Journal of Neuroscience Methods 228 (2014) 1–14

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Computational Neuroscience

Single-trial classification of EEG in a visual object task using ICA and machine learning

Andrew X. Stewart[a,*], Antje Nuthmann[b], Guido Sanguinetti[c]

[a] Neuroinformatics Doctoral Training Centre, Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, UK
[b] Psychology Department, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, UK
[c] Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, UK

HIGHLIGHTS

- We consider machine learning in assessing information in different EEG data.
- We train SVM classifiers using EEG data from a visual object stimuli task.
- New data can be correctly labelled with 'object present' state well above chance.
- Using one channel of ICA data as input increases classification accuracy to 87%.
- We discuss how this method and IC sources might help studies of visual cognition.

ARTICLE INFO

Article history:
Received 27 November 2012
Received in revised form 23 February 2014
Accepted 24 February 2014

Keywords:
EEG
ICA
Classification
SVM
Single-trial

ABSTRACT

Presenting different visual object stimuli can elicit detectable changes in EEG recordings, but this is typically observed only after averaging together data from many trials and many participants.

We report results from a simple visual object recognition experiment where independent component analysis (ICA) data processing and machine learning classification were able to correctly distinguish presence of visual stimuli at around 87% (0.70 AUC, $p < 0.0001$) accuracy within single trials, using data from single ICs.

Seven subjects observed a series of everyday visual object stimuli while EEG was recorded. The task was to indicate whether or not they recognised each object as familiar to them. EEG or IC data from a subset of initial object presentations was used to train support vector machine (SVM) classifiers, which then generated a label for subsequent data. Task-label classifier accuracy gives a proxy measure of task-related information present in the data used to train.

This allows comparison of EEG data processing techniques – here, we found selected single ICs that give higher performance than when classifying from any single scalp EEG channel (0.70 AUC vs 0.65 AUC, $p < 0.0001$). Most of these single selected ICs were found in occipital regions. Scoring a sliding analysis window moving through the time-points of the trial revealed that peak accuracy is when using data from +75 to +125 ms relative to the object appearing on screen. We discuss the use of such classification and potential cognitive implications of differential accuracy on IC activations.

1. Introduction

Electroencephalography (EEG) allows neuroimaging with high temporal resolution. This can be used for investigation of the properties of cognitive neuroscience processes, like the speed of early visual processing in humans. For example, EEG Event-Related Potential (ERP) differences can be detected in animal versus non-animal images within 150 ms of presentation (Thorpe et al., 1996) and faces versus shapes within 85 ms (Mouchetant-Rostaing and Giard, 2000).

However, ERP averaged over many subjects and many trials is not the totality of the information present in EEG recordings. The use of ERP alone has been criticised (Rousselet and Pernet, 2011; Gaspar et al., 2011) for masking effects within grand averaging and not using all available information.

* Corresponding author at: 2.53 Informatics Forum, Edinburgh EH8 9AB, UK.
E-mail address: andrew.x@ed.ac.uk (A.X. Stewart).
URL: http://www.anc.ed.ac.uk/dtc/find.php?andrewstewart (A.X. Stewart).

http://dx.doi.org/10.1016/j.jneumeth.2014.02.014

Advanced data mining of EEG data (Makeig et al., 2004) has been suggested as an alternative approach, taking into account other aspects, such as spectral power, phase, trial-to-trial consistency and ICA activations. It may be more difficult to interpret information integrated from these multiple aspects than it is to observe grand-averaged ERP plots. How to compare these techniques is then a relevant question.

1.1. Machine learning with EEG data

One approach to assess the use of these techniques is to use machine learning on EEG data. Training examples from different experimental states can be used to train machine learning classifiers, and the ability to correctly and robustly identify new examples can be assessed.

This is done in the adjacent field of brain–computer interfacing, where EEG signals are used to control motor prostheses (Wolpaw et al., 2000; Donoghue, 2008). For EEG motor prostheses, the experimental states of interest might be presence or absence of prompted hand movement and success of the classifier would be to robustly identify this motor movement in subsequent EEG data.

We hypothesise a similar approach could be applied to EEG data recorded in visual object presentations. Instead of identifying motor movement such as 'grab' from resting baseline with no movement, the experimental states could be the presence of a given visual object stimulus (e.g. 'spoon') and resting baseline with no visual stimulus.

In EEG motor prostheses, there is a desire for accurate and fast classification. Consequently, many analysis methods have been attempted (Sajda et al., 2003; Lotte et al., 2007). Use of data preprocessing and machine learning tools have proved effective in improving the ability to use EEG data to predict movement intent (Müller et al., 2008).

The use of machine learning on visual EEG experiments is not as well studied as EEG data in motor movement. The EEG response elicited from different object stimuli is likely less regular, more subtle and likely has a lower signal-to-noise ratio than EEG from different motor movements, but we suggest recent advances from motor brain–computer interfaces might also apply to improving learning from EEG data in visual experiments.

Therefore, we designed a simple visual object presentation experiment where we labelled EEG data (as 'object present' or 'object absent'), trained an SVM classifier with this data and labels, and assessed that model's accuracy at labelling subsequent unseen EEG data correctly. We used classifier task-labelling accuracy as a metric of task-relevant information in the data used to train the classifier. In the present context, this allows comparison of EEG data processing methods.

1.2. EEG features and ICA

In the field of machine learning, the input channels of the data that are used to train classifiers are termed 'features'. To improve the accuracy of brain–computer interfaces, a useful technique has been to focus on a small spatial subset of relevant features from EEG. An example of this is identifying 'common spatial features' (Müller et al., 2008; Wang and Jung, 2012). It is well known that these motor signals are highly localised in the motor strip and premotor areas. This made identification of reduced subsets of relevant features tractable (Blankertz et al., 2008). A comparable feature extraction procedure has not yet been accomplished for visual EEG data.

In order to give a comparable decomposition that could be relevant in vision, independent component analysis (ICA) was used on the EEG data in the present study. Each of the IC sources found from the EEG can be considered a reduced subset of the activity in the EEG. In essence, it separates the main independent generators of variance within a signal (Bell and Sejnowski, 1995). An illustrative example of ICA is the 'cocktail-party problem' where, given data from three microphones in a room and three overlapping voices, ICA can separate the three individual voices as three distinct sources (Hyvärinen and Oja, 2000). For EEG data then, the detectors of EEG electrodes act as the microphones and the varying electrical patterns as the voices.

ICA is typically used for identifying and removing noisy electrodes, blinks and other artefacts to 'clean up' EEG data (Luck, 2005) before proceeding with conventional ERP analysis. Here, we instead use ICA as another way of describing the EEG data, to give subsets of data that may be both more interpretable and give higher classifier performance. The ICA components provide an estimation of possible 'sources' of generated activity. This can be particularly advantageous in EEG data analysis, where much of the signal (and noise) is shared across all channels (Onton et al., 2006). In summary, some of the identified IC sources may be useful descriptions of subsets of variation within the EEG, as 'common spatial features' have been in motor EEG.

1.3. Support vector machines classifiers

We sought to determine whether EEG data in our experiment can be automatically classified using machine learning tools. To this end, we used support vector machines (SVMs) – a flexible and powerful statistical learning tool (Burges, 1998; Cortes and Vapnik, 1995). This technique has given particularly good results in a wide range of domains, including cancer classification (Furey et al., 2000) and face detection (Osuna et al., 1997). In classification of EEG, SVMs have shown good performance in many contexts (Lotte et al., 2007). SVMs were the most commonly used technique for highest accuracy in an EEG classification competition (Blankertz et al., 2006). Thus, we used SVMs as our classifier here.

Specifically, we used SVMs to classify the EEG data into two classes according to the presence or absence of specific visual stimuli. The underlying principle of SVM classification is to solve a (non-linear) classification problem by transforming it into a linear classification problem in a different, higher dimensional space (or *feature space*).

This is achieved by introducing a non-linear map (*feature map*) into the feature space, which can often be an infinite dimensional space of function. The important aspect of this procedure is that, for many widely used algorithms, one is only interested in the scalar products between pairs of feature vectors; these scalar products can be computed by means of a *kernel function* which depends only on the original (non-transformed) data points. Therefore, the need to work explicitly in the high dimensional space is removed and the feature map is defined implicitly by a choice of kernel function.

A commonly adopted kernel is the so-called *Radial Basis Function* (RBF) kernel; the scalar product of the images of two data points x and y under the feature map implied by the RBF kernel is computed as:

$$k(x, y) = \exp[-\gamma||x - y||^2] \quad (1)$$

where γ is a tunable parameter.

Once a kernel function is selected, the SVM algorithm works by identifying a hyperplane in feature space that optimally separates the two classes in the training data, giving the maximum margin between the images in feature space of the points in the two classes. Often it is desirable to allow a few misclassifications in order to achieve a wider margin of separation; this trade-off is controlled by another parameter, called the training error cost, and usually denoted by C.

1.4. Classification of visual object presence from EEG data

Previous studies of visual perception using EEG have demonstrated significant task-specific ERP differences within 150 ms of stimuli appearing (Thorpe et al., 1996). In their study, human subjects were able to perform a rapid decision on whether or not a visual scene contained an animal. Notably, examination of the EEG data was able to show the evolution of the difference in ERP on animal and non-animal trials.

Researchers interested in visual perception might ask what the source of such ERP differences is. To what extent can EEG signatures be used to decode observed images? What parts of the EEG signatures are useful for this?

The current study aims to extend this examination of EEG differences upon the presentation of single images by quantifying classification performance from different transforms of the EEG data, and EEG data from different times.

In designing a classifier of visual stimuli given EEG data, we first sought to determine if a visual object presentation could be distinguished at all. EEG data recorded in each visual object presentation trial was used along with more EEG data from resting baseline (blank screen), to train a machine learning SVM classifier. This was used to predict the correct label (as object present or absent) of additional unseen EEG data.

If the EEG data used to train the classifier contains information that is indicative of one state or another, it might allow high trial classification accuracy. We can choose what parts of the data are used for training and classifying, and so search for those parts of the EEG data that are most relevant for distinguishing trials.

As ICA gives an estimation of sources within the signal, we can use this classification procedure to assess the trial-classifying performance of each of these IC sources.

1.5. Hypothesis

Here we propose a method for classifying visual object presentation from EEG readings using machine learning. We test whether, using SVMs, we can classify the presence of visual object stimuli at above chance levels.

Further, we wish to compare scalp EEG data to EEG data transformed into ICs. Since ICA is a data reduction technique, the expectation is that some ICs might give higher accuracy.

2. Methods

2.1. Participants

Seven participants (five women, two men; median age of 25) took part in the study and were each compensated £7 per hour. Ethical approval was granted by the University of Edinburgh Psychology Research Ethics Committee. Participants self-reported normal or corrected-to-normal vision and gave informed written consent.

2.2. Materials and design

Participants were instructed to observe a series of object images presented to them on a computer screen. Fifty colour photographs of common real-world objects were selected from the Bank of Standardized Stimuli (BOSS) provided by Brodeur et al., 2010. Multiple presentation trials of each object were required for classifier training and testing. There were five blocks of 50 trials, where each object was shown once in a given block, in random order.

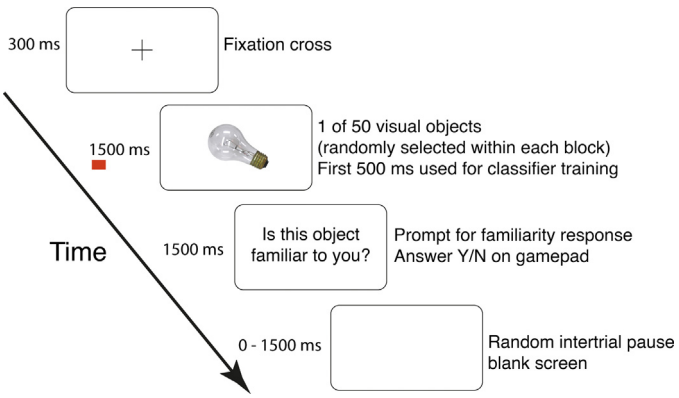

Fig. 1. Experimental design of the visual object presentation protocol. The timecourse of the events is shown. Participants were shown a fixation cross before stimulus presentation of an everyday visual object – in this example, a light-bulb – compiled from a published standardised image bank (Brodeur et al., 2010). There were 250 of these trials – five presentations each of the 50 stimuli, ordered randomly. The red bar indicates the 500 ms after presentation that was used as object onset data to train classifiers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.3. Apparatus

The experimental script was written in MATLAB 2009b (Mathworks, Inc., Natick, MA, USA). Extensions from Psychophysics Toolbox 3 were used for better timing precision (Brainard, 1997; Kleiner et al., 2007). Stimuli were displayed on a fast-refresh 22 inch Samsung SM2233RZ monitor at 1000×1000 pixels centred on a 1680×1050 pixel display, with the participants leaning on a chin-rest 72 cm away, giving an object size of 15–22° in the visual field. The complete screen was around 38° by 24°.

2.4. Experimental procedure

Each trial began with a central fixation cross for 300 ms, followed by a randomly selected object image that was presented for 1.5 s (Fig. 1). An on-screen text prompt then asked the subject if they had recognised this object as something familiar to them. This was implemented to avoid passive viewing of the stimuli. Subjects responded using buttons on a gamepad – one marked 'yes' if this object was recognised as familiar and one marked 'no' otherwise. The text was replaced by a blank screen on button response or after 1.5 s. There was then a random intertrial interval of 0–1.5 s, where the screen remained blank until the next trial began. The experiment was performed in a single session per subject of around 1.5 h, in which stimuli presentation and breaks lasted around 25 min.

2.5. EEG recording

EEG was recorded from 64 head electrodes and six support electrodes using a Biosemi ActiveTwo amplifier at a sample rate of 1024 Hz. These 64 electrodes were placed according to the standard 10/20 EEG electrode system, and held in place using a Biosemi electrode cap of appropriate size for the participant. Triggers generated from the experimental code were recorded on the EEG device to allow timing synchronisation. All electrode offsets were below 20 mV. The six support electrodes were placed on two mastoids, two temples, and above and below the right eye, for better EOG detection. The experiment took place in a shielded experimental room.

2.6. EEG data processing and ICA

The processing of the raw EEG data was performed using custom code that included use of standard EEG processing functions from the EEGLAB v12 toolbox (Delorme and Makeig, 2004) in MATLAB. Biosemi data was loaded using left mastoid reference, and re-referenced to an average reference later (Luck, 2005). A Hamming-windowed FIR band-pass filter of 0.1–80 Hz was applied, using 'eegfiltnew' in EEGLAB v12. The 1024 Hz recording was down-sampled to a sample rate of 256 Hz using EEGLAB functions.

Data rejection was performed using standard EEGLAB functions. Noise of four times that of the median was chosen as a threshold. Electrodes with values exceeding this criterion were removed from subsequent analysis. This resulted in a median on 20 channels rejected (range: 11–22 for the seven subjects), leaving a median of 49 electrodes. Most of the head remained well represented as no large areas were left without any electrode coverage.

Independent component analysis (ICA) was applied to the whole dataset from each experimental session, after artefact channel rejection, using the Infomax ICA algorithm (Delorme et al., 2007). Infomax ICA here returns one IC for each electrode, giving a median of 49 ICs for each subject. The timing triggers were used to label the time-points in which each image was presented.

After rejection and ICA, but before classifier training, each data vector was normalised by subtracting the mean of each channel and dividing data by it's standard deviation. This scales the data for the SVM classifiers.

2.7. Classifier training

Separate classifiers were used for each subject. In the training phase, a 'one-versus-one' SVM model was trained for each of the 50 objects presented, where the labels were 'object appears on screen' or 'no object present'. The classification task was to best apply this label appropriately to subsequent data, given this training data. That is, rather than using a single multi-class classifier that would give output of 1 of 50 labels, we used 50 binary classifiers each labelling 'object stimuli present' or 'object absent' for their respective object stimuli. Both the voltage time-points from EEG and the independent component (IC) transform activations were normalised (Fig. 2). The kernel used with the SVM was a radial basis function (Keerthi and Lin, 2003) from the Matlab implementation of libSVM (Chang and Lin, 2011, software available at the libSVM website).

To find appropriate values for the SVM hard margin training error cost 'C' and the radial basis function kernel parameter 'γ', we used the libSVM parameter sweep tool. A grid search was performed on 18 parameter values between $C = [10^{-2}$ to $10^{10}]$ and $\gamma = [10^{-4}$ to $1]$ on data from two subjects. This suggested values of $C = 1$ and $\gamma = 1/$number features, and those parameter values were used for all models.

2.8. Classifier input

Each object had five presentation trials. Classifiers were trained using four of these and tested using their accuracy at predicting the appearance of each object on the fifth trial.

The training data was EEG data beginning at the time of object presentation until some time after (initially 500 ms). For each object, a 'positive training data' label was applied to data from trials where that object was shown. Resting baseline EEG data, in which no visual stimuli was displayed (blank screen), was termed 'negative training data', where the correct label is 'object absent'. These training data were labelled as '1' and '0' respectively when used with our classifier.

For each object classifier, positive 'object present' input data came from the 0.5 s after each presentation of that object. As we used four trials, with each trial giving 0.5 s of data in which an object is present, we acquired 2 s of image-present data for each object. As a sample rate of 256 Hz was used, this gave a positive example training vector of 512 data points × number of channels.

For the negative data, 'object absent' input data came from 19.5 s of randomly selected intertrial data. Classifiers were trained using a proportion of approximately 10% 'object present' data and 90% 'object absent' data – 2 s of data to 19.5 s of data. Test data also used this approximate proportion of data.

Test data came from the one presentation trial of each object that was not used for training (rotated as explained below) along with further negative examples of random intertrial data in which no object was presented.

2.9. Cross-validation

To improve both the robustness of the classifiers and their ability to generalise to new data, five-fold (leave-one-out) cross-validation was used (Efron and Gong, 1983). This process reduces the likelihood of erroneous results, as multiple splits of the data are considered. Cross-validation was performed on data from each object by dividing the data into five splits, with a single object presentation trial in each split. We then iterated through five separate SVM models – each training on four of five trials. The remaining fifth trial was then used as a blind test. As we iterated through the cross-validation, each trial was used once as a test data.

Classification test results then came from the accuracy on classifying the respective unseen fifth, and the score was averaged across the five splits. All classification results reported are from this average of all five cross-validation splits.

2.10. Performance metrics

The success of a classifier can be given simply as percentage correctly classified. This can be valid in many contexts, but does not clearly show that performance depends on both 'sensitivity' (true positives) and 'specificity' (true negatives). A receiver–operator characteristic (ROC) plot (Mason and Graham, 2002; Hand, 2009) illustrates both sensitivity and specificity – with the area under the curve (AUC) of the ROC of 0.5 signifying random chance prediction and one being perfect prediction. This was relevant here as over 90% of data in both training and test situations belonged to the 'object absent' class. If a classifier were to predict 'object absent' everywhere, it might get 90% accuracy despite conveying no useful information. AUC, however, would correctly score that classification as no-better-than-chance performance.

Plotting a ROC curve can be particularly useful when sensitivity and specificity are being manipulated separately, but here we simply use area under this curve (AUC) as a concise metric of both classifier sensitivity and specificity.

2.11. ERP waveforms

After data processing, we confirmed that our visual stimuli presentation elicited a visually evoked potential change. EEG trace data was visually inspected (Fig. 3) to check recording quality for EEG artefacts. Fig. 3 (top) shows EEG trace from one trial on 10 EEG channels. No obvious EEG artefacts are present. The thin blue line around 1.1 and 4.6 s into the recording indicates presentation times of two objects, with the times labelled indicated with the red bars.

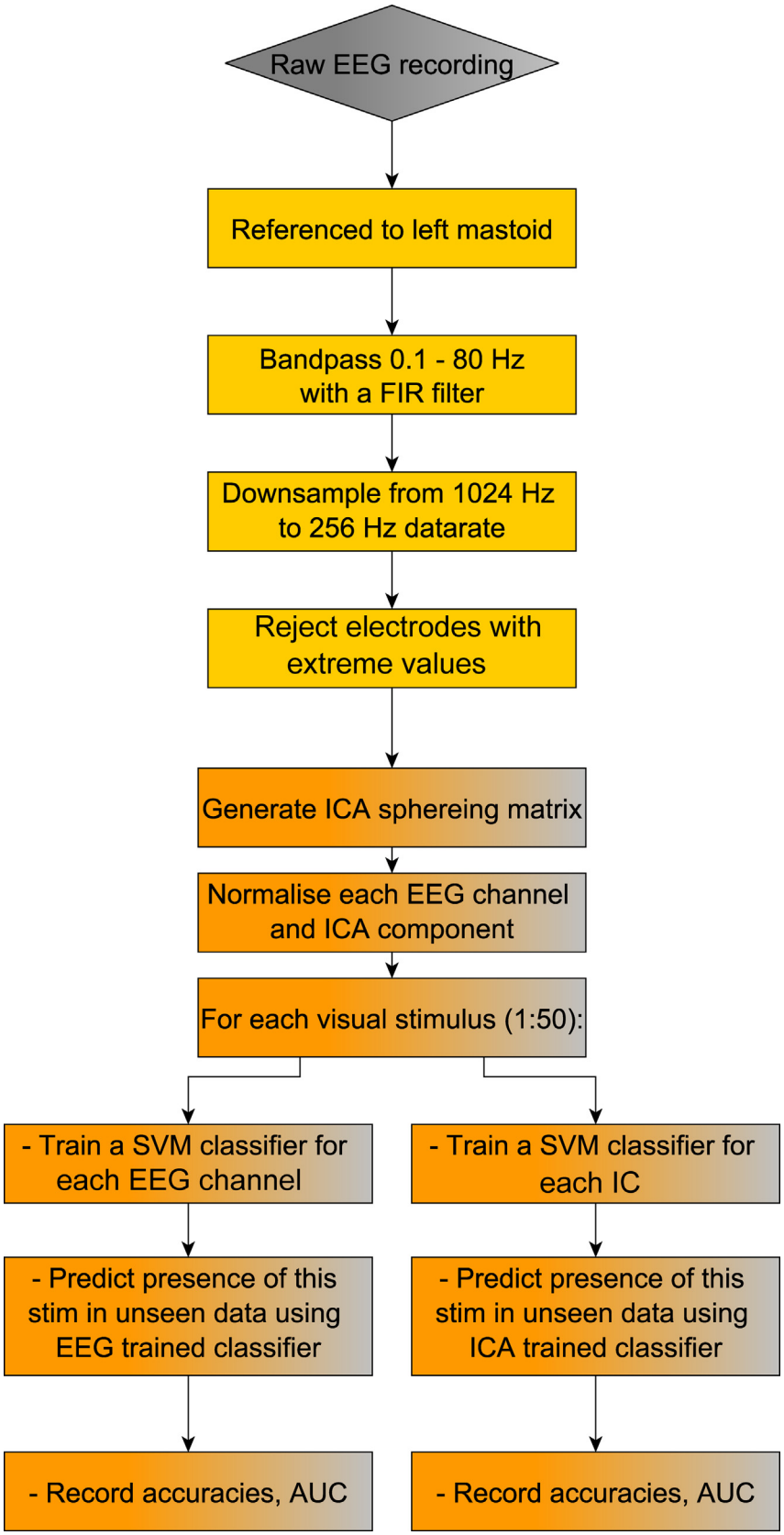

Fig. 2. Data flow for EEG processing. Data processing steps are shown in yellow and machine learning steps in orange. The input here was the scalp EEG recording, which was referenced to the 'mastoid' electrodes behind the left ear. A FIR filter was used to reduce signal outside 0.1–80 Hz. The sample rate was then reduced from 1024 to 256 Hz to speed subsequent analysis, and electrodes with very high electrodes noise were rejected. The ICA was run and SVM classifiers were trained for all objects using EEG channels data (left) and IC activation data (right).

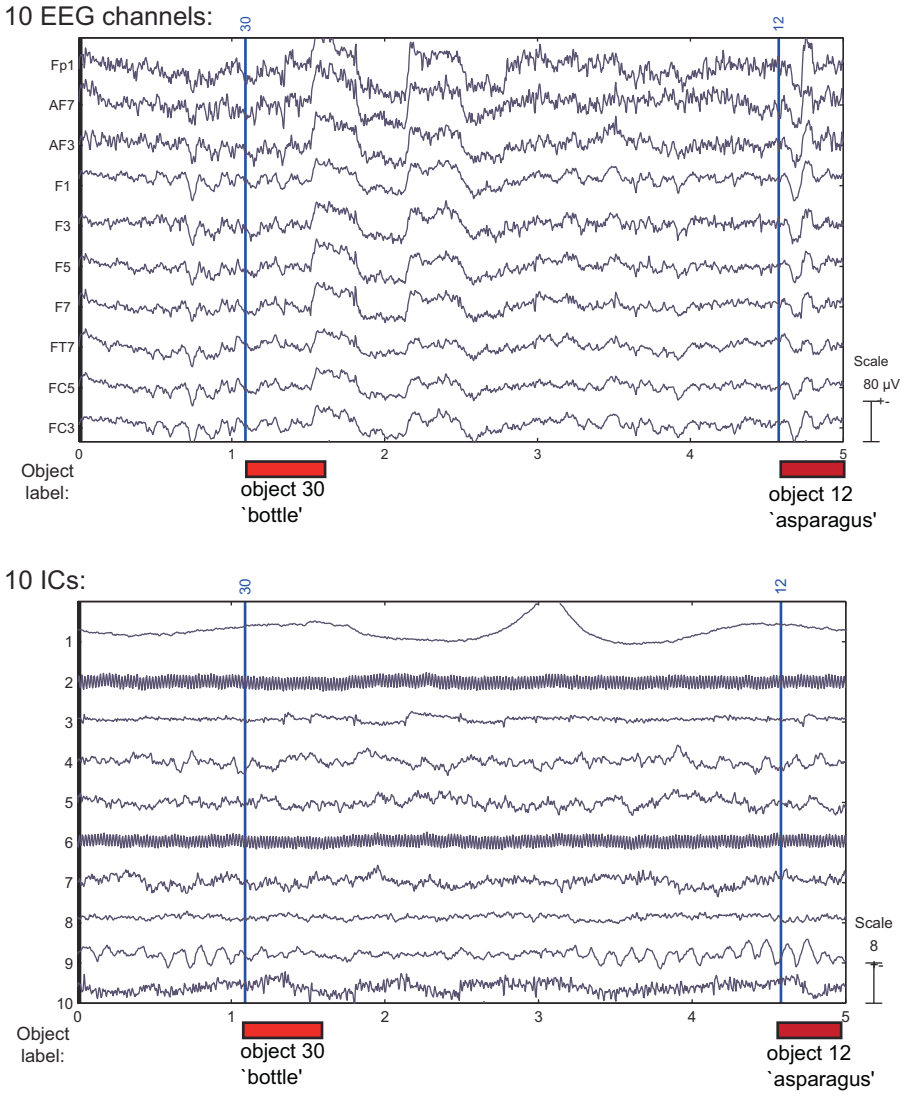

Fig. 3. EEG trace and IC activations in an example trial. The top panel displays example EEG trace data from 10 channels for 5 s around the first visual object presentation (subject 1, trial 1). The lower panel shows the data from the same period in data from 10 ICs. The blue vertical lines show object presentation time (object 30, 'bottle', shown at around 1 s and object 12, 'asparagus' shown at 4.6 s) and the red bars indicates the approximate time labelled as 'object present' used for training classifiers. The intertrial time was labelled as 'resting baseline' (no object present). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Results

We present results from classifying new EEG as either 'object present' or 'object absent', based on SVM models built on prior EEG data. We first report results using 500 ms of data and from classifiers using all EEG data, classifiers using each of the individual channels of EEG data, and classifiers where only one selected channel is used. We repeat this with IC data.

In addition, we report details and dynamics of high-accuracy ICs. Furthermore, we consider classification from EEG channels in which artefactual ICs are removed. Timecourse analysis is examined in classification on single EEG channels and IC data in 500 and 50 ms 'sliding windows'.

3.1. Classification of visual trials from EEG data

Training SVM classifiers based on a data vector of all channels simultaneously (median number of 49 channels) appeared to give low accuracy (around 0.51 AUC). When training classifiers on data vectors from a single individual EEG channel, mean accuracy was found to be 0.51 AUC (see Fig. 4, leftmost bar).

3.2. Finding a single selected EEG channel/IC

The analyses here considers EEG channels and ICs that gave the highest accuracy. A small subset of channels consistently gave much higher classification accuracy.

To objectively identify these, we used an automated selection procedure, based on data separate from the test data. These high accuracy channels can be selected automatically by further partitioning the data (prior to testing) in a training and validation set. In this way, the generalisation ability of each individual channel can be assessed on the validation set, and then confirmed (or not) on the test set. We refer to these as 'selected channels'.

Specifically, data from 10 of the 50 objects was separated. On this data, classifiers were trained using data from each individual channel. That is, around 49 separate classifiers were trained for each object, using data from each of the data channels (median number of 49). The single data channel that gave highest AUC on these

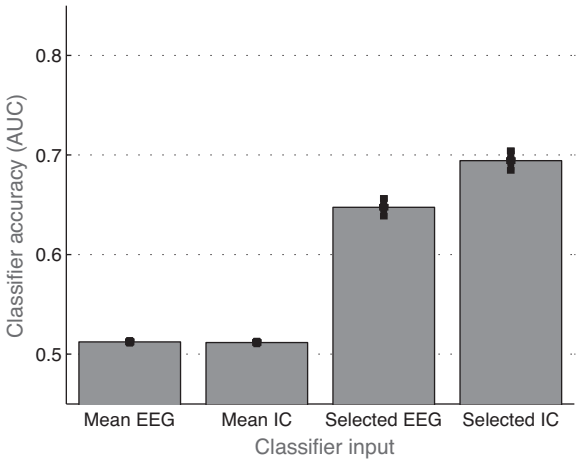

Fig. 4. Mean classifier performance across all subjects and objects using EEG (mean of all channels), ICA (mean of all channels), and the selected single channel for EEG and ICA. Classification performance is measured in successful classifications area under a ROC curve. The error bars show the standard error of the mean. With 50 objects shown five times each to the seven subjects, and models trained and tested for each scalp electrodes or IC, the mean is an average of thousands of classifications. The 'selected' EEG channel or IC was a single channel selected on the basis of performance on separate data, with the selected IC here giving 0.7 AUC.

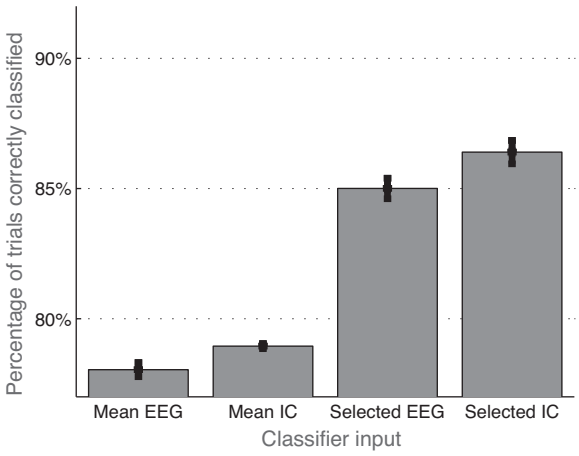

Fig. 5. Mean classifier performance in terms of percentage of test single-trials correctly labelled. The error bars show the standard error of the mean. The selected IC gave 87% of single-trials classified correctly, across all subjects and objects, which corresponds to 0.7 AUC.

first 10 objects was then used as the single 'selected' channel. The selected channel was used as the best input to classifiers for training SVMs on the remaining 40 objects. Note that this data selection procedure was performed once on each participant for both EEG and IC data (see below).

In the case of using scalp EEG data, this selection of one channel input increased classifier performance from 0.51 AUC mean to 0.65 AUC (Fig. 4, bars 1 and 3). We show the results of selected EEG channel for each subject in Table 1. Percentage of trials correctly classified is shown in Fig. 5.

3.3. Classification of visual trials using ICA data

We wished to assess the classification performance on EEG data when transformed with ICA into single ICs. This IC activation data was acquired by running ICA on each participant's entire EEG recording.

As with the mean of all scalp EEG channels, the mean of ICs was low, around 0.51 AUC. Repeating the same 'selection procedure' as in EEG data above, we found single ICs gave greatly increased accuracy of 0.70 AUC (see Fig. 4).

Classification accuracy was significantly higher when using the selected IC as opposed to selected scalp EEG channel using a Wilcoxon signed rank test ($p < 0.0001$, means 0.70 AUC and 0.65 AUC).

These results can also be summarised in terms of mean percentage of trials correctly classified (Fig. 5). We find that these agree with the AUC reported in Fig. 4, with the selected IC giving highest raw average classifier performance of 87% trials correctly classified.

The above results demonstrate high accuracy (0.70 AUC, 87%) at correctly labelling 500 ms of EEG data from single ICs with 'object present' or 'object absent'. We can now consider the properties of these high accuracy ICs, examine their accuracy over time, and why they might offer higher accuracy.

Table 1
Accuracy of visual object classification (AUC) on each subject, using selected classifier input data.

Input	EEG selection (AUC)	IC selection
Subject 1	0.61	0.61
Subject 2	0.54	0.62
Subject 3	0.53	0.71
Subject 4	0.56	0.54
Subject 5	0.74	0.78
Subject 6	0.82	0.91
Subject 7	0.73	0.68
Mean	0.65	0.70

3.4. Properties of the selected EEG channels

As shown in Fig. 6, the selected EEG channel was different in each subject, although mostly occipital (O1, O2) or parietal (P3, P5) electrodes. On presentation of visual object stimuli, we found the mean grand-average ERP (Fig. 6) with a prominent positive voltage deflection around 100–200 ms after presentation on central and parietal electrodes. This is in agreement with ERPs observed elsewhere – e.g. visual object presentation in Rousselet et al., 2007.

3.5. Properties of the selected ICs

Component activations of the selected ICs in each subject are shown in the lower section of Fig. 6, along with accuracy of each IC. A 'topoplot' shows the spatial distribution of each IC on a 2D headmap. The majority (5/7) of the ICs appear concentrated in occipital regions. All those ICs with accuracy of above 0.62 AUC were found to be in this region.

Extended details of the properties of a selected IC and a low predictive power component are compared in Fig. 7. The IC activations were mapped to spatial locations on the head (Delorme and Makeig, 2004). This mapping is shown on a standard 3D head model, for both a high and a low predictive IC. Also shown is the average ERP and the trial-by-trial 'ERP image' heat-map. The ERP image is a coloured plot in which every horizontal line shows activity from a single trial (Makeig et al., 2004). In this way, the variation in individual trials can be shown alongside the averaged ERP. Here, IC 3 from subject 6 is the 'high predictive' IC, and has a complex activation pattern over time, but seems to have a 12 Hz oscillation that flattens around 100 ms after object presentation. In contrast, the activation of the 'low predictive' IC 14 is not consistent across trials and does not have consistent timing in relation to the stimulus appearing, and appears to be an irrelevant frontal IC source.

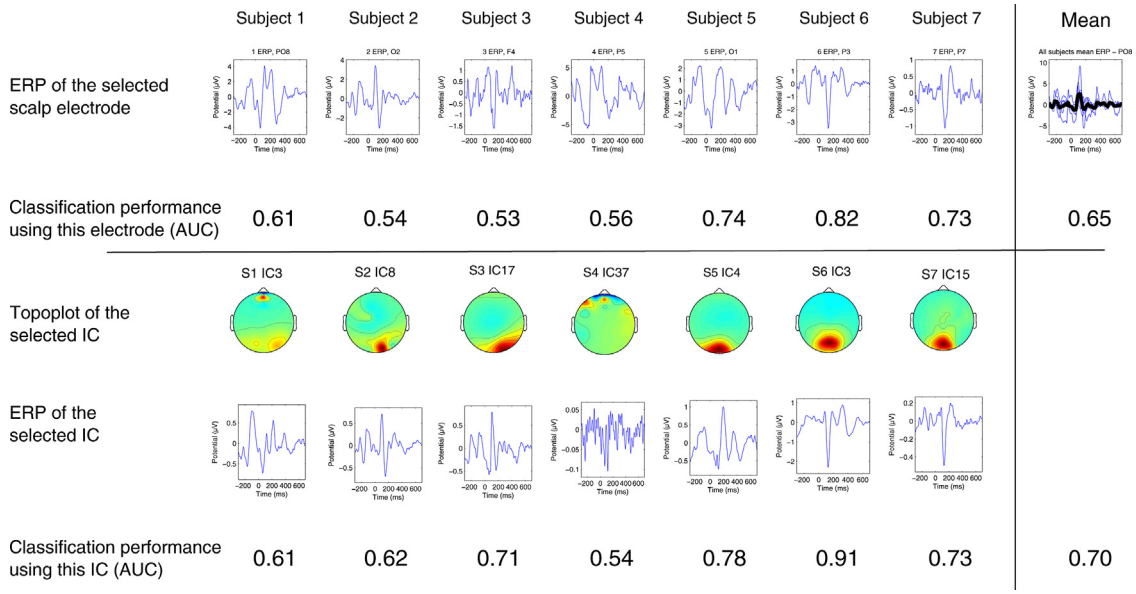

Fig. 6. Top Scalp EEG ERPs from the selected electrodes in each of the seven subjects, along with the object classification performance when using data from that electrode. Thus, electrode PO8 is the selected electrode in subject 1, and when data from PO8 is used as input for SVM classifiers, object presence is correctly identified at 0.61 AUC. In subject 6, P3 is found to be the better electrode and it gives 0.82 AUC in that subject. **Bottom** The lower section shows the selected IC for each subject. The 2D headplot (using EEGLAB's 'topoplot') shows the spatial distribution of each of the selected ICs. Note that all ICs above 0.62 AUC (5 of 7) appear in the occipital region. The IC ERP dynamics are shown to indicate the mean activity of a given IC across all trials. The figure also displays the classification performance from Table 1, with the selected IC in subject 6 having the highest AUC (0.91).

3.6. Scalp EEG data with IC artefacts removed

The above results suggest improved classification when using data from activations of components identified by ICA with the SVM classifiers rather than using scalp EEG data, at least when a selected higher accuracy data channel is selected. A possible explanation for this might be that relevant neural signals are overwhelmed by electrical artefacts that are not relevant to the visual state and these are present on all electrodes. Then, when ICA is run, the contributions of those artefactual signals are 'captured' in single ICs. In order to assess this, we now remove the contribution of each of those probable artefactual ICs from the scalp EEG data and test the visual state prediction again. This effectively removes the contribution of these ICs from the EEG data. The 'autorej' function of EEGLAB v12 was used to identify ICs with extreme values. Each component with activity above a rejection threshold of values ±50 standard deviation was tagged as a possible artefact and the contribution it gave to that subject's EEG was removed. This resulted in a mean of 4.5 ICs being removed (range of 1–12). Visual classification using this IC-cleaned scalp EEG data is shown in Fig. 8. Using ICA to remove the contribution of obvious artefacts from the scalp EEG data improved the accuracy on this 'IC cleaned scalp EEG' over that from scalp EEG alone ($p < 0.0001$, on means of 0.68 AUC and 0.65 AUC). Mean classification from a single selected ICA channel remained higher than IC cleaned scalp EEG data (0.70 AUC to 0.68 AUC), but this was not significant ($p = 0.1$). From this, we conclude that selected ICA data or selected IC cleaned scalp EEG data gives better accuracy in classifying vision from EEG data using this kind of SVM protocol.

3.7. Assessing accuracy over time with a sliding analysis window

For the previous results, classification was performed using data from the first 500 ms after each object was shown – from $t = 0$ ms to $t = +500$ ms, relative to the object appearing on the screen. In order to compare data within this time period, a 'sliding window' approach (Maris and Oostenveld, 2007) was then used, with windows of 500 and 50 ms. We use this sliding window approach to test the sensitivity of our approach to both the start of the data window, and the amplitude of the data window.

This sliding window protocol was performed in two conditions – first with the length of the analysis window remaining fixed at 500 ms, as before, and also with a narrower temporal window of 50 ms. At a window length of 500 ms, classification performance of scalp EEG data and IC data could be assessed when 0–500, 25–525, 50–550 ms, and so on up to 500–1000 ms, were used as positive training data. This data is shown in Fig. 9. Data from the one selected IC (black circles) gave highest accuracy at 0.72 AUC when using data from the interval within 25–525 ms. The one selected IC gave classification performance higher than other data sources at all time window intervals, greatly so until 100–600 ms, although performance drops steadily to around 0.54 AUC at 400–900 ms. Data from selected EEG channel follows this classification performance, but at around 0.1 AUC lower than IC data at 25–525 ms.

A smaller analysis window allows more precise timing resolution. At a window length of 50 ms, classification performance of scalp EEG data and IC data could be assessed when 0–50, 25–75, 50–100 ms, and so on up to 500–550 ms, were used as positive training data. This is shown in Fig. 10. Here, peak accuracy is no longer in the earliest intervals after the object appears, but instead in the slightly later intervals of 75–125 or 100–150 ms. Selected IC data has slightly higher average accuracy than selected EEG channel data. Selected EEG data has as peak accuracy at 0.65 AUC at the 100–150 ms interval. With training data consisting of 20 single-trials of just 50 ms, we find above-chance classification and the accuracy is considerably higher using data from 75 to 175 ms into the trial than at other intervals.

With only a tenth of the data (that from 50 ms rather than 500 ms), peak classification accuracy does drop from 0.72 AUC using 25–525 ms to 0.65 AUC using 100–150 ms. These data support conclusions that much of the task related EEG variance occurs in the 100–150 ms time interval, as models trained on that period have 0.65 AUC object classification accuracy. This is in agreement with related work showing a peak in 'decodability' at this time interval (Carlson et al., 2013), who also report this in

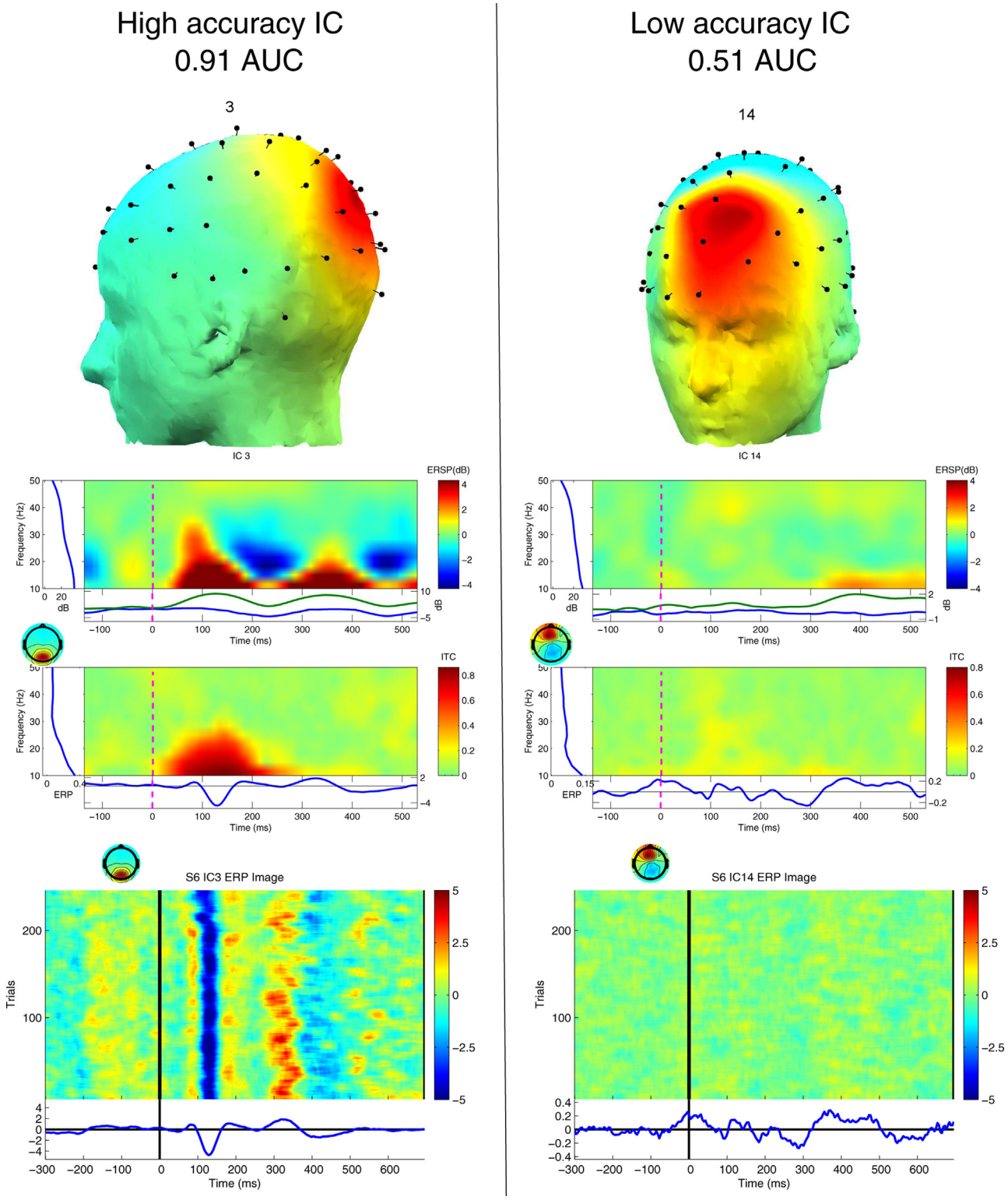

Fig. 7. Individual IC activation details. Activations of two example ICs in subject 6 – IC 3 (left) and IC 14 (right). **Top** standard 3D head models with EEG electrode positions as black pins and colours showing the spatial distribution of two ICs. In this example, IC 3 is an occipital component and IC 14 is a frontal component. **Mid** EEGLAB displays of frequency dynamics (event-related spectral perturbation) and plots of inter-trial coherence. **Bottom** corresponding IC ERP-image plots showing the 250 individual object presentation trials, with red indicating higher activation of this IC at that time. The lower plot is the standard IC ERP averaged over 250 trials. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

MEG object category discrimination. The results here agree with Carlson et al., 2013 in that object decoding performance peaks at around 100–150 ms, and that even later data from 375–425 ms gives higher object decoding performance than the early 25–75 ms interval. The sliding window data above shows average classification performance across all seven subjects, using data from different EEG channels or ICs. Sliding window results separated for each subject and data source are shown for three random subjects

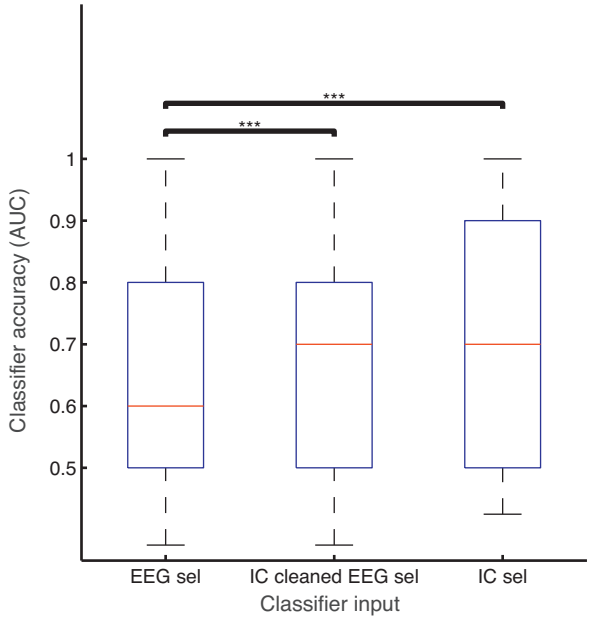

Fig. 8. Trial classification performance using EEG data, EEG data in which noisy ICs have been removed, and IC data. The box plot shows median accuracy of each data source in red, with the edges of the blue box indicating 25th and 75th percentiles. Asterisks (***) indicate $p < 0.0001$, as found using a Wilcoxon signed-rank test. An average of 4.5 (of around 49 ICs total) were removed for the 'IC cleaned EEG data'. The increase in the median accuracy in the IC-cleaned EEG data suggests that these few artefactual ICs hindered ability to classify trials from scalp EEG data. The single selected IC is 'isolated' from the contribution from the artefactual IC noise. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

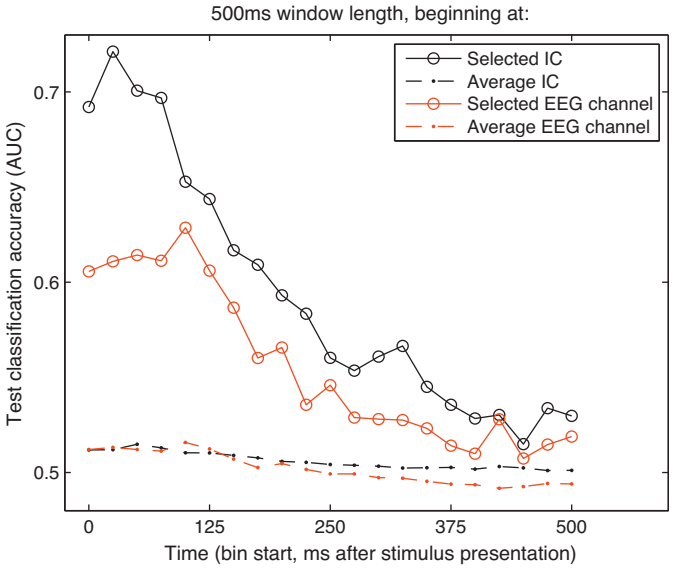

Fig. 9. A 'sliding window' analysis, to assess accuracy over trial time. Accuracy using data from 21 different time windows were used, all with length of 500 ms, first starting at 0–500 ms, and stepping through the trial times in 25 ms increments. Peak accuracy occurs here when using data from the selected IC (black circles), and models are trained to classify based on data from 25 to 525 ms after the object appears on screen, with 0.72 AUC. At the later time bin starting at 325 ms, the selected IC drops to 0.56 AUC, suggesting less trial-predictive information there, but is still above chance.

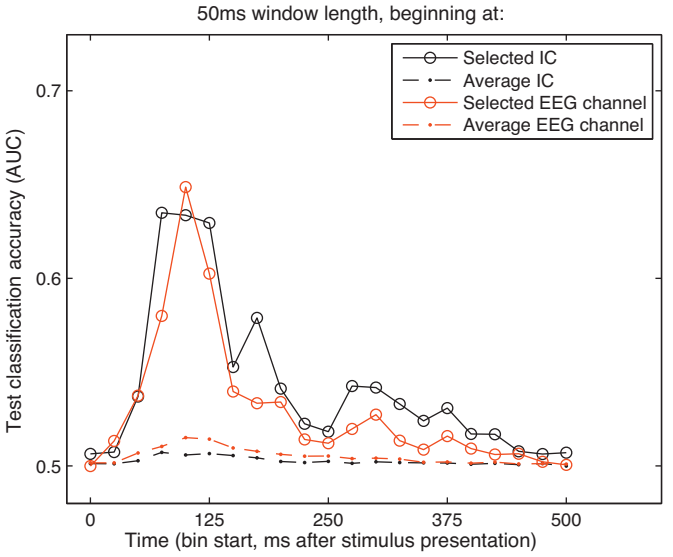

Fig. 10. The 'sliding window' is repeated, with a window length reduced to 50 ms of data is used for training models rather than 500 ms. We can see that data from 75 to 125 ms is sufficient to give 0.64 AUC accuracy. Accuracy falls off sharply around this period. Data from 0–50 ms and 25–75 ms (the first two time bins) are all at chance, suggesting no useful information for our classifiers at this time. Data from after 400 to 450 ms are also back at chance.

in Fig. 11. For each subject, two plots show colour-coded AUC for each electrode and each IC over the different latencies. High accuracy is concentrated in a few ICs, and those time bins starting before 150 ms appear to give higher accuracy.

3.8. Classification performance on each image

We report the success rate of classifying each object in Fig. 12. This shows the classification performance on each of the 50 objects used as visual stimuli using the selected EEG and IC data, averaged across the seven subjects. Highest accuracy was obtained when using the selected IC data on the 'log.jpg' with 0.89 AUC.

4. Discussion

We detail a method for classifying visual state from EEG using machine learning. We demonstrate that this can distinguish data from visual object presentation trials from data without object presentation at 87% accuracy.

We found single-trial visual classification accuracy well above chance when using data from a single selected EEG channel or IC, with selected IC giving higher accuracy.

Further, we report single channel accuracy at 75–125 ms gives higher accuracy that other time bins, when classifying with these selected ICs (Fig. 10). This scoring of classifier accuracy automatically identifies electrodes, ICs, and time periods with activity that may be more relevant for the given trial.

4.1. Improving accuracy with selected input

The finding that input of a subset of a single selected data channel outperformed using all channels together deserves consideration. This may be due to classifier 'overfitting' – where the model parameters fit the properties of the training data too rigidly and so do not best generalise to classifying new data (Babyak, 2004). The overfitting when using all data could be due to suboptimal model parameters of the training error cost parameter 'C' and the radial-basis function kernel parameter 'γ'. However, these were the best values provided by a parameter search.

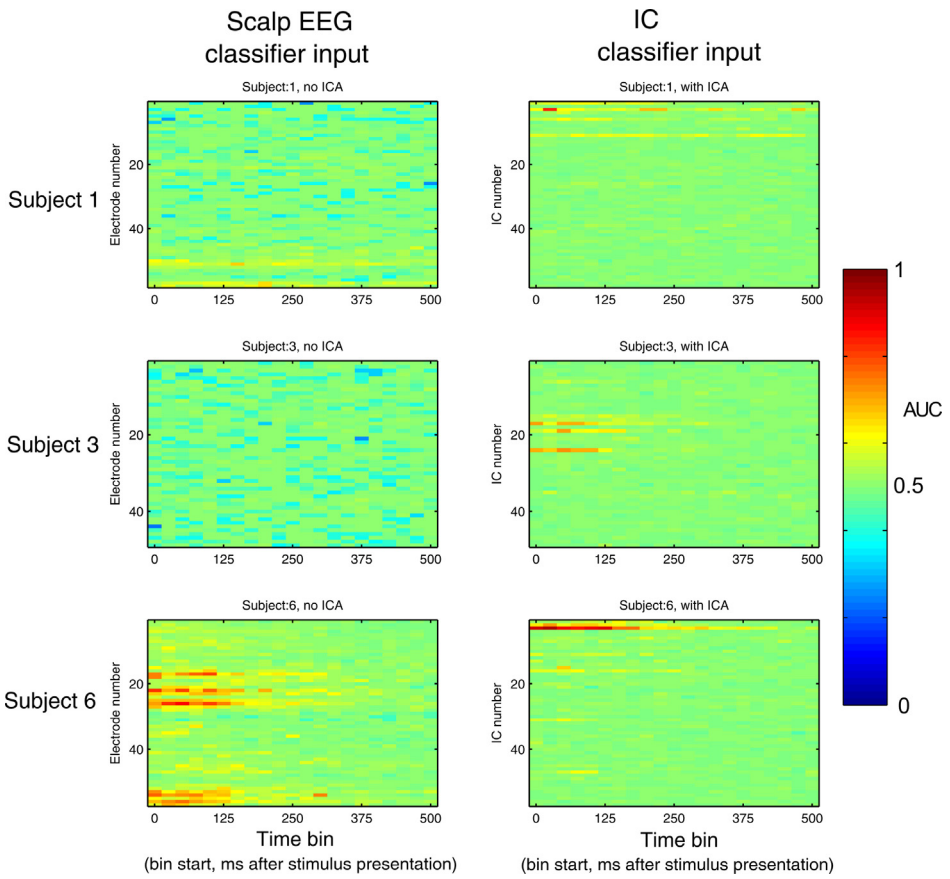

Fig. 11. A heatmap plot of the classifier performance (AUC) for a random three subjects using every electrode (left) and IC (right) as classifier input data, when using data from different timepoints in the trial. Outside the first 150 ms, accuracy is low. Peak accuracy is higher in IC data, with our classifier better able to separate classes in that data type. (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

Generalisation was improved by using appropriate dimensions of the classifier input data. Using one selected channel of input data rather than all channels or typical single channel gave increased mean accuracy. Notably, this considerably increased reliability (from 0.51 AUC to 0.65 AUC), with classification accuracy well above chance. This could be as a result of our implementation, but does indeed still show that a single selected channel of EEG input data can be a consistent and concise source of input data for high-accuracy classification.

4.2. The use of ICA

When classifying presence of visual images, using classifier input of the IC activations from a single selected IC gave highest accuracy in all seven subjects at 500 ms window. Although the average IC input classification was approximately equal to the average single EEG channel input, those ICs that performed best on the initial 10 objects continued to outperform EEG channel input in further test data (0.70 AUC using selected IC data, compared to 0.65 AUC in the selected EEG channel). This corresponds to a significant increase in timepoints correctly labelled.

In trying to improve performance of neuroimaging classifiers, there are two domains commonly focussed upon: feature extraction and the classification itself (Farquhar and Hill, 2006). EEG data is intrinsically noisy, contains highly correlated features and has much variance both between different subjects and within the same subject over time. This suggests that considering decomposition of possible EEG sources could be useful when assessing features to extract from EEG (van Gerven et al., 2009). ICA does this by decorrelating the inputs and attempting to minimise mutual information in forming the components.

In previous studies, ICA has been shown to improve performance of a classifier in a simple auditory task (Hill et al., 2004). To our knowledge, the present report is the first to document ICA giving such an improvement in a visual object classification task.

For five of seven subjects, the single selected IC was found in occipital regions (Fig. 6). As expected, artefact related ICs had low predictive power to discriminate stimuli. Several of the components seemed to be related to eye movements, muscle artefacts or electrical noise, as identified by back-projected location and activation properties (Delorme et al., 2007). Nonetheless, we have shown here that several ICA components have a greater degree of information content that can be used to predict perceptual processes than unprocessed EEG data had.

ICA appeared to separate sources of noise into some ICs, and sources of task-related neural activity into other ICs. This resulted in the single-trial classification performance in many ICs being low, but a few ICs giving higher accuracy than that of any EEG channel. Thus, ICA may be thought of as concentrating source signals into denoised ICs (as shown in the fewer, but more peaky, bands of yellow and red in Fig. 11), and so our SVM classifiers can give higher accuracy using this input. Examining IC input also has the advantage of assessing a source-space deconvolution (Bell and Sejnowski, 1995).

This view was also supported by an improvement in trial classification accuracy from EEG data where probable-artefacts have been removed (see Fig. 8).

Within subjects, many EEG channels had relatively similar predictive power. While the magnitude of the averaged ERP might be quite different on electrodes across the scalp, at a single-trial

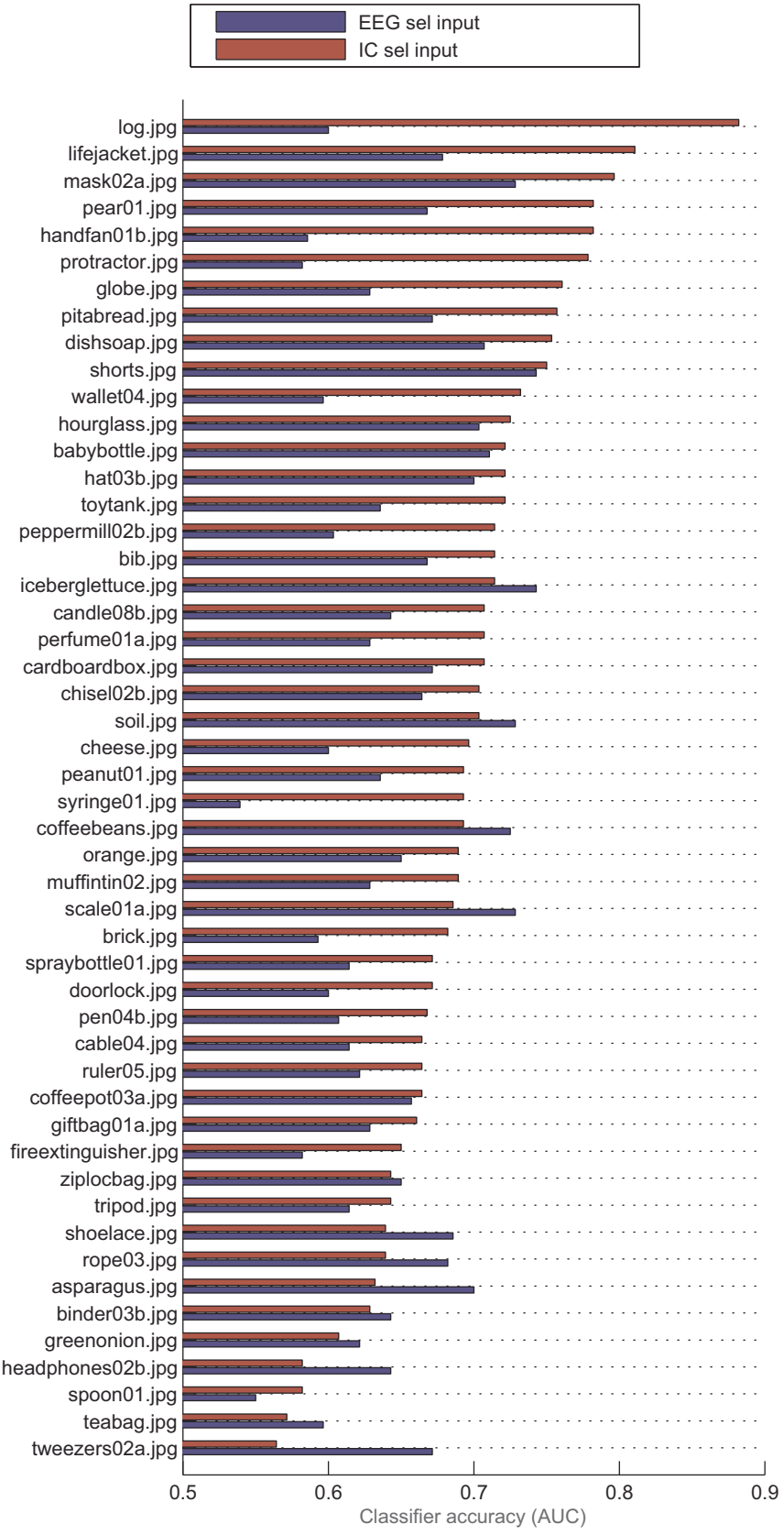

Fig. 12. Classification performance shown by each object of our 50 visual object stimuli, sorted by IC AUC. This is averaged across the seven subjects, using the selected IC data (shown in dark red) and selected EEG channel data (dark blue). Single-trials with pictures of the log, life-jacket and mask were labelled correctly at high accuracy, and the IC outperformed scalp electrode data for most objects. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

level much of the signal present at any electrode is also present on neighbouring electrodes (Luck, 2005) albeit at a different scale. As the SVM classifier here is separating classes in a high-dimensional hyperplane, it can remain sensitive to small (but consistent) separating features. Thus, features that would not be obviously visually apparent on a plotted ERP can still be captured.

We suggest that the task predictive power of an IC has the potential to inform theories of visual cognition. For example, an IC that is localised in neck muscles and also gives low accuracy when used to classify the task, would be an unlikely target for further investigation. In contrast, an IC that is spatially localised in visual cortex and that gives high accuracy of classifying the visual state under specific conditions may be a fruitful target to profile.

4.3. Visual object processing

Although EEG studies have shown stimuli-specific ERP separation within 150 ms (Thorpe et al., 1996; Johnson and Olshausen, 2005; Mouchetant-Rostaing and Giard, 2000), there appears to be some uncertainty whether these differences represent low-level visual properties or higher-level cognitive categorisation of visual stimuli. Related work in macaque monkeys using multi-unit neuron cell recordings seems to indicate clear object- and category-specific information in primate inferior temporal cortex at 100–125 ms after presentation (Hung et al., 2005). Cell firing was used as input for a classifier to correctly identify the presented object stimulus at 70% accuracy using only the spiking behaviour of a few neurons in inferior temporal cortex.

This neuron spiking data strongly indicates that the brain activity is present at these time-scales, but that, of course, does not necessitate that we could also observe this within EEG data, which would allow more probing in humans. Clearly, the ability of averaged ERP EEG analysis alone is limited in addressing this (Makeig et al., 2004; Rousselet and Pernet, 2011). We suggest that our approach of ICA and SVM machine learning is a more suitable one.

4.4. Timing of accurate trial classification throughout a trial

With use of the sliding window moving though data from different periods of the time into a trial (Figs. 9 and 10), we found that data from the time period at 75–150 ms was most useful for classifier performance. This is in agreement with studies of ERPs in visual object presentation (Thorpe et al., 1996), where the difference in the ERP from different objects is small before 75 ms and after 350 ms.

This profile of accuracy using these selected occipital ICs over time is in agreement with related studies of early visual object processing listed above, with the current procedure automatically finding these task-relevant ICs.

We suggest that this reveals how the timing of relevance of the information contained in different EEG channels, processing methods and IC sources can be assessed by how well they can be used to predict experimental conditions with these kinds of SVM classifiers.

4.5. Cognitive implications and future work

The identification of EEG sources that have robust task-related predictive power would allow cognitive science experimental designs that target that specific IC. Getting a profile of IC localisation, task dependence, temporal and spectral properties may be much more accessible in probing possible underlying neural processes in greater detail than using EEG data alone. Observing increased accuracy when using IC activation input suggests that they are a promising target, although the extent to which specific ICs might relate to underlying neural processes is not yet well known (Onton et al., 2006).

While standard ERP components such as the N170 and P300 are frequently reported landmarks in EEG data for visual perception (Joyce and Rossion, 2005) and oddball responses (Polich and Kok, 1995), such ERPs are limited in being averages of electrode traces, and so may lose useful information (Rousselet and Pernet, 2011).

Other components have been suggested as descriptions for EEG data. Philiastides and Sajda, 2006 report their generation of components to best decode trials in which either cars or faces are shown, and find one such component that resembles properties of the N170. This demonstrates an alternate description that may underly these ERP components, and so give another way of studying them.

In the current study, we also report components that give high decoding performance. Here ICA was used to generate the ICs, and so represents possible sources of generated EEG activity, obtained blindly from the data. The automated scoring procedure we used identified a group of occipital IC components that represent the best data sources for classifying visual object presence in single trials.

Other studies of visual perception have considered the similarity of fMRI activity in inferior temporal cortex in response to objects of different categories (Kriegeskorte et al., 2008). The current study used 50 images from the BOSS image database (Brodeur et al., 2010), all of which are likely to be considered 'natural objects' or 'artificial objects' in category. We studied a range of single images and did not examine object categories, but future studies could target discrimination of categories with relevant training stimuli.

Fast, high accuracy classification of visual state also allows real-time detection within a recording session, and so experiments involving EEG feedback or manipulation could be performed. Do any ICs have activations associated with borderline visual percepts, or is the task-related representation we see here downstream of an all-or-nothing percept trigger?

These components – along with our SVM prediction models – can classify new trials of visual stimulation well, but could likely be used to probe other steps of visual processing and perception. We suggest that profiling the activity of these components identified by ICA in different tasks, quantified with machine learning, might be fruitful conceptual targets for future visual processing experiments. Might any of their classifiers respond to imagined objects rather than presented objects? Given several objects presented simultaneously to a subject, might some classifiers respond to the attention of the subject and some respond only to the immediate visual input, regardless of context? With a visual task of noisy stimuli where object perception is only sometimes reported, will some classifiers predict onset of perception even before perception occurs?

4.6. Limitations

The metric used for 'prediction power' gives an indication of task-related information, but it will also have some dependencies on the properties of the classifier used (Meyer et al., 2003). Nonetheless, the prediction power of those ICs that likely capture artefactual signal is low, suggesting that task prediction power can be a valid proxy for task-related information content.

Models using data from all channels simultaneously (as in Section 3.1) gave low average accuracy. This suggests this classifier model was not fitting the signal-to-noise of this data well, and overfitting (Chang and Lin, 2011). Using one selected 'best' IC resolved this, but at the expense of not utilising all available input data.

It may be the case that ICA was particularly useful here as little other preprocessing was used. Prior research suggests that use of tailored data extraction can increase accuracy on classifying EEG data (Blankertz et al., 2006), where frequencies of interest, channels of interest and input ranges are manually specified. We suggest that in cases where the anticipated activity is relatively unknown, use of the more automated ICA might avoid misuse of tailored data extraction when it is not appropriate.

We considered classifying the presence or absence of single images. It could be possible to extend this to instead classify each image against the other images: as 'image1 vs image2' rather than 'image1-present' versus 'image1-not-present', or classifying multiple images simultaneously through use of a multi-class classifier, but we have not addressed this here.

4.7. Conclusion

We presented a method for automatically scoring task-related information present within EEG transforms using SVM classifiers trained on that data. We found 'object presence' was classified at 0.70 AUC (87%) when using data from a single selected IC. Further, the use of a sliding window analysis revealed that the time window spanning 75–150 ms gave the highest accuracy, when using these selected ICs. We suggest that this method of machine learning and independent 'source separation' might allow detailed probing of information content within EEG data.

Acknowledgements

This work was supported in part by grants EP/F500385/1 and BB/F529254/1 for University of Edinburgh Neuroinformatics, from EPSRC, BBSRC and MRC.

References

Babyak MA. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. Psychosom Med 2004;66(3):411–21.
Bell AJ, Sejnowski TJ. An information-maximization approach to blind separation and blind deconvolution. Neural Comput 1995;7(6):1129–59.
Blankertz B, Müller KR, Krusienski DJ, Schalk G, Wolpaw JR, Schlögl A, et al. The BCI competition. III: validating alternative approaches to actual BCI problems. IEEE Trans Neural Syst Rehabil Eng 2006;14(2):153–9.
Blankertz B, Tomioka R, Lemm S, Kawanabe M, Müller Kr. Optimizing spatial filters for Robust EEG single-trial analysis. IEEE Signal Process Mag 2008;25(1):41–56.
Brainard DH. The psychophysics toolbox. Spat Vis 1997;10(4):433–6.
Brodeur MB, Dionne-Dostie E, Montreuil T, Lepage M. The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. PLoS One 2010;5(5).
Burges C. A tutorial on support vector machines for pattern recognition. Data Min Knowl Discov 1998;167:121–67.
Carlson T, Tovar DA, Alink A, Kriegeskorte N. Representational dynamics of object vision: the first 1000 ms. J Vis 2013;13(10):1–19.
Chang C, Lin C. LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2011;2(3):1–27.
Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20(3):273–97.
Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 2004;134(1):9–21.
Delorme A, Sejnowski T, Makeig S. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. Neuroimage 2007;34(4):1443–9.
Donoghue JP. Bridging the brain to the world: a perspective on neural interface systems. Neuron 2008;60(3):511–21.
Efron B, Gong G. A leisurely look at the bootstrap, the jackknife, and cross-validation. Am Stat 1983;37(1):36–48.
Farquhar J, Hill J. In: Learning optimal EEG features across time, frequency and space,in NIPS 2006 Workshop Current Trends Brain-Computer Interfacing; 2006. http://www.kyb.mpg.de/publications/attachments/nips2006_4262%5B0%5D.pdf
Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics 2000;16(10):906–14.
Gaspar CM, Rousselet GA, Pernet CR. Reliability of ERP and single-trial analyses. Neuroimage 2011;58(2):620–9.
van Gerven M, Farquhar J, Schaefer R, Vlek R, Geuze J, Nijholt A, et al. The brain–computer interface cycle 2009;6(4):041001.
Hand DJ. Measuring classifier performance: a coherent alternative to the area under the ROC curve. Mach Learn 2009;77(1):103–23.
Hill N, Lal T, Bierig K. Attention modulation of auditory event-related potentials in a brain–computer interface. In: IEEE Int Work Biomed Circuits Syst; 2004. p. 3–6.
Hung CP, Kreiman G, Poggio T, DiCarlo JJ. Fast readout of object identity from macaque inferior temporal cortex. Science 2005;310(5749):863–6.
Hyvärinen A, Oja E. Independent component analysis: algorithms and applications. Neural Networks 2000;13(4–5):411–30.
Johnson J, Olshausen B. The earliest EEG signatures of object recognition in a cued-target task are postsensory. J Vis 2005;5(4):299–312.
Joyce C, Rossion B. The face-sensitive N170 and VPP components manifest the same brain processes: the effect of reference electrode site. Clin Neurophysiol 2005;116(11):2613–31.
Keerthi S, Lin C. Asymptotic behaviors of support vector machines with Gaussian kernel. Neural Comput 2003;15(7):1667–89.
Kleiner M, Brainard D, Pelli D, Ingling A. What's new in Psychtoolbox-3. In: Percept ECVP Abstr Suppl; 2007. p. 14.
Kriegeskorte N, Mur M, Ruff Da, Kiani R, Bodurka J, Esteky H, et al. Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 2008;60(6):1126–41.
Lotte F, Congedo M, Lécuyer A, Lamarche F, Arnaldi B. A review of classification algorithms for EEG-based brain–computer interfaces. J Neural Eng 2007;4(2):R1–13.
Luck S. An introduction to the event-related potential technique. Cambridge, MA: MIT Press; 2005.
Makeig S, Debener S, Onton J, Delorme A. Mining event-related brain dynamics. Trends Cogn Sci 2004;8(5):204–10.
Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 2007;164:177–90.
Mason S, Graham N. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: statistical significance and interpretation. Q J R Meteorol Soc 2002;125(584):2145–66.
Meyer D, Leisch F, Hornik K. The support vector machine under test. Neurocomputing 2003;55(1–2):169–86.
Mouchetant-Rostaing Y, Giard M. Early signs of visual categorization for biological and non-biological stimuli in humans. Cogn Neurosci 2000;11(11):2521–5.
Müller KR, Tangermann M, Dornhege G, Krauledat M, Curio G, Blankertz B. Machine learning for real-time single-trial EEG-analysis: from brain–computer interfacing to mental state monitoring. J Neurosci Methods 2008;167(1):82–90.
Onton J, Westerfield M, Townsend J, Makeig S. Imaging human EEG dynamics using independent component analysis. Neurosci Biobehav Rev 2006;30(6):808–22.
Osuna E, Freund R, Girosit F. Training support vector machines: an application to face detection. In: Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit; 1997. p. 130–6.
Philiastides MG, Sajda P. Temporal characterization of the neural correlates of perceptual decision making in the human brain. Cereb Cortex 2006;16(4):509–18.
Polich J, Kok A. Cognitive and biological determinants of P300: an integrative review. Biol Psychol 1995;41(2):103–46.
Rousselet GA, Husk JS, Bennett PJ, Sekuler AB. Single-trial EEG dynamics of object and face visual processing. Neuroimage 2007;36(3):843–62.
Rousselet GA, Pernet CR. Quantifying the time course of visual object processing using ERPs: it's time to up the game. Front Psychol 2011;2:1–6.
Sajda P, Gerson A, Müller KR, Blankertz B, Parra L. A data analysis competition to evaluate machine learning algorithms for use in brain–computer interfaces. IEEE Trans Neural Syst Rehabil Eng 2003;11(2):184–5.
Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. Nature 1996;381(6582):520–2.
Wang Y, Jung TP. Improving brain–computer interfaces using independent component analysis. In: Allison BZ, Dunne S, Leeb R, Del R, Millán J, Nijholt A, editors. Towar. Pract. Brain–Computer Interfaces. Berlin, Heidelberg: Springer; 2012, Biological and Medical Physics, Biomedical Engineering.
Wolpaw JR, Birbaumer N, Heetderks WJ, McFarland DJ, Peckham PH, Schalk G, et al. Brain–computer interface technology: a review of the first international meeting. IEEE Trans Rehabil Eng 2000;8(2):164–73.