

# Closing the loop in cortically-coupled computer vision: a brain–computer interface for searching image databases

Eric A Pohlmeier<sup>1</sup>, Jun Wang<sup>2</sup>, David C Jangraw<sup>1</sup>, Bin Lou<sup>1</sup>,  
Shih-Fu Chang<sup>2</sup> and Paul Sajda<sup>1</sup>

<sup>1</sup> Department of Biomedical Engineering, Columbia University, New York, NY 10027, USA

<sup>2</sup> Department of Electrical Engineering, Columbia University, New York, NY 10027, USA

E-mail: [psajda@columbia.edu](mailto:psajda@columbia.edu)

Received 5 November 2010

Accepted for publication 22 March 2011

Published 12 May 2011

Online at [stacks.iop.org/JNE/8/036025](http://stacks.iop.org/JNE/8/036025)

## Abstract

We describe a closed-loop brain–computer interface that re-ranks an image database by iterating between user generated ‘interest’ scores and computer vision generated visual similarity measures. The interest scores are based on decoding the electroencephalographic (EEG) correlates of target detection, attentional shifts and self-monitoring processes, which result from the user paying attention to target images interspersed in rapid serial visual presentation (RSVP) sequences. The highest scored images are passed to a semi-supervised computer vision system that reorganizes the image database accordingly, using a graph-based representation that captures visual similarity between images. The system can either query the user for more information, by adaptively resampling the database to create additional RSVP sequences, or it can converge to a ‘done’ state. The done state includes a final ranking of the image database and also a ‘guess’ of the user’s chosen category of interest. We find that the closed-loop system’s re-rankings can substantially expedite database searches for target image categories chosen by the subjects. Furthermore, better reorganizations are achieved than by relying on EEG interest rankings alone, or if the system were simply run in an open loop format without adaptive resampling.

 Online supplementary data available from [stacks.iop.org/JNE/8/036025/mmedia](http://stacks.iop.org/JNE/8/036025/mmedia)

## 1. Introduction

The ever growing capabilities of both computer processing and storage have led to an abundance of data which can easily lead to information overload, as our speed in accessing data far outweighs our ability to process it. This makes identification of only the most useful information a significant problem. Examples range from searching product databases of tens of thousands of entries to using services such as Flickr and Google images to find particular images amongst the billions accessible. These search scenarios become even more challenging if the database metadata is not relevant to the search (or is nonexistent), or if the search can only be vaguely defined (e.g. looking for ‘interesting’ entries).

Computer vision (CV) has had a major focus on automating the identification, annotation, recognition or search of imagery. However, while CV has shown notable successes, these typically involve highly constrained situations in which the search can be well defined, and the usefulness of CV systems for general purpose or subjective search is limited. These limits are particularly notable in comparison to human vision (HV), which remains unmatched in regards to robust and general purpose object recognition. HV can parse a scene and easily recognize objects despite wide variations in scale, lighting, pose, background clutter, etc, and can easily direct its attention to look for images that are ‘interesting’ to a particular individual. Furthermore, HV can do so in as little as a few hundred milliseconds [1], easily getting the ‘gist’ of a scene

with only brief glances [2]. However, despite these abilities to extract general information from a single image quickly, CV is still superior to HV for factors such as processing speed, data throughput and lack of fatigue, being able to process large batches of imagery much more quickly.

Cortically-coupled computer vision (C3Vision) refers to a particular class of brain–computer interface (BCI) meant to combine the complementary strengths of CV and HV to provide robust image search and retrieval in high throughput tasks [3–5]. C3Vision is largely inspired by the notable successes of BCIs that have shown that noninvasive EEG recordings can be used to control systems as varied as communication systems, computer cursors, mobile devices (e.g. wheelchairs), and artificial muscle stimulation [6–18]. In particular, C3vision is related to those BCIs that rely on some combination of a target detection and the attentional orienting response. These natural phenomena result from a person being attentive to certain, relatively rare, stimuli, and produce measurable scalp EEG signals, in particular event-related potentials (ERPs) such as the P300 (also termed the P3b) and the P3a [19–21]. Numerous BCIs, such as P300 spellers, utilize the ‘oddball’ response, typically by relying on a user’s neural response to rare visual stimuli, as a method of controlling communication systems [9, 22–25]. Notably though, these BCI systems tend to have low information transfer rates (unsurprisingly given the low SNR characteristics of EEG recordings), and, as such, have been meant primarily as assistive devices for handicapped users. Like these systems, C3Vision focuses on users’ ERP responses when they observe images for which they currently have an interest. C3Vision diverges from many BCI systems in that it is part of a growing effort to create BCIs meant to assist able-bodied users in rapidly finding objects of interest within either large images or image databases [4, 5, 26–28]. By combining CV and HV, C3Vision can help users locate objects of interest contained within rapidly presented streams of images (often as fast as 10 images s<sup>-1</sup>) [3–5].

In this paper, we describe and evaluate a closed-loop variation of C3Vision. In this configuration, the human and computer vision components of the system interact (unlike our previous open-loop systems which operated in a serial fashion, with either CV preceding HV or vice versa [3–5, 28]). This interaction offers significant potential for improved performance, as other BCI studies have found that closed-loop systems (where feedback to the user facilitates learning and adaptation) have better overall performance than their open-loop counterparts [29–34]. In the case of C3Vision, a closed-loop system architecture allows the CV system to query the user for more information as necessary to help ensure that it can identify target images, while reducing the number of images that must be directly inspected by the user. Specifically, in our tests we explore how effectively a closed-loop C3Vision (CL-C3Vision) system can be used to find target images in a moderately-sized image database whose images fall into well-defined categories. Initially a fraction of the database images are rapidly presented to the user, and are tagged with ‘interest scores’ by decoding the user’s EEG. The images with the highest interest scores are flagged and sent to

a CV module. The CV decides whether it can infer a set of consistent properties between the flagged images, and whether it can use that information to effectively search the full image database. If not, the CV acquires additional information by resampling the database to select another small set of images (using the currently identified images of interest) to present to the user. In this way, the CV can aggregate information over a growing subset of the database’s full contents to refine its target knowledge from the user’s EEG data, rather than relying exclusively on improving SNR through repeatedly showing the same images. This closed-loop interaction continues with the CV adaptively resampling the database to present additional images to the user until the CV is ready to perform a final scoring. This last scoring includes all the database images, including those unseen by the user (which can conceivably extend to millions of images). The success of the final scoring is quantified by how well it can be used to reorganize the database to expedite a search for the target images. As a final test, the CV also attempts to explicitly infer the user’s intent by identifying the specific image category which had been chosen as the search target (the target either having been assigned by the experimenter or freely chosen by the subject).

## 2. Methods

### 2.1. System overview

The closed-loop C3Vision (CL-C3Vision) system we have developed is aimed at assisting a user in searching through a large database of imagery and locating images of ‘interest’, i.e. target imagery. The architecture involves three primary components: an image database that the user wishes to search, an EEG interest detector, and a CV system. Figure 1 illustrates the connections between these components, and algorithm 1 summarizes their closed-loop interaction. The search is started by a random sample of the database imagery being presented to the user. An EEG interest detector (created previously during a brief training phase) ranks the user’s interest in the images. The highest ranked images are flagged and sent to the CV module. The CV module uses these flagged images to accomplish three linked purposes. First, it assesses the examples to determine whether they are highly similar to each other (‘Connectivity Test’ in figure 1). If so, the CV reorganizes the full database (including all the imagery unseen by the user) to expedite the user’s search for other images of interest (in a cyclical process termed ‘Self-Tuning’), and ‘guesses’ the target category of the user’s search. Alternatively, the CV can request more information by selecting an additional set of images for the user to review (‘No Self-Tuning’ in figure 1, see section 2.5 below for more details regarding the CV and self-tuning). Specifically, the CV selects a small subset of images from the database that are most likely to be target related. These images are then reviewed by the user, creating a closed-loop interaction that continues until the CV is ready to perform the final reorganization of the full database.

### 2.2. Image database

Testing images were taken from the Caltech-101 database [35]. To prevent fluctuations in image size during the RSVP

**Algorithm 1.** Processing pipeline of the closed-loop C3Vision system.**Offline processing:**

Construct TAG affinity graph for all images in database

Estimate threshold value for TAG connectivity

Train subject-specific EEG interest detector

Estimate threshold value for indicating significant interest from EEG score

Initialize RSVP sequence with a random sample of images from database

**Online looping****repeat**

Display RSVP sequence and compute EEG score to estimate subject's interest

Flag interesting images using EEG score and pre-computed subject interest threshold

Create training image set for TAG using flagged images from current and previous loops

Use flagged images to assign pseudo-labels to TAG graph nodes

Compute average connection strength between pseudo-labeled nodes

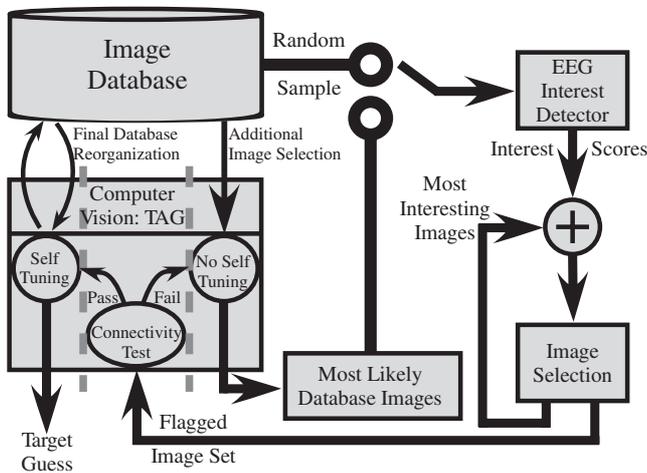
Propagate pseudo-labels and assign ranking scores to every node

Update RSVP sequence: select images based on TAG ranking results

**until** average connection strength of TAG input images exceeds TAG connectivity threshold**System output:**

Initiate TAG self-tuning

Compute final ranking scores for images in database



**Figure 1.** Closed-loop C3Vision architecture. A sample of images from the database is randomly selected and presented to the user in a RSVP. Interest scores are assigned to each RSVP image based on the user's EEG response. These scores are subsequently used to flag the most 'interesting' images. A computer vision module, based on a transductive graph-based model (TAG), is used to measure how similar these flagged images are via their connectivity in the TAG graph. If sufficiently connected, the TAG is used to fully re-sort the image database and 'guess' the target category. Otherwise, the TAG selects another set of RSVP images to be shown to the user to acquire additional search target information, see algorithm 1.

from influencing subjects' visual responses, only a subset of the Caltech-101 images were used. Specifically, 62 image categories (a category being Caltech-101 images of a given type, e.g. 'elephants', 'crayfish') were selected because their images could be rescaled to uniform dimensions during visual presentations with negligible distortion. This provided a total of 3798 images (42% of the Caltech-101 images), with 31–128 images/category (mean = 61, STD = 22, supplementary table S1 available at [stacks.iop.org/JNE/8/036025/mmedia](http://stacks.iop.org/JNE/8/036025/mmedia) lists all the categories).

### 2.3. Rapid serial visual presentation

Images were presented to the user using a rapid serial visual presentation (RSVP) paradigm [36]. Specifically, images were

shown in blocks of 100, with images within each block being presented at 5 Hz. Each image was centered on the computer monitor. A fixation cross appeared just prior to each block to allow the users to center their gaze on the images during the RSVP sequences. For these tests, the initial RSVP sequence consisted of 500 images (five blocks) drawn randomly from the database. Whenever the CV system requested more information, a two-block (200 image) RSVP sequence of CV-selected images would be presented (see below). Users were allowed a self-paced rest period (typically a few seconds) between each RSVP block.

### 2.4. EEG interest detector

The users' EEG responses were used to quantify their interest in each image shown during the RSVP. EEG data were recorded using an ActiveTwo Biosemi (Amsterdam, The Netherlands) system with 64 electrodes positioned according to the 10–20 international system. The EEG data were acquired at 2048 Hz, with 60 Hz notch and 0.5 Hz high-pass filters. The method used to calculate the EEG interest scores has been previously described [3, 5]. Briefly, it is based on a linear model,

$$y_n = \sum_i w_i x_{in} \quad (1)$$

where  $x_{in}$  is EEG activity at the data sample  $n$  measured by the electrode  $i$ , and  $w$  is a set of spatial (i.e. over electrode) weights. Weights are learned from a set of training data such that  $y$  maximally discriminates between target and non-target images. The spatial distribution of this activity is assumed to change over time with a temporal resolution ( $T$ ) of 100 ms. Thus, weight vectors,  $w_{ki}$ , are found for several 100 ms windows following each image presentation ( $k$  is the time window index and  $F_s$  is the sampling rate):

$$y_k = \left(\frac{1}{N}\right) \sum_n \sum_i w_{ki} x_{i[(k-1)N+n]}, \quad n = 1, 2, \dots, N, \quad N = T/F_s. \quad (2)$$

The resulting values for each of the separate time windows ( $y_k$ ) are then combined in a weighted average to provide a final interest score ( $y_{IS}$ ) for each image:

$$y_{IS} = \sum_k v_k y_k. \quad (3)$$

The EEG detector used the 1200 ms of data immediately following each image, although the first 100 ms were ignored (i.e. only time windows  $k = 2, 3, \dots, 12$  were used). Fisher linear discriminant (FLD) analysis was used to calculate the spatial coefficients,  $w_{ki}$ , and a logistic regression was then used to create the temporal coefficients,  $v_k$  [4].

To analyze the spatial topology of the learned EEG components contributing to the interest score, we computed a forward model for each spatial component. The forward model for the component  $k$  is given by

$$\mathbf{a}_k = \frac{\mathbf{X}_k \mathbf{y}_k}{\mathbf{y}_k^T \mathbf{y}_k} \quad (4)$$

where  $\mathbf{a}_k$  is a vector having a length equal to the number of EEG channels (in this case 64),  $\mathbf{X}_k$  is the EEG data for the temporal window on which the component was estimated, having dimensions channels by trials, and  $\mathbf{y}_k$  is the vector of the  $k$ th component response for all trials. The forward model can be seen as the normalized correlation between the component and the measured EEG, and thus can be interpreted as a first-order approximation of what that component would look like if measured at the scalp.

During the real-time tests of the system, the EEG interest scores (equation (3)) assigned to each RSVP image were used to identify a set of images that were of significant interest to the user. Specifically, these ‘flagged’ images had interest scores that were two or more standard deviations (STDs) above the mean interest scores for the subject’s training data (see section 2.6 below). If fewer than five images passed this threshold, the threshold was relaxed to 1.64 or 1.0 deviations above the mean, as necessary, until five or more images from the RSVP were flagged. Additionally, only images that passed the two-STD threshold were added to a list of ‘most interesting’ images. In a search in which the CV requested additional closed-loop RSVPs, these ‘most interesting’ images were added to the list of images and scores obtained from the most recent RSVP (if an image was listed multiple times, its scores were averaged). This was done prior to the image scores being compared to the interest threshold and the interesting images being flagged for transmission to the CV (represented by the feedback prior to the ‘Image Selection’ in figure 1).

## 2.5. Computer vision: TAG

The computer vision system used is a method termed transductive annotation by graph (TAG), a semi-supervised learning technique which uses a small portion of labeled images from a database as example images for pattern discovery, and which then propagates the results through a weighted and undirected graph to score all the images in the database [37]. In our experiments, these examples were the images flagged by the EEG interest detector. Details of the TAG can be found in [38] and [28] (in the latter, its use in

an open-loop C3Vision system is described in detail). Briefly though, TAG uses an affinity graph to quantify the pairwise similarity between images in a database via their visual similarity and any underlying subspace structures determined from a predefined feature space. Thus, when given a noisy set of target images (noisy meaning that some of the example images may be false positives and thus likely to be outliers in terms of visual similarity), the TAG produces prediction scores ( $f$ ) for all images in the database using a pre-computed graph [38]. This allows novel images that are similar to the example image set to be located. As EEG can be very noisy, the presence of false positives among the flagged images is quite likely. As such, the TAG’s ability to work with noisy labels is essential, as incorrect labels will degrade performance [39]. In the current system, gist features were used together with a b-Matching method to construct the affinity graph [40]. The specific gist feature parameters were chosen in a general setting (i.e. blindly to the specific images of the Caltech database) to generate a 512-dimensional vector, as suggested in [41].

The TAG can often improve its performance in the presence of a noisy labeled input set by increasing the visual consistency among the example set prior to generating  $f$ . This is accomplished by replacing some of the example images with more likely candidates from the larger image database [28]. This process of ‘self-tuning’ uses the strength of the TAG graph connections (or ‘connectivity’) to identify outlier images in the example set. These outliers are dropped from the example set and replaced with other images from the database (which are selected via their connectivity with the remaining example images).

These abilities of the TAG make it well suited to perform the three CV functions in the CL-C3Vision architecture described previously in section 2.1. For the first task, i.e. assessing the visual consistency of the example image set, the TAG uses the connectivity value (the same value as used by the TAG when self-tuning [39]) between the example images to quantify their similarity. Specifically, the average connectivity of the example images is compared to the average connectivity found in simulations wherein ten example images were randomly selected 10 000 times. Because connectivity in the TAG graph is related to the quantity of example images, the simulation connectivity values are always normalized to match the number of inputs that were passed to the CV module. If the input set connectivity clearly exceeds the corresponding random simulations value (in this case more than 3 STD from the mean), the CV module exits the loop and does the final database reorganization (while predicting the identity of the user’s selected target category).

The difference between the second and third CV tasks lies in the use of self-tuning by the TAG. For the second task, in which the CV module requests more information, the TAG identifies additional images to be shown to the user. A basic TAG scoring without self-tuning is performed, and the 200 images with the highest scores are then shown to the user in an additional RSVP sequence. In cases where the CV module decides that the system should exit the closed loop and perform the final reorganization, the TAG uses self-tuning to refine the example image set, and then scores all the images in

the database. Given that this calculation will not be performed until the example set already passed the connectivity check for image similarity, only 25% TAG self-tuning is used, i.e. one quarter of the image set images that are furthest from the others in the TAG graph are eliminated and replaced prior to the final TAG scoring.

## 2.6. Experimental protocol

The parameters for the EEG interest detectors ( $w_{ki}$  and  $v_k$  in equations (2) and (3)) were estimated for each user using a set of training data collected at the beginning of each experimental session (training data were thus completely separate from the later testing datasets). During the training period, the subject was shown a sequence of RSVP blocks (the same format as the later testing sequences, i.e. blocks of 100 images, 5 Hz presentation rate), having been instructed to pay attention to images containing baseball gloves (subjects were shown several example baseball glove images in preparation). Each training RSVP block contained 98 distractor and two target images, and the interest detector's parameters and current performance were continually updated as each image was shown. Each subject was shown 25–35 such blocks (i.e. 2500–3500 images), depending on performance. All training images (target and distractor) were taken from a subset of imagery from the Caltech-256 database [42], so that there was no overlap between the training and testing image databases.

The CL-C3Vision's effectiveness (in both identifying what kind of image a subject was interested in and then reorganizing the image database accordingly) was then tested in a series of searches in which the EEG interest detector and computer vision modules described above interacted in real time. The target images for which the subjects searched always corresponded to one of the 62 Caltech-101 categories in the image database. For example, for the first search each subject was instructed to search for images showing 'brain'. The BCI process described above was then implemented. While the metadata containing the Caltech category of each image was *not* available to the BCI during each experimental search, it was used to evaluate the success of the final database reorganization after the BCI had concluded and exited the loop. Specifically, it was used to calculate the average precision (AP) for every category of images in the database. The category with the highest AP was presented to the user at the end of each search to see if the BCI had successfully inferred the user's intent and 'guessed' the identity of the target category. AP, a standard performance metric used in image search [43], estimates a value between 0 and 1 that approximates the area under the precision/recall curve. It goes through the ranked list of retrieval results, computes the precision value whenever a new true positive result is found, and then computes the average when the list is exhausted [44].

This entire process was then repeated for one or more subsequent searches. In these searches the subject was allowed free selection of any of the other 62 database categories as the search target. To ensure that each subject had at least two search targets in common, a final search was then performed in which each subject was instructed to look for grand pianos

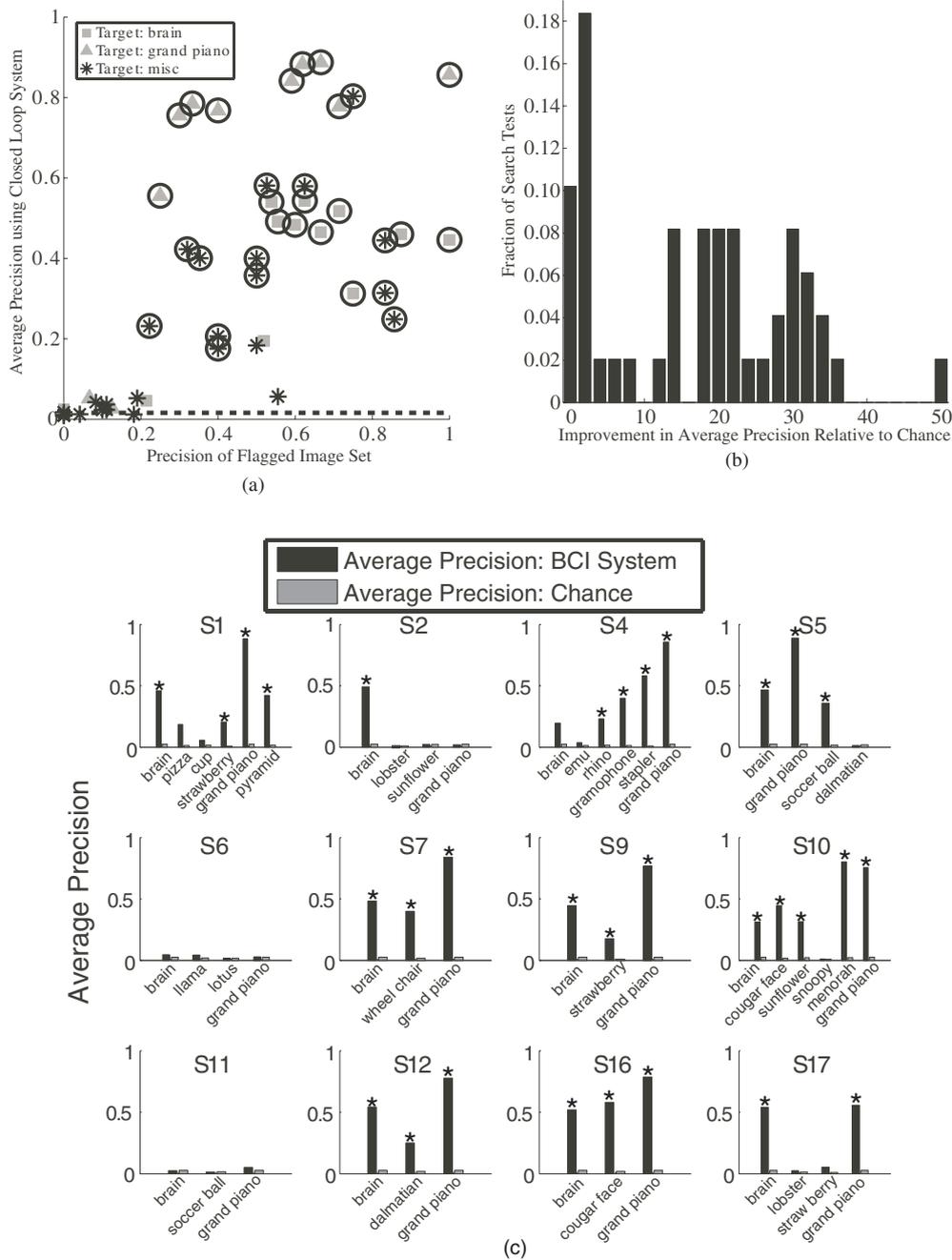
(assuming that the subject had not already chosen this category as a target on their own initiative). Prior to every search sequence, the subject was shown several random example images from the target category (15% of the images from that category). In some experiments, the BCI loop was exited and the final database reorganization performed even if the images given to the CV module did not pass the connectivity criteria. These included three instances in which the next set of images recommended by the TAG for presentation did not include any target images (typically discovered shortly after the fact when the subjects reported that they had not observed any targets in the RSVP), and two cases in which five RSVP sequences had been presented with the connectivity requirement still remaining unsatisfied.

Seventeen volunteers participated in the current study. None reported any history of neurological problems and informed consent was obtained from all participants in accordance with the guidelines and approval of the Columbia University Institutional Review Board. Five subjects were excluded because they showed a larger probability of blinking or making other eye movements (consciously or unconsciously) following the appearance of target images compared to distractors during the training session. While these movements did not appear to systematically impact the performance of the EEG interest detectors either positively or negatively (see the supplementary material, table S2 and figure S1, available at [stacks.iop.org/JNE/8/036025/mmedia](http://stacks.iop.org/JNE/8/036025/mmedia)), the subjects were excluded to emphasize the BCI aspect of the study and avoid the possibility of EOG artifacts influencing the results. Each of the remaining 12 subjects (mean age: 22, STD: 2.2; five females) searched for three to six different targets (mean: 4.1, STD: 1.2), for a total of 49 different target searches and 20 different target categories. Four of these subjects reported having previously participated in EEG studies, but also reported that these other studies were not reminiscent of the current BCI study.

## 3. Results

### 3.1. Overall system performance

The CL-C3vision BCI was very effective in increasing the AP of the target category when reorganizing the image database. Figure 2(a) shows the AP of the target category after the final database reorganization relative to the fraction of target images within the final flagged image set sent to the CV module. The dashed line at the bottom of figure 2(a) shows the average prevalence of the different image categories in the database. A category's chance AP corresponds to its prevalence; thus, the improvement of the system over chance can clearly be seen (one-sided *t*-test,  $n = 49$ ,  $p \ll 0.001$ ). In only 10% of the searches did the final AP fall below chance values, and in 65% of the searches the final AP exceeded chance by more than a factor of 10 (figure 2(b)). Furthermore, the system successfully 'guessed' the target category in 31 of the 49 searches (shown by the embossed circles in figure 2(a)), a correct guess meaning that the target category had the highest AP of any image category, making its images

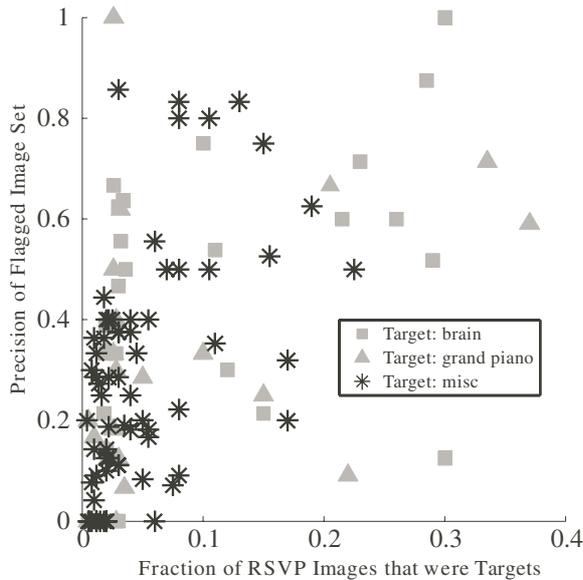


**Figure 2.** CL-C3Vision consistently reorganized the image database to emphasize the target images. (a) The final AP achieved by the system was heavily affected by the fraction of target images in the flagged image set produced by the EEG interest detector, but exceeded chance levels (the dashed line shows the average chance prevalence between all categories) in 44/49 searches, and did so more than tenfold in 32 cases. The data points embossed by a circle reflect searches in which the system correctly identified the selected target category, i.e. that category’s images were those most emphasized by the final database reorganization; (b) shows the AP improvement over chance for all the searches; (c) shows how results (final AP versus chance) varied between subjects, with three subjects (S2, S6, S11) showing markedly less effective system performance. The stars indicate searches in which the system correctly identified the target category. Success also appeared somewhat related to the specific target as well, as improvement over chance was consistently stronger when the target category was ‘grand piano’ than when it was ‘brain’ (paired one-sided  $t$ -test,  $p = 0.01$ ,  $n = 12$ ).

the most emphasized (easiest to find) in the final database reorganization.

Figure 2(c) shows how search successes varied between subjects and for different search targets. Notably, three subjects (S2, S6, S11) accounted for 10 of the 18 cases in which the system failed to correctly guess the target category.

During the tests for these three subjects, the interest detector was notably less successful in flagging target images, resulting in low-precision image sets being sent to the CV (observable in figure 2(a)). Specifically, the average target precision in the image sets given to the CV following the initial (randomly selected) RSVP sequences was 0.10 (STD = 0.16,  $n = 11$ )



**Figure 3.** Impact of the number of target images in the RSVP on the EEG interest detector performance. Shown is the precision of target images in the image sets flagged as ‘interesting’ by the interest detector (for every search RSVP between all 12 subjects) as well as the fraction of the RSVP that was composed of target images. Typically sets of images of higher precision were flagged when the RSVP also had a higher prevalence of target images, allowing the CV to aid the interest detector by increasing the number of target images in successive RSVP sequences.

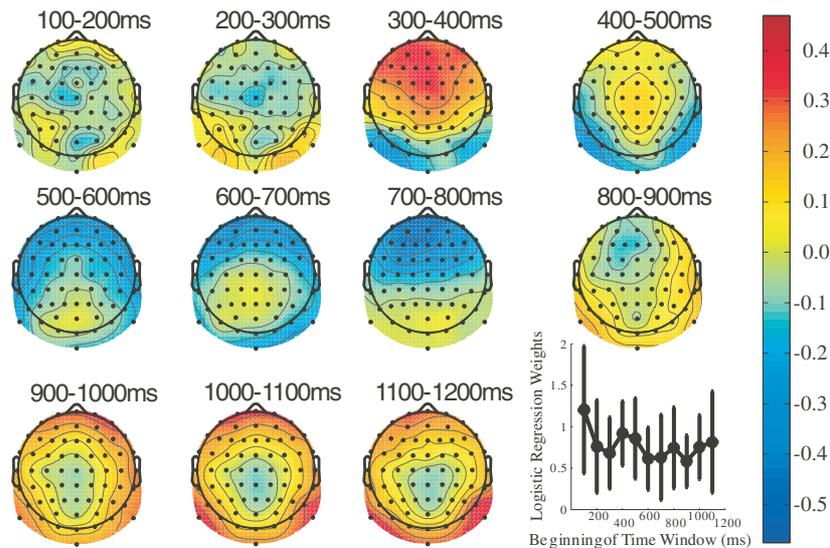
for these three subjects versus 0.34 (STD = 0.22,  $n = 38$ ) for the others.

How the interaction between the CV and EEG interest detector modules impacted overall system performance can be seen in figures 2(a) and 3. Figure 2(a) shows how the

CV module performed better when the EEG module provided flagged image sets of higher target precision, in particular when the precision of the flagged image sets exceeded 0.2. Similarly, figure 3 shows how the subjects usually generated higher precision flagged image sets when there were more target images in the RSVP, and increasing target prevalence beyond chance levels depended on the CV module’s ability to resample the database appropriately for subsequent RSVP sequences, even in cases where it was unsure of the specific target type. Interestingly, the performance gains in the interest detector are less consistent when target prevalence increases beyond 10–15%. This is reminiscent of P300 oddball studies that have found the strength of the P300 response increases for more infrequent target appearances [45–49], and which commonly employ target probabilities of 20% or less. While the interdependence between the HV and CV modules could theoretically lead to situations in which each module waited for the other to first achieve high performance, the closed-loop aspect of this system allowed the modules to gradually reinforce each other. Nonetheless, both modules’ independent capabilities were obviously important to overall performance.

### 3.2. EEG interest detector module performance

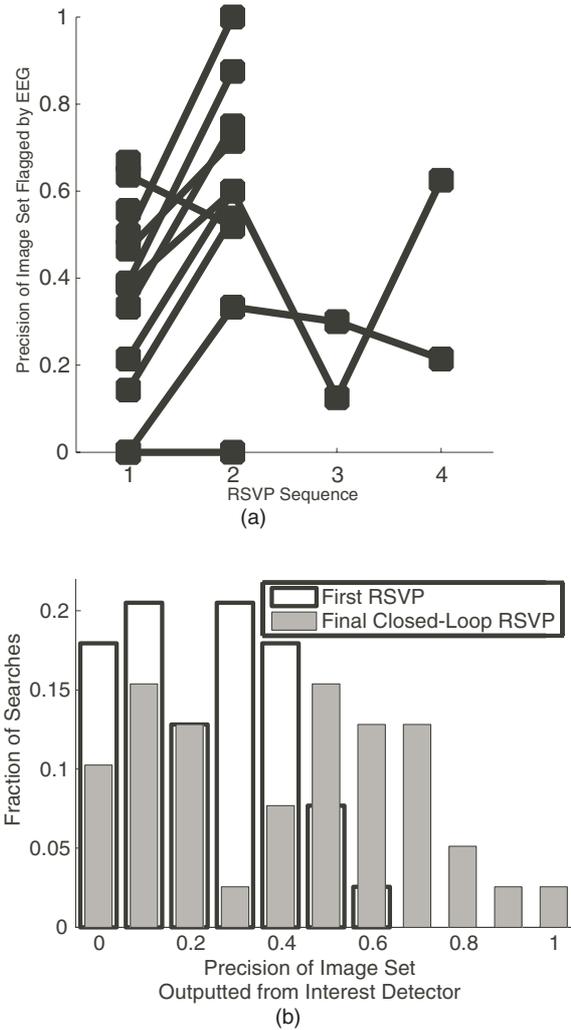
The interest detector (equations (2) and (3)) successfully detected patterns of EEG activity that specifically followed the appearance of target images in the RSVP sequences. Figure 4 shows the forward model (averaged between all subjects) of the interest detector’s 11 spatial filters (each for a 100 ms time window) with the inset showing the temporal weighting of these spatial components. Interestingly, the spatial maps and temporal weights suggest that three EEG



**Figure 4.** Scalp maps of the mean forward model (averaged between all subjects) indicate the temporal and spatial contribution of useful discriminatory information obtained from the EEG electrodes. Information between 300 and 500 ms (suggesting contributions by the P300 response) as well as early and longer latency activity was very useful to the EEG classifier, though the subjects showed considerable variation in precisely which time windows were most heavily weighted by the logistic regression (the inset figure shows the mean and STD of the temporal weights, equation (3), between subjects).

components make up the activity contributing to the EEG-score. The first occurs 100–200 ms post-stimulus (notable in the logistic regression weights), with a sign and a distributed topology indicative of a visual N100. The N100 is an ERP that has been observed for both visual and auditory stimulus paradigms. It is present for unexpected stimuli independent of tasks demands, and for visual stimuli it often reflects vigilance and attention [50–52]. The second component is the late positive complex, which appears to include both a P3a (fronto-central activity) transitioning to a P300/P3b topology (central-parietal). This is consistent with our previous findings [5] and likely indicates that the composite EEG-score includes both target- and novel-related activity—i.e. some of the images, though targets, are completely new to the subject and may evoke a strong attentional orientating response. As we reported previously [5], we also see late activity with a central topography. Since there is no behavioral response in our experimental paradigm and the activity is not lateralized, it is very unlikely that this is related to a covert motor process. This late component has a topology that is somewhat consistent with self-monitoring activity, such as the error-related negativity (ERN), which is often associated with perceived error and/or conflict [53, 54]. Interestingly, our experiments are such that subjects often ‘realize’ that they saw or missed a target a few images after the target image is presented, and thus there are at least some reports of a self-monitoring or decision conflict during the task. What is striking about this component is that there is no overt/behavioral response—e.g. the ERN is usually observed following a response. However, the RSVP nature of the task potentially leads to a relatively precise time-locking of neural processes underlying self-monitoring of decisions which can be detected by our algorithm. Nonetheless, additional experiments are needed to determine if this self-monitoring component is indeed what we are observing.

The EEG interest detector module’s performance was improved through its interaction with the CV module. If the system had been run in an open-loop configuration, only the first (500 random images) RSVP sequence would have been utilized for each database search. The average area under the ROC curve ( $A_z$ ) value for these open loop EEG scores (all subjects, all searches) was 0.82 (STD = 0.13,  $n = 49$ ), significantly above chance (chance value is 0.5, one-sided  $t$ -test,  $p \ll 0.001$ ). However, despite these high  $A_z$  scores, there were often still significant numbers of false positives in the image sets that were flagged under these conditions. Through the closed loop, the interest detector module could improve the precision of targets in the flagged image sets it outputted. Figure 5(a) illustrates this when the search target was ‘brain’. While in two searches only a single RSVP was necessary for the system to identify the target category, for eight of the other ten searches the final precision of target images in the flagged image set was higher after using the closed loop. Figure 5(b) shows overall (all subjects, all searches involving more than a single RSVP) how the precision of targets in the flagged image set under open loop conditions was significantly lower than the final image set of the closed-loop system (0.24 versus 0.42, one-sided  $t$ -test,  $p < 0.001$ ,  $n = 39$ ). As such, the system typically took advantage of the

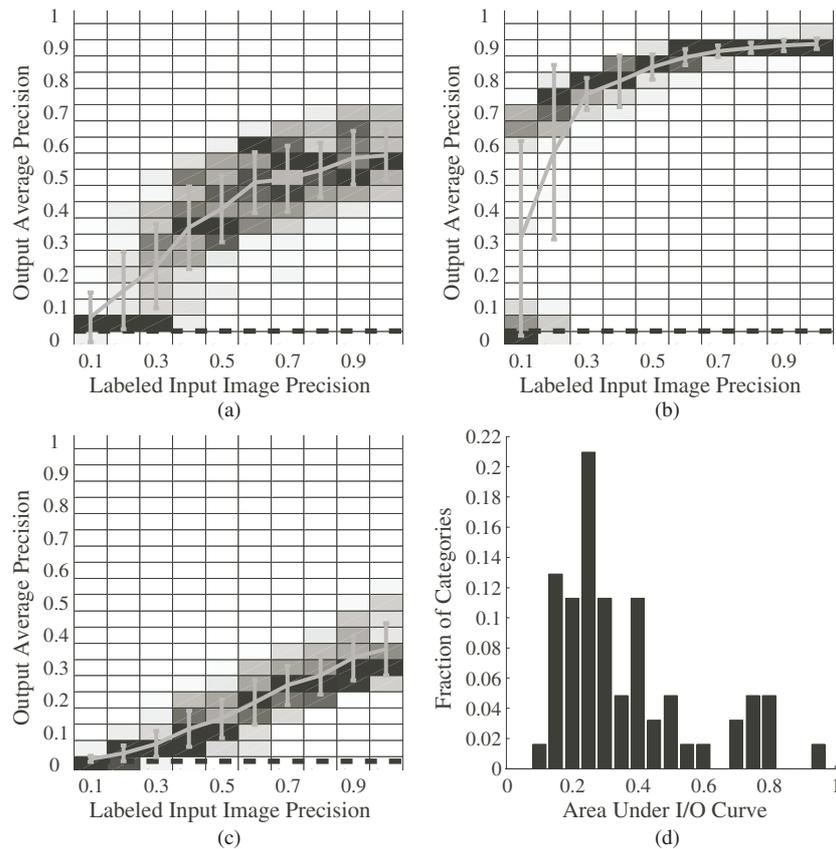


**Figure 5.** Performance of the EEG interest detector typically improved across successive RSVP sequences. This allowed the closed-loop system to flag sets of interesting images of higher precision than if the system had been operating in an open-loop configuration; (a) shows for all the subjects how the precision of targets in the image sets flagged by the interest detector varied across each successive RSVP (target category: ‘brain’). As can be seen, eight subjects (of the ten who needed more than a single RSVP sequence) increased the precision of their flagged image set using the closed-loop system; (b) shows the distributions of the precisions of the flagged image sets after the first RSVP (hollow, mean = 0.24) compared to the image sets when the system exited the CL-C3Vision loop (solid, mean = 0.42) for all subjects and targets in which there was more than one RSVP for the search.

closed-loop architecture (average number of RSVP sequences: 2.4, STD = 1.2,  $n = 49$ ), with only ten searches involving a single RSVP.

### 3.3. Computer vision (TAG) module performance

The TAG module was effective for both gradually growing the prevalence of target images between subsequent RSVP sequences, which did not involve self-tuning, and for the final database reorganization, in which self-tuning was utilized. The effectiveness of the TAG in both these capacities depended not only on the precision of the example image sets provided by the

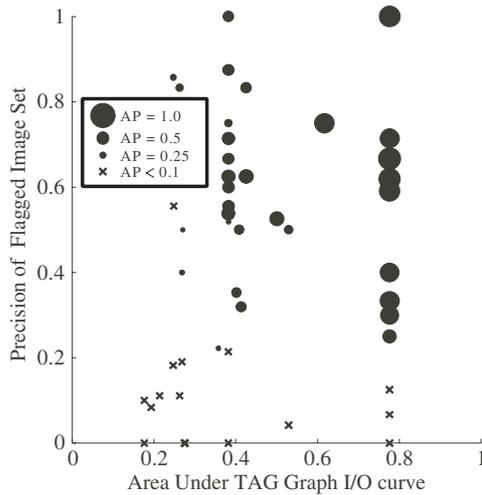


**Figure 6.** How well categories were represented in the TAG affinity graph directly impacted system performance. To quantify how well the TAG graph captured different image categories, simulations in which 20 images (of varying precision of the labeled input images) were given to the TAG (100 simulations for each precision level, self-tuning of 25%) for all 62 target categories. The resulting AP for the target category was used to quantify the TAG performance for each simulation. (a–c) Three examples (‘brain’, ‘grand piano’, and ‘lobster’ respectively) of how well the TAG reorganized the image database depending on input precision. The vertical columns in each of the plots are histograms of the simulation results (target AP resulting from the TAG image ranks) for each input precision level. The solid lines indicate the mean and STD of the APs between all the simulations for a given input precision, and the dashed lines show chance APs. (b) A relatively common phenomenon in which the TAG performance for that image category depended on a distinct input set precision threshold. The area under the simulation I/O curves was used to quantify the effectiveness of the TAG graph in capturing each image category (for (a–c) the areas under the I/O curves were 0.38, 0.78, and 0.18, respectively). (d) The distribution of the area under the I/O plots for all 62 image categories.

EEG interest detector, but also on how well the TAG affinity graph captured the various image categories. To quantify how well image categories were represented in the TAG graph, we ran a series of simulations in which we assessed how useful TAG rankings were (when using 25% self-tuning) when the fraction of target images within the example image set given to the TAG (i.e. the input precision) varied. For each category, 100 simulations were run for each of several different input precisions (with the target and distractor images within the example sets being randomly selected for a total of 20 images), and the resulting TAG ranks were used to calculate the simulation output AP. Figures 6(a)–(c) show examples of the resulting input/output (I/O) curves that reflect how well the categories were represented by the TAG graph. ‘Brain’ is an average category, ‘grand piano’ is a category well represented by the graph, and ‘lobster’ is poorly captured (but we note that even given small precision input sets for lobster, the TAG would still often provide above chance results). The area under each category’s I/O simulation curve summarizes how well that category was represented in the TAG graph (mean =

0.36,  $STD = 0.21$ ,  $n = 62$ ), with the distribution shown in figure 6(d). Notably, each category carried a better representation than chance (chance value of the area under the I/O curve corresponding to a category’s prevalence in the database, with the average prevalence between categories being 0.016). Figure 7 demonstrates how a category’s representation in the TAG graph affected the performance of the BCI system, as categories with stronger representations did not require as high a precision of the input set to achieve good search results.

The use of TAG (without self-tuning) to boost the number of target images in the RSVP sequences helped ensure that the image set outputted by the EEG module was of adequate precision for the final database reorganization. Not utilizing self-tuning for the TAG was appropriate in this context. This is because when using self-tuning, the TAG output AP can be very low when the input set precision is quite low (as can be seen in figure 6), and this is not an unexpected occurrence in early RSVP sequences, particularly for low-prevalence categories or for subjects for whom the performance of

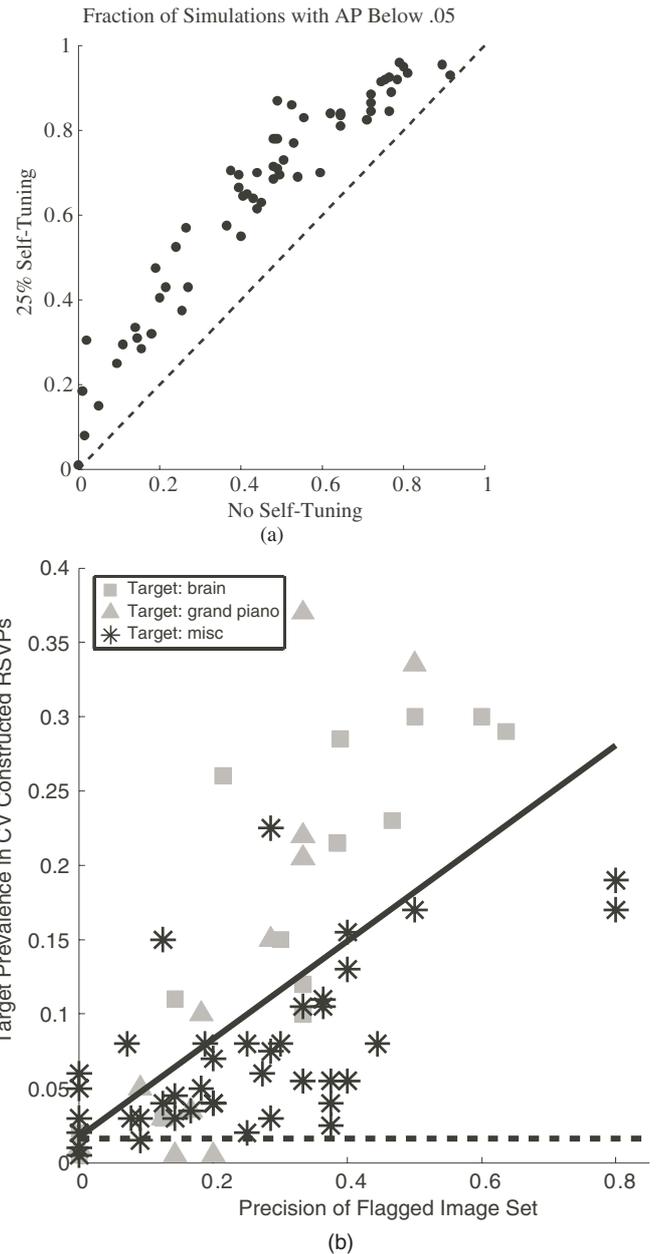


**Figure 7.** The TAG graph representation of the target category and the precision of the flagged image set outputted by the EEG interest detector both affected overall system performance. The size of each mark reflects the magnitude of the final AP for each of the search tests (x's are used for APs below 0.1). The vertical axis gives the precision of the flagged image set provided by the interest detector during the final RSVP in each search, while the horizontal axis is a measure of how well the target categories of each search were captured by the TAG graph (in terms of the area under the TAG simulation I/O curves, see figure 6). When searches involved categories that were weakly represented in the TAG graph, higher precisions were necessary in the input image set for the system to achieve good search results.

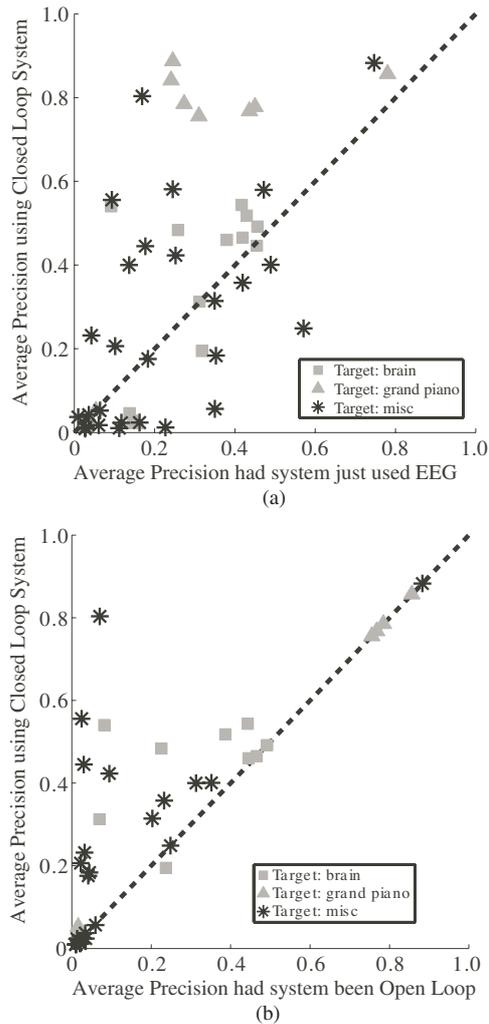
the EEG interest detector was only mediocre. Conversely, figure 8(a) shows how in the TAG graph simulation tests, the simulations not using self-tuning had many fewer instances of the TAG output approaching chance (here AP values  $\leq 0.05$  are considered chance levels) compared to self-tuning when the input precision was low (see the supplementary material available at [stacks.iop.org/JNE/8/036025/mmedia](http://stacks.iop.org/JNE/8/036025/mmedia), figures S2 and S3). Consequently, in the actual test searches, the TAG (with no self-tuning) worked similarly to a local analysis query expansion in information retrieval searches [55], as it constructed the closed-loop RSVPs around images directly related in the TAG graph to the flagged image sets. Figure 8(b) shows how this resulted in more target images in the closed-loop RSVP sequences than would be expected by chance, so long as the EEG interest detector module had outputted some base level of performance (typically a flagged image precision of 10% or better). This growth in target prevalence during the RSVPs greatly enhanced the benefit of the closed-loop aspect of the BCI architecture, as it made it easier for the EEG interest detector to flag example input sets of higher precision (figure 3). This in turn improved the self-tuning TAG implementation when it was used for the final database reorganization.

**4. Discussion**

The BCI presented here differs from previous C3Vision systems in that it includes a closed-loop interaction between the computer and HV components. The resulting



**Figure 8.** TAG was effective for boosting the number of target images in the closed-loop RSVP sequences. The precision of targets in the image sets flagged by the interest detector is likely to be low following early search RSVPs; thus, it is better to use TAG with no self-tuning when creating the closed-loop RSVP sequences. The benefit in not using self-tuning is shown in (a), which plots, for each category, the fraction of the TAG graph simulation results (see figure 6) that had a final output AP below .05 (i.e. the TAG reorganization was approaching chance levels) when the precision of the labeled input set was  $\leq 0.2$ ; (a) shows how simulations that used 25% self-tuning were more likely to have poor outputs compared to simulations in which no self-tuning was employed (the hatched line shows the 1:1 relationship); (b) then shows how during the search tests, TAG without self-tuning did provide closed-loop RSVP sequences in which the prevalence of the target images was clearly above chance (the hatched line indicates chance, i.e. the mean target prevalence between all categories), so long as the flagged image set had at least a few target images. The solid line is the best-fit line (correlation coefficient = 0.66 with slope significantly nonzero: *t*-test,  $p \ll 0.001$ ).



**Figure 9.** Combining the EEG interest detector and the CV in a closed-loop configuration offered better performance than operating each component serially in an open-loop, or simply using EEG interest detection alone; (a) shows how the CL-C3Vision system gave significantly higher APs (paired  $t$ -test  $p = 0.004$ ,  $n = 49$ ) compared to simply using the EEG interest detector to rank images (mean APs of 0.35 versus 0.26); in those tests, the AP achieved by the EEG interest detector for the initial RSVP of 500 randomly selected images was considered representative of how the detector would have ranked the full database had all 3798 images actually been shown to the users; (b) shows how the C3Vision system reorganized the image database better when used in the full closed-loop system than if it had been operating in an open-loop fashion (mean APs of 0.34 versus 0.23, paired  $t$ -test  $p < 0.001$ ,  $n = 39$ , only searches in which there were more than a single RSVP were compared). The open-loop results were obtained by performing the final database reorganization based on the EEG scores obtained from only the first RSVP. In both plots the hatched lines show a 1:1 relationship.

CL-C3Vision system is not necessarily intended to act as an assistive device for individuals with disabilities (although nothing prevents it from being applied to such goals), but rather is meant to improve an able-bodied user's ability to quickly locate images of interest in a large image database. Figure 9 illustrates both how combining CV with EEG interest measures improves the detection of interesting images, and

how the closed-loop aspect helps the system take better advantage of the CV component. Furthermore, the CL-C3Vision system boosts the benefit from those closed-loop RSVPs by using the CV to adaptively resample the database to emphasize images that are likely to be targets in the additional RSVPs. This increases the information gained in each RSVP, thus limiting the number of images that the user must view and speeding the search task. It also allows the system to refine its target information by accruing data from a small but growing subset of images rather than relying on averaging EEG information through repeated presentations of the same images. This can help speed the search, as time is not spent on repetitively showing a few images that may turn out to be irrelevant.

#### 4.1. Benefits of a closed-loop cortically-coupled computer vision BCI in practical search tasks

CL-C3Vision systems could be useful for a variety of different search tasks. While we tested the system using a search of a relatively modestly sized image database, it could be applied to much larger image databases, or even databases of different types (such as video). Additionally, CL-C3Vision could be of particular use if the search target were poorly defined, for example, a search of a database of flowers for not a specific type of flower, but rather simply for a flower that the user would find attractive in a centerpiece (e.g. 'I'm just looking for some flowers that have attractive petals'). All that is really required to use CL-C3Vision advantageously is a large, possibly poorly organized or annotated (relative to the search goal) database, from which samples can be reviewed at a high rate, and for which it would be impractical or overly time consuming to manually inspect each entry.

The full benefits of using a CL-C3Vision system would then be dependent on both the size and the complexity of the database, as well as the required parameters of the search task. For example, in most cases, the user would likely want to make a final review of the most highly ranked images to select specific images. As the CL-C3Vision output is simply a ranked list of the likelihood of each image being of interest, the quantity of images included in this second pass would depend on the specific goals of the search itself. If the user were simply inspecting a database to find a few desirable images (such as the flower search above), the second pass might simply involve reviewing a small subset of database images whose ranks fell well above the mean, similar to reviewing only the first few pages of an internet search engine's results. This could allow the user to search a database of considerable size and find essentially what they are looking for with a significant time savings. Alternatively, a larger second pass would be necessary if a more comprehensive search were desired in which any image even remotely related to the flagged images must be located. For example, in the current experiments, output APs of about 0.5 or above (typically reflecting an over tenfold improvement above chance) were often achieved. In such cases, if the users had then needed to identify 50% of the target images in these tests, they would have required a second pass that included roughly 2.1% of the database images. This

second pass would then grow to include as much as 38% of the database if 90% of the images from the target category needed to be manually reviewed by the subject following the use of the CL-C3Vision.

#### 4.2. Refining the CL-C3Vision system

Several factors resulted in the current system periodically failing to identify the correct target category. The primary factor was the EEG interest detector repeatedly outputting flagged image sets with few target images. For the three unsuccessful subjects previously noted in figure 2, this appeared to be a consistent issue in system performance. Among the other subjects' eight failed searches, two also appeared primarily related to the interest detector not flagging target images adequately. In contrast, in another three, it appeared that (even though the EEG interest detector produced flagged image sets of reasonable precision) the TAG either decreased the number of targets in the subsequent RSVPs, or apparently focused the final database reorganization around a different target category (or both). In two failed search cases, the performances of both CV and EEG modules appeared weak. Thus, excluding the three subjects that showed multiple failures, there was a fairly even mix of the system suffering from either a weakness in the EEG module, the TAG module, or both when it failed to ID the target category. In the last ID failure, in the tests presented here, the subject had chosen to look for lobsters (S17 in figure 2), but instead seemed to respond most strongly to images of crayfish (as well as a few lobsters, which crayfish closely resemble), with the final BCI prediction reflecting that tendency. This last case does not really reflect a failure of the system, but rather highlights its recognition of the user's subjective response, rather than any objective property of the images being viewed.

The above weaknesses suggest several means by which the system can be improved. In particular, despite the benefits of the closed-loop interaction between the EEG and CV modules, their individual capabilities are still quite important. While small numbers of individuals do seem to simply have more difficulty in controlling BCI systems that use P300 ERPs [25], a refined EEG interest detector that outputted better precision image sets for more subjects would have avoided most cases of incorrect target identification in our tests. Such refinements could include better algorithms for single trial detection of user interest, or perhaps focusing the current algorithm to search for information specifically related to P300 events. This latter approach could include restricting time windows or electrodes to those more likely to be relevant to P300 to reduce the number of parameters, or perhaps using specific P300 related variables (such as amplitude or latency) in the interest detector. Alternatively, longer or better training regimes should be attempted to explore whether the current detector could become advantageous to more users. For example, a 5 Hz presentation rate was used. While, anecdotally, individuals well versed in the system can achieve good results at faster rates (even 10 Hz), the 5 Hz rate was felt to be reasonable for most naive subjects. However, half of the subjects in this study reported that they felt that the

rate was a bit fast (although half of those subjects noted that they became more comfortable with the rate over time). Using subject-specific rates, especially in training when the rates could be started slowly and then gradually increased, may be useful. The effectiveness of using different quantities and/or patterns of electrodes could also be explored. For example, we compared the AP for the first RSVP of each subject (target: 'brain') if data from only nine electrodes (locations: POz, Fz, Cz, C3, C4, F3, F4, P3, P4) had been used in the classifier instead of the full set of 64 electrodes. The average drop in AP when switching to nine electrodes was only 13%, with four subjects having better results with 9 electrodes than 64. Also, accounting for blinking in individual subjects may improve system performance. While in this work subjects who appeared to generate eye movements preferentially following target images were eliminated, no attempt was made to address the consequences of the other subjects blinking and thus missing a target, or perhaps missing a target due to an attentional blink [56] resulting from the appearance of a previous target image. Future studies should consider these issues, particularly if the RSVP rate is increased.

In addition to improving the EEG interest detector, the computer vision module could be improved. While it appears that only three failures in this work resulted from the target category being weakly represented in the TAG graph (the mean value of the TAG simulation I/O metric for those three categories was 0.3, while that for successful search target categories was 0.5), as many as 35 categories could be challenging to identify (i.e. have I/O metrics  $\leq 0.3$ , figures 6 and 7) if chosen as search targets. The TAG module could be improved by using an affinity graph specific to the image database (the current graph was constructed indifferently to the types of images known to be in the database), or by making a more studied use of TAG self-tuning (rather than using a single fixed value of 25%). Additionally, entirely different computer vision techniques could be explored, possibly with different CV methods being used for one or more of the tasks currently assigned to the TAG (assessing flagged image similarity, boosting target presence in subsequent RSVPs, and performing the final database ranking) for an overall improvement in the performance of the CV module.

## 5. Conclusions

In this paper, we have presented a BCI that uses cortically-coupled computer vision (C3Vision) to find images of interest in a large image database. This system modifies previous C3Vision systems by including a closed-loop interaction between the human and computer vision system components. This allows the system to query the user for more information when necessary to effectively search for the target image category. Furthermore, the adaptive resampling used for those queries helped maximize the information gained from each closed-loop interaction, reducing the number of images reviewed by the user and speeding the search task. The resulting system proved effective in reorganizing the test database to expedite searches for a diverse set of image

categories that were either assigned to or freely chosen by the BCI users as search targets. In many cases these benefits exceeded tenfold improvements over chance, and the system proved capable of inferring the user's intent and identifying the specific image category that had been chosen as the search target. While there are still many aspects of the CL-C3Vision system that could be improved, this study suggests that such closed-loop BCI systems will potentially be useful for assisting individuals to deal with the deluge of imagery they are confronted with in their everyday lives.

## Acknowledgments

This work was supported by DARPA under contracts N10PC20050 and NBCHC080029. Paul Sajda is a co-founder of Neuromatters, LLC, a neurotechnology company developing applications of C3Vision technology.

## References

- [1] Thorpe S, Fize D and Marlot C 1996 Speed of processing in the human visual system *Nature* **381** 520–2
- [2] Oliva A 2005 Gist of the scene *Encyclopedia of Neurobiology of Attention* (San Diego, CA: Elsevier) pp 251–6
- [3] 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
- [4] Parra L C, Christoforou C, Gerson A D, Dyrholm M, Luo A, Wagner M, Philiastides M G and Sajda P 2008 Spatiotemporal linear decoding of brain state: application to performance augmentation in high-throughput tasks *IEEE Signal Process. Mag.* **25** 95–115
- [5] Sajda P, Pohlmeier 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
- [6] Iturrate I, Antelis J M, Kubler A and Minguez J 2009 A noninvasive brain-actuated wheelchair based on a P300 neurophysiological protocol and automated navigation *IEEE Trans. Robot.* **25** 614–27
- [7] Bentley A S J, Andrew C M and John L R 2008 An offline auditory P300 brain-computer interface using principal and independent component analysis techniques for functional electrical stimulation application *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2008** 4660–3
- [8] Daly J J and Wolpaw J R 2008 Brain-computer interfaces in neurological rehabilitation *Lancet Neurol.* **7** 1032–43
- [9] Farwell L A and Donchin E 1988 Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials *Electroencephalogr. Clin. Neurophysiol.* **70** 510–23
- [10] Leuthardt E C, Schalk G, Roland J, Rouse A and Moran D W 2009 Evolution of brain-computer interfaces: going beyond classic motor physiology *Neurosurg. Focus* **27** E4
- [11] Lotte F, Congedo M, Lécuyer A, Lamarche F and Arnaldi B 2007 A review of classification algorithms for EEG-based brain-computer interfaces *J. Neural Eng.* **4** R1–13
- [12] McFarland D J, Sarnacki W A and Wolpaw J R 2010 Electroencephalographic (EEG) control of three-dimensional movement *J. Neural Eng.* **7** 036007
- [13] Millán J D R et al 2010 Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges *Front. Neurosci.* **4**
- [14] Pfurtscheller G, Müller G R, Pfurtscheller J, Gerner H J and Rupp R 2003 'Thought'-control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia *Neurosci. Lett.* **351** 33–6
- [15] Millán J D R, Renkens F, Mourifio J and Gerstner W 2004 Noninvasive brain-actuated control of a mobile robot by human EEG *IEEE Trans. Biomed. Eng.* **51** 1026–33
- [16] 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
- [17] Wolpaw J R and McFarland D J 2004 Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans *Proc. Natl Acad. Sci. USA* **101** 17849–54
- [18] Rebsamen B, Guan C, Zhang H, Wang C, Teo C, Ang M H and Burdet E 2010 A brain controlled wheelchair to navigate in familiar environments *IEEE Trans. Neural Syst. Rehabil. Eng.* **18** 590–8
- [19] Linden D E J 2005 The P300: where in the brain is it produced and what does it tell us? *Neuroscientist* **11** 563–76
- [20] Polich J 2007 Updating P300: an integrative theory of P3a and P3b *Clin. Neurophysiol.* **118** 2128–48
- [21] Sutton S, Braren M, Zubin J and John E R 1965 Evoked-potential correlates of stimulus uncertainty *Science* **150** 1187–8
- [22] Lenhardt A, Kaper M and Ritter H J 2008 An adaptive P300-based online brain-computer interface *IEEE Trans. Neural Syst. Rehabil. Eng.* **16** 121–30
- [23] Krusienski D J, Sellers E W, McFarland D J, Vaughan T M and Wolpaw J R 2008 Toward enhanced P300 speller performance *J. Neurosci. Methods* **167** 15–21
- [24] Klobassa D S, Vaughan T M, Brunner P, Schwartz N E, Wolpaw J R, Neuper C and Sellers E W 2009 Toward a high-throughput auditory P300-based brain-computer interface *Clin. Neurophysiol.* **120** 1252–61
- [25] Guger C, Daban S, Sellers E, Holzner C, Krausz G, Carabalona R, Gramatica F and Edlinger G 2009 How many people are able to control a P300-based brain-computer interface (BCI)? *Neurosci. Lett.* **462** 94–8
- [26] Poolman P, Frank R M, Luu P, Pederson S M and Tucker D M 2008 A single-trial analytic framework for EEG analysis and its application to target detection and classification *Neuroimage* **42** 787–98
- [27] Bigdely-Shamlo N, Vankov A, Ramirez R R and Makeig S 2008 Brain activity-based image classification from rapid serial visual presentation *IEEE Trans. Neural Syst. Rehabil. Eng.* **16** 432–41
- [28] Wang J, Pohlmeier E, Hanna B, Jiang Y-G, Sajda P and Chang S-F 2009 Brain state decoding for rapid image retrieval *MM '09: Proc. 17th ACM Int. Conf. on Multimedia* (New York: ACM) pp 945–54
- [29] Fetz E E 2007 Volitional control of neural activity: implications for brain-computer interfaces *J. Physiol.* **579** (Pt 3) 571–9
- [30] Wessberg J, Stambaugh C R, Kralik J D, Beck P D, Laubach M, Chapin J K, Kim J, Biggs S J, Srinivasan M A and Nicolelis M A 2000 Real-time prediction of hand trajectory by ensembles of cortical neurons in primates *Nature* **408** 361–5
- [31] Taylor D M, Tillery S I H and Schwartz A B 2002 Direct cortical control of 3D neuroprosthetic devices *Science* **296** 1829–32
- [32] Schwartz A B, Cui X T, Weber D J and Moran D W 2006 Brain-controlled interfaces: movement restoration with neural prosthetics *Neuron* **52** 205–20
- [33] Koyama S, Chase S M, Whitford A S, Velliste M, Schwartz A B and Kass R E 2010 Comparison of brain-computer interface decoding algorithms in open-loop and closed-loop control *J. Comput. Neurosci.* **29** 73–87
- [34] Carmena J M, Lebedev M A, Crist R E, O'Doherty J E, Santucci D M, Dimitrov D F, Patil P G, Henriquez C S

- and Nicolelis M A L 2003 Learning to control a brain-machine interface for reaching and grasping by primates *PLoS Biol.* **1** E42
- [35] Fei-Fei L, Fergus R and Perona P 2004 Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories *Proc. Conf. Computer Vision and Pattern Recognition Workshop CVPRW '04* p 178
- [36] Potter M C and Levy E I 1969 Recognition memory for a rapid sequence of pictures *J. Exp. Psychol.* **81** 10–5
- [37] Wang J and Chang S-F 2008 Columbia TAG system—transductive annotation by graph version 1.0 *Technical report*, Columbia University, NY, USA
- [38] Wang J, Jebara T and Chang S-F 2008 Graph transduction via alternating minimization *ICML '08: Proc. 25th Int. Conf. on Machine Learning* (New York: ACM) pp 1144–51
- [39] Wang J, Jiang Y-G and Chang S-F 2009 Label diagnosis through self tuning for web image search *Proc. IEEE Conf. Computer Vision and Pattern Recognition CVPR 2009* pp 1390–7
- [40] Jebara T, Wang J and Chang S-F 2009 Graph construction and bmatching for semi-supervised learning *Int. Conf. on Machine Learning* pp 441–8
- [41] Oliva A and Torralba A 2001 Modeling the shape of the scene: a holistic representation of the spatial envelope *Int. J. Comput. Vis.* **42** 145–75
- [42] Griffin G, Holub A and Perona P 2007 Caltech-256 object category dataset *Technical Report 7694* California Institute of Technology, CA, USA
- [43] Smeaton A F, Over P and Kraaij W 2006 Evaluation campaigns and TRECVID *MIR '06: Proc. 8th ACM Int. Workshop on Multimedia Information Retrieval* (New York: ACM) pp 321–30
- [44] Buckley C and Voorhees E M 2000 Evaluating evaluation measure stability *Proc. 23rd Annu. Int. ACM SIGIR Conf. on Research and Development in Information Retrieval, SIGIR '00* (New York: ACM) pp 33–40
- [45] Polich J and Margala C 1997 P300 and probability: comparison of oddball and single-stimulus paradigms *Int. J. Psychophysiol.* **25** 169–76
- [46] Polich J, Ellerson P C and Cohen J 1996 P300, stimulus intensity, modality, and probability *Int. J. Psychophysiol.* **23** 55–62
- [47] Heinrich S P and Bach M 2008 Signal and noise in P300 recordings to visual stimuli *Doc. Ophthalmol.* **117** 73–83
- [48] Cohen J and Polich J 1997 On the number of trials needed for P300 *Int. J. Psychophysiol.* **25** 249–55
- [49] Bonala B, Boutros N N and Jansen B H 2008 Target probability affects the likelihood that a P300 will be generated in response to a target stimulus, but not its amplitude *Psychophysiology* **45** 93–9
- [50] Vogel E K and Luck S J 2000 The visual N1 component as an index of a discrimination process *Psychophysiology* **37** 190–203
- [51] Luck S J, Woodman G F and Vogel E K 2000 Event-related potential studies of attention *Trends Cogn. Sci.* **4** 432–40
- [52] Wascher E, Hoffmann S, Sanger J and Grosjean M 2009 Visuo-spatial processing and the N1 component of the ERP *Psychophysiology* **46** 1270–7
- [53] Scheffers M K and Coles M G 2000 Performance monitoring in a confusing world: error-related brain activity, judgments of response accuracy, and types of errors *J. Exp. Psychol. Hum. Percept. Perform.* **26** 141–51
- [54] Botvinick M M, Cohen J D and Carter C S 2004 Conflict monitoring and anterior cingulate cortex: an update *Trends Cogn. Sci.* **8** 539–46
- [55] Xu J and Croft W B 1996 Query expansion using local and global document analysis *Proc. 19th Annu. Int. ACM SIGIR Conf. on Research and Development in Information Retrieval* pp 4–11
- [56] Raymond J E, Shapiro K L and Arnell K M 1992 Temporary suppression of visual processing in an RSVP task: an attentional blink? *J. Exp. Psychol. Hum. Percept. Perform.* **18** 849–60