUROC for the three IA subjects (0.88). It interesting to see that even though we trained the classifier for IA Subjects 2 and 3 on planes and used it to classify helipad targets, system performance was consistent with the earlier pilot experiments using only plane targets in both Training and Testing sessions.

1.10 Statistical Power

Clearly, repeated sessions on a larger number of RSVP-experienced subjects will be required in Phase 2 to statistically estimate the robustness of image clip classification that could be expected in operational use.

Lessons Learned

1.10.1 Unforeseen Challenges Encountered During the Conduct of the Experiments

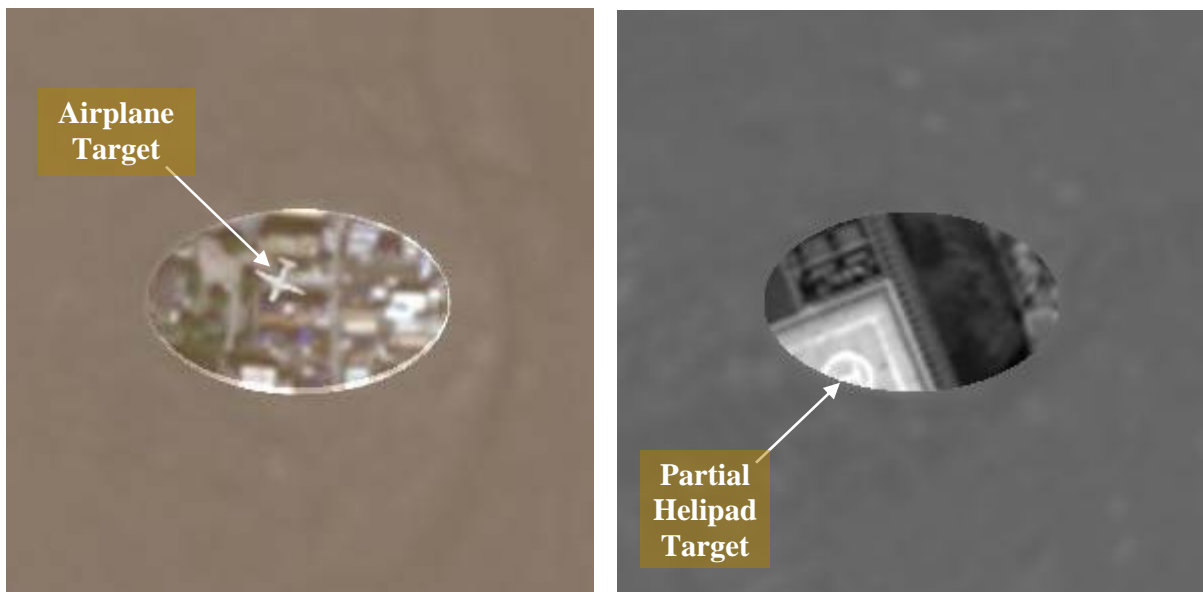
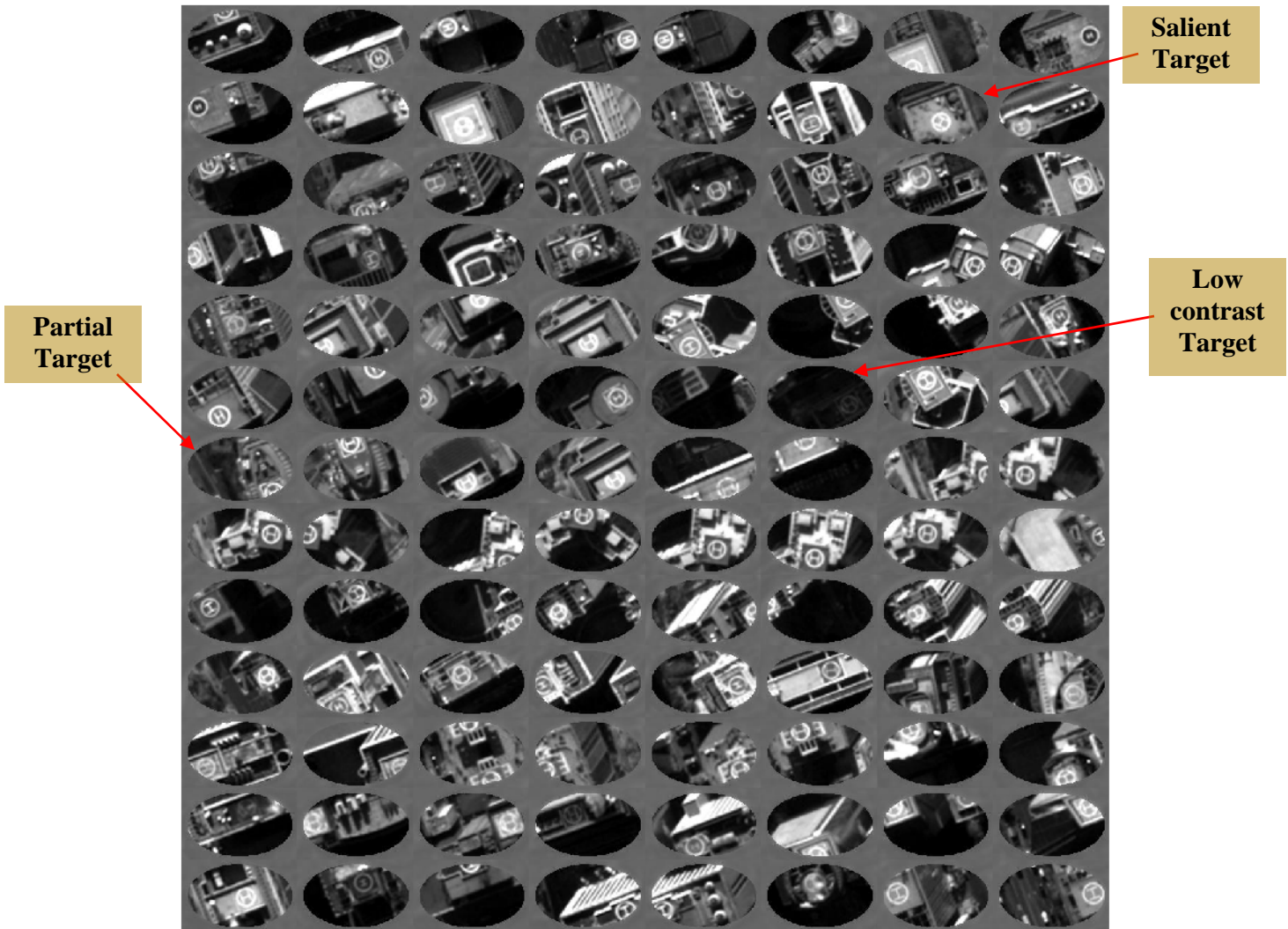


Figure 35. Sample salient airplane target clip (left) and a partially obscured helipad target image clip (right)

Target centering. For our non-IA subjects, we used salient airplane targets artificially placed near the center of the image clip to invoke a clear target detection and concomitant EEG response. For our metric experiments on the three IA subjects, we decided to design experiments coming as close as possible to the operational situation. This meant we could not use any information regarding target positions in the broad-area image while creating the image clips. We elected to use relatively small image clips to maximize behavioral detection, and the same relatively high (12-Hz) RSVP rate we used in our earlier plane experiments. An undesired side effect of small image clip size is that some targets may appear only partially in the foveal image area, as illustrated in Figure 35 (left).



**Figure 36. Target image clips from the IA subject
Test sessions**

A simple solution to the partial-target problem, the one adopted by us in the Metric experiments, is to present image clips with overlapping fovea. An undesired side-effect of this strategy, however, is to increase the time required for exhaustive broad-area search, since some portion of the image much be shown twice, further reducing the number of new image territory

presented by each image clip. The opposite strategy, to increase the foveal area of each image clip would, however, have the undesirable side-effect of reducing detection probability, a surmise supported by preliminary psychophysical results presented by PI Makeig at the final Phase I PI meeting.

Response latency variation. In response to image clips containing partially hidden targets, subjects may tend to be uncertain about the presence of a target in the image clip, and might also require significantly more time to make a target detection decision. A similar decision delay might also occur if the target image is not partial but instead unclear or difficult to detect. In these cases the strength and timing of the EEG signature could be expected to shift accordingly and become less similar to that following quick recognition of more salient targets. As our classifier relies on more or less precise time-locking of the EEG response signature to target detections (given the short 80-ms delay between consecutive images in 12-Hz RSVP), the appearance of partial targets or targets with varying degrees of saliency can only adversely affect classification performance. Figure 26 (above) shows a number of helipad target image clips from IA test experiment, clearly revealing a wide range of saliencies of the helipad target images.

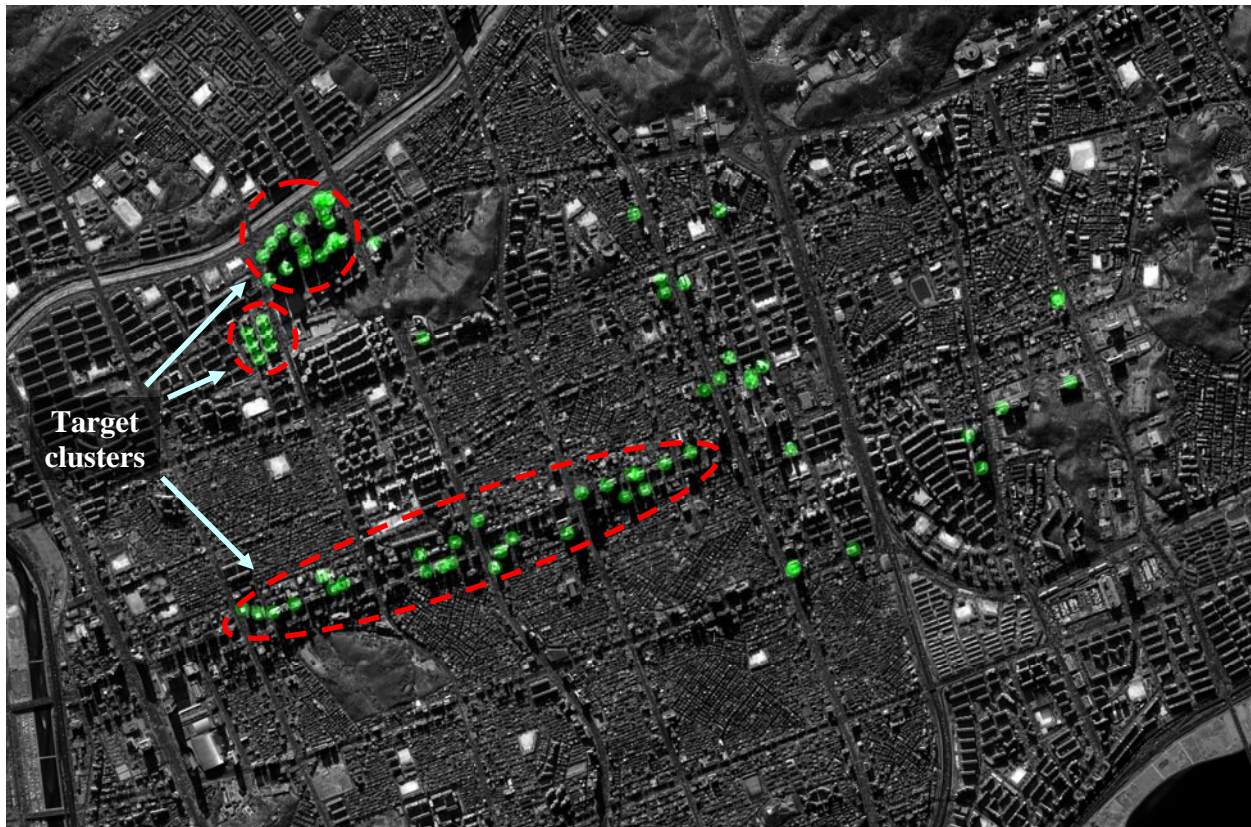


Figure 37. Targets colored in green in IA broad area test image

Target clustering. Another classification challenge is related to the number of targets presented in one burst. Although overall target frequency was low in the IA Subject Test session images (less than %1), they appeared in tight clusters in a few portions of the image (Figure 37). When targets appear in tight clusters, any image clip extraction strategy that preserves locality in the

RSVP search path will necessarily produce a number of bursts containing several to many targets. The EEG signature of detecting multiple targets in a short time (for example, 7 targets in a 4-second burst) might be quite different from our more idealized Training session experiment in which a maximum of only one target could appear in a burst.

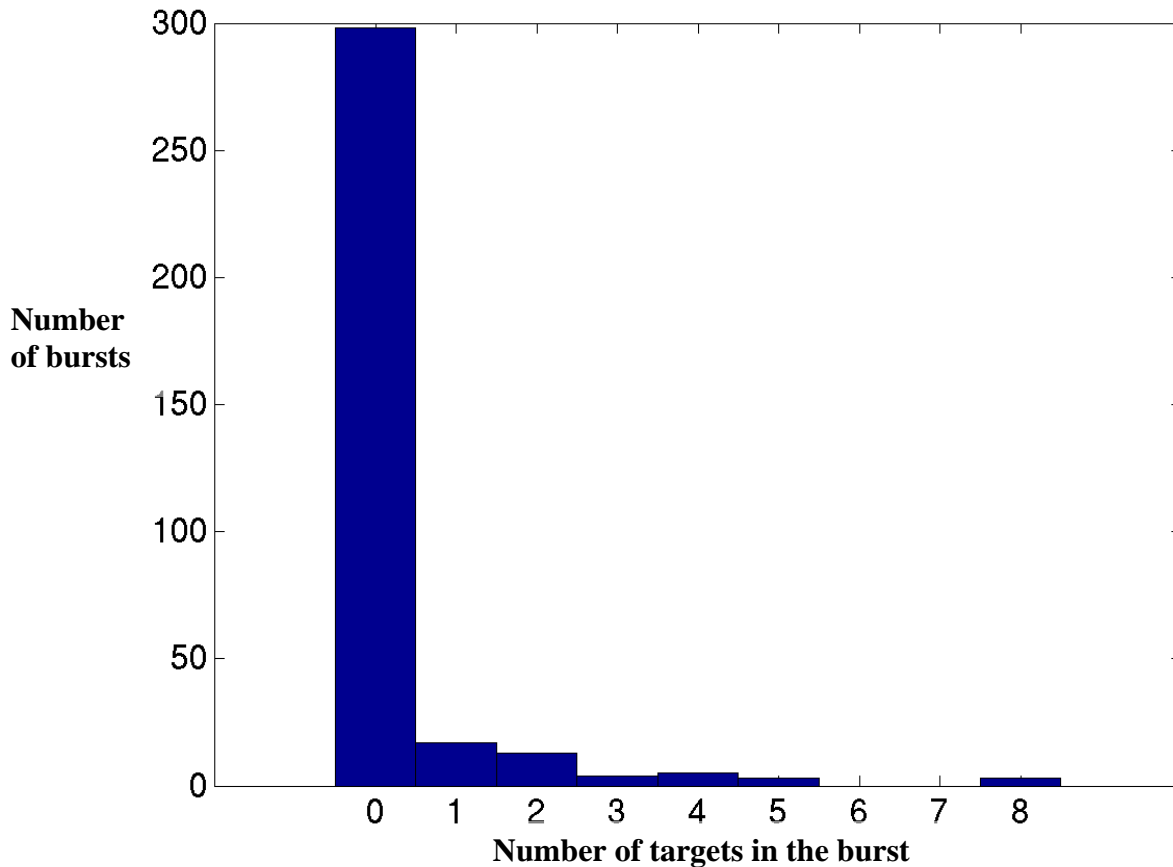


Figure 38. Histogram of number of targets in each Test session RSVP burst

Since our heptunx search method preserves locality, 11 out of 45 target bearing bursts in the IA Subject experiments contained more than three targets (Figure 38).

Tentative conclusions: More widely-varying target saliency, the larger number of partially-obscured targets, and the larger number of bursts containing many targets are probably the main reasons that classification performance in our IA Test sessions was lower than in our pilot experiments using isolated, salient, and unobscured airplane targets.

1.10.2 Findings Outside the Scope of the Program

We are continuing our efforts to improve our response classification methods. Very recent results indicate that performance can be increased by building a separate linear RVM classifier for each 1st-level ICA component, then combining the posterior probabilities of these classifiers using naïve Bayesian fusion. For a selected cross-session numerical experiment (exp18→exp21) using time domain features from bursts in which targets were correctly detected by *the* subject, this method improved area under the ROC curve from 0.92 (by standard LDA on top 100 2nd-level ICA features from all components) to 0.94 (a 25% improvement). The key to this result is the

high degree of independence among posterior probabilities calculated from the different independent components. It enables us to combine these probabilities simply by multiplication (naive Bayes). This is not possible when channel data are used instead of component activities; using channel data, the cross-session AUROC decreased to 0.85. The reason for this decrease is that nearby channels contain very similar information. This leads to highly dependent posterior probabilities for classifiers constructed from channel data that cannot be effectively combined by simple multiplication.

Future Development Plans

Networked EEG Response Classification system enhancements

Further improvements in response classification.

In Phase 2, we propose to continue to explore and implement more effective classification methods in the real-time classification system. This includes the committee of experts scheme discussed above.

- **Electrode re-positioning.** Our system requires one model Training session per subject. This approach creates an accurate model and also saves time, as training does not to be repeated at the beginning of each Testing session. The downside of this approach is that in each Testing session the electrode positions on the scalp may be slightly different. A straightforward partial solution to this ever-present problem for neurotechnology is to increase the accuracy with which we place the EEG cap by introducing additional, yet convenient measurements. A more elegant solution is to adapt our ICA model to find the same independent component sources used in the classifier based on data gathered in the Testing session. Since the locations of the independent brain sources used in the classifier should be constant across sessions involving similar tasks, they provide a coordinate system for signal representation that is not dependent on the locations of the individual electrodes. In comparison, a classifier that uses individual channel data does not have access to this natural coordinate system.

- **Adaptive ICA training.** A UCSD collaborator in a parallel DARPA dry-electrode project, Gert Cauwenberghs, is developing an On-line Recursive Independent Component Analysis (ORICA) method. This algorithm offers incremental estimation of independent components during on-line data collection while retaining the convergence properties of batch-mode ICA. By further developing and testing this algorithm, we plan to apply the ICA model learned from the Training session to data gathered in the beginning of each Testing session. This may automatically compensate for small changes in electrode positions across sessions, and/or for differences in electrode impedance, by fine-tuning the previously learned ICA unmixing matrix.

- **Dynamic image sequencing.** To reduce the error in evaluating the probability of a certain image being a target, we can present it again in a succeeding RSVP burst and average the probability estimates across presentations. To minimize the overhead caused by this repetition, we can select for repetition only images with marginal initial target probability estimates. Most software components needed to implement this dynamic sequencing method are already in place in our system.

- **Visual surprise-based image sequencing.** As explained in an earlier session, manipulating the degree of visual low-level surprise in an RSVP sequence can improve subject target detection performance. We propose to explore both sequence selection and surprise-based image filtering to minimize the documented effects of attentional blink-like effects of visual surprise on RSVP detection performance.

- **Incorporation of dry electrodes.** We propose the use of the NERC system for testing the functionality and operational potential of the dry-electrode caps we are developing in a parallel DARPA project.

- **EEG-based feedback for analyst training.** Our real-time system can be fairly easily adapted to operate in continuous instead of burst mode, and to provide immediate feedback about classifier output to the subject, a mode of operation recently referred to as ‘neurofeedback.’ By this method, the subject may be able to modify their EEG responses to maximize target classification by the system. For example, after training the classifier on initial Training session data, subsequent Testing sessions can also serve as feedback training sessions. In these sessions, some known target images may be introduced into the RSVP sequences. The system can then ‘feed back’ to the subject that strength of their target recognition response (for example, the estimated target probability, as a number of in some other visual or auditory format). The subject will be asked to attempt to keep these feedback signals, indicating successful target response detection, as high or strong as possible. Online or offline updating of the subject neural signature model could further tighten the human-machine loop and further optimize classifier performance. Because of the modular structure of our NERC system code, only limited modifications will be necessary to implement feedback or neurofeedback capabilities.

Further directions in neurotechnology research

Neurobiology and brain source localization.

Most studies of visual target detection have focused on the averaged EEG signals at the sensor level. This has made it difficult to determine which neuronal networks truly contribute to the measured differences since many brain regions could be activated in parallel and since the averaging process eliminates any non-phase-locked dynamics that could be relevant for target detection. Our response classification approach bypasses both of these problems by using ICA to extract source-specific single trial dynamics. However, to fully characterize the physiology of the networks contributing to automatic target response classification, more sophisticated source localization techniques, such as our recent multi-scale geodesic sparse Bayesian learning (SBL) algorithm, would have to be used in conjunction with more realistic BEM or FEM forward models based on the subject’s MRI (Ramirez et. al., 2005, 2006). Importantly, nonlinear source localization techniques including SBL may improve the performance of the classifier if voxel-wise source dynamical features are used to relax the ICA assumption of spatial stationarity. Also source imaging will clarify the neurophysiological relevance of theta responses in target detection and will help us discover new source signatures that may contain more precise information about the mental states of analysts and the images they perceive. These new methods, applied to NERC system data from 20 subjects with MRI head images, should allow

unprecedented understanding of the neural dynamics of categorical perception and decision-making that underlie successful EEG-based response classification.

Neural dynamics and training of visual search.

Research is needed to better understand the brain dynamics accompanying normal visual search of a large static image or scene using 3-4/s saccadic eye movements. The research could suggest a neurotechnology system that uses efficient, assisted visual (saccadic) search in place of RSVP search. Might an EEG-based feedback approach, applied to EEG responses to search saccades in normal visual search, help train analysts to perform more efficient visual search? The answer to that question may have practical importance, since training of imagery analysts is currently a lengthy and difficult process.

Neural dynamics of video search.

A fruitful area of applied research in this direction could be to assist analysts in reviewing overhead or other intelligence video.

Feedback to maintain alertness and attention.

The ability of analysts to make use of other forms of EEG-based feedback concerning their current levels of alertness and attention to the presented images also merits careful investigation.

Other possible intelligence applications.

The NERC system might also be used to evaluate the familiarity of a set of images to a subject. These images might be the subject of intelligence interest. For example, by showing pictures of possible collaborators to a suspect and classifying his/her EEG responses, the probability of an association between the subject and the viewed suspects others whose pictures were viewed might be assessed. In a more common 'lie detection' mode, one might show pictures of objects from certain crime scene mixed with non-related objects in same categories to a suspect. Classifying suspect's single-trial EEG responses might reveal the degree of familiarity of suspect to objects from the crime scene and provide important clues about his/her knowledge of the crime or of the crime scene. The real-time classification system could use dynamic sequencing (explained above) to efficiently arrange the sequence of presented images, re-presenting images with marginal estimates of probability of subject recognition on first presentation, thus increasing confidence in the results.

References

- Carneirom G and Vasconcelos, N, "Formulating Semantic Image Annotation as a Supervised Learning Problem", *Technical Report SVCL-TR-2004-03*, December 2004 (available from <http://www.svcl.ucsd.edu>).
- Chun, M.M. "Types and tokens in visual processing: a double dissociation between the attentional blink and repetition blindness." *Journal of Experimental Psychology. Human Perception and Performance*, 23(3), 738-755, 1997.
- Cox, R. T. Probability, frequency and reasonable expectation. *Am. J. Phys.* 14 , 1-13, 1964.
- Delorme A, Makeig S. "EEG changes accompanying learned regulation of 12-Hz EEG activity." *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2(2):133-136, 2003.
- Delorme A, Makeig S, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, 134:9-21, 2004.
- DiRusso, F., Martinez, A., and Hillyard, S.A. "Source analysis of event-related cortical activity during visuo-spatial attention," *Cerebral Cortex* **13**: 486-499, 2003.
- W. Einhäuser, C. Koch, and S. Makeig The duration of the attentional blink in natural scenes depends on stimulus category. *Vision Research*, in press [see proof sttached]
- Evans, K.K., & Treisman, A. "Perception of objects in natural scenes: is it really attention free?" *Journal of Experimental Psychology. Human Perception and Performance*, 31(6), 1476-1492, 2005.
- Einhäuser, W., Koch, C., Makeig, S. "The duration of the attentional blink in natural scenes depends on stimulus category." *Vision Research*, in press
- Einhäuser, W., Martin, K.A.C. and König P. "Are switches in perception of the Necker cube related to eye-position?" *Eur. J. Neurosci.* **20**(10): 2811-2818, 2004.
- Körding, K.P., Kayser, C., Einhäuser, W. and König, P. "How are complex cell properties adapted to the statistics of natural stimuli?" *J. Neurophysiol.* **91**(1): 206-212, 2004.
- Einhäuser, W., C. Kayser, K.P. Körding and König. P., "Learning distinct and complementary feature-selectivities from natural colour videos," *Rev. Neurosci.* **14**: 43-52, 2003.
- Einhäuser, W., and König, P. "Does luminance-contrast contribute to a saliency map for overt visual attention?" *Eur. J. Neurosci.* **17**(5): 1089-1097, 2003.
- Einhäuser, W., Kayser, C., König, P. and Körding, K.P. "Learning the invariance properties of complex cells from their responses to natural stimuli," *Eur. J. Neurosci.* **15**(3): 475-486, 2002.
- Fabre-Thorpe M, Delorme A, Marlot C, Thorpe S. A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *J Cogn Neurosci.* 2001 Feb 15;13(2):171-80.

- Fecteau, J. H. & Munoz, D. P. Exploring the consequences of the previous trial. *Nat Rev Neurosci* 4, 435-443, 2003.
- Fletcher, P. C., Anderson, J. M., Shanks, D. R. et al. Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. *Nat Neurosci* 4, 1043-1048, 2001.
- Galambos, R., Makeig, S. and Talmachoff, P. "A 40 Hz auditory potential recorded from the human scalp," *Proc Natl Acad Sci USA* 78:2643-2647, 1981.
- Guittou D, Douglas RM, Volle M. "Eye-head coordination in cats." *J Neurophysiol.* 52(6):1030-50, 1984.
- Hupe, J-M, James, A, Girard, P, Lomber, SG, Dayne, BR. "Feedback connections act on the early part of the responses in monkey visual cortex." *J Neurophysiol* 85:134-145, 2001.
- Jean-Michel Hupé, Andrew C. James, Pascal Girard, Stephen G. Lomber, Bertram R. Payne, and Jean Bullier Feedback Connections Act on the Early Part of the Responses in Monkey Visual Cortex. *J Neurophysiol.* 85: 134-145
- Itti, L. Quantifying the Contribution of Low-Level Saliency to Human Eye Movements in Dynamic Scenes, *Visual Cognition*, 2005. (in press)
- Itti, L. & Baldi, P., A Principled Approach to Detecting Surprising Events in Video, In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2005.
- Itti, L, & Koch, C. Computational Modeling of Visual Attention, *Nature Reviews Neuroscience*, Vol. 2, No. 3, pp. 194-203, Mar 2001
- Jaynes, E. T. Probability Theory. *The Logic of Science* (Cambridge University Press, 2003).
- Johnson JS, Olshausen Bam Timecourse of neural signatures of object recognition. *J Vis.* 2003;3(7):499-512. Epub 2003 Sep 2.
- Jung T-P, Makeig, S., M. Stensmo and T. Sejnowski, "Estimating alertness from the EEG power spectrum," *IEEE Transactions on Biomedical Engineering* 44:60-69, 1997.
- Kanwisher, N.G. "Repetition blindness: type recognition without token individuation." *Cognition*, 27(2), 117-143, 1987.
- Klopp J, Marinkovic K, Chauvel P, Nenov V, Halgren E. Early widespread cortical distribution of coherent fusiform face selective activity. *Hum Brain Mapp.* 2000 Dec;11(4):286-93.
- Knight, R. Contribution of human hippocampal region to novelty detection. *Nature* 383, 256-259, 1996.
- Kreutz-Delgado, K., Murray, J.F., Rao, B.D., Engan, K., Lee, T-W., and Sejnowski, T.J. "Dictionary Learning Algorithms for Sparse Representation," *Neural Computation* 15: 349-396, 2003
- Li, FF, VanRullen, R, Koch, C, and Perona, P, "Rapid natural scene categorization in the near absence of attention," *Proc Natl. Acad. Sci. USA* 99: 9596-9601, 2002.
- Li, S., Cullen, W. K., Anwyl, R. & Rowan, M. J. Dopamine-dependent facilitation of LTP induction in hippocampus by exposure to spatial novelty. *Nat Neurosci* 6, 526-531, 2003.
- Luce R.D. *Response Times*. Oxford University Press: Oxford, 1986.

- Makeig, S., Elliott, F.S., and Postal, M., *First Demonstration of an Alertness Monitoring/Management System*, Technical Report 93-36, Naval Health Research Center, San Diego, CA, 1993.
- Makeig, S., and Galambos, R., "The CERP: Event-related perturbations in steady-state responses," *Brain Dynamics Progress and Perspectives*, (pp. 375-400), ed. E. Basar and T.H. Bullock, 1989.
- Makeig, S., "Event-related dynamics of the EEG spectrum and effects of exposure to tones," *Electroencephalogr clin Neurophysiol* 86:283-293, 1993.
- Makeig, S. and Inlow, M., "Lapses in alertness: coherence of fluctuations in performance and the EEG spectrum," *Electroencephalogr clin Neurophysiol*, 86:23-35, 1993.
- Makeig, S. and Jung, T-P., "Changes in alertness are a principal component of variance in the EEG spectrum," *NeuroReport* 7:213-216, 1995.
- Makeig, S., Bell, A.J., Jung, T-P, and Sejnowski, T.J., "Independent component analysis of electroencephalographic data," In: D. Touretzky, M. Mozer and M. Hasselmo (Eds). *Advances in Neural Information Processing Systems* 8:145-151 MIT Press, Cambridge, MA, 1996.
- Makeig, S. and T-P. Jung, "Tonic, phasic, and transient EEG correlates of auditory awareness during drowsiness," *Cognitive Brain Research* 4:15-25, 1996.
- Makeig, S., T-P. Jung, D. Ghahremani, A.J. Bell & T.J. Sejnowski, "Blind separation of auditory event-related brain responses into independent components." *Proc Natl Acad Sci USA* 94:10979-10984, 1997.
- Makeig, S., M. Westerfield, J. Townsend, T-P. Jung, E. Courchesne, T. J. Sejnowski., "Functionally independent components of the early event-related potential in a visual spatial attention task." *Philosoph Trans Royal Soc: Biolog Sci* 354:1135-44, 1999.
- Makeig, S., Marissa Westerfield, Tzyy-Ping Jung, James Covington, Jeanne Townsend, Terrence J. Sejnowski and Eric Courchesne. "Functionally independent components of the late positive event-related potential during visual spatial attention." *J. Neurosci.* 19:2665-2680, 1999.
- Makeig S, Jung T-P. and Sejnowski TJ. "Awareness during drowsiness: Dynamics and electrophysiological correlates." *Can J Exp Psychol.*, 8:208-211, 2000.
- Makeig S, Enghoff S, Jung T-P, Sejnowski TJ, "A natural basis for efficient brain-actuated control." *IEEE Trans Rehab Eng*, 8:208-211, 2000.
- Makeig S, Westerfield M, Jung T-P, Enghoff S, Townsend J, Courchesne E, Sejnowski TJ. "Dynamic brain sources of visual evoked responses." *Science* 295:690-694, 2002.
- Makeig S, Delorme A, Westerfield M, Jung T-P, Townsend J, Courchesne E, Sejnowski TJ "Electroencephalographic brain dynamics following manually responded visual targets." *PLOS Biology*, 2(6):747-762, 2004.
- Makeig S, Debener S, Onton J, Delorme A. "Mining event-related brain dynamics." *Opinion, Trends in Cognitive Science*, 8(5):204-210, 2004.
- Noesselt, T., et al. "Delayed striate cortical activation during spatial attention," *Neuron* 35: 575-587, 2002.

- Onton J, Delorme A, Makeig S, "Frontal midline EEG dynamics during working memory." In Press, *Neuroimage*, 2005.
- Osterberg, G. "Topography of the layer of rods and cones in the human retina." *Acta Ophthalmol.*, suppl. 6, 1-103, 1935.
- Peters, RJ, Iyer, A, Itti, L & Koch, C. Components of bottom-up gaze allocation in natural images, *Vision Research*, 2005. (in press)
- Potter, M.C., & Levy, E.I. "Recognition memory for a rapid sequence of pictures." *Journal of Experimental Psychology*, 81, 10-15, 1969.
- Potts, GF. An ERP index of task relevance evaluation of visual stimuli. *Brain Cogn.* 56(1):5-13, 2004.
- Ramírez, R.R., and Makeig, S. (2006). Neuroelectromagnetic Source Imaging Using Multiscale Geodesic Neural Bases and Sparse Bayesian Learning. Human Brain Mapping Conference 2006, Florence, Italy.
- Ramírez, R.R., Weber, D.L., Delorme, A., Dale, C.L., Simpson, G.V., Makeig, S. (2005) Cortical networks involved in directed spatial attention and target detection: recursive adaptive minimum-norm ICA source imaging. Society for Neuroscience Conference 2005, Washington DC, USA.
- Ranganath, C. & Rainer, G. Neural mechanisms for detecting and remembering novel events. *Nat Rev Neurosci* 4 , 193-202, 2003.
- Raymond, J.E., Shapiro, K.L., & Arnell, K.M. "Temporary suppression of visual processing in an RSVP task: an attentional blink?" *Journal of Experimental Psychology*. Human Perception and Performance, 18, 849-860, 1992.
- Reddy, L, Wilken, P and Koch, C, Face-gender discrimination is possible in the near-absence of attention. *J. Vision* 4: 106-117, 2004.
- Rousselet, G. A., Fabre-Thorpe, M. and Thorpe, SJ, *Nat Neurosci* 5, 629-630, 2002.
- Savage, L. J. The foundations of statistics (Dover, New York, 1972). (First Edition in 1954).
- Schapire, R, "A brief introduction to boosting", *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, 1999.
- Schroeder CE, Foxe JJ. The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Brain Res Cogn Brain Res.* 14:187-98, 2002.
- Schultz, W. Predictive reward signal of dopamine neurons. *J Neurophysiol.* 80(1):1-27., 1998
- Schultz, W. & Dickinson, A. Neuronal coding of prediction errors. *Annu Rev Neurosci* 23, 473-500, 2000.
- Stern, C. E., Corkin, S., Gonzalez, R. G. et al. The hippocampal formation participates in novel picture encoding: evidence from functional magnetic resonance imaging. *Proc Natl Acad Sci U S A* 93, 8660-8665, 1996.
- Thorpe, S, Fize, D, Marlot, C. Speed of processing in the human visual system. *Nature.* 6;381(6582):520-2, 1996.

- Van Orden, K, Jung, T-P, Makeig, S, "Combined eye activity measures accurately estimate changes in sustained visual task performance," *Biological Psychology*, 52:221-40, 2000.
- VanRullen, R., and Thorpe, S. "The time course of visual processing: From early perception to decision making," *J. Cogn. Neurosci.* 13: 454-461 (2001).
- Viola, P. and Jones, M. "Rapid object detection using a boosted cascade of simple features." In *Proc. CVPR*, pages 511–518, 2001.
- Viola, P and Jones, M, "Robust Real-time Object Detection", *International Journal of Computer Vision*, Volume 57, Issue 2, pp. 137 – 154, 2004.
- Wundt, W. (1913) *Grundriss der Psychologie* (Engelmann, Leipzig: Germany).

List of Acronyms

RSVP = rapid serial visual presentation

NERC system = our 'networked EEG response classification' system

RVM = relevance vector machine

LDA = linear discriminant analysis

ROC = receiver operating characteristic

AUROC = area under the ROC curve

IA = intelligence analyst

ICA = independent component analysis

IC = independent component

heptunx = a maximally compact grouping of 7 elements in a hexagonal grid

image clip, clip = Part of large broad area image presented in one frame of RSVP sequence

image chipping = Extracting small image clips from a large image


Program Effort – Publications To Date

VR 4962
13 January 2007 Disk Used


ARTICLE IN PRESS

No. of Pages 11, Model 5-
Jayalakshmi (GE) / Gnanasekar (TE)

Available online at www.sciencedirect.com



ELSEVIER



ScienceDirect

Vision Research xxx (2007) xxx–xxx

Vision
Research

www.elsevier.com/locate/vires

The duration of the attentional blink in natural scenes depends on stimulus category

Wolfgang Einhäuser ^{a,*}, Christof Koch ^{a,b}, Scott Makeig ^c

^a Division of Biology, California Institute of Technology, Pasadena, CA, USA
^b Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA, USA
^c Swartz Center for Computational Neuroscience, University of California at San Diego, San Diego, CA, USA

Received 6 March 2006; received in revised form 26 November 2006

Abstract

Humans comprehend the “gist” of even a complex natural scene within a small fraction of a second. If, however, observers are asked to detect targets in a sequence of rapidly presented items, recognition of a target succeeding another target by about a third of a second is severely impaired, the “attentional blink” (AB) [Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, 18, 849–860]. Since most experiments on the AB use well controlled but artificial stimuli, the question arises whether the same phenomenon occurs for complex, natural stimuli, and if so, whether its specifics depend on stimulus category. Here we presented rapid sequences of complex stimuli (photographs of objects, scenes and faces) and asked observers to detect and remember items of a specific category (either faces, watches, or both). We found a consistent AB for both target categories but the duration of the AB depended on the target category.

© 2007 Published by Elsevier Ltd.

Keywords: Attentional blink; RSVP; Natural scenes; Faces; Categorization

1. Introduction

When processing complex natural stimuli, humans grasp the “gist” of a scene within a small fraction of a second. This remarkable capability has often been probed using rapid serial visual presentation (RSVP) tasks. In an early demonstration, Potter and Levy (1969) presented series of images at rates between 0.5 and 8 Hz. After each of these RSVP sequences, subjects were asked to look through a set of images and to decide for each image whether it had been presented in the sequence. While the ability of subjects to recollect the scenes dropped with presentation speed, they still performed above chance at the highest tested rate (8 Hz). Biederman (1981) demonstrated that subtle violations of natural relations—such as a fire-hydrant standing on top of a mail box—are detectable in scenes, presented as briefly as 150 ms and followed by a mask. Coarse categorization of objects (e.g., animal vs. non-animal) in natural scenes is possible for stimuli displayed for only 20 ms (unmasked), though in these experiments the earliest category-dependent signal in the event-related potential (ERP) began about 150 ms after stimulus onset (Thorpe, Fize, & Marlot, 1996). All these findings highlight the remarkable processing speed of the human visual system, especially for complex natural stimuli.

When observers are instructed to respond to or remember a particular item (“target”) in an RSVP sequence, the detection of a second target (T2) is impaired if it is presented in close succession (about 200–600 ms) after the first target (T1). This impairment, the so-called “attentional blink” (AB), is absent if T2 appears directly after T1 (Raymond, Shapiro, & Arnell, 1992). In their original report of the AB, Raymond et al. (1992) defined T1 by its color (a white letter in a sequence of black letters) and T2 by the

0042-6989/\$ - see front matter © 2007 Published by Elsevier Ltd.
doi:10.1016/j.vires.2006.12.007

Please cite this article in press as: Einhäuser, W. et al., The duration of the attentional blink in natural scenes depends ..., *Vision Research* (2007), doi:10.1016/j.vires.2006.12.007

1. [Please right click and select *Acrobat Document Object* → *Open* to see the full article in press, above. This reports preliminary psychophysical data from our collaboration].