IOP Publishing Journal of Neural Engineering
J. Neural Eng. **11** (2014) 046003 (12pp) doi:10.1088/1741-2560/11/4/046003

Neurally and ocularly informed graph-based models for searching 3D environments

David C Jangraw[1], **Jun Wang**[2], **Brent J Lance**[3], **Shih-Fu Chang**[2,4] **and Paul Sajda**[1,2,5]

[1] Department of Biomedical Engineering, Columbia University, New York, NY 10027, USA
[2] Department of Electrical Engineering, Columbia University, New York, NY 10027, USA
[3] Translational Neuroscience Branch, Human Research and Engineering Directorate, Army Research Laboratory, Aberdeen Proving Ground, MD 21001, USA
[4] Department of Computer Science, Columbia University, New York, NY 10027, USA
[5] Department of Radiology, Columbia University, New York, NY 10027, USA

E-mail: psajda@columbia.edu

Received 10 September 2013, revised 4 March 2014
Accepted for publication 26 March 2014
Published 3 June 2014

Abstract

Objective. As we move through an environment, we are constantly making assessments, judgments and decisions about the things we encounter. Some are acted upon immediately, but many more become mental notes or fleeting impressions—our implicit 'labeling' of the world. In this paper, we use physiological correlates of this labeling to construct a hybrid brain–computer interface (hBCI) system for efficient navigation of a 3D environment. *Approach.* First, we record electroencephalographic (EEG), saccadic and pupillary data from subjects as they move through a small part of a 3D virtual city under free-viewing conditions. Using machine learning, we integrate the neural and ocular signals evoked by the objects they encounter to infer which ones are of subjective interest to them. These inferred labels are propagated through a large computer vision graph of objects in the city, using semi-supervised learning to identify other, unseen objects that are visually similar to the labeled ones. Finally, the system plots an efficient route to help the subjects visit the 'similar' objects it identifies. *Main results.* We show that by exploiting the subjects' implicit labeling to find objects of interest instead of exploring naively, the median search precision is increased from 25% to 97%, and the median subject need only travel 40% of the distance to see 84% of the objects of interest. We also find that the neural and ocular signals contribute in a complementary fashion to the classifiers' inference of subjects' implicit labeling. *Significance.* In summary, we show that neural and ocular signals reflecting subjective assessment of objects in a 3D environment can be used to inform a graph-based learning model of that environment, resulting in an hBCI system that improves navigation and information delivery specific to the user's interests.

Keywords: brain–computer interface, EEG, eye tracking, pupillometry, computer vision

S Online supplementary data available from stacks.iop.org/JNE/11/046003/mmedia

(Some figures may appear in colour only in the online journal)

1. Introduction

Most brain–computer interface (BCI) research endeavors to help disabled users navigate and interact with the world (Wolpaw *et al* 2002). For paralyzed users, BCIs have been used to drive wheelchairs (Galan *et al* 2008, Leeb *et al* 2007), operate robotic arms (Hochberg *et al* 2006), and navigate assistive robots (Perrin *et al* 2010); for 'locked-in' patients, BCIs can be used to type messages (Sellers and Donchin 2006). The goal of these BCIs is to restore, at least in part, some function of the human body that has been lost, and this goal limits the user base to a small group possessing certain

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
03 JUN 2014		**00-00-2014 to 00-00-2014**

4. TITLE AND SUBTITLE
Neurally and ocularly informed graph-based models for searching 3D environments

5a. CONTRACT NUMBER
5b. GRANT NUMBER
5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER
5e. TASK NUMBER
5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Columbia University,Department of Computer Science,116th Street and Broadway,New York,NY,10027

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)
11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Objective. As we move through an environment, we are constantly making assessments judgments and decisions about the things we encounter. Some are acted upon immediately, but many more become mental notes or fleeting impressions?our implicit ?labeling? of the world. In this paper, we use physiological correlates of this labeling to construct a hybrid brain?computer interface (hBCI) system for efficient navigation of a 3D environment. Approach. First, we record electroencephalographic (EEG), saccadic and pupillary data from subjects as they move through a small part of a 3D virtual city under free-viewing conditions. Using machine learning, we integrate the neural and ocular signals evoked by the objects they encounter to infer which ones are of subjective interest to them. These inferred labels are propagated through a large computer vision graph of objects in the city, using semi-supervised learning to identify other, unseen objects that are visually similar to the labeled ones. Finally the system plots an efficient route to help the subjects visit the ?similar? objects it identifies. Main results. We show that by exploiting the subjects? implicit labeling to find objects of interest instead of exploring naively, the median search precision is increased from 25% to 97%, and the median subject need only travel 40% of the distance to see 84% of the objects of interest. We also find that the neural and ocular signals contribute in a complementary fashion to the classifiers? inference of subjects? implicit labeling. Significance. In summary, we show that neural and ocular signals reflecting subjective assessment of objects in a 3D environment can be used to inform a graph-based learning model of that environment, resulting in an hBCI system that improves navigation and information delivery specific to the user?s interests.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**12**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

disabilities. Although BCIs for healthy users have long been the subject of speculation and science fiction, traditional BCI inputs like motor imagery and the P300 remain slower, less reliable substitutes for physical input methods like the mouse and keyboard (Zander *et al* 2010).

The prospect of BCIs for able-bodied users offers an opportunity to vastly expand the BCI audience and expose the field to the benefits of increased scale (including monetary resources, rigorous testing and the support of a large community of users), so interest in this objective has grown in recent years (Pfurtscheller *et al* 2010, Wang *et al* 2009b, Allison 2010, Lance *et al* 2012). One approach is a shift from explicit inputs, which the user must generate for the purpose of operating the BCI, to naturally evoked ones, produced without the intent of computer control. These can be brain signals, like theta power (Grimes *et al* 2008), or other physiological signals, like galvanic skin response (Allanson and Fairclough 2004). Naturally evoked signals offer the distinct advantage of requiring little to no user effort or remapping of thought to action (Zander *et al* 2010), but BCIs using these signals are limited in the scope of applications they can address since the signals must be produced instinctively or even subconsciously. To achieve the best of both worlds, some 'hybrid BCI' (hBCI) systems have begun to fuse multiple modalities that use naturally evoked signals like heart rate or pupil accommodation speed in concert with explicit control signals like motor imagery or SSVEP (Lee *et al* 2010, Pfurtscheller *et al* 2010). These systems use multiple modalities of input to create multi-dimensional control signals or correct for errors.

Even so, healthy users have physical input alternatives available, and are therefore unlikely to tolerate the number of incorrect classifications produced by even the most accurate hBCI. But the study of realistic stimuli and scenarios, an important step towards 'mobile' BCIs for healthy users (Bayliss and Ballard 2000, Healy and Smeaton 2011, Brouwer *et al* 2013), has introduced an opportunity to use environmental context to further improve BCI results. If the user is searching for a consistent type of object, a graph-based semi-supervised computer vision (CV) system can use measures of visual similarity to reject false positives and find other, unseen objects that might also be of interest (Wang *et al* 2009b, Pohlmeyer *et al* 2011). In this way, CV's broad awareness of environmental context can be used to classify a multitude of objects based on hBCI output, but without requiring the user to view all of them.

The emerging research area of fixation-related potentials (FRPs), initially used in the context of reading (Dimigen *et al* 2011, Hutzler *et al* 2007), has revealed visual search as a strong opportunity for an intuitive, context-conscious hBCI. Recent studies have shown that fixations on a target stimulus initiate EEG responses similar to the P300 elicited by a flashed target stimulus (Dandekar *et al* 2012a, Kamienkowski *et al* 2012). Several studies have successfully classified fixations as being on a target or distractor stimulus (Brouwer *et al* 2013, Healy and Smeaton 2011, Luo and Sajda 2009). In similar visual search paradigms, studies have shown that subjects tend to fixate longer on targets than on distractors, and this tendency has been exploited for computer control (Jacob 1991). Pupil size, which has long been known to correlate with interest and mental effort (Hess and Polt 1960, 1964, Kahneman and Beatty 1966), also changes with memory load during visual search (Porter *et al* 2007). Thus, the single act of visual search naturally evokes both neural and ocular signals that are distinct for targets and distractors. But whether these signals remain informative in a naturalistic, dynamic scenario—and whether they are productive to include in a classifier together, or are merely redundant indicators of the same internal state—remains unclear. Our study investigates whether each modality can provide information that is independent from the others to an hBCI, and whether CV can be used to further improve classification when visual search is conducted in a realistic environment.

The system we present in this paper employs a user's naturally evoked EEG, eye position and pupil dilation to construct a hybrid classifier capable of distinguishing objects of interest from distractors as the user moves past objects in a three-dimensional (3D) virtual environment. We show that the hybrid classifier is more accurate than one trained using any one of the three modalities alone. The system also uses a CV graph to reject anomalies in the hybrid classifier's predictions and find other, visually similar objects in the environment, including new objects that the user has not yet visited. We show that using CV increases the precision and size of the predicted target set well beyond that of the hybrid classifier. Finally, the system plots an efficient route to assist the user in visiting the targets it predicts. Our study provides insight into how naturally evoked neural and ocular signals can be simultaneously exploited and integrated with environmental data to enable augmented search and navigation in an hBCI application.

2. Methods

2.1. System overview

The system is designed to plot an efficient route to search for objects of interest (or 'targets') in a large mapped environment. The environment contains many objects, and limited data about each object is available—in this case, visual features extracted from the object and its position in the environment, but no text tags or human labeling.

A system block diagram is shown in figure 1. The user explores a small fraction of the environment looking for targets, and her EEG, eye position and pupil size are tracked as she explores. Artifacts are removed from the data, and potentially discriminatory features are extracted. Using a 2-stage classifier, these features are used to produce a set of hBCI predicted targets. A CV system then tunes this set to reject false positives and extrapolates it to find other visually similar objects in the environment. The most visually similar objects are labeled as CV predicted targets, and the system plots an efficient route to visit them. By traversing this route, the user should see more targets per unit of distance traveled than if she explored the environment without the system.

This study implements and tests a proof-of-concept version of this system in a 3D virtual environment. Subjects

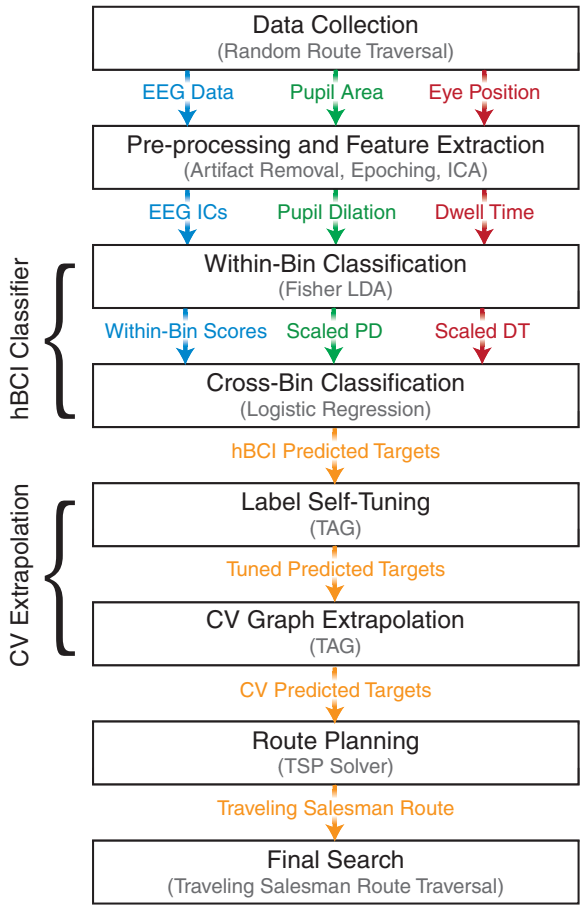

Figure 1. Modular framework for the proposed system. Each box represents a stage of processing described in detail in subsections 2.3–2.10 of the text. The stage's general function appears in black at the top of the box. The method(s) used to serve that function in the current study appear below it in gray. Arrows between the boxes represent the EEG (blue), pupil (green), gaze (red) and multimodal (orange) inputs/outputs passed between the stages. ICA = independent component analysis, LDA = linear discriminant analysis, PD = pupil dilation, DT = dwell time, hBCI = hybrid brain–computer interface, TAG = transductive annotation by graph, CV = computer vision, TSP = traveling salesman problem.

are navigated through a grid of streets and asked to count objects of a target category while also performing a secondary driving-related task. The signals naturally evoked by this task, along with the CV features of our virtual stimuli, are used as input to our system, which identifies an efficient route to find predicted targets in unexplored parts of the environment. The next subsection outlines our virtual environment, and the following subsections correspond to the sequential stages shown in figure 1.

2.2. Virtual environment

The system was tested in a 3D virtual environment, making it possible to present a realistic yet consistent background to subjects while randomizing the stimuli. The environment was constructed using Unity 3D game development software (Unity Technologies, CA). It consisted of a grid of streets with two alleys on each block, one on either side. The subject's viewpoint was automatically navigated down the streets as if

Figure 2. Screenshot of the virtual city as viewed by the subject. The subject was instructed to (1) count the number of billboards (right) that belonged to a certain target category, and (2) press a button whenever the leading car (center) illuminated its brake lights and slowed down. A video of the subject's view is presented in supplementary movie 1 (available from stacks.iop.org/JNE/11/046003/mmedia).

riding in a car. The environment was displayed to the subject on a 30″ Apple Cinema HD display with a 60 Hz refresh rate, and subtended approximately 30 × 23 visual degrees.

In each pair of alleys, a square 'billboard' object was placed so that the object gradually became visible as the subject passed it (see figure 2 and supplementary movie 1) (available from stacks.iop.org/JNE/11/046003/mmedia). The image on the billboard was selected from a subset of images from the CalTech101 database (Fei-Fei *et al* 2007). The subset consisted of 50 images in each of four categories: car sides, grand pianos, laptops and schooners. These categories were chosen because they were photos (not drawings) and were well represented by the CV system (see simulations in Pohlmeyer *et al* (2011), supplementary information). The identity of the image, and the side of the subject's viewpoint on which it appeared, was randomized (with replacement). Subjects were asked to count objects of one category (targets) and ignore the others (distractors) but make no physical response. They were allowed to move their eyes freely during the task. The subjects saw 20 objects per block, and each block lasted approximately 100 s. At the end of each block, they were asked to verbally report the number of target objects they had seen. 13–15 blocks were recorded so that each subject observed 260–300 objects, about 25% of which were targets. Each object was in view for approximately 1160 ms. Although the luminance of the stimuli was not standardized, the target category was randomly assigned for each subject.

To keep the subjects engaged and make the driving simulation more realistic, subjects were also asked to press a button when a car in front of them illuminated its brake lights and slowed to half its normal speed. When the button was pressed, this 'leading car' would speed back up to return to its default distance in front of the subject. The time between braking events was randomly selected with a uniform distribution between 5 and 10 s. This secondary task also served to default the subjects' gaze to the center of the screen.

2.3. Data collection

Ten healthy volunteer subjects were recruited for this study (ages 19–42, 3 female, 1 left-handed). All reported normal or corrected-to-normal vision. Informed consent was obtained in writing from all participants in accordance with the guidelines and approval of the Columbia University Institutional Review Board. Each subject was provided with a set of written task instructions and was shown a small subset of the stimuli before the first block. If the subject failed to press the button in response to the leading car braking (which took place in the first block of two subjects), that block was aborted and removed from analysis.

EEG data were amplified with a gain of 1000 and collected at 1000 Hz from 77 Ag/AgCl electrodes (selected from a 10–10 montage) using a Sensorium DBPA-1 Amplifier (Sensorium Inc., VT). Recordings were referenced to the left mastoid with a forehead ground. All electrode impedances were less than 50 kΩ, while the amplifier has an input impedance of 100 GΩ. The amplifier applied high-pass and low-pass analog filters with cutoffs at 0.01 and 500 Hz, respectively.

An EyeLink 1000 eye tracker (SR Research, Ontario, Canada) was used to collect eye position and pupil area data from one eye at 1000 Hz. The tracker was a 'tower mount' with chin and forehead rests to stabilize the subject's head. A 9-point validation was performed before each block, and if the validation was unsatisfactory the eye tracker was re-calibrated. Just before each screen update, Unity's record of the bounding box surrounding any object on the screen was sent to the EyeLink computer for recording via a dedicated TCP/IP connection. The recording setup is described further in Jangraw and Sajda (2011).

To synchronize the data, a parallel port pulse was sent from the eye tracker computer to the EEG amplifier every 2 s. The time at which the parallel port pulses were sent and received were used to synchronize the eye tracker and EEG data. The discrepancy between the two systems' records of the time between pulses was never more than 2 ms.

2.4. Pre-processing and feature extraction

Saccades and fixations were detected using the EyeLink online parser. Eye position and pupillometry data were analyzed using MATLAB (The MathWorks Inc., MA). Some blocks were found to have a large, constant eye position drift, and so a post-hoc drift correction was performed. The median eye position from each block was calculated, and the eye position for that block was shifted so that the median fell on the center of the screen (see supplementary figure S1).

Using the frame-by-frame record of each object's bounding box and the drift-corrected record of eye position, the first fixation on each object (when the fixation start position fell within 100 pixels, or 3.0°, of the object's bounding box) was identified. The first fixation away from the object (when the fixation start position fell more than 100 pixels outside the bounding box) was also determined. The 'dwell time' for each object was defined as the time between these two fixations. If the subject did not fixate on an object, that object was removed from further analysis (on average, 4.7% of objects were removed).

The subject's pupil area during each blink was estimated using linear interpolation. Each subject's pupil area data were then divided by the mean across that subject's blocks and multiplied by 100, so that the units could be interpreted as a percentage of the mean. The pupil area data were epoched from 1000 ms before to 3000 ms after the first fixation on each object. A baseline of −1000 to 0 ms was subtracted from each epoch to calculate the pupil dilation evoked by each object.

EEG data were analyzed using the EEGLAB toolbox (Delorme and Makeig 2004). The signals were band-pass filtered from 0.5 to 100 Hz, notch filtered at 60 Hz, and down-sampled to 250 Hz. All blocks were concatenated, and excessively noisy channels were removed by visual inspection (on average, 3.5 channels per subject were removed). To define a fixation onset well synchronized with EEG, we computed an average ERP locked to the first fixations on all objects, identified the peak time of the saccadic spike (a negative peak in posterior regions), and defined this point as time zero. This is similar to the method of Brouwer *et al* (2013), but we used a single timing correction for each subject rather than trial-by-trial realignment.

Components related to blink and horizontal electrooculographic (HEOG) artifacts were determined using the maximum power and maximum difference methods described in Parra *et al* (2005), but artifact-contaminated data from the task was used instead of a dedicated 'calibration paradigm' in which the subject is instructed to produce artifacts by blinking or moving his eyes. In our analysis, the blink component was the component with maximum power in periods marked by the eye tracker as blinks, and the HEOG component was the component that was maximally different when the subject happened to fixate on the far left side of the screen and the far right side of the screen. These components were projected out of the EEG data using the 'interference subtraction' method described by Parra *et al* (2005), in which the activity of each noise source is estimated from the data, projected back into sensor space, and then subtracted from the signal.

Epochs were extracted from the first 1000 ms of data after the first fixation on the object, and a post-saccadic baseline of 0 to 100 ms was subtracted as in Hutzler *et al* (2007). A voltage threshold of 75 μV was applied as in Kamienkowski *et al* (2012): if fewer than five electrodes exceeded the threshold at any point in the epoch, those electrodes were interpolated from all remaining electrodes using the inverse distance between electrodes as weights. If more than five electrodes exceeded the threshold, the epoch was discarded (on average, 1.5% of trials were discarded). The 0–100 ms baseline was subtracted again after interpolation.

To reduce the dimensionality of the feature space and avoid rank deficiency issues, principal component analysis (PCA) was performed on each subject's epoched EEG, and only the top 20 PCs were retained. Temporal independent component analysis (ICA), which identifies components whose temporal patterns of activity are statistically independent from one another, was then performed

on the data using the Infomax ICA algorithm (Bell and Sejnowski 1995, Makeig *et al* 1996). The resulting IC activities were used as features in the classifier (see supplementary figure S2 for components from one subject).

2.5. Within-bin classification

A hierarchical classifier was adapted from the hierarchical discriminant component analysis (HDCA) described in Gerson *et al* (2006), Pohlmeyer *et al* (2011) and Sajda *et al* (2010) to accommodate multiple modalities. To learn each subject's classifier, the EEG data from 100 to 1000 ms after the first fixation on the object were separated into nine 100-ms bins. A set of 'within-bin' weights across the ICs was determined for each bin using Fisher linear discriminant analysis (FLDA):

$$\mathbf{w}_j = (\mathbf{\Sigma}_+ + \mathbf{\Sigma}_-)^{-1}(\mu_+ - \mu_-) \quad (1)$$

where $\mathbf{w}_j$ is the vector of within-bin weights for bin j, μ and $\mathbf{\Sigma}$ are the mean and covariance (across training trials) of the data in the current bin, and $+$ and $-$ subscripts denote target and distractor trials, respectively. The weights $\mathbf{w}$ can be applied to the IC activations $\mathbf{x}$ from a separate set of evaluation trials to get one 'within-bin interest score' z_{ji} for each bin j in each trial i:

$$z_{ji} = \mathbf{w}_j^T \mathbf{x}_{ji}. \quad (2)$$

The within-bin interest scores from the evaluation trials will serve as part of the input to the cross-bin classifier. The use of an evaluation set ensures that if the within-bin classifier over-fits to the training data, this over-fitting will not bias the cross-bin classifier towards favoring these features.

In order to maintain a consistent sign and scale for all the inputs to the cross-bin classifier, we processed the pupil dilation and dwell time data similarly to the EEG data. The pupil dilation data from 0 to 3000 ms were separated into six 500-ms bins and averaged within each bin (the shortest time between saccades to objects was 3272 ms). For each bin, this average was passed through FLDA to create a discriminant value whose 'sign' was the same as the EEG data's (so that targets $>$ distractors). The dwell time data were also passed through FLDA. The scale of each EEG, pupil dilation and dwell time feature was then normalized by dividing by the standard deviation of that feature across all evaluation trials. A second-level feature vector $\mathbf{z}_i$ was created for each evaluation trial i by appending that trial's rescaled EEG, pupil dilation and dwell time features into a single column vector.

To examine the scalp topography of the EEG data contributing to the discriminating components, we calculated forward models for each EEG bin. For each bin j, we appended the z_{ji} values across trials into a column vector $\mathbf{z}_j$ and the $\mathbf{x}_{ji}$ vectors into a matrix $\mathbf{X}_j$. This allowed us to calculate the forward model $\mathbf{a}_j$ as follows:

$$\mathbf{a}_j = \frac{\mathbf{X}_j \mathbf{z}_j}{\mathbf{z}_j^T \mathbf{z}_j}. \quad (3)$$

This forward model can be viewed as a scalp map and interpreted as the coupling between the discriminating component and the original EEG recording.

2.6. Cross-bin classification

To classify the second-level feature vectors from each trial ($\mathbf{z}_i$), 'cross-bin' weights $\mathbf{v}$ (across temporal bins and modalities) were derived using logistic regression, which maximizes the conditional log likelihood of the correct class:

$$\mathbf{v} = \arg\min_{\mathbf{v}} \left(\sum_i \log\{1 + \exp[-c_i \mathbf{v}^T \mathbf{z}_i]\} + \lambda ||\mathbf{v}||_2^2 \right) \quad (4)$$

where c_i is the class (+1 for targets and -1 for distractors) of trial i and $\lambda = 10$ is a regularization parameter introduced to discourage overfitting. These weights can be applied to the within-bin interest scores from a separate set of testing trials to get a single 'cross-bin interest score' y_i for each trial:

$$y_i = \mathbf{v}^T \mathbf{z}_i. \quad (5)$$

The effectiveness of the classifier lies in its ability to produce cross-bin interest scores y_i that are higher for targets than for distractors. The area under the receiver operating characteristic (ROC) curve (AUC) was therefore used as a figure of merit. Trials with cross-bin interest scores more than 1 standard deviation above the mean were identified as 'hBCI predicted targets'. For comparison purposes, single-modality (EEG only, pupil dilation only and dwell time only) and dual-modality (each pair of modalities) classifiers were developed. These classifiers use the same process described above, but they classify using only the within-bin scores of one or two modalities.

The use of an evaluation set (which was not used in the original HDCA) is essential in the hybrid case to avoid overly weighting the EEG bins, since the first-level EEG classifiers have a much higher dimensionality than the ocular features and are thus more prone to overfitting (Duin 2002). Training, evaluation and testing sets were generated using nested 10-fold cross-validation. That is, for each of ten 'outer folds', one tenth of the trials were left out and placed in the testing set. Then, in each of ten 'inner folds', one tenth of the remaining trials were left out and placed in the evaluation set, and the rest were assigned to the training set. In generating the ten sets to be left out in the ten different folds, trials were grouped chronologically.

2.7. Label self-tuning

A CV system called transductive annotation by graph (TAG) was used to tune the hBCI predicted target set for each subject and extrapolate the results of our hybrid classifier to the rest of the objects in the environment. The TAG constructed a 'CV graph' containing all the images on billboard objects in the environment, using their similarity to determine connection strength (Wang *et al* 2008, 2009a). The graph employs 'gist' features (low-dimensional spectral representations of the image based on spatial envelope properties, as described in Oliva and Torralba (2001)). The similarity estimate for each pair of objects is based not only on the features of that pair, but also on the distribution of features across all objects represented in the CV graph.

The TAG performed 'label self-tuning' on the hBCI predicted target set by removing images that did not resemble

the set as a whole and replacing them with images that did (Sajda *et al* 2010, Wang *et al* 2009a, 2009b). Conceptually, the image in the hBCI predicted target set least connected to the others was deemed most likely to be a false positive. It was removed from the set and replaced with the image not in the set that was most connected to the set[6]. This process was repeated one time for every image in the hBCI predicted target set. Images in the resulting set were called 'tuned predicted targets'.

2.8. CV graph extrapolation

The tuned predicted target set was propagated through the CV graph to determine a 'CV score' for each image in the environment, such that the images with the strongest connections to the tuned predicted target set were scored most highly. A cutoff was determined by fitting a mixture of two Gaussians to the distribution of CV scores (labels were not used, but ideally one represented the distribution of targets and the other that of distractors) and finding the intersection point of the Gaussians that falls between their means. The images with CV scores above the cutoff were identified as 'CV predicted targets'. Because each image is paired with a billboard object in virtual environment space, these CV predicted targets represent our system's predictions of the places in the environment that are most likely to contain objects that the subject would like to visit.

2.9. Route planning

A traveling salesman problem (TSP) solver using the 2-opt method (Croes 1958) was modified to allow only routes on the environment's grid. This solver employed a distinct graph-based model of the environment that contained the same nodes as the CV graph (i.e., billboard objects) but different edge strengths (based on physical proximity instead of visual similarity). The TSP solver was used to produce an efficient 'traveling salesman route' (in the form of a text file list of waypoints) that the user could take to visit all the CV predicted targets in the virtual environment.

2.10. Final search

The list of waypoints can be fed back into the Unity software and traversed to view the CV predicted targets efficiently (see supplementary movie 2) (available from stacks.iop.org/JNE/11/046003/mmedia). To provide insight into the efficiency of search using the output of this system, the distance traveled and number of targets seen by following this route can be compared with a brute-force search (the route the TSP solver would recommend to see all the objects in the environment). More efficient searches will visit more targets per unit of distance traveled.

[6] In practice, images were added or removed from the predicted target set in order to maximize an objective function. This function incorporates the smoothness of the CV predicted label function across the graph and the fitting of the CV predicted labels with the hBCI-derived labels (see Wang *et al* (2009a) for more details).

3. Results

3.1. Feature averages

System testing afforded us an opportunity to observe neural and ocular signals during free viewing of a realistic environment. The subject's gaze sometimes moved to the (task-irrelevant) background of roads and buildings, not just the stimuli we had placed in the environment. Peripheral vision could be employed in the task as well. We expected to see a P300, longer dwell times and larger pupil dilations for targets, but the size and constancy of these trends in our dynamic, free-viewing scenario were unknown.

Mean target and distractor FRPs across subjects are plotted in figure 3(a). These FRPs are somewhat consistent with those reported in other target detection tasks (Brouwer *et al* 2013, Dandekar *et al* 2012a, Healy and Smeaton 2011, Kamienkowski *et al* 2012), and a P3b-like separation between target and distractor fixations is apparent on electrode Pz (Polich 2007).

The mean subject-median target and distractor timecourses of pupil dilation (median across trials, mean across subjects) are plotted in figure 3(b). The pupil contraction preceding fixation onset is likely related to motor preparation before and during the preceding saccade (Jainta *et al* 2011). The pupil dilations of target and distractor trials begin to diverge within the first second after fixation and remain separated long after the object has disappeared from view. A cumulative histogram of dwell times is shown in figure 3(c). Subjects tended to have higher dwell times for targets than for distractors, but the distributions overlap considerably.

3.2. hBCI classifier performance

The average forward models and weights learned by the hybrid classifier are shown in figure 4. EEG forward models and temporal weights correspond roughly to the P300 (Brouwer *et al* 2013, Healy and Smeaton 2011, Pohlmeyer *et al* 2011). Earlier components sometimes implicated in target detection BCIs do not appear to be influential in this classifier, perhaps due to our post-saccadic baseline. Pupil dilation is weighted highly after 1000 ms, peaking between 1500 and 2000 ms post-fixation. Dwell time is weighted more highly than any individual EEG or pupil dilation bin (but less than the sum of those bins).

The hBCI classifier's AUC for each subject are shown in figure 5 alongside those of the single-modality classifiers. Sorting subjects in descending order of EEG AUC score highlights an important quality of the hybrid classifier: when EEG classification is better than the other two modalities, hybrid classifier performance closely tracks EEG classifier performance. When another modality is superior, it tends to track that modality's performance instead. Many subjects produce strong classifiers in one area and weak classifiers in another, and the hybrid classifier's ability to rely on the best modality appears to be its greatest advantage. But in cases where more than one modality provides good information (e.g., subjects 1, 4 and 9), the hybrid classifier also tends to receive an extra boost above the best classifier. A similar

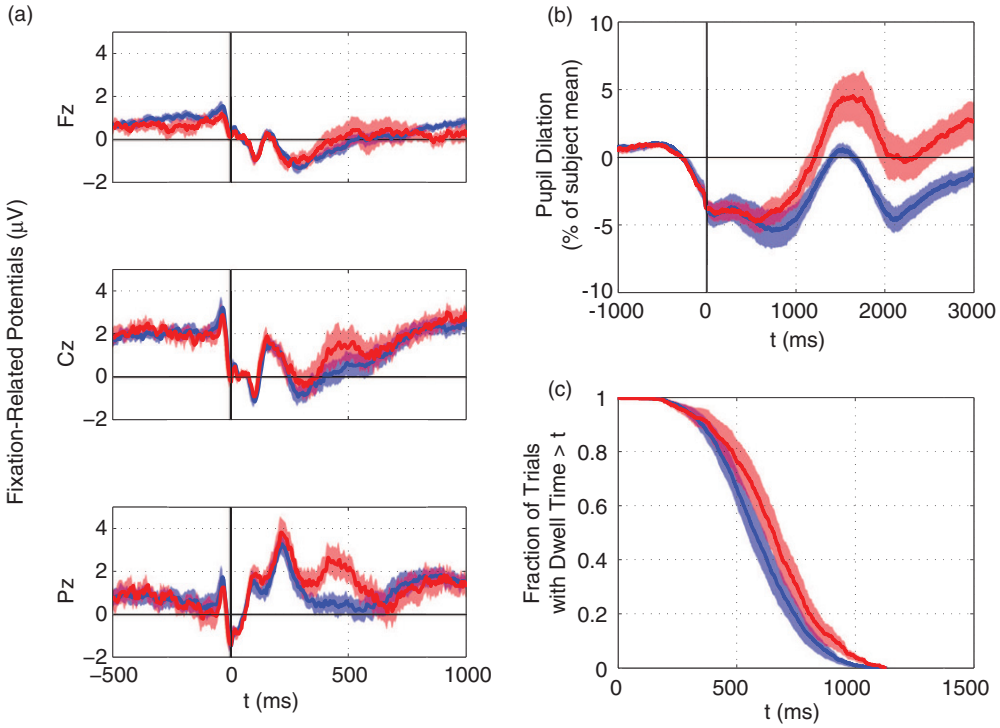

Figure 3. Average features for target (red) and distractor (blue) trials, over time (where time $t = 0$ is the start of the subject's first fixation on the object). Translucent patches represent standard error across subjects ($N = 10$). (a) Grand average fixation-related potentials at midline electrodes. Note that pre-fixation values differ from zero because a post-saccade baseline was used. (b) Mean subject-median pupil dilation (as a percentage of each subject's mean pupil area). (c) Inverse cumulative histogram of dwell times. This can be interpreted as the chance that a subject's gaze remains on the object at the given time.

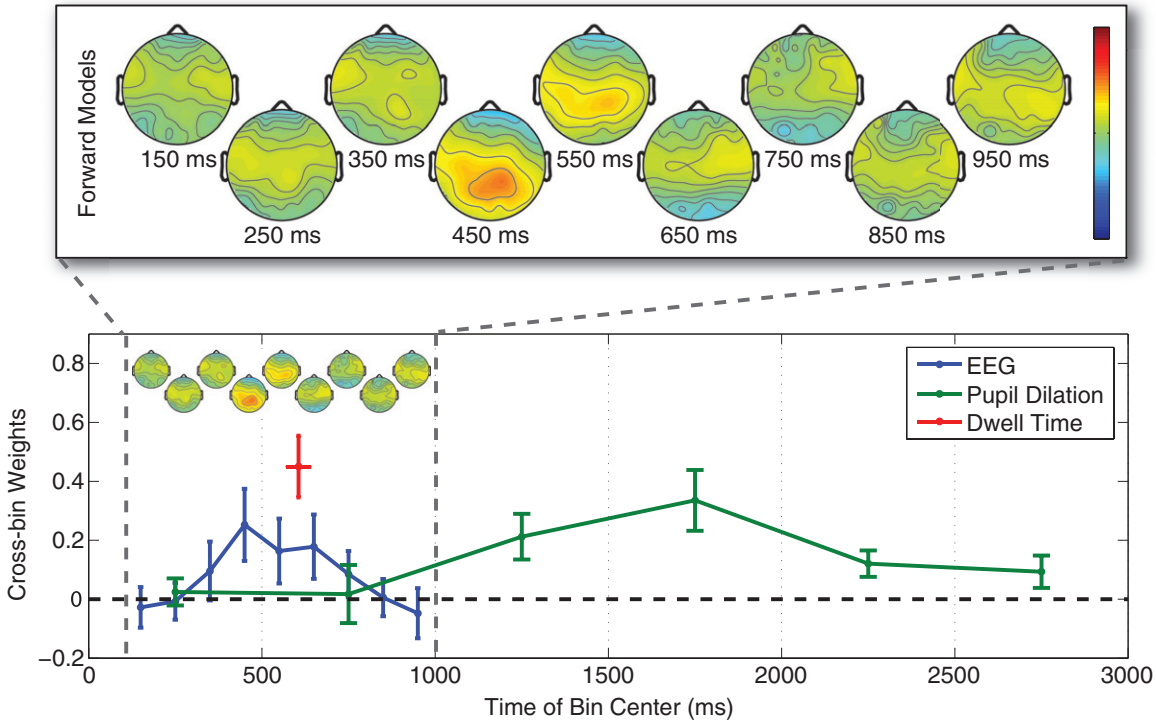

Figure 4. Forward models and weights produced by the hybrid classifier. Top: the forward models calculated using the within-bin weights from each EEG bin appear above the time of the center of that bin (all times are relative to the onset of the subject's first fixation on the object). The mean across all 10 × 10 nested cross-validation folds and 10 subjects is shown. Bottom: the cross-bin weights for each modality and bin as learned by the hybrid classifier (mean across folds, mean ± standard error across subjects, $N = 10$). The dwell time weight's horizontal position and error bars represent the mean ± standard error of the subjects' mean dwell times.

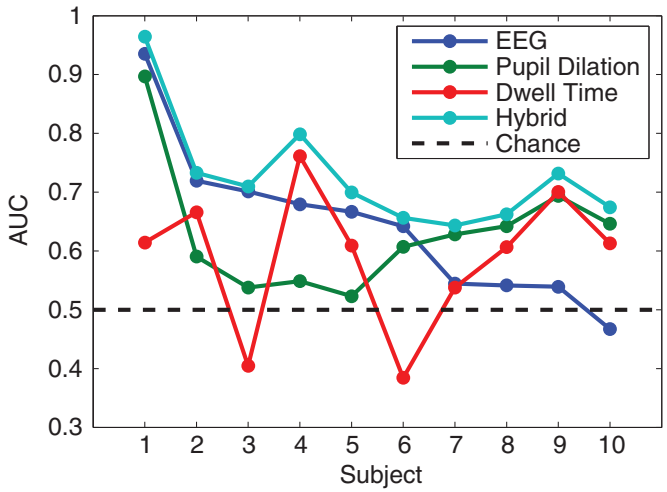

Figure 5. Performance of hybrid and single-modality classifiers. The area under the ROC curve (AUC) is used as a threshold-free figure of merit. Subjects are sorted in descending order of their EEG classifier's AUC score. For all ten subjects, the hybrid classifier performs better than any one of the single-modality classifiers.

plot of hybrid classifier performance relative to dual-modality classifiers (EEG + pupil dilation, EEG + dwell time, and pupil dilation + dwell time) is shown in supplementary figure S3. To test against the null hypothesis that single- or dual-modality classifiers produce AUC values greater than or equal to those from the multimodal classifier, while pairing the results for each subject but not assuming parametric distributions, we used a one-sided Wilcoxon signed-rank test for significance testing. This test shows that the AUC values are significantly higher for the hybrid classifier than for any of the single-modality classifiers ($p < 0.001$) or any of the dual-modality classifiers ($p < 0.05$).

3.3. CV classifier performance

The CV system has two goals: to increase the precision of the predicted target set and to increase the size of the predicted target set beyond what the subject could see in her limited exploration. In figure 6(a), we see that the first goal is clearly accomplished. The median precision of the hBCI predicted target set is 51%, while that of the CV predicted target set is 97%, a significant increase (one-sided Wilcoxon signed-rank test, $p < 0.005$). In figure 6(b), we see that the second goal is also accomplished, as the median percentage of true targets identified increases from 9.5% to 84% (a significant increase, $p < 0.005$). The hBCI predicted target set is very small because the subject only views a fraction of the environment, but the CV system has information about all objects in the environment. The CV predicted target set is both much larger and (usually) higher precision than the hBCI predicted target set, and so it stands to reason that it will identify many more of the true targets in the environment.

The CV system decreased the precision of the predicted target set for one subject (S8). This subject was instructed to look for laptops, and this category was not captured by the TAG system quite as well as the others were (see simulations in Pohlmeyer *et al* (2011)). S1 was the only other subject instructed to search for laptops. For S1, hBCI classification was good enough that the system still performed very well, but for S8, the CV system latched onto other image categories, pushing the precision of the CV predicted target set below chance.

3.4. Overall system performance

To accomplish its overall goal of an efficient target search, the system must provide a short route to visit the predicted targets. We see in figure 6(c) that by following the route produced by the system instead of visiting all objects in the environment naively, the median subject will travel 40% of the distance to reach 84% of the targets, more than twice as many targets per unit distance traveled. This represents a significant improvement in search efficiency (one-sided Wilcoxon signed-rank test on targets seen per unit distance traveled, $p < 0.005$). A sample of the hBCI, CV and TSP outputs for a single subject, plotted in environment space, are shown in figure 7. A comparison of overall system performance using single-modality hBCI classifiers and the full hybrid classifier is shown in supplementary figure S4.

4. Discussion

4.1. Advantages of hybrid classification

The addition of an eye tracker to a BCI system may induce material and calibration costs, but the results of this study indicate significant benefits as well. We used the output of the eye tracker to remove EOG artifacts without a separate training paradigm, time-lock epochs reliably, and, most importantly, enhance classification. Our results demonstrate that eye position, pupil size and EEG can each provide independent information to allow a hybrid classifier to produce more accurate output. Whether this is because they originate from distinct neural processes—or because they are co-varying measures of the same internal arousal signal combined with independent sources of noise—is a matter of some debate (Linden 2005, Murphy *et al* 2011). For the purposes of a user-centric system for able-bodied people, we only assert that if all three signals can be measured, it is advantageous to include them in the classifier.

The successful use of graph-based CV in the system speaks to an important consideration in BCIs for healthy users: most BCIs, including our hybrid classifier, generate a quantity of false positives that healthy users are unlikely to tolerate (Zander *et al* 2010). The label self-tuning step represents a way to reduce the cost of these false positives by removing them before they influence the output provided to the user. In future iterations of the system, the CV graph could delay its final extrapolation step until it can verify the consistency of the hBCI predicted target set, as in Pohlmeyer *et al* (2011).

4.2. Modularity of system

The system demonstrated in this paper is just one manifestation of the modular framework described in figure 1. Additional features could be incorporated, such as heart rate or galvanic

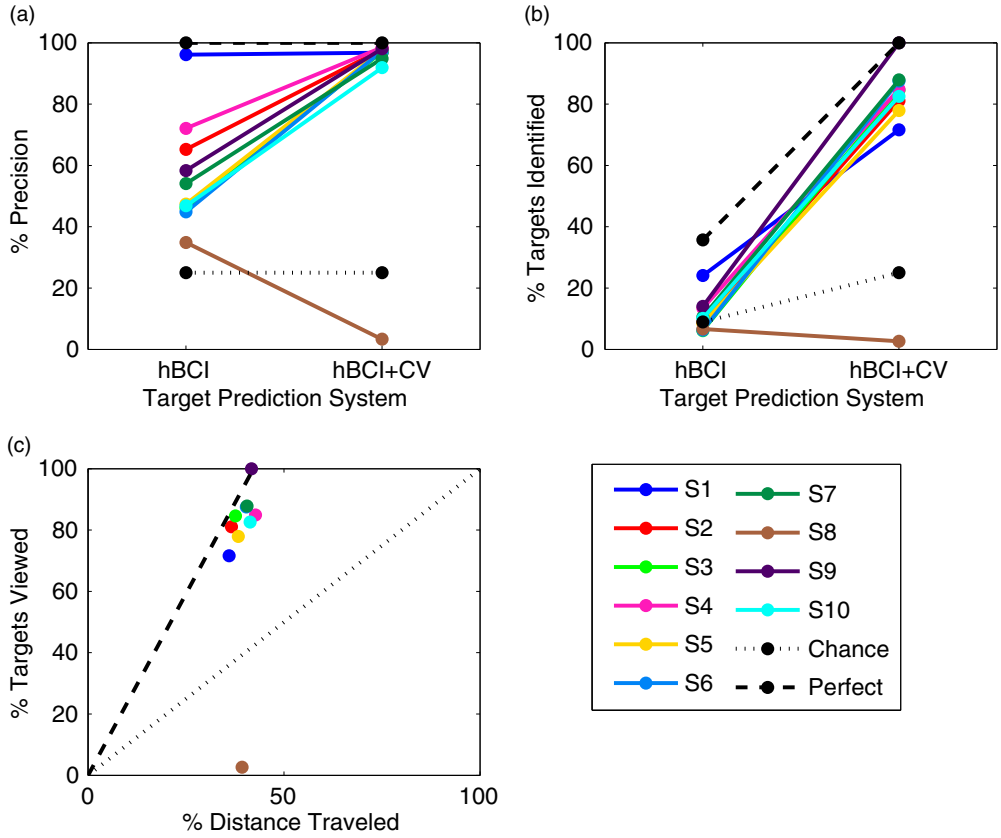

Figure 6. System performance metrics for all subjects ('S1' = subject 1, subject numbers match those in figure 5). (a) The precision of the hBCI predicted target set is already above chance (black dotted line), but that of the CV predicted target set is greatly increased for every subject except S8 (see text for possible explanation). (b) The subject viewed a small fraction of the environment during the exploration phase, so the hBCI classifier only identified a small percentage of all the targets in the environment. The CV graph included many objects that the subject had not seen, and so the CV predicted target set included a much higher percentage of the true targets. Note that S1 and S8 were asked to look for laptops, which were not as well captured by the CV graph as the other categories were. (c) The distance traveled (as a percentage of that needed to see all the objects) is plotted against the number of targets seen (as a percentage of all the targets in the environment). By following the system's route, all subjects except S8 would view significantly more targets per unit distance traveled (slope of line from origin to dot) than if they had explored the environment naively (black dotted line), and much closer to if they had explored with perfect efficiency (traveling salesman route between true targets, black dashed line). These system performance metrics compare favorably with those using single-modality hBCI classifiers, as seen in supplementary figure S4.

skin conductance. If outliers are anticipated, Fisher LDA could be replaced with a more robust regression method. If complex relationships between the features are uncovered, any number of classifier combination rules could be used for cross-bin classification (Duin 2002, Kittler *et al* 1998). The CV feature set used here is one of many ways to enforce consistency in the hBCI predicted target set: facial recognition software, object metadata, audio and video, and direct user input could also be used as features in the graph-based model of the environment. The TSP solver could be replaced with a suggestion of a single 'best match' based on a combination of classifier certainty and physical proximity. These choices should change based on the state of current knowledge about the relevant signals and the specific needs of the user.

4.3. Relation to recent studies

Our intention to develop an hBCI system for able-bodied people drove our choice of applications to address. Pfurtscheller *et al* (2010) proposed a similar hBCI whose classification of explicit motor imagery could be augmented with associated heart rate changes (although to our knowledge, the device has not been implemented). As in our study, multiple signals were generated by a single action in a virtual environment, but unlike our study, the goal was navigation for a tetraplegic user, and the elevated heart rate highlights the difficulty of producing the control signals. We chose to help healthy users search their environment because it achieves a common goal using signals easy for the user to produce.

Our focus on motion and exploration also drove our study's association of graph nodes with locations. Unlike earlier studies combining BCI and CV to speed image search (Pohlmeyer *et al* 2011, Wang *et al* 2009b), nodes in our CV graph correspond with points in the physical (or virtual) space that the user is exploring. The use of environmental awareness to narrow the continuum of navigational destinations to this discrete set of waypoints is akin to the assistive robot controller described by Perrin *et al* (2010). But by selecting from a large set of waypoints concurrently rather than making a binary selection at each intersection, we are able to address the needs of an able-bodied user who navigates with ease but lacks our

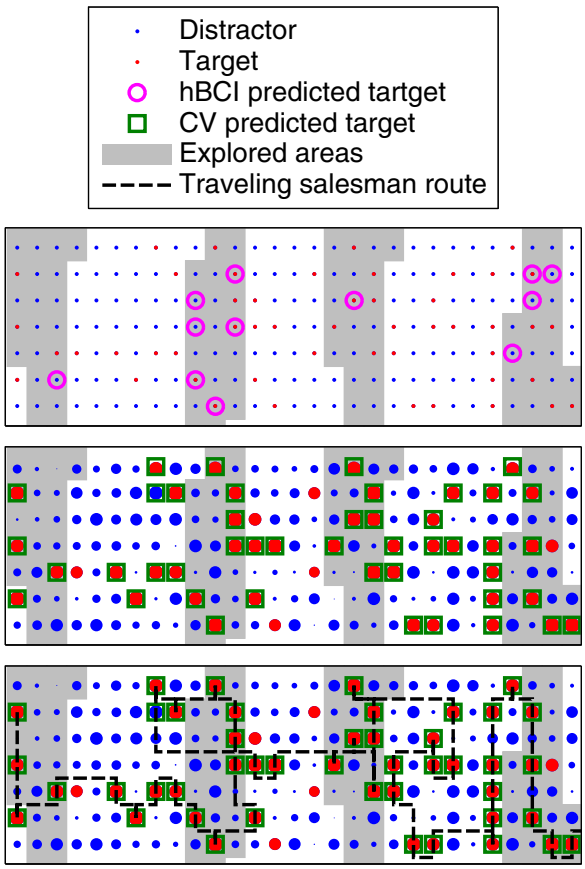

Figure 7. Sample system outputs. These birds-eye views of about 1/4 of subject 7's environment superimpose the predictions and outputs of the system onto the locations of the objects (represented by red/blue dots). Top: hBCI classification. After the subject explored the areas shaded in gray, the hBCI classifier was able to label some of those objects as hBCI predicted targets (magenta circles). Middle: CV extrapolation. The TAG system tunes the hBCI predicted target set and extrapolates it through the graph to give each object a CV score (red/blue dot size $\propto$ CV score). The objects with the highest CV scores are labeled as CV predicted targets (green squares). Note that this includes objects in unexplored areas as well as explored areas. Bottom: route planning. Finally, the TSP solver generates an efficient route that the user can traverse to visit all the CV predicted targets in the environment (black dashed line). Note that these views are zoomed in for visual clarity, and the traveling salesman route for the whole environment is continuous. A movie of the final route traversal for another subject is shown in supplementary movie 2 (available at stacks.iop.org/JNE/11/046003/mmedia).

CV system's ability to efficiently choose targets from a large database.

The desire to build a classifier using multiple naturally evoked signals locked to one user-initiated event led to this study's novel stimulus presentation paradigm. In contrast to other studies of FRPs in visual search (Brouwer *et al* 2013, Healy and Smeaton 2011, Kamienkowski *et al* 2012), we used natural images as stimuli, placed them in a dynamically explored 3D environment, allowed multiple fixations on each object, and did not eliminate peripheral vision. Unlike other target response studies in virtual reality (Bayliss and Ballard 2000), our target response signals were locked to user-generated fixations rather than experimenter-defined stimulus onsets, and the user did not need to respond physically. These choices facilitated a natural exploration of an environment that might elicit signals similar to those we could expect in the real world. This allowed us to use dwell time as a naturally evoked control signal and not an explicit one, unlike most gaze-controlled interfaces (Lee *et al* 2010) but similar to some used in reading (Rotting *et al* 2009).

The use of naturally evoked signals means that our system could be referred to as an 'opportunistic' BCI (Lance *et al* 2012), since it could provide a benefit without requiring additional effort from the user. Our system also takes a step towards the integration of such BCIs with pervasive computing technologies: since navigation is a key goal of the system, real-world implementations could interface with mobile devices like GPS trackers and head-mounted displays.

4.4. Outlook for future advances

As low-cost, mobile EEG hardware continues to advance (Lin *et al* 2009, Liao *et al* 2012) and artifact rejection becomes more and more sophisticated (Gwin *et al* 2010, Lau *et al* 2012, Lawhern *et al* 2012), mobile BCIs for the able-bodied user are becoming technically feasible. We believe that our system presents an effective way to address many of the barriers to BCI for healthy users, including training time, attentional costs, accuracy, reliability and usability (Allison 2010). These barriers have come into focus during recent discussions of 'passive' BCIs (Cutrell and Tan 2008, Rotting *et al* 2009, Zander *et al* 2010), which encourage the use of naturally evoked signals as part of a more user-centric design process. Our system combines the conscious control ability of a reactive BCI with the complementarity, composability and cost-control of a passive BCI (see Zander *et al* (2010) for a discussion of these terms).

Still, obstacles remain. In real-world scenarios, the environment is much more stimulus-rich, and subjects would sometimes explore multiple objects within the time span of the classifier presented here. The stimuli in this task were viewed far enough apart in time that target responses would not be expected to overlap. Other studies have excluded short fixations to eliminate such overlap (Brouwer *et al* 2013, Kamienkowski *et al* 2012). Although Dandekar *et al* (2012b) showed that target responses are present in overlapping FRP signals if they can be teased apart, extracting these signals from individual FRPs when the target/distractor classes are unknown remains a challenge for future research. At the system level, the use of a virtual environment allowed us to bypass steps that could require significant development in a real-world application, including building a database of object features and locations; fusing object location, subject location, head and eye tracking; and the identification of CV features that are reliable across a wide array of objects and outdoor scenes.

5. Conclusion

In this study, we demonstrated a complete system that helps users efficiently search for objects of interest in a large 3D

environment, while requiring very little conscious effort from the user. To do this, we incorporated a neural signal and two ocular signals that are all produced by the same act of fixating on an object of interest. We demonstrated that each of these signals contributes to improved classification across subjects. To increase the precision and scope of the predicted target set, we employed a graph-based computer vision model of the environment to reject false positives and extrapolate hBCI results. We then plotted an efficient search route in the 3D environment, providing an output useful to our anticipated user base of able-bodied individuals. We have applied lessons from machine learning, passive BCI, computer vision, ergonomics and reading research to address this multidisciplinary problem, and we believe that multidisciplinary efforts will continue to bring an effective real-world mobile BCI application closer to reality.

Acknowledgments

The authors would like to thank Ansh Johri and Meron Gribetz for their invaluable work refining the virtual environments and developing the route traversal scripts. Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0022. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory of the US Government. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

Allanson J and Fairclough S H 2004 A research agenda for physiological computing *Interact. Comput.* **16** 857–78

Allison B Z 2010 *Brain–Computer Interfaces: The Frontiers Collection* ed B Graimann, G Pfurtscheller and B Allison (Heidelberg: Springer) pp 357–87 (Towards Ubiquitous BCIs)

Bayliss J D and Ballard D H 2000 A virtual reality testbed for brain–computer interface research *IEEE Trans. Rehabil. Eng.* **8** 188–90

Bell A J and Sejnowski T J 1995 An information-maximization approach to blind separation and blind deconvolution *Neural Comput.* **7** 1129–59

Brouwer A M, Reuderink B, Vincent J, van Gerven M A and van Erp J B 2013 Distinguishing between target and nontarget fixations in a visual search task using fixation-related potentials *J. Vis.* **13**

Croes G 1958 A method for solving traveling-salesman problems *Oper. Res.* **6** 791–812

Cutrell E and Tan D 2008 BCI for passive input in HCI *Proc. ACM CHI 2008, Workshop on Brain–Computer Interfaces for HCI and Games* **8** 1–3

Dandekar S, Ding J, Privitera C, Carney T and Klein S A 2012a The fixation and saccade p3 *PLoS One* **7** e48761

Dandekar S, Privitera C, Carney T and Klein S A 2012b Neural saccadic response estimation during natural viewing *J. Neurophysiol.* **107** 1776–90

Delorme A and Makeig S 2004 EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis *J. Neurosci. Methods* **134** 9–21

Dimigen O, Sommer W, Hohlfeld A, Jacobs A M and Kliegl R 2011 Coregistration of eye movements and EEG in natural reading: analyses and review *J. Exp. Psychol. Gen.* **140** 552–72

Duin R P 2002 The combining classifier: to train or not to train? *Proc. 16th Int. Conf. on Pattern Recognition* vol 2 (Quebec: IEEE) pp 765–70

Fei-Fei L, Fergus R and Perona P 2007 Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories *Comput. Vis. Image Underst.* **106** 59–70

Galan F, Nuttin M, Lew E, Ferrez P W, Vanacker G, Philips J and Millan J R 2008 A brain-actuated wheelchair: asynchronous and non-invasive brain–computer interfaces for continuous control of robots *Clin. Neurophysiol.* **119** 2159–69

Gerson A D, Parra L C and Sajda P 2006 Cortically coupled computer vision for rapid image search *IEEE Trans. Neural Syst. Rehabil. Eng.* **14** 174–9

Grimes D, Tan D S, Hudson S E, Shenoy P and Rao R P 2008 Feasibility and pragmatics of classifying working memory load with an electroencephalograph *Proc. SIGCHI Conf. on Human Factors in Computing Systems ACM (Florence, Italy)* pp 835–44

Gwin J T, Gramann K, Makeig S and Ferris D P 2010 Removal of movement artifact from high-density EEG recorded during walking and running *J. Neurophysiol.* **103** 3526–34

Healy G and Smeaton A F 2011 Eye fixation related potentials in a target search task *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2011** 4203–6

Hess E H and Polt J M 1960 Pupil size as related to interest value of visual stimuli *Science* **132** 349–50

Hess E H and Polt J M 1964 Pupil size in relation to mental activity during simple problem-solving *Science* **143** 1190–2

Hochberg L R, Serruya M D, Friehs G M, Mukand J A, Saleh M, Caplan A H, Branner A, Chen D, Penn R D and Donoghue J P 2006 Neuronal ensemble control of prosthetic devices by a human with tetraplegia *Nature* **442** 164–71

Hutzler F, Braun M, Vo M L, Engl V, Hofmann M, Dambacher M, Leder H and Jacobs A M 2007 Welcome to the real world: validating fixation-related brain potentials for ecologically valid settings *Brain Res.* **1172** 124–9

Jacob R J K 1991 The use of eye movements in human–computer interaction techniques: what you look at is what you get *ACM Trans. Inf. Syst.* **9** 152–69

Jainta S, Vernet M, Yang Q and Kapoula Z 2011 The pupil reflects motor preparation for saccades—even before the eye starts to move *Front. Human Neurosci.* **5** doi: 10.3389/fnhum.2011.00097

Jangraw D C and Sajda P 2011 A 3-d immersive environment for characterizing EEG signatures of target detection *5th Int. IEEE/EMBS Conf. on Neural Engineering (NER)* pp 229–32

Kahneman D and Beatty J 1966 Pupil diameter and load on memory *Science* **154** 1583–5

Kamienkowski J E, Ison M J, Quiroga R Q and Sigman M 2012 Fixation-related potentials in visual search: a combined EEG and eye tracking study *J. Vis.* **12** 4

Kittler J, Hatef M, Duin R P and Matas J 1998 On combining classifiers *IEEE Trans. Pattern Anal. Mach. Intell.* **20** 226–39

Lance B J, Kerick S E, Ries A J, Oie K S and McDowell K 2012 Brain–computer interface technologies in the coming decades *Proc. IEEE* **100** 1585–99

Lau T M, Gwin J T, McDowell K G and Ferris D P 2012 Weighted phase lag index stability as an artifact resistant measure to detect cognitive EEG activity during locomotion *J. Neuroeng. Rehabil.* **9** 47

Lawhern V, Hairston W D, McDowell K, Westerfield M and Robbins K 2012 Detection and classification of subject-generated artifacts in EEG signals using autoregressive models *J. Neurosci. Methods* **208** 181–9

Lee E C, Woo J C, Kim J H, Whang M and Park K R 2010 A brain–computer interface method combined with eye tracking for 3d interaction *J. Neurosci. Methods* **190** 289–98

Leeb R, Friedman D, Muller-Putz G R, Scherer R, Slater M and Pfurtscheller G 2007 Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: a case study with a tetraplegic *Comput. Intell. Neurosci.* **8** 79642

Liao L D, Lin C T, McDowell K, Wickenden A E, Gramann K, Jung T P, Ko L W and Chang J Y 2012 Biosensor technologies for augmented brain–computer interfaces in the next decades *Proc. IEEE* **100** 1553–66

Lin C T, Ko L W, Chang M H, Duann J R, Chen J Y, Su T P and Jung T P 2009 Review of wireless and wearable electroencephalogram systems and brain–computer interfaces—a mini-review *Gerontology* **56** 112–9

Linden D E 2005 The p300: where in the brain is it produced and what does it tell us? *Neuroscientist* **11** 563–76

Luo A and Sajda P 2009 Do we see before we look? *Proc. 4th Int. IEEE/EMBS Conf. on Neural Engineering (NER)* pp 230–3

Makeig S, Bell A J, Jung T P and Sejnowski T J 1996 Independent component analysis of electroencephalographic data *Adv. Neural Inform. Process. Syst.* vol 8 pp 145–51

Murphy P R, Robertson I H, Balsters J H and O'Connell R G 2011 Pupillometry and p3 index the locus coeruleus-noradrenergic arousal function in humans *Psychophysiology* **48** 1532–43

Oliva A and Torralba A 2001 Modeling the shape of the scene: a holistic representation of the spatial envelope *Int. J. Comput. Vision* **42** 145–75

Parra L C, Spence C D, Gerson A D and Sajda P 2005 Recipes for the linear analysis of EEG *Neuroimage* **28** 326–41

Perrin X, Chavarriaga R, Colas F, Siegwart R and Millan J R 2010 Brain-coupled interaction for semi-autonomous navigation of an assistive robot *Robot. Auton. Syst.* **58** 1246–55

Pfurtscheller G, Allison B Z, Brunner C, Bauernfeind G, Solis-Escalante T, Scherer R, Zander T O, Mueller-Putz G, Neuper C and Birbaumer N 2010 The hybrid BCI *Front. Neurosci.* **4** 30

Pohlmeyer E A, Wang J, Jangraw D C, Lou B, Chang S F and Sajda P 2011 Closing the loop in cortically-coupled computer vision: a brain–computer interface for searching image databases *J. Neural Eng.* **8** 036025

Polich J 2007 Updating p300: an integrative theory of p3a and p3b *Clin. Neurophysiol.* **118** 2128–48

Porter G, Troscianko T and Gilchrist I D 2007 Effort during visual search and counting: Insights from pupillometry *Q. J. Exp. Psychol.* **60** 211–29

Rotting M, Zander T, Trosterer S and Dzaack J 2009 *Industrial Engineering and Ergonomics* (Heidelberg: Springer) pp 523–36 (Implicit Interaction in Multimodal Human–Machine Systems)

Sajda P, Pohlmeyer E, Wang J, Parra L C, Christoforou C, Dmochowski J, Hanna B, Bahlmann C, Singh M K and Chang S F 2010 In a blink of an eye and a switch of a transistor: cortically coupled computer vision *Proc. IEEE* **98** 462–78

Sellers E W and Donchin E 2006 A p300-based brain–computer interface: initial tests by ALS patients *Clin. Neurophysiol.* **117** 538–48

Wang J, Jebara T and Chang S F 2008 Graph transduction via alternating minimization *Proc. 25th Int. conf. on Machine Learning* pp 1144–51

Wang J, Jiang Y G and Chang S F 2009a Label diagnosis through self tuning for web image search *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* pp 1390–7

Wang J, Pohlmeyer E, Hanna B, Jiang Y G, Sajda P and Chang S F 2009b Brain state decoding for rapid image retrieval *Proc. 17th ACM Int. Conf. on Multimedia ACM (Beijing, China)* pp 945–54

Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain–computer interfaces for communication and control *Clin. Neurophysiol.* **113** 767–91

Zander T, Kothe C, Jatzev S and Gaertner M 2010 *Brain–Computer Interfaces (Human–Computer Interaction Series)* ed D S Tan and A Nijholt (London: Springer) pp 181–99 (Enhancing Human–Computer Interaction with Input from Active and Passive Brain–Computer Interfaces)