Brain activity-based image classification from rapid serial visual presentation

Nima Bigdely-Shamlo, Andrey Vankov, Rey R. Ramirez, Scott Makeig

Abstract — We report the design and performance of a brain-computer interface (BCI) system for real-time single-trial binary classification of viewed images based on participant-specific dynamic brain response signatures in high-density (128-channel) electroencephalographic (EEG) data acquired during a rapid serial visual presentation (RSVP) task. Image clips were selected from a broad area image and presented in rapid succession (12/s) in 4.1-s bursts. Participants indicated by subsequent button press whether or not each burst of images included a target airplane feature. Image clip creation and search path selection were designed to maximize user comfort and maintain user awareness of spatial context. Independent component analysis (ICA) was used to extract a set of independent source time-courses and their minimally-redundant low-dimensional informative features in the time and time-frequency amplitude domains from 128-channel EEG data recorded during clip burst presentations in a training session. The naive Bayes fusion of two Fisher discriminant classifiers, computed from the 100 most discriminative time and time-frequency features respectively, was used to estimate the likelihood that each clip contained a target feature. This estimator was applied online in a subsequent test session. Across 8 training/test session pairs from 7 participants, median area under the receiver operator characteristic (ROC) curve, by tenfold cross-validation, was 0.97 for within-session and 0.87 for between-session estimates, and was nearly as high (0.83) for targets presented in bursts that participants mistakenly reported to include no target features.

***Index Terms*—Real-time systems, electroencephalogram, EEG, independent component analysis, ICA, rapid serial visual presentation, RSVP, target detection, classification, brain computer interface, BCI.**

INTRODUCTION

The prospect of 'reading the mind' of participants engaged in some task by interpreting their non-invasively recorded brain signals has long seemed intriguing. Recent advances in computer technology make sufficient computing power readily available to collect data from a large number of scalp electroencephalographic (EEG) sensors and to perform sophisticated spatiotemporal signal processing in near-real time. A primary focus of recent work in this direction is to create brain-computer interface (BCI) systems that might allow completely paralyzed individuals direct brain-actuated control of communication or other devices. Another approach to building 'neurotechnology' to extract useful information from EEG signals is to model the brain dynamics of a participant performing some task, then to use this model to obtain subsequent, near real-time estimates of some features of their task performance or environment from their EEG activity, an approach first successfully demonstrated for monitoring of operator alertness [1].

Sajda et al. [2], [14], [16] have recently demonstrated another use for EEG-based neurotechnology, to monitor participants' brain responses while they are engaged in visual search through a broad-area image. The intended goal was to enhance the performance of satellite imagery analysts by, in effect, performing a rapid 'triage' selection of a small number of image features for later, more detailed inspection. To accomplish this, they proposed to detect EEG dynamic features associated with brain detection of attended target features in a stream of image 'clips' drawn from a larger search image and presented using rapid serial visual presentation (RSVP). Here, we report the design and performance of a system for this purpose using spatial filtering by independent component analysis (ICA) of high-density EEG signals immediately following image clips presented at a relatively high rate (12/s), and discuss its possible generalizations for purposes including communication with 'locked-in' persons.

STIMULI

Overlapping small image clips from a publicly available satellite image of London, with superimposed small target airplane images, were shown to 8 participants in RSVP mode. Each image clip focused on an elliptical ~26 × 16 m^2 ground area. Broad-area map context for the image clip was added to each clip by warping the broad area image around the magnified image clip focal center, followed by Gaussian blurring and contrast reduction (Fig. 1). Image clips were presented in 12/s in 4.1-s bursts comprised of 49 image clips each showing an enlargement of a small, elliptical clip of the satellite image plus a surrounding warped, blurred, and contrast-reduced transform of the surrounding area.

Clear airplane targets were added to some of these clips such that each burst contained either zero (40%) or one (60%) target clip (Fig. 1d). To prevent interference from EEG responses to burst edges, clips containing targets were never presented closer than 500 ms from burst onset or offset. To clearly distinguish target and non-target clips, only complete

Manuscript received October 15, 2007. The experiments and paper preparation were supported by National Science Foundation USA (NSF 0613595) and by the Swartz Foundation (Old Field, NY).

airplane target images were added, though they could appear anywhere and at any angle in the focal clip region.

Fig. 2 shows a burst trial timeline. After fixating a cross (left) for 1 s, the participant viewed the RSVP burst and was then asked to indicate whether or not he/she had detected a plane in the burst clips, by pressing one of two (yes/no) finger buttons. In Training sessions only, visual error/correct feedback was provided. Training and Test sessions were each comprised of 504 RSVP bursts organized into 72 bouts with a self-paced break after each bout. In all, each session thus included 290 target and 24,104 non-target image presentations.

METHODS

Image preprocessing and experimental protocol

To maintain information about the broader spatial context during detailed RSVP-based search, we developed a scene warping method of constructing image clips that allows continuous display of different spatial scales of a high-resolution image around a small 'focal' search area in the center of each image clip (Fig. 1). By this means, the participant may retain basic knowledge and visual memory of the relation of the focal clip area to the broader image context during RSVP-paced visual search. Retinal receptor distribution and oculomotor range are anisotropic, extending further in the horizontal than in the vertical direction. Also, saccadic eye movements are elliptically distributed, with a constant aspect ratio of about 1.6 (horizontal/vertical) for a variety of naturalistic tasks and display sizes [3], [4], [11]. On the supposition that an elliptical field mirroring the natural geometry of human vision may allow more efficient target detection at RVSP rates too high (> 2-4 Hz) to allow natural saccade-based inspection, we therefore tested use of elliptical clip foci with a 1.6 (width/height) eccentricity (Fig. 1d)

Choice of an RSVP image clip path through the set of image clips dissecting a broad area image should both a) maximize the viewer's sense of spatial context by increasing image sequence locality (thus avoiding long linear search path segments), and b) maximize participant acceptability by minimizing any sense of vertigo (e.g., thus avoiding frequent long jumps and 'typewriter' or spiral search paths). To these ends we tested use of 'heptunx search,' a path-generation algorithm implemented on a 2-D hexagonal (rather than rectangular) grid decomposed into a nested hierarchical set of roughly self-similar patterns with hexagonal symmetry that tile the grid (Fig. 3), as noted a generation ago by H. William Gosper, and therefore referred to as 'Gosper snowflakes.' Gosper noted the existence of a ('flowsnake') path starting at the largest image scale and burrowing down into smaller scales, similar in spirit to Peano's Cartesian space-filling path. By contrast, we developed an RSVP search path strategy that, beginning at any point in the lowest level set, exhaustively explored each level in the hierarchy recursively until every focal area in the broad-area image was visited.

In this heptunx search, each image clip focal area is an ellipse centered on a hexagonal grid node. The basic 'heptunx' pattern (or level-1 heptunx) is a hexagon with a seventh node in the center (after 'quincunx', the five-dot pattern used on dice, from the Latin word for 5/12). In the heptunx RSVP search path, the 7 elliptical focal areas from each level-1 heptunx are presented consecutively in random spatial order. The search then proceeds hierarchically. After displaying image clips obtained for a level-N heptunx pattern, the search path algorithm moves to the superordinate level (N+1). This scale pattern consists of 7 level-N heptunx patterns arranged in a similar filled-hexagon or heptunx pattern. Its level-N heptunx are again explored in random order (Fig. 3).

The multi-scale heptunx hierarchy can 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 at each scale in the broad-area image. The resulting heptunx search path mimics qualities of gaze paths used by participants in 2-D visual search tasks **[**5**].** It is designed to maximize the temporal stability of near-circular area context during search while minimizing the number of large jumps in the search path that may introduce radical changes in image context and visual texture (for example avoiding frequent jumps from desert to urban areas). Such sudden alterations in visual texture might induce EEG responses to visual surprise [10] overlapping those induced by perception of visual targets, thereby interfering with EEG target-response classification.

In our experiments, we searched a level-5 heptunx pattern superimposed on a public-domain image of urban London [12], using 21 random heptunx paths through the image space beginning at different starting points on the hexagonal grid. No path was presented to any participant more than once.

Eight volunteer participants (seven female and one male, aged 19-29 years), recruited from the university community, participated in two two-hour RSVP recording sessions conducted under the aegis of a University human use review board. In each session, the participant sat in a standard office chair in a dimly-lit room viewing a CRT computer screen on which a series of 4.1-s image-clip bursts were presented, each cued by a participant button press. Following each burst, the participant pressed one of two finger buttons, upon visual cue, to indicate whether or not he/she had detected a target airplane in any of the burst image clips. During Training sessions, visual (Correct/Error) feedback was given after each button press. Fig. 4 illustrates schematically the real-time stimulation, acquisition, data processing and interactive display system set-up for the Training and Test sessions.

Data acquisition

EEG was collected using a BIOSEMI Active View 2 system with 256 electrodes mounted in a whole-head elastic electrode cap (E-Cap, Inc) with a custom near-uniform montage across the scalp, neck, and bony parts of the upper face. Computer data acquisition was performed via USB using a customized acquisition driver at a 256-Hz sampling rate with 24-bit digitization. We used 128 channels for real-time classification; data from all 256 channels were stored for later re-analysis[1].

[1] The data from this study can be obtained for method development and comparison purposes from nima@sccn.ucsd.edu.

Data preprocessing

Preprocessing of Training session data was performed in the Matlab (Mathworks, Inc.) environment using EEGLAB [13] functions and data structures. For Test sessions, the same preprocessing steps were replicated in the Borland C++ environment using low-level CPU-specific vector optimization provided by the MaxVec library (Dew Research, LLC).

First, the scalp channel data were re-referenced from the active-reference Biosemi EEG data to an electrode over the right mastoid. Next, frequency-domain filtering was performed on the EEG data in three steps: (1) IIR High-pass filtering above 2 Hz (using a Butterworth filter of order 2) to remove low-frequency trends. (2) IIR low-pass filtering below 50 Hz (order-3 Butterworth) to remove high-frequency noise. (3) IIR Band-reject filtering between 40 Hz and 80 Hz (order-3 Butterworth) to minimize 60-Hz line noise. Infinite impulse-response (IIR) filters were used to achieve the desired frequency response using shorter filter lengths, thereby reducing computation time and eliminating the need to place incoming data in a long buffer, thus minimizing the classification delay produced by filtering. Following Training sessions, 7.1-s EEG data epochs each centered on a 4.1-s RSVP burst were extracted from the continuous EEG and concatenated.

Channel interpolation

Following Training sessions, noisy EEG channels were identified by calculating the maximum absolute correlation of each channel with other scalp channels within a 5-cm radius using 1-s non-overlapping time windows. If this value fell below 0.5 for more than 1% of the experiment duration, the channel was labeled noisy and discarded. The following algorithm was used to replace the activities of noisy channels for EEG classification in Test sessions:

1. Eliminate the rows of the ICA mixing matrix M associated with the 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.

$$\mathbf{M}_{trunc} = \mathbf{M}_{train}; \quad \mathbf{M}_{trunc}(noisy,:) = [] \tag{1}$$

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

$$\mathbf{U}_{trunc} = \mathbf{M}^{+}_{trunc} \tag{2}$$

3. Obtain a new set of activation time series A using the truncated unmixing matrix and the Test session data D.

$$\mathbf{A}_{trunc} = \mathbf{U}_{trunc}\mathbf{D}_{test} \tag{3}$$

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

$$\mathbf{D}_{virtual} = \mathbf{M}_{train}\mathbf{A}_{trunc} \tag{4}$$

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

$$\mathbf{D}_{denoised}(noisy,:) = \mathbf{D}_{virtual}(noisy,:) \tag{5}$$

We found this algorithm to give better classification performance than channel replacement using interpolation using nearest-neighbor channels or spherical splines. Importantly, this method can be successfully applied in cases even when a local group of channels covering a large area becomes noisy, since the projections to the scalp surface via volume conduction of independent components accounting for brain EEG activity are typically quite broad. Typically, about 5 to 10 of the 128 channel signals were interpolated during on-line response classification.

Independent component analysis

Since the duration of recorded Training session data (~1 hour) imposed a limit on the number of ICA components that could be learned accurately from the data, a uniformly spaced channel subset with roughly 128 channels was selected from the non-rejected data channels. This channel subset of the Training data epochs was then decomposed using extended-infomax ICA [8],[6],[9] to obtain about 128 maximally independent components (ICs). ICA learned spatial filters in the form of an unmixing matrix separating EEG sensor data into temporally maximally independent processes, most appearing to predominantly represent the contribution to the scalp data of one brain EEG or non-brain artifact source, respectively. In Test sessions, the same ICA unmixing matrix was imported by the real-time classifier and used to obtain online estimate of IC time courses (or activations).

Near-real time image classification

After observing significant changes in EEG activity pattern near the end of each burst due to transition from RSVP to a blank screen and degradation of performance on the last two images in the sequence caused by this, we decided not to include these last images in the classification. In practice, these two images might always be presented again in subsequent bursts.

Fig. 6. summarizes the classification process. Information in both time-domain and frequency domain features of the EEG signals following image clip presentations was exploited. Time-course templates were selected by a novel method that involved, for every IC, performing a second-level deconvolutive ICA decomposition on the time-domain IC activations following image clip presentations. To do this, we constructed matrices for each IC of size lags (205) by images presented (~ 24000) from concatenating columns consisting of IC activation time series in the 800 ms following presentation of each image. By applying principal component analysis (PCA) we then reduced the row number of this matrix from 205 to 50. A second-level ICA was then performed to find 50 independent time course templates (ITs) for each IC. These features form a natural spatiotemporal decomposition of the EEG signal with low interdependency, making them natural candidate features for an EEG classifier.

Time-frequency amplitude ITs were found similarly by decomposing the single-trial spectrograms of the IC activations following image clip onsets, which were vectorized and concatenated into a (frequencies×latencies by images) matrix that was reduced to its first 50 principal dimensions by PCA before ICA decomposition.

Observing that, as expected, the weights for a number of these time and time-frequency domain dimensions contained robust information for target/non-target image classification,

we used the 100 most informative ITs of each domain from all ICs (as determined by computing individual area under the ROC curves for the target/non-target discrimination) as input to a Fisher discriminant linear classifier. For time-frequency domain features, we used the frequency-domain independent template weights for each image as input. For time-domain features, we used columns of the transpose of the second-level ICA mixing matrix as matched filters (Fig. 5), since in preliminary testing this gave slightly better results.

We found that the number of retained features that maximized the area under the ROC curve (by five-fold validation) for within-session target/non-target classification on the Training session data alone did not determine the optimum number of features for between-session classification. However, computing the area under the ROC curve for between-session classification for all pairs of experiments using different numbers of selected time-domain classifier features showed that the performance of the classifier was not highly dependent on the number of features; retaining 70-150 features gave near-equivalent results.

Stimulus delivery and online response classification, schematized in Fig. 4, was programmed on a small personal computer (PC) local area network (LAN) running Windows XP using software built on a DataRiver software platform for online multi-data acquisition, synchronization, and computation (A. Vankov). Online classification of the 49 image clips in each 4.1-s burst required less than 2.5 s on the Pentium 4 CPU (3.2 GHz) PC.

Results

Table I shows the area under the ROC curve for ten-fold cross validation using different feature types for classification. Combining results from classifiers acting on time-domain and frequency domain features by naïve Bayes (multiplication of their posterior probabilities) gave best classification performance. Within-session classification was at least 0.9 for all participants. Table II shows the area under the ROC curve for the actual real-time across-sessions classification. Overall classification was 0.78-0.95 for 7 of the 8 participants (median, 0.88). Columns show, for each participant, the area under the ROC curve for all bursts (trials), bursts correctly assessed by the participant, those in particular that did include a clip containing a target airplane, for bursts incorrectly responded, for bursts containing a target that the subject missed, and finally the number of such bursts. Fig. 7 shows ROC curves with true and false positive rates for all bursts in cross session classification.

Interestingly, for most participants (6 out of 8) the performance of the EEG classifier in trials that they missed reporting the presence of a target was relatively high. Median area under ROC curve for those behaviorally-missed bursts was 0.83, far above chance (0.5). Thus, the real-time classifier was quite often able to detect an EEG signature consistent with brain target detection even when it was not followed by behavioral detection.

Neural signals

Fig. 8 (left panel) shows, for a representative participant training session, the envelopes of the average event-related potential (ERP) difference between target and non-target activations of the maximally projecting brain ICs, back-projected into channel space and superimposed on the outline or 'envelope' of the ERP difference for all channels, after removal of ICs accounting primarily for eye movements and muscle artifacts. The difference in non-artifact (left) following presentation of an image clip containing a target feature is largest between 180 and 380 ms and also near 500 ms after clip onset. For this participant, target feature recognition was associated with (left) partial EEG phase locking in posterior visual areas as well as in central midline and left frontal EEG activities.

Differences in mean activities of ICs accounting for electrooculographic (EOG) activities are shown in the right panel of Fig. 8. The largest post-target versus non-target difference in mean eye movement-related activity occurs between 600 and 700 ms. Target feature detection was also associated, on average, with eye-movement related ICs were included in the real-time classifier when they contained useful information about target detection. However, offline testing of classifier performance omitting eye-movement ICs resulted in less than 1% degradation in performance (ROC area under the curve), demonstrating that nearly all the information used in the classifier was derived from the activities of ICs *not* accounting for non-brain artifacts. The decision to include or exclude eye-movement related ICs in a working EEG-based classification system should best depend on the relative reliability of the two types of information, which could only be determined by more extended within-subject testing.

Fig. 9 panels (a) show scalp maps and equivalent dipole models for each of six example ICs used in the real-time classifier for the same participant and Training session as in Fig. 8. We may define the 'envelope' of a multi-channel time waveform to be its maximum and minimum mean potential values, across channels, at each time point. Panels (b) plot the envelopes of the mean difference between IC activations following targets and non-target clips, back-projected onto the scalp channels (red), and outline the envelope of the difference summed overall classifier ICs (black traces). Panels (d) show the envelope of the summed linear classifier template for each ICs, back-projected onto the scalp channels (blue fill), and the envelope of the back-projected classifier template for all ICs used in the classifier (black traces).

While the envelopes of the mean target minus non-target difference (c) and the summed classifier templates (d) are similar, clear differences are apparent for some ICs. None of the component features is concentrated solely at one response peak (for example at the so-called 'P300' response peak near 350 ms), and all include unexpected oscillatory patterns with a peak frequency between 16 and 32 Hz. Only one (IC2) of these peaks is at a harmonic of the stimulus presentation rate (12 Hz). They would appear to distinguish the phase of activity at or near these frequencies during the response period. As expected, replacing the (blue filled) learned templates with the (red filled) mean IC response differences in offline testing reduced classifier performance.

Both time-domain and time/frequency-domain features were included in the online classifier on the basis of pilot results showing improved performance on pilot data sessions from their combined use. In Fig. 10, similar differences between the mean target minus non-target event-related

spectral power differences, and the summed time/frequency IC template, are shown for two of the ICs. For both component processes, target presentations are associated with increases theta band energy near 350 ms (left), the learned templates (right) being more tightly focused in time and frequency, and including a number of smaller time/frequency contrasts not seen in the (target – non-target) differences (left).

Conclusions

Currently, reliable computer vision methods for real-world detection of many features of interest in satellite images are not yet available. The proposed system for RSVP-based detailed inspection of satellite or other large image reported here demonstrates that relatively high-quality information is available in the human scalp EEG about brain 'flickers of recognition' that follow perception of attentively searched target image features (e.g., as here, airplane targets in satellite images), even at RSVP rates as high as 12 images per second. Such a system might thus be useful for accomplishing an orderly search of a portion of a large high-resolution image for some class of image features for which current computer vision algorithms are not effective [15]. The method might be most useful for thorough searches of large, relatively featureless image areas (e.g. oceans, deserts), conditions under which not multi-scale context clues are available and unaided human visual search may be inefficient.

As classifier performance may not be expected to be perfect, its output might be used to select (or to 'triage') image locations for later, more detailed search, as first proposed by Sajda et al. [2]. An advantage of this approach may be that, whereas a computer vision algorithm would have to be trained for each new target feature, an EEG-based system might be target-type invariant if, as we expect, brain responses are specific only to the detection of an attended target and relatively independent of the specific target feature. If this were the case, the flexibility of the expert participant brain's attentional tuning and feature recognition system might avoid the necessity to train an individualized classifier for each new target type.

Although the classification performance results we report here seem robust, direct quantitative comparison with results of other EEG response classification methods is unfortunately not possible, since any observed differences might arise from differences in stimulation parameters (for example, rate, size, and target location). We plan to make our data available on request to anyone interested in performing post hoc classification method comparisons.

We here successfully demonstrated brain response classification using an RSVP rate (12/s) far above the human saccade rate (3-4/s). However, our experience suggests that the overall speed of RSVP-based search should be limited in practice by at least three factors. First, since high-resolution vision occurs only in the fovea, the maximum effective size of the focal areas of the image clips yielding successful visual or EEG-based detection of difficult image feature-detection targets is small, particularly when RSVP rates are too high (> ~2/s) to allow integrative saccadic eye movements. While the size of the image clip foci we used appeared, in preliminary psychophysical testing, to be near optimal for the RSVP rate we used, more detailed study would be required to determine the optimal combination of image size and RSVP rate. Possibly, also, the optimal target sizes to use for behavioral and psychophysiological classification might differ.

Second, as the size of the attended target features may be a significant fraction of the optimum image clip focus, many such features should be expected to occupy the boundaries of two or more adjacent image-clip foci. Therefore, to achieve good detection performance the image-clip foci must be partially overlapping (as in Fig. 3e), further increasing the time required to search the image space by RSVP.

Third, the need to use a large number of relatively small image-clip foci in RSVP-based search may produce viewer visual or mental fatigue, necessitating frequent rest breaks and further slowing the search process.

Thus, RSVP-based search even using a relatively high RSVP rate (as here) might prove to be slower than expert saccade-based visual search using software with appropriate visual zoom and point of gaze history tracking capabilities. In particular, saccadic search may be preferable when low spatial frequency information in the search area (mountains, water features, etc.) act as visual landmarks and/or allow experts to exclude areas unlikely to contain targets, for example deep water areas during a search for motor vehicles.

It should be of interest to determine whether EEG-aided response classification might also improve the efficiency of saccade-based search, a process for which the human visual system and brain are likely optimized. The system reported here might be readily adapted to RSVP presentation of single letters to paralyzed, so called 'locked-in' subjects attending delivery of target letters spelling out a word they wish to communicate. The system might also be readily extended to perform classification of more than two classes of EEG response dynamics, or adapted to perform linear or nonlinear regression to estimate continuous behavioral or perceptual variables. This would allow uses for online EEG monitoring such as monitoring of alertness [1], [7], monitoring of participant attentiveness during computer-based learning, or classification of brain responses to warning signals by operators engaged in system monitoring. The ability of ICA to isolate and segregate many classes of non-brain artifacts associated with eye and head muscle activity might also allow implementation in mobile operator environments.

ACKNOWLEDGMENT

We acknowledge the programming assistance of Sashikanth Madduri**,** and valuable advice received during discussions with Nuno Vasconcelos, Hamed Masnadi-Shirazi**,** Wolfgang Einhaueser-Treyer**,** Jason Palmer, Tzyy-Ping Jung, and Laurent Itti. The system described was first developed under contract HM1582-05-C-0044 from the Defense Advanced Research Project Agency (DARPA), USA. Final system development, the testing reported here, and preparation of this report were supported by a grant from the National Science Foundation USA (NSF 0613595) and by a gift from the Swartz Foundation (Old Field, NY).

REFERENCES

[1] S. Makeig and M. Inlow, "Lapse in alertness: coherence of fluctuations in performance and EEG spectrum" *Electroencephalogr. And Clin. Neurophysiolog* .vol. 86, January 1993, pp. 23-35.

[2] P. Sajda, A. Gerson and L. Parra, "High-throughput image search via single-trial event detection in a rapid serial visual presentation task" in *Proc. 1st Inter. IEEE EMBS Conf. on Neural Engineering*, Capri Island, Italy, 2003.

[3] G. A. Osterberg, "Topography of the layer of rods and cones in the human retina," *Acta Ophthalmol.*6(Suppl.):1, 1935.

[4] D. Guitton, R. M. Douglas, and M. Volle, "Eye-head coordination in cats," *J. Neurophysiol.*, vol. 52, pp. 1030-50, 1984.

[5] A. L. Yarbus, *Eye Movements and Vision.* Plenum Press, New York, NY, USA, 1967.

[6] T. Lee, M. Girolami, T. J. Sejnowski, "Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Subgaussian and Supergaussian Sources," *Neural Computation*, Feb. 15, 1999, vol. 11, no. 2, pp. 417-441.

[7] T.P. Jung, S. Makeig, M. Stensmo, and T. J. Sejnowski, "Estimating Alertness from the EEG Power Spectrum," *IEEE Transactions on Biomedical Engineering*, 1997, vol. 44, pp. 60-69.

[8] A. Bell and T. Sejnowski. "An Information Maximization Approach to Blind Separation and Blind Deconvolution," *Neural Computation*, July 1995, vol. 7, pp. 1129-1159.

[9] A. Delorme, S. Makeig, "EEG changes accompanying learning regulation of the 12-Hz EEG activity," *IEEE Transactions on Rehabilitation Engineering*, 2003, vol. 11, no.2, pp. 133-136.

[10] W. Einhuser , T. N. Mundhenk, P. Baldi, C. Koch, and L. Itti, "A bottom-up model of spatial attention predicts human error patterns in rapid scene recognition," *J. of Vision*, 2007, vol.7 no.10, pp.1-13.

[11] W. Einhauser-Treyer, California Institute of Technology, Pasadena, CA, personal communication.

[12] London satellite map. Available: http://www.mapmart.com/DownloadArea/MapMart_Samples/Quickbird/cssample.zip

[13] Available: http://sccn.ucsd.edu/eeglab

[14] A. Gerson, L. Parra and P. Sajda, "Cortically-coupled computer vision for rapid image search," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2006, vol. 14 (2), pp. 174-179.

[15] DARPA NIA project: http://www.darpa.mil/dso/thrusts/trainhu/nia/

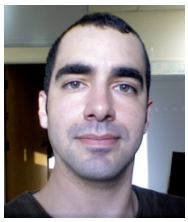
[16] L.C. Parra, C. Christoforou, A. D. Gerson, M. Dyrholm, A. Luo, M. Wagner, M. G. Philiastides, and P. Sajda, "Spatio-temporal linear decoding of brain state: Application to performance augmentation in high-throughput tasks," *IEEE Signal Processing Magazine*, Jan. 2008, vol. 25 no. 1, pp. 95-115.

Nima Bigdely-Shamlo was born in Esfahan, Iran in 1980 and graduated with a bachelor's in physics from Sharif University of Technology, Tehran, Iran in 2003. He received a Masters degree in Computational Science from San Diego State University in 2005.

He is currently a programmer/analyst at the Swartz Center for Computational Neuroscience of the Institute for Neural Computation, UCSD, working on real-time EEG classification and on new methods for analysis of simultaneous human motion capture and EEG dynamics.

Andrey Vankov was born in Sofia, Bulgaria in 1951. He received a M.D. degree from Medical University, Varna, Bulgaria in 1976 and Ph.D. degree in Neurophysiology from Bulgarian Academy of Sciences in 1985. He has worked as scientist and senior scientist at the Bulgarian Academy of Sciences, at the University Pierre and Marie Curie, Paris, France, and from 1994 at the University of California San Diego. In 2002 he joined the Swartz Center for Computational Neuroscience, at which he is presently Senior Development Engineer.

Dr. Vankov is a member of the Bulgarian Physiological Society, the Society for Neuroscience, and the Society for Experimental Biology and Medicine.



Rey R. Ramirez was born in Santurce, Puerto Rico in 1974, and graduated with a bachelor's degree in Biological Sciences from Rutgers University in 1996. In 2005 he received a Ph.D. in Neuroscience and Physiology from New York University School of Medicine, developing new inverse methods for magnetoencaphalography.

He is currently a Postgraduate Researcher at the Swartz Center for Computational Neuroscience of the Institute for Neural Computation, UCSD, working on methods for the analysis of multimodal brain imaging data, and the application of these methods for cognitive and clinical neuroscience.

Dr. Ramírez is a member of the Society for Neuroscience, the Organization for Human Brain Mapping, and the Cognitive Neuroscience Society.

Scott Makeig was born in Boston, MA, USA in 1946 and completed a bachelor's degree, 'Self in Experience,' at the University of California Berkeley in 1972. He received a Ph.D., 'Music psychobiology,' from the University of California San Diego (UCSD) in 1985.

After a year in Ahmednagar, India as a American India Foundation research Fellow, he became a Psychobiologist at UCSD, and then a Research Psychologist at the Naval Health Research Center, San Diego. In 1999, he became a Staff Scientist at the Salk Institute, La Jolla, and moved to UCSD as a Research Scientist in 2002 to develop the Swartz Center for Computational Neuroscience, which he now directs. Dr. Makeig is a member of the Society for Neuroscience, the Cognitive Neuroscience Society, and the Society for Human Brain Mapping.

TABLE I
AREA UNDER ROC CURVE FOR TENFOLD WITHIN-SESSION CROSS VALIDATION

Participant	Time and Frequency domain features	Time domain features	Frequency domain features
10	0.96	0.94	0.91
19	0.97	0.94	0.95
20	0.97	0.94	0.90
22	0.97	0.95	0.94
23	0.99	0.98	0.96
24 [a]	0.91	0.90	0.83
25	0.90	0.89	0.75
21	0.98	0.96	0.94

[a]Participant 24 participated in no Test session.

TABLE II
AREA UNDER ROC CURVE FOR BETWEEN-SESSION CLASSIFICATION

Participant	All Bursts	Correct Response	Bursts Containing a Target, Correct Response	Incorrect Response	Bursts Containing a Target, Incorrect Response (Behaviorally Missed)	Number of Behaviorally Missed Bursts
10	0.89	0.93	0.93	0.81	0.81	93
10 [a]	0.92	0.97	0.96	0.86	0.87	153
19	0.78	0.78	0.78	0.64	0.61	15
20	0.84	0.88	0.88	0.80	0.80	155
22	0.86	0.87	0.87	0.84	0.84	48
23	0.95	0.96	0.96	0.91	0.91	39
25	0.56	0.61	0.61	0.49	0.50	90
21	0.88	0.89	0.88	0.87	0.84	192

[a]Participant 10 participated in two Test sessions.

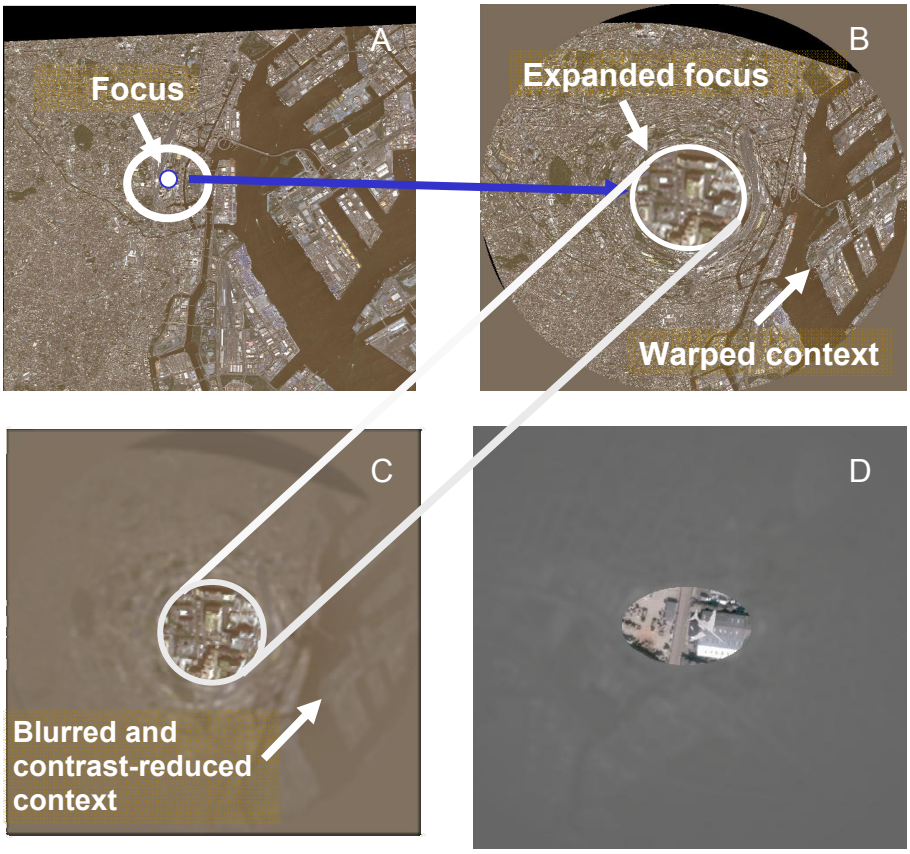


Fig. 1. Image clip extraction and scene-warping: (A) Selecting the clip focus (center dot) (B) Expanding the focal area and warping the large-scale image around it (C) Blurring and contrast-reducing the broad area surround. (D) Example of a warped, blurred, and contrast reduced satellite image clip containing a superimposed target airplane image. Instead of using circular foci, as illustrated above for simplicity, the clip foci used in the experiments were elliptical to better resemble the human visual field and the eye movement pattern employed by participants during natural saccade-based search.

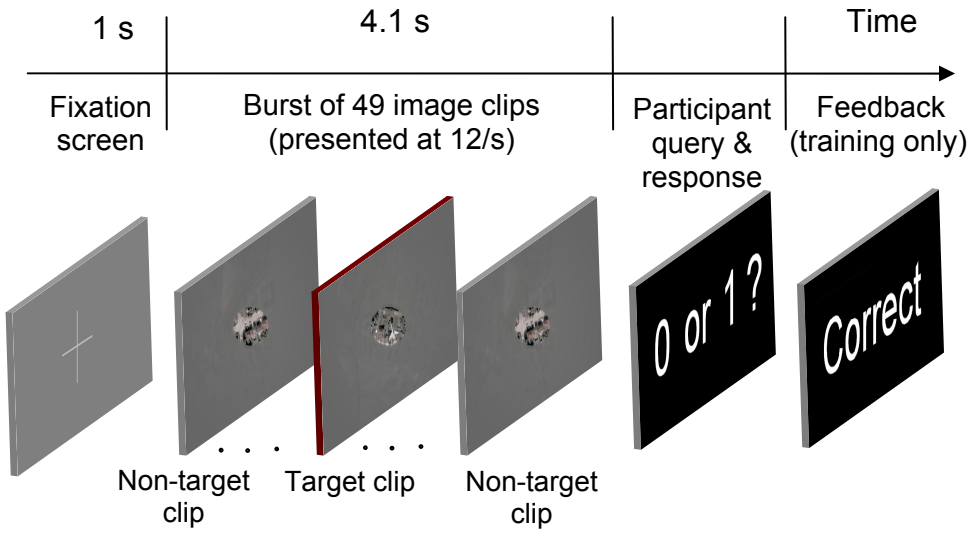


Fig. 2. Time-line of each RSVP burst. Participant response feedback ('Correct' or 'Incorrect') was delivered only during Training sessions (rightmost panel).

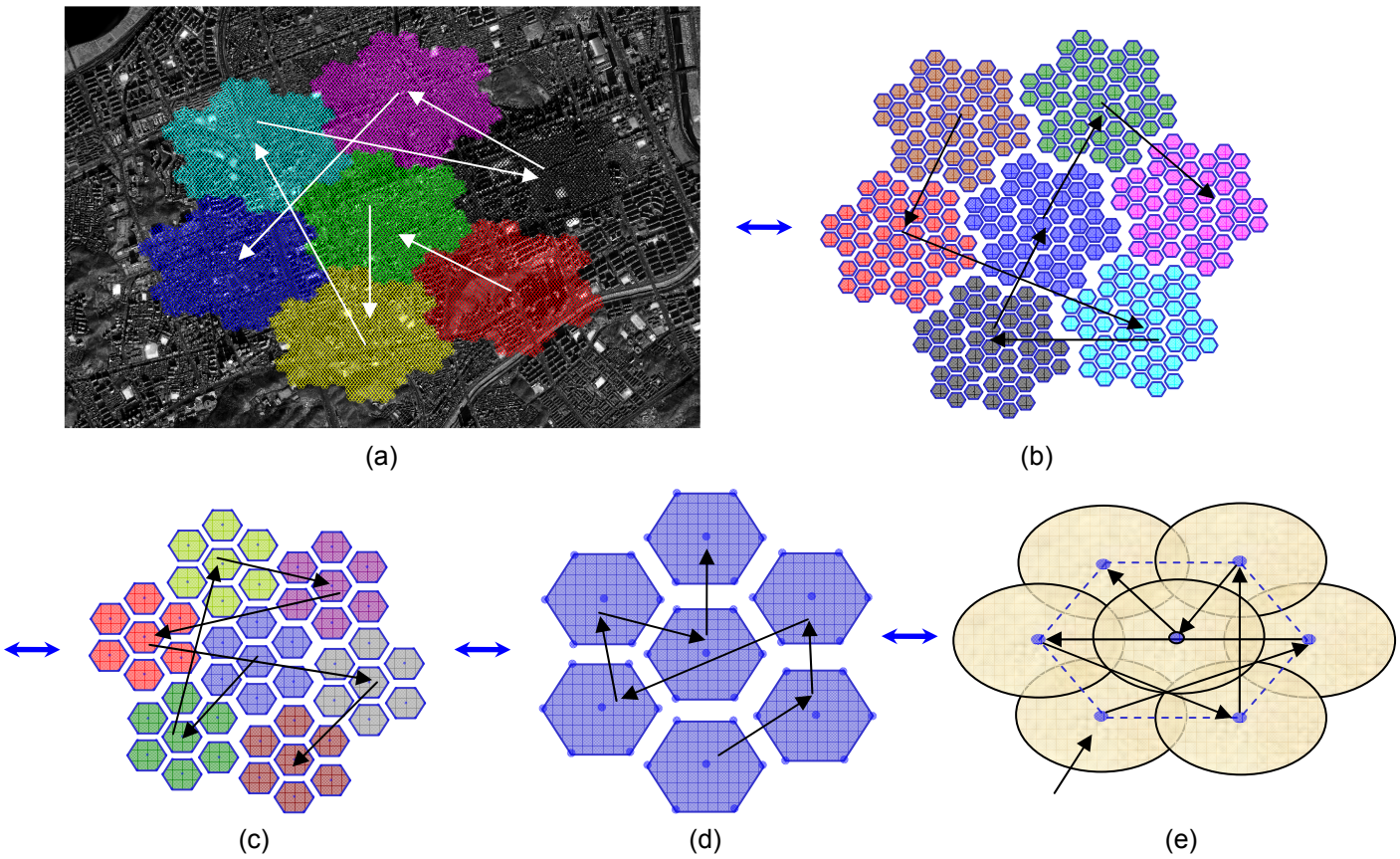


Fig. 3. Heptunx search path construction: (a) A random level-5 heptunx search path through seven level-4 heptunx placed on a large satellite image. Note the employed horizontal stretching of the heptunx grid. (b)-(d) Schematic level-4 through level-2 heptunx grids and sample search paths. (e) Elliptical image clip focal areas superimposed on a level-1 heptunx (dashed hexagon), with a sample random search path (arrows). Heptunx search proceeds from level-1 through successively higher level heptunx until the desired search space has been covered.

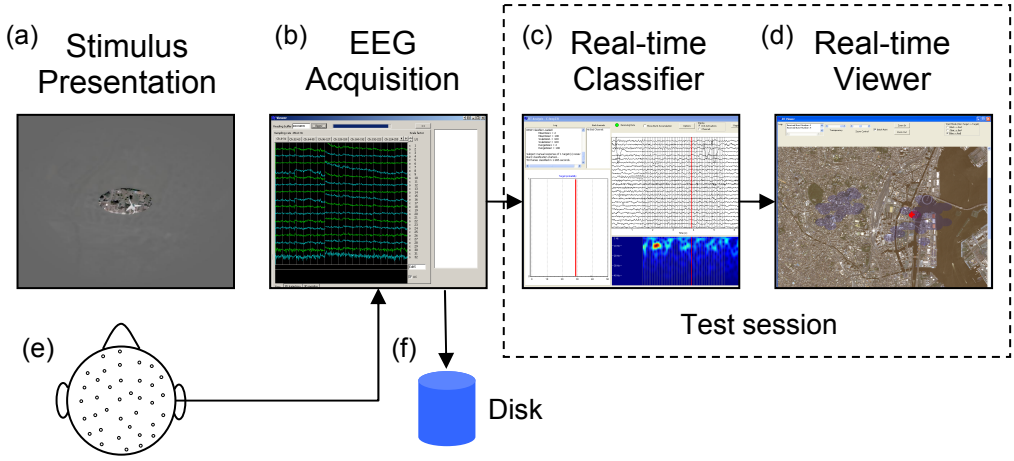


Fig. 4. Schematic of the EEG-based response classification system operation. The participant (e) wears an electrode cap while viewing bursts of image clips presented in rapid succession on a CRT screen (a). EEG is amplified, converted to digital format and sent by optical fiber to an EEG acquisition computer (b) that displays the data to the experimenter, and also (f) saves the raw data to disk. During Test sessions, (c) other computer nodes compute and visualize for the experimenter the information about the progress of the classifier. (d) After each image burst, the computed target probabilities are painted on the broad-area image from which the image clips are drawn using an interactive display application. The processes shown above are coordinated by a suite of interacting software daemons operating on three CPUs on a local area network (LAN).

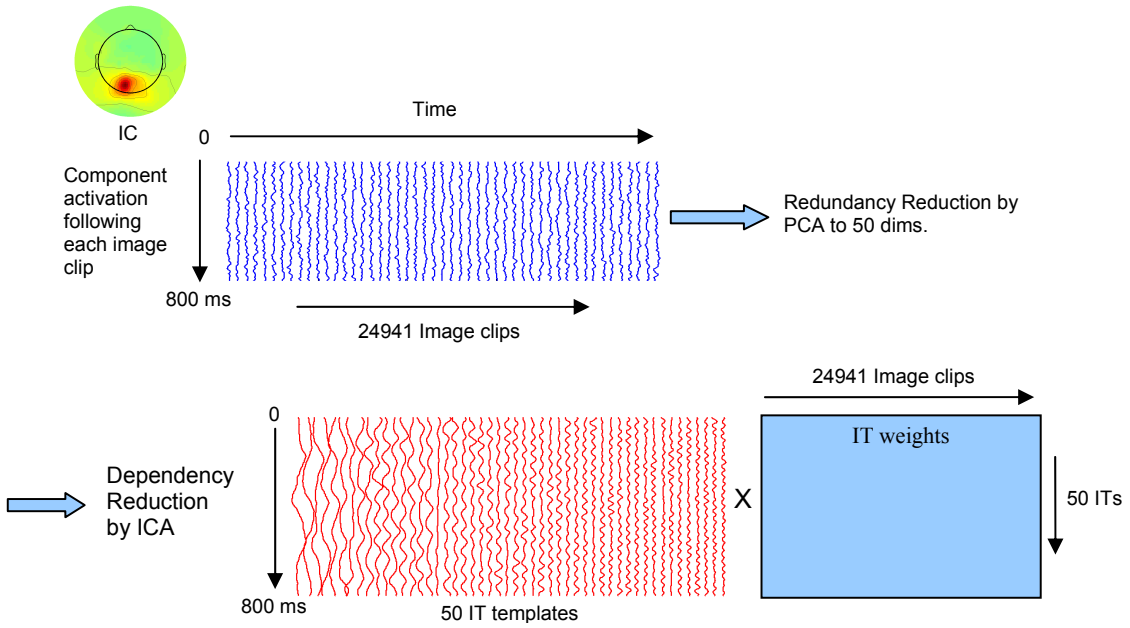


Fig. 5. Constructing independent time-domain templates (ITs) from the activity of an independent component (IC) following image clip presentations.

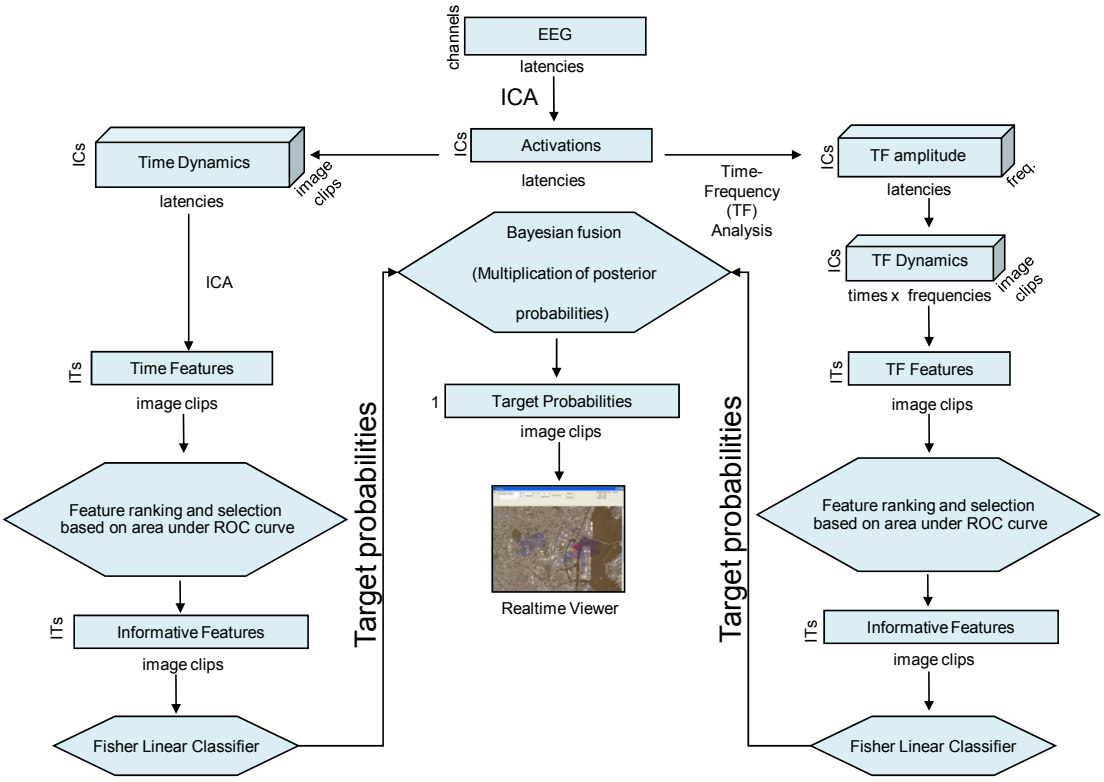


Fig. 6. Flowchart of the real-time classifier system. Both (left) temporal activity patterns (phase-locking following stimulus onsets) and (right) event-related changes in spectral amplitude are used by the classifier (center). Results are painted onto the broad-area map from which the individual image clips are drawn; color coded for the probability that the image clip area contains a target feature.

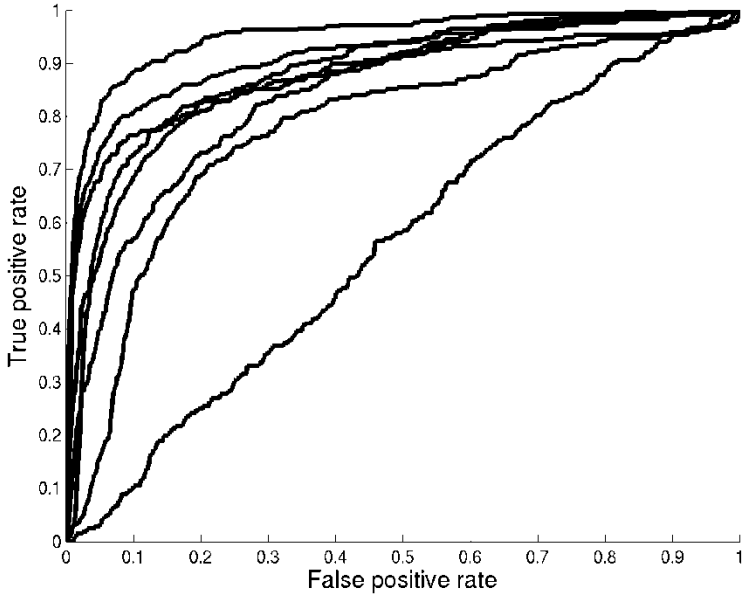


Fig. 7. Receiver operating characteristic (ROC) curves for all 8 test sessions of 7 subjects (one subject participated in two test sessions). The area under the ROC curve is above 0.78 for all but one subject.

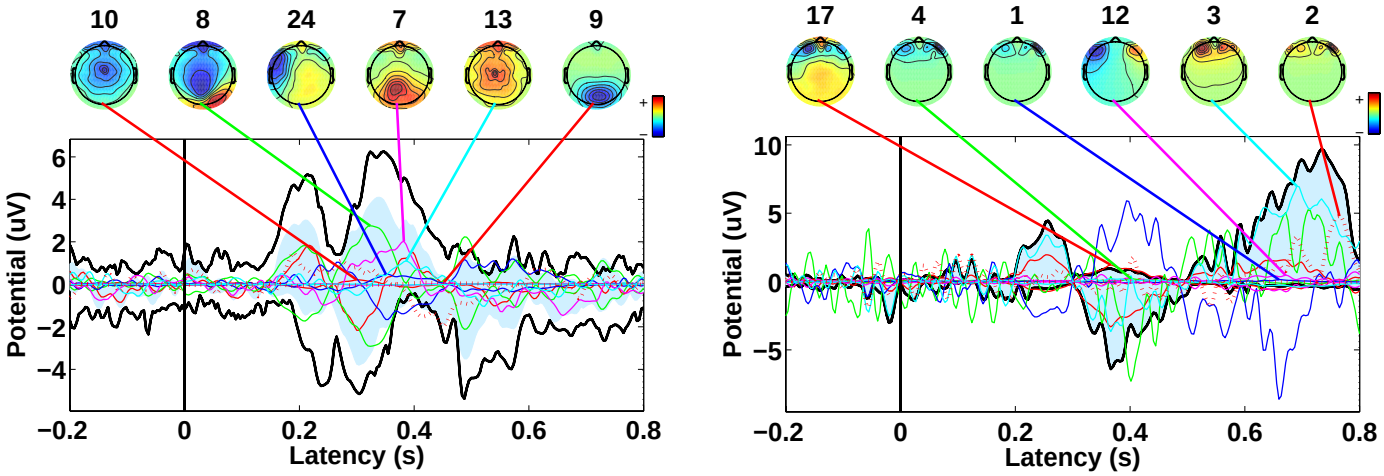


Fig. 8. Mean time-domain target-related activity for one session. Black traces show the 'envelopes' of mean event-related potential (ERP) differences between target and non-target responses, across all channels, defined as the maximum and minimum mean potential values, across channels, at each time point. Complementary left and right panels show (left) the mean response difference with eye movement-related ICs removed, and (right) summing the activities of the eye movement-related ICs only. The colored traces show the envelopes of the data accounted for by the ICs contributing the most to the mean difference (the IC indices given above their respective scalp maps). Light blue areas show the envelope of the summed activities of the highlighted ICs.

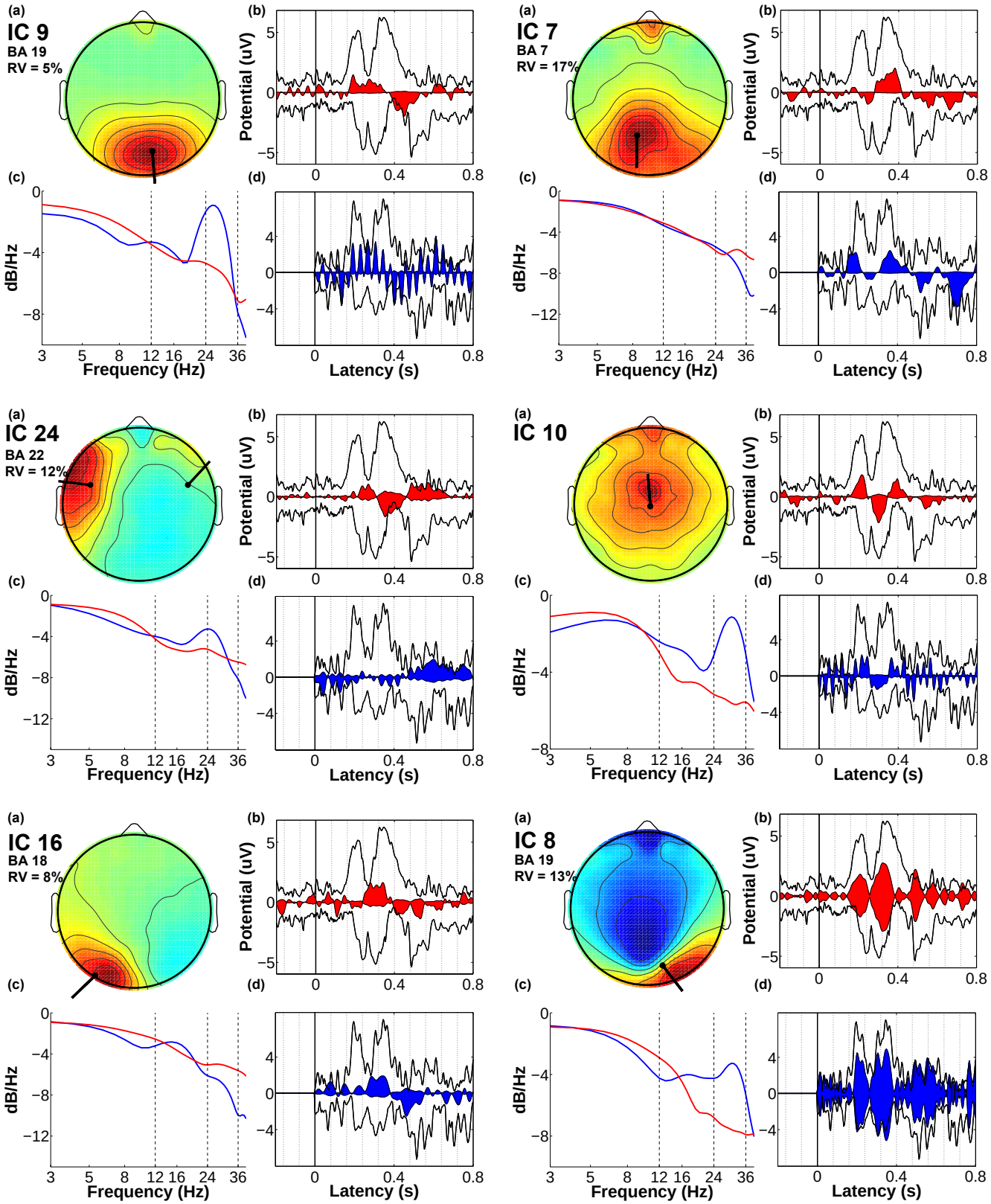


Fig. 9. Time-domain information used in the classification. Highly informative independent components (IC) from the same session as Fig. 8, giving (a) the IC scalp maps, equivalent dipole models, the nearest cortical Brodmann area (BA), and residual variance (RV) in the IC scalp map not accounted for by the model. One example is shown (IC2) for which a dual-symmetric equivalent dipole model better accounted for the IC scalp map. (For IC 10 the best-fitting single equivalent dipole was localized to below the cortex). (b) Black traces show the envelope of the mean target minus non-target difference at all channels (as in Fig. 8, left), while red areas fill the envelope of the mean difference for the activity accounted for by the IC. (c) Mean spectra of the IC activation (red), and the summed IC classifier template (blue). (d) (blue) Filled envelope of the summed linear classifier template for this IC, back-projected onto the scalp channels, against (black traces) the summed back-projected classifier template from all ICs used in the classifier. Note the similarities and differences between (b) and (d).

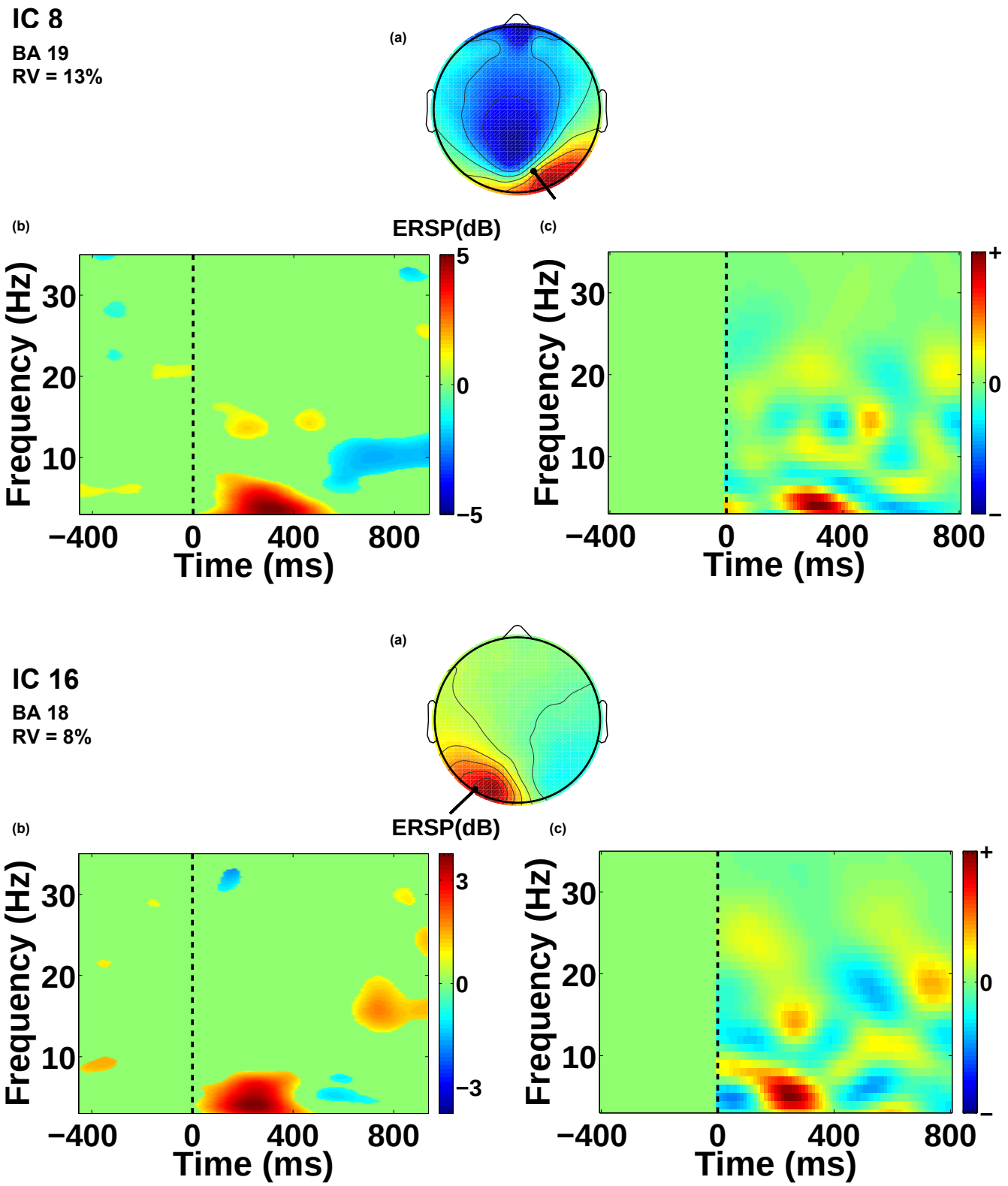


Fig. 10. Time-frequency domain information used in the classification.(a) Independent Component (IC) scalp map and equivalent dipole model, plus nearest cortical Brodmann area (BA) and unaccounted residual variance (RV) (as in Fig. 9). (b) Mean event-related spectral perturbation (ERSP) difference between EEG following target and non-target images (masked to show significant values ($p<0.02$) as non-green). (c) Summed time-frequency domain linear classifier template for the IC. Note similarities and differences between (b) and (c).