Title
Discriminative dimensionality reduction for analyzing EEG data

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 36(36)

ISSN
1069-7977

Authors
Noh, Eunho
de Sa, Virginia

Publication Date
2014

Peer reviewed

Discriminative dimensionality reduction for analyzing EEG data

Eunho Noh (eunoh@ucsd.edu)
Department of Electrical and Computer Engineering, University of California, San Diego
9500 Gilman Drive, La Jolla, CA, USA

Virginia R. de Sa (desa@cogsci.ucsd.edu)
Department of Cognitive Science, University of California, San Diego
9500 Gilman Drive, La Jolla, CA, USA

Abstract

We propose a novel way to use discriminative analysis to project high-dimensional EEG data onto a low-dimensional discriminative space for visualization, analysis, and statistical testing. This multivariate analysis directly controls for the multiple comparison problem (MCP) by effectively reducing the number of test variables. A major advantage of this approach is that it is possible to compare the brain activity across conditions even when the trial count is low, provided that a sufficient number of trials are used to establish the initial hyperplane(s), meaning that error conditions and conditions that divide subtle behavioral differences can be readily compared. Currently these data are either ignored or lumped with other data thereby losing the ability to reveal the neural mechanisms underlying subtle behavioral differences. The proposed method provides a powerful tool to analyze conditions with relatively small numbers of trials from high-dimensional neural recordings.

Keywords: MCP; statistical testing; pattern classification; multivariate analysis

Introduction

Electroencephalography (EEG) is extensively used in cognitive neuroscience and related fields to reveal the neural responses associated with specific sensory, cognitive, and motor operations. In a standard EEG experiment, subjects are given a number of trials representing different experimental conditions (i.e. type of stimulus presented, type of response elicited) while the EEG signal is recorded. One way to analyze the EEG signals is to use event related potentials (ERPs). ERPs are computed by averaging over the EEG in response to a given stimulus across numerous trials. ERP analysis reveals the time- and phase-locked brain response to a stimulus, which would not be visible in a single trial of EEG. Another way to analyze EEG signals is to conduct a frequency-domain analysis. By computing the spectral power changes over time and averaging over multiple trials, we can produce a two-dimensional map often referred to as event-related spectral perturbations (ERSPs) (Scott Makeig, 2004).

In order to determine whether a certain effect is statistically significant based on the hypothesis underlying the experiment, statistical testing methods are applied to compare between different experimental conditions (or factors). However EEG recordings are high-dimensional in nature due to the spatio-temporal structure of the data. Hence, we often come up against the multiple comparison problem (MCP) when conducting statistical analysis to compare the brain activities from different conditions. It is usually difficult to control the family-wise error rate (FWER) using standard statistical procedures that operate at the level of a single comparison since typical experiments consist of thousands of dependent variables (Groppe, Urbach, & Kutas, 2011).

When comparing ERP components where the latency of the component is known a priori (such as the P300 or N400), the standard approach is to compare the amplitudes of the component from the different conditions by averaging over the samples within a given time window (Luck, 2005). The spatial locations of interest are often restricted based on prior knowledge. This simple approach effectively increases the SNR of the signal while avoiding the comparisons across adjacent time samples. The component of interest is then tested for significance using a statistical approach such as ANOVA or t-tests. Many cognitive experiments deal with interactions in a cross factorial design so ANOVA is a common choice. However, the test will fail to reveal any unexpected effects lying outside of the temporal, spectral, or spatial analysis window.

Another common method used to conduct statistical analysis on EEG data is a non-parametric randomization test using cluster-based correction (Maris & Oostenveld, 2007). Rather than averaging over a pre-defined window, the cluster-based method figures out the time/spatial/spectral cluster with the most significant activity from the data. First, the test statistic between the two test conditions (stimulus type, behavioral response, etc.) is calculated for each variable (time sample, frequency bin, or electrode position). Clusters are then identified by finding adjacent variables with significant difference between the two conditions (below a certain p-value, e.g. $p < 0.05$). The cluster-level statistic is calculated by summing up these differences for each cluster and selecting the cluster with the maximum value. This result is compared to the cluster-based statistic of the permutation distribution generated from a large number of random permutations of the trial labels. This approach allows the researcher to directly solve the MCP but the sensitivity of the test depends on the threshold used to select significant variables.

In order to increase the statistical power of a test, it is desirable to have many observations in a given condition. This means that conditions with few trials (e.g. sub-categorical conditions or error trials) are often not analyzed or combined with other conditions due to low statistical power. These rare conditions however are likely to reveal critically interesting information about how neural activity gives rise to complex behavior.

Data from two highly discriminative behavioral conditions can be used to train a classifier to learn the discriminative hyperplane between the two conditions (or *classes*). The pattern classifier efficiently combines the temporal, spatial and spectral features from the EEG data and projects the data onto a vector which is perpendicular to the discriminative hyperplane defined by the two discriminative conditions. For example, if the two conditions were remembered vs. forgotten trials from a recognition memory experiment, the projection values would be related to encoding success. The observations from the remaining conditions which have similar cognitive components can be projected onto this vector and the resulting values can be compared across conditions. This classifier-based method can be considered a multivariate analysis which directly controls for the MCP. We show that this approach can reveal significant differences between conditions without requiring prior assumptions based on existing data or theory. The proposed method also gives high sensitivity even for low trial count conditions, provided that a sufficient number of trials are used to establish the initial hyperplane(s) and that the conditions are related to the training conditions. This method was applied to a simulated dataset and compared to the conventional t-test and cluster-based test to investigate its effectiveness.

Methods

Let us consider a dataset of EEG recordings from a cognitive experiment where there are two highly discriminative behavioral conditions (e.g. recollected vs. forgotten trials in a recognition memory experiment). We can formulate a supervised learning problem with two classes using this dataset. Let $\mathbf{S}_T = \{(\overline{x_i}, y_i)\}_{i=1}^{N}$ be the training set with N trials. Each pair consists of the recorded signal $\overline{x_i}$ and the class (or condition) $y_i \in \{1,2\}$ from one of the two discriminative conditions. Also let $\mathbf{S}_R = \{(\overline{x_i}, y_i)\}_{i=N+1}^{N+M}$ be the set of observations from the remaining conditions ($y_i \notin \{1,2\}$) which share some cognitive components with the discriminative conditions. The goal is to construct a function $h : \overline{x_i} \to p_i$ where p_i is a projection of trial i onto the vector perpendicular to the discriminative hyperplane. In this paper, we map p_i to q_i which is a value between 0 and 1. The composed function can be considered as a mapping of the EEG signal onto the $[0,1]$ interval where a value closer to 0 (1) implies that the given trial shows features which are more similar to those from training condition 1 (2).

The trials from classes 1 and 2 are then projected onto a vector which is perpendicular to the discriminative hyperplane. Careful attention is required when projecting the trials from the conditions used to train the pattern classifier. The trials used as the training set should not be used to evaluate the classifier performance or the significance of an experimental variable. In other words, the trial being projected onto the vector should always be isolated from the classifier training procedure in order to eliminate any overfitting (Bishop, 2006). This can be achieved via cross-validation or by pre-designating an evaluation set which is excluded from *any* training procedure. The trials from the two highly discriminative conditions would ideally have significantly different outputs when projected onto the vector provided that there is discriminative information in the neural signals and appropriate feature extraction and classification methods are used. Specific feature extraction and classification methods for EEG signals will not be discussed here since it is not under the scope of the current paper.

The observations from the remaining conditions ($y_i \notin \{1,2\}$, e.g. trials with low confidence responses in an episodic memory experiment) can be projected onto this vector in the same manner but without risk of overfitting. The average projection values for the different conditions can be compared using conventional statistical methods such as t-tests or ANOVA. Here we only consider comparing the features in the temporal domain, but the proposed method can be applied to spatial, spectral, or a combination of features in a similar manner.

Analysis on a simulated dataset

The effectiveness of the proposed algorithm was evaluated using artificial datasets. One evaluation set was designed to test sensitivity with different means across the different conditions and the other set was designed to test specificity with equal means. The sensitivity and specificity of the current method was compared to that of two conventional methods, namely the t-test and the cluster based statistical test (Maris & Oostenveld, 2007).

Simulated dataset

Artificial datasets were generated from real EEG data to simulate EEG from a single channel during memory retrieval. ERPs were calculated using data from an actual recognition memory experiment (see Experiment 2 from Mollison and Curran (2012) for details of the experimental procedures) from two conditions which typically show the parietal old/new effect (class 1: source correct old trials, class 2: correctly rejected new trials). Data from the left posterior superior electrodes were averaged to acquire the ERP templates for each class. As illustrated in Figure 1, the parietal old/new effect can be observed between 500-800 ms after item presentation (noted as 0 ms).

Gaussian noise was added to each of the ERP templates to generate the training set (50 observations per class) where the standard deviation of the noise was 2 times as large as the standard deviation of the ERP templates (the same noise parameters were used for the evaluation sets). Note that this training set was only used to train the classifier for the classifier-based method. Two evaluation sets were generated where each evaluation set had two classes. The evaluation sets were designed to represent intermediate conditions (e.g. sub-conditions with smaller effects) between the two training conditions. The sensitivity evaluation set consisted of two classes with different means while the specificity evaluation set consisted of two classes with equal means. The

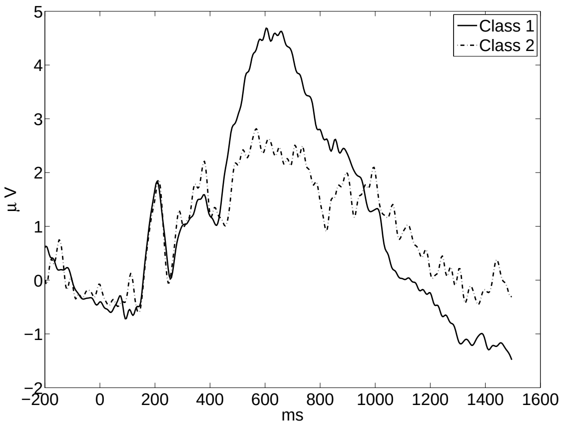


Figure 1: The ERP templates used to generate the training set for the simulation analysis. Class 1 is an average of source correct old trials and class 2 is an average of correctly rejected new trials from Experiment 2 described in (Mollison & Curran, 2012).

two conditions in the sensitivity evaluation set were generated by adding Gaussian noise to the weighted averages of the ERP templates used to generate the training set (0.75 $\times$ template 1 + 0.25 $\times$ template 2 for one condition and 0.25 $\times$ template 1 and 0.75 $\times$ template 2 for the other condition). The two classes in the specificity evaluation set were generated by adding Gaussian noise to the average of templates 1 and 2.

Analysis procedure

The evaluation sets were tested for significance in three different time segments (500-800/500-1200/500-1500 ms). The motivation for this was to compare the different methods when the tests were performed within an optimal/sub-optimal region. Note that the best window width/location should not be determined from the full dataset as that would result in overfitting and too high a false alarm rate.

The following analysis was conducted for multiple numbers of trials per class (ranging from 5 to 50) on both evaluation sets for each of the time segments. In order to evaluate the performance of the classifier-based method, we compared its results to two other statistical tests often used for EEG analysis. The first method involved a t-test conducted on the average over the segment between the two classes in each evaluation set. The second method was a non-parametric randomization test based on cluster-based correction (Maris & Oostenveld, 2007). Clusters were identified by finding adjacent time points with t-statistic lower than 0.05. The largest cluster was selected to be the true cluster statistic. This true cluster statistic was compared to the permutation distribution computed from 1000 random permutations of the class labels.

For the classifier-based method, the linear classifier trained with the training set described above was used to project the evaluation trials onto a value between 0 and 1. Before training the classifier, the dimensionality of all trials were reduced by averaging over 100 ms non-overlapping windows of the given time segments. Linear discriminant analysis (LDA) with probability outputs was selected for classification and trained using the feature vectors from the training set. This classification approach is known to be effective at classifying temporal features of the EEG data (Blankertz, Lemm, Treder, Haufe, & Müller, 2011). A t-test was conducted on the classifier outputs to examine whether the mean projections were significantly different between the two classes.

In order to reduce the noise from the randomness of the evaluation sets, 1000 random evaluation sets were assessed for each trial count for each of the three methods. The number of times a significant effect was observed ($p < 0.05$) out of the 1000 runs were identified and the ratio of observing a significant effect was computed.

Results

The ratio of observing a significant effect ($p < 0.05$) across the 1000 random runs conducted on the sensitivity evaluation set (different mean condition) is given in Table 1 for each set of the parameters in the simulations. The results for the different trial counts are given in separate columns. The t-test and classifier-based results gave p-values below 0.05 for the majority of the simulations when the analysis was restricted to 500-800 ms. However, the sensitivity of the t-test decreased as the analysis window increased. It was found that the t-test only found significant effects 10 to 65 % of the time for the 500-1500 ms analysis window. The classifier-based method found significant effects over 95 % of the time when the trial counts per condition were 10 or more for all window sizes and for all trial counts when the ROI was 500-1500 ms. The cluster-based method gave consistent results across all window sizes, but only gave reliable results for the cases where there were at least 40 trials per condition. The ratio of observing a significant effect ($p < 0.05$) across the 1000 random runs conducted on the specificity evaluation set gave false alarm rates close to 5 % for all three methods which is expected given the common use of that p-value.

In order to check whether the improvement in sensitivity resulted from the discriminative nature of the approach or from dimensionality reduction, we compared the classifier-based approach to principal component analysis (PCA). To conduct a fair comparison, the PCA projections were also learned from the training conditions and only the projections onto the first principal component was used. The results from the previous sensitivity evaluation set gave similar results for both the classifier- and PCA-based approach. However when the noise amplitude was increased by 25 %, the classifier-based approach gave more reliable results compared to the PCA-based approach as given in Table 2.

Discussion

The proposed method found significant effects in the sensitivity evaluation set even when the number of test observations

Table 1: The statistical test results for the sensitivity evaluation set(different means between the two conditions). The values in the table represent the ratio of *observing* a significant effect ($p < 0.05$) out of a total of 1000 random simulations for a given number of trials per condition. Values above 0.95 are given in bold.

ms	#trials/cond.	5	10	15	20	25	30	40	50
500-800	t-test	0.849	**1**	**1**	**1**	**1**	**1**	**1**	**1**
	cluster	0.139	0.385	0.611	0.736	0.852	0.918	**0.994**	**0.998**
	classifer	0.911	**0.998**	**1**	**1**	**1**	**1**	**1**	**1**
500-1200	t-test	0.441	0.863	**0.978**	**0.998**	**0.999**	**0.999**	**1**	**1**
	cluster	0.074	0.327	0.412	0.596	0.794	0.885	**0.977**	**0.999**
	classifer	0.934	**1**	**1**	**1**	**1**	**1**	**1**	**1**
500-1500	t-test	0.101	0.148	0.254	0.298	0.363	0.431	0.574	0.642
	cluster	0.083	0.252	0.479	0.6	0.791	0.884	**0.973**	**0.999**
	classifer	**0.961**	**1**	**1**	**1**	**1**	**1**	**1**	**1**

Table 2: The statistical test results for the sensitivity evaluation set with higher noise shows that non-discriminative dimensionality reduction methods (e.g. PCA) may not be as effective as discriminative approaches. The values in the table represent the ratio of *observing* a significant effect ($p < 0.05$) out of a total of 1000 random simulations for a given number of trials per condition. Values above 0.95 are given in bold.

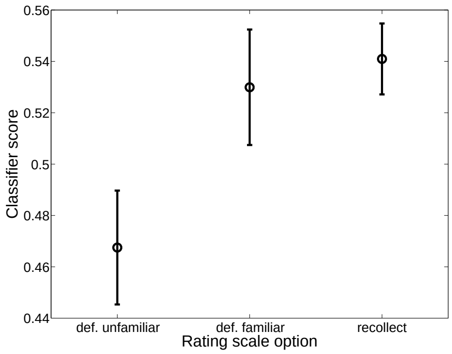
ms	#trials/cond.	5	10	15	20	25	30	40	50
500-800	PCA	0.68	0.935	**0.983**	**0.996**	**0.996**	**1**	**1**	**1**
	classifer	0.782	**0.98**	**0.997**	**1**	**1**	**1**	**1**	**1**
500-1200	PCA	0.639	0.91	**0.977**	**0.991**	**0.998**	**0.999**	**1**	**1**
	classifer	0.829	**0.99**	**0.997**	**1**	**1**	**1**	**1**	**1**
500-1500	PCA	0.7	0.93	**0.984**	**0.999**	**1**	**1**	**1**	**1**
	classifer	0.865	**0.992**	**1**	**1**	**1**	**1**	**1**	**1**

per condition was as low as 5 (for the 500-1500 ms window). The t-test and the classifier-based method gave comparable results when the evaluation was conducted within the time period where the old/new effect was evident (500-800 ms). The sensitivity of the t-test decreased as the ROI increased while the proposed method was not affected by the change. The increase in the analysis window was disadvantageous for the t-test because the cross-over between the ERPs decreased the size of the effect. The reason that the PCA-based approach performed relatively well is due to the fact that the time segments with the largest variance were in fact informative in distinguishing between the two training conditions. Hence the PCA and classifier weights may have been similar. However, PCA can easily be misled if the high variance features are uninformative in distinguishing between the two conditions.

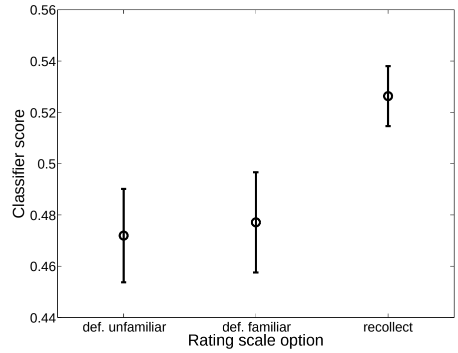
The advantage of the classifier-based method appears to be achieved with the cost of specificity in the feature space. However, the relevance of an individual feature to a given comparison can be identified by examining the activation patterns corresponding to the projection weights (Haufe et al., 2014). Nevertheless, a direct comparison between the EEG signals better reveals the characteristics (e.g. latency, size, duration, location) of the effect identified by the classifier. Hence, a post-hoc analysis using the raw EEG features should always be conducted in order to better understand sources of the effect.

The cluster-based method only gave comparable results to the other methods after the trial count per condition was at least 40 per class for the sensitivity evaluation set. The statistical power of the cluster-based method depends on how well the permutation distribution represents the null hypothesis. However, in a low trial count condition there are only a small number of possible permutations to estimate the permutation distribution. For example if there are 5 trials available for each condition, the number of possible permutations of the labels is only 252 ($=10!/(5! \times 5!)$). This may result in an inaccurate estimation of the significance of a cluster. The cluster-based method gave consistent results across the different evaluation windows. Since the cluster-based method chooses the cluster with the maximum effect it is less susceptible to the cross-over between the two conditions. It was found that the average end time of the clusters was approximately 800 ms even when analysis was conducted on the 500-1500 ms window.

In a recent collaborative project (Noh et al., 2014), we utilized this pattern classification method to replicate previous findings on the subsequent memory effect (Sanquist, Rohrbaugh, Syndulko, & Lindsley, 1980; Paller & Wagner, 2002; Otten, Quayle, Akram, Ditewig, & Rugg, 2006; Otten, Quayle, & Puvaneswaran, 2010; Park & Rugg, 2010; Guderian, Schott, Richardson-Klavehn, & Duezel, 2009; Fell et al., 2011). Moreover, the single-trial analysis revealed interesting findings regarding the neural mechanisms related to recollection and familiarity. The classifier was trained on the recollected vs. unfamiliar trials which projected the high-dimensional EEG data onto a discriminative vector which represented encoding strength. Note that the subjects were instructed to give recollect responses only when they had a conscious recollection of learning the item in the study phase (i.e., they remembered the context of learning the item). The analysis on the classifier score revealed that the trials with



(a) Definitely familiar and ***definitely unfamiliar*** trials were significantly different ($p < 0.004$). Recollect and definitely unfamiliar trials were significantly different ($p < 10^{-5}$).



(b) Definitely familiar and ***recollect*** trials were significantly different ($p < 10^{-3}$). Recollect and definitely unfamiliar trials were significantly different ($p < 10^{-3}$).

Figure 2: A reproduced version of the results from Noh et al. (2014). The average projection values and the standard errors of the three conditions given by the classifiers trained on the alpha (7-12 Hz) power between (a): 400-800 ms after stimulus presentation and (b): 1000-1400 ms after stimulus presentation. The classifiers were trained using only the recollected vs. unfamiliar trials.

definitely familiar responses (which were not involved in the training procedure) were mapped closer to the *recollected* trials early in the epoch (400-800 ms) while the same trials were mapped closer to the *definitely unfamiliar* trials later in the epoch (1000-1400 ms) when classification was conducted on the spectral information in the alpha band (7-12 Hz). Post-hoc analysis showed that the alpha desynchronization between 400-800 ms was weaker for the recollected trials in the left central electrodes while the 1000-1400 ms showed stronger desynchronzation for the recollected trials in the posterior electrodes. These results (illustrated in Figures 2 and 3) suggest that the brain activity represented by the alpha band may shift from encoding of the stimulus to also encoding the contextual information of that trial.

In fMRI studies, pattern classifiers have been used as multivoxel pattern analysis (MVPA) methods for detecting and analyzing cognitive states (Norman, Polyn, Detre, & Haxby, 2006). This approach has been proven to be successful in characterizing neural coding and information processing in many domains of cognitive neuroscience (Watanabe et al., 2011; Yoo et al., 2012). In EEG, pattern classifiers have mostly been used for the purpose of classifying brain signals for brain-computer interfaces (Lotte, Congedo, Lécuyer, Lamarche, & Arnaldi, 2007). Our results suggest that multivariate analysis techniques also can be beneficial when testing for significance in EEG data particularly for revealing significant effects in conditions with small numbers of trials. They are also useful when prior hypotheses do not precisely specify the ROI for the given experimental manipulation.

This multivariate analysis directly controls for the MCP by effectively reducing the number of test variables. A major advantage of this approach is that it is possible to compare the brain activities across low trial count conditions, provided that a sufficient number of trials are used to establish the initial hyperplane(s). Hence conditions that divide subtle behavioral differences can be readily compared. The strength of this method comes from the fact that the information in the observations used to train the classifier can be exploited to restrict the comparisons of the test observations. Another way to look at this method is as a boot-strapping method where we utilize conditions which are not currently being compared to estimate the characteristics of the conditions of interest. The application of this method is not restricted to EEG data but can be applied to other high-dimensional neural data such as MEG or single-unit recordings. It should also be noted that careful cross-validation procedures or partition of the training and evaluation set is required when directly comparing the conditions used to train the classifiers. The classifier outputs may overfit the training data and results should only be used on held out data.

Acknowledgments

This work was supported by the NSF (grants SMA 1041755, IIS 1219200) and an Innovative Research Grant from the Kavli Institute of Brain and Mind. We would like to thank Dr. Tim Curran and Dr. Grit Herzmann for helpful comments on the work. We would also like to thank Matthew V. Mollison for providing us with the EEG dataset and valuable suggestions.

References

Bishop, C. M. (2006). *Pattern recognition and machine learning (information science and statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Blankertz, B., Lemm, S., Treder, M. S., Haufe, S., & Müller, K.-R. (2011). Single-trial analysis and classification of ERP components – a tutorial. *NeuroImage*, *56*, 814–825.

Fell, J., Ludowig, E., Staresina, B. P., Wagner, T., Kranz, T.,

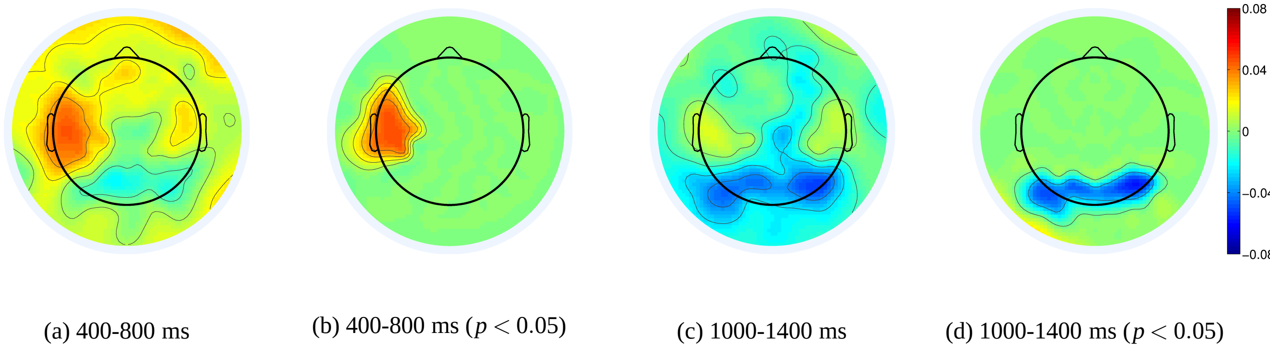


Figure 3: A reproduced version of the results from Noh et al. (2014). Difference in alpha (7-12 Hz) power between the recollected vs. unfamiliar trials for the different time segments ($\log(\mu V^2)$). The spatial pattern in (b) and (d) are masked by the most significant cluster resulting from cluster-based analysis ($p < 0.05$).

Elger, C. E., et al. (2011). Medial temporal theta/alpha power enhancement precedes successful memory encoding: evidence based on intracranial EEG. *Journal of Neuroscience*, *31*(14), 5392-5397.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*, 1711-1725.

Guderian, S., Schott, B. H., Richardson-Klavehn, A., & Duezel, E. (2009). Medial temporal theta state before an event predicts episodic encoding success in humans. *Proceedings of the National Academy of Sciences*, *106*(13), 5365-5370.

Haufe, S., Meinecke, F., Grgen, K., Dhne, S., Haynes, J.-D., Blankertz, B., et al. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, *87*(0), 96 - 110.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for eeg-based brain-computer interfaces. *Journal of Neural Engineering*, *4*(2), R1-R13.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA, USA: MIT Press.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177-190.

Mollison, M. V., & Curran, T. (2012). Familiarity in source memory. *Neuropsychologia*, *50*, 2546-2565.

Noh, E., Herzmann, G., Curran, T., & de Sa, V. R. (2014). Using single-trial eeg to predict and analyze subsequent memory. *NeuroImage*, *84*, 712-723.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fmri data. *Trends Cognitieve Science*, *10*, 424–430.

Otten, L. J., Quayle, A. H., Akram, S., Ditewig, T. A., & Rugg, M. D. (2006). Brain activity before an event predicts later recollection. *Nature Neuroscience*, *9*(4), 489-491.

Otten, L. J., Quayle, A. H., & Puvaneswaran, B. (2010). Prestimulus subsequent memory effects for auditory and visual events. *Journal of Cognitive Neuroscience*, *22*(6), 1212-1223.

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, *6*(2), 93-102.

Park, H., & Rugg, M. D. (2010). Neural correlates of encoding within- and across-domain inter-item associations. *Journal of Cognitive Neuroscience*, *9*, 2533-2543.

Sanquist, T. F., Rohrbaugh, J. W., Syndulko, K., & Lindsley, D. B. (1980). Electrocortical signs of levels of processing: perceptual analysis and recognition memory. *Psychophysiology*, *17*(6), 568–576.

Scott Makeig, J. O. A. D., Stefan Debener. (2004). Mining event-related brain dynamics. *Trends in Cognitive Sciences*, *8*, 204-210.

Watanabe, T., Hirose, S., Wada, H., Katsura, M., Chikazoe, J., Jimura, K., et al. (2011). Prediction of subsequent recognition performance using brain activity in the medial temporal lobe. *NeuroImage*, *54*(4), 3085-3092.

Yoo, J. J., Hinds, O., Ofen, N., Thompson, T. W., Whitfield-Gabrieli, S., Triantafyllou, C., et al. (2012). When the brain is prepared to learn: Enhancing human learning using real-time fMRI. *NeuroImage*, *59*(1), 846–852.