

MIT Open Access Articles

*Sparse dictionary methods for EEG
signal classification in face perception*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Shariat, Shahriar et al. "Sparse Dictionary Methods for EEG Signal Classification in Face Perception." IEEE, 2010. 331–336. Web. ©2010 IEEE.

As Published: <http://dx.doi.org/10.1109/MLSP.2010.5589166>

Publisher: Institute of Electrical and Electronics Engineers

Persistent URL: <http://hdl.handle.net/1721.1/70068>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



SPARSE DICTIONARY METHODS FOR EEG SIGNAL CLASSIFICATION IN FACE PERCEPTION

Shahriar Shariat¹, Vladimir Pavlovic¹, Thomas Papathomas², Ainsley Braun³, and Pawan Sinha³

¹Dept. Computer Science, Rutgers University, Piscataway, NJ

²Dept. of Biomedical Engineering and RuCCS/Lab of Vision Research, Rutgers University, Piscataway, NJ

³Dept. of Brain and Cognitive Sciences, MIT, Cambridge, MA

ABSTRACT

This paper presents a systematic application of machine learning techniques for classifying high-density EEG signals elicited by face and non-face stimuli. The two stimuli used here are derived from the vase-faces illusion and share the same defining contours, differing only slightly in stimulus space. This emphasizes activity differences related to high-level percepts rather than low-level attributes. This design decision results in a difficult classification task for the ensuing EEG signals. Traditionally, EEG analyses are done on the basis of signal processing techniques involving multiple instance averaging and then a manual examination to detect differentiating components. The present study constitutes an agnostic effort based on purely statistical estimates of three major classifiers: L1-norm logistic regression, group lasso and k Nearest Neighbors (kNN); kNN produced the worst results. L1 regression and group lasso show significantly better performance, while being able to identify distinct spatio-temporal signatures. Both L1 regression and group lasso assert the saliency of samples in 170ms, 250ms, 400ms and 600ms after stimulus onset, congruent with the previously reported ERP components associated with face perception. Similarly, spatial locations of salient markers point to the occipital and temporal brain regions, previously implicated in visual object perception. The overall approach presented here can provide a principled way of identifying EEG correlates of other perceptual/cognitive tasks.

1. INTRODUCTION

The brain mechanisms underlying the ability of humans to process faces have been studied extensively in the last two decades. Brain imaging techniques, particularly fMRI (functional Magnetic Resonance Imaging) that possesses high spatial resolution but limited temporal resolution, are advancing our knowledge of the spatial and anatomical organization of face-sensitive brain areas (for review, see [1, 2]). At the other end are EEG recording techniques with high temporal resolution but poor spatial resolution. They reveal

event-related potentials (ERPs) that serve as correlates for various aspects of facial processing. The best-established ERP marker of face perception is N170, a negative component that occurs roughly 170 ms after stimulus onset [3]. Other markers, such as N250, N400 and P600, which occur later than N170, are considered to contribute to face recognition and identification [4]. In a typical fMRI or ERP study, signals are recorded while the subject is exposed to a face or a non-face stimulus to isolate the brain areas that respond differentially to faces (fMRI), and the temporal intervals that exhibit major differences between face and a non-face responses in the data (ERP). Early studies in both fMRI and ERP employed simplistic signal processing techniques, involving multiple instance averaging and then a manual examination to detect differentiating components. More advanced statistical signal analysis techniques were first applied to fMRI signals (for a review see [5]). Systematic analyses generally lagged behind in the ERP domain. A recent principled approach is by Moulson et al. [2], in which the authors applied statistical classification in the form of Linear Discrimination Analysis (LDA) to face/non-face ERPs and obtained classification accuracies of about 70%.

There are two major goals for the current ERP study: First, to emphasize higher brain areas of visual processing at the expense of early visual areas by using a strategically designed control non-face stimulus. Second, and more importantly, this study seeks to apply systematic machine learning and pattern recognition techniques for classifying face and non-face responses in both the spatial and temporal domains. Toward the first major goal, we use a face (Figure 1c) and a non-face (Figure 1b) stimulus derived from the well-known vase/faces illusion (Figure 1a). The key feature of the face and non-face stimuli is that they share the same defining contours, differing only slightly in stimulus space. The defining contours in Figure 1a are attributed to whatever forms the figure; as the percept alternates spontaneously, these contours are attributed alternately to the faces or to the vase. This attribution is biased, using minimal image manipulations, toward the vase in Figure 1b and the faces



Fig. 1. Face/vase illustrations

in Figure 1c, but the same contours are used in both cases. These shared contours result in relatively similar responses for the faces and vase stimuli and produce a difficult classification task for the ensuing ERP signals. With this choice of face and non-face stimuli, we bias the analysis towards detecting signal differences that are elicited more by the high-level percepts (face or non-face) rather than low-level image differences.

With respect to the second major goal, we note that several types of classifiers have been used in previous EEG classification studies: kNN [6], logistic regression and multi layer perceptron [7], support vector machines [8] and LDA [2, 9]. In [10] authors have used group penalty for lasso regression in frequency domain for EEG signal classification and showed the utility of grouping in that domain. The present study extends this systematic use of classifiers by using the ERP signals obtained with the faces/vase stimuli to test three major classification schemes: kNN, L1-norm logistic regression, and group lasso logistic regression. We perform the classification analysis between two classes of face and vase responses agnostically, based on purely statistical estimates, without favoring any sensors or temporal intervals. Our goal is to use the data to point to salient spatio-temporal ERP signatures most indicative of the stimuli classes. Obviously, ERP classification can provide important applications in neuroscience, such as in brain-computer interfaces (BCI) and in detecting pathologies such as epilepsy.

The main results of our tests with the three major classifier schemes are: First, kNN produced the worst performance, having to rely on all, potentially noisy, features. Second, the other two schemes were able to classify the signals with an overall accuracy of roughly 80%. Finally, the learned weights of L1-norm logistic regression and group lasso were able to locate the salient features in space (electrode position) and time in close agreement with the accepted wisdom of previous studies, confirming various markers of face perception such as N170, N400 and P600.

2. PRIOR WORK

EEG signals are very challenging to analyze due to the noisy nature of sampling, cross-talk among different channels and,

most importantly, "artifacts" of routine brain activities such as blinking or breathing. Independent Component Analysis [11], coupled with other dimensionality reduction methods such as the PCA, is often used as a tool to recover the important underlying signals while removing such artifacts. Nevertheless, the unsupervised reduction methods are typically insufficient for identification of features needed for accurate and robust classification of EEG signals as the large variance of input signals does not always warrant its classification significance.

In contrast to unsupervised data-driven feature extraction, common EEG signal analysis approaches focus on fixed dictionaries of wavelet [6], short term Fourier transform [12], or other well-established non-stationary signal representations. However, variability in temporal occurrence of important EEG events, typically exhibited across different subjects and trials, makes the use of such features challenging, requiring temporal alignment. Moreover, fixed dictionary representations, not adapted to data, can be dense, affecting their robustness and making them less attractive for classification settings.

In the Brain-Computer Interface (BCI) literature Common Spatial Patterns (CSP) have shown very promising results [13] as the data-driven means for characterizing temporal patterns that can be used for signal classification. CSP is a linear transformation that maps the original signal into a space where the data shows a high class relative variance. However, CSP's applicability is typically restricted to sparse EEG signals, with well defined temporal boundaries, making such patterns less appropriate for settings of dense signals, such as those in our dataset.

In this work we focus on identification of sparse, data-driven spatio-temporal EEG dictionaries that directly reflect our classification objective. Our goal is not only to identify those patterns as the means of efficient and robust classification but to also assert their relationship with established EEG perceptual signatures such as P100, N170, or P250.

3. PREDICTIVE MODELS

Our goal is to design robust and accurate predictors of the stimulus classes that can account for high levels of noise/artifacts in the input signal as well as inter-subject variability. We also seek to identify a small subset of features of the EEG signal, the so-called predictive signatures, that contribute to these predictions.

To achieve this goal we focus on a recently proposed family of sparse logistic regression models. Sparse logistic regression models are a class of probabilistic parametric classifiers whose goal is to model the predictive process while minimizing the number of parameters used in this prediction. In particular, consider the binary response variable $y \in \{0, 1\}$, the predictor (feature) vector $\mathbf{X} \in \mathbb{R}^p$. For in-

stance, in the EEG setting y_i can be the class of response (face/vase) while the k -th feature X_k could be the measured EEG signal in channel c at time t . Furthermore, consider the set of N training points $\mathcal{D} = \{(y_i, \mathbf{X}_i)\}_{i=1}^N$. The logistic regression model represents the class-conditional probabilities through a sigmoidal function of the linear predictors,

$$\log \frac{Pr(y_i = 1 | \mathbf{X}, \beta_0, \beta)}{Pr(y_i = 0 | \mathbf{X}, \beta_0, \beta)} = \beta_0 + \mathbf{X}^T \beta. \quad (1)$$

The classifier uses the Bayes rule, $f(\mathbf{X} | \beta_0, \beta) = \arg \max_y Pr(y | \mathbf{X}, \beta_0, \beta)$. To find the optimal classifier one can consider a number of objectives, such as the square loss [14] or logistic loss [15], such that loss of prediction on this set of points \mathcal{D} is minimized, while keeping the cardinality k of the utilized features small, $k \ll p$. For instance, for logistic loss we seek to

$$\min_{\beta_0, \beta} \frac{1}{N} \sum_{i=1}^N \log P(y_i | \mathbf{X}_i, \beta_0, \beta) \quad (2)$$

$$\text{s.t. } \text{card}(\beta) \leq k, \quad (3)$$

where $\text{card}(\cdot)$ denotes the cardinality or the number of non-zero entries in vector β (equivalently, the L_0 norm of β).

Unfortunately, this task is, in general, computationally intractable. One typically instead focuses on a tractable relaxation known as the L_1 or lasso regression [14] (also known as sparse coding), reflected in the Lagrangian:

$$\min_{\beta_0, \beta} \frac{1}{N} \sum_{i=1}^N \log P(y_i | \mathbf{X}_i, \beta_0, \beta) + \lambda \|\beta\|_1 \quad (4)$$

where $\|\beta\|_1$ is the L_1 norm $\sum_{i=1}^p |\beta_i|$. Some more recent work has quantified conditions under which the optimization in (4) is guaranteed to lead to the same solutions as the original objective (3). Nevertheless, in practice the lasso objective leads to models with few non-zero coefficients that focus on those aspects of the feature vector most responsible for distinction between 0/1 classes of inputs. The objective in (4) is convex in β , leading to several efficient gradient based algorithms, see c.f., [15].

Lasso regression focuses on identification of individual features. In many tasks, such as the EEG signal analysis, individual features (values of voltage measured by electrode e at time t) may be insufficiently strong or too variable to lead to robust predictions. Instead, groups of spatio-temporally proximal features may serve as better predictors (e.g. short spatio-temporal signal forms). A typical approach taken in the community is to design such features using fixed dictionaries, such as the Fourier, wavelet, or other bases. In contrast, we seek to find those features directly from data, i.e., identify compact data-dependent spatio-temporal dictionaries of EEG signals. We use Group lasso [16], an extension of lasso, to accomplish this task.

Formally, consider the partitioning of the weight vector β into a set of groups $\beta_j, j = 1, \dots, J$, where each group contains a subset of coefficients β (or, equivalently features \mathbf{X}). Let $\|\eta\|_K$ be the K norm of vector $\eta \in \mathbb{R}^d, d \geq 1$, $\|\eta\|_K = (\eta^T K \eta)^{\frac{1}{2}}$ where K is a symmetric positive-definite matrix. The group lasso estimate is defined as the minimizer of

$$\frac{1}{N} \sum_{i=1}^N \log P(y_i | \mathbf{X}_i, \beta_0, \beta_1, \dots, \beta_J) + \lambda \sum_{j=1}^J \|\beta_j\|_{K_j}, \quad (5)$$

where

$$\log \frac{Pr(y_i = 1 | \mathbf{X})}{Pr(y_i = 0 | \mathbf{X})} = \beta_0 + \sum_{j=1}^J \mathbf{X}_{i,j}^T \beta_j. \quad (6)$$

In an extension to the lasso regression, Group lasso leads to predictors with few non-zero or *active groups of coefficients*. Here we have considered K to be diagonal.

4. EXPERIMENTS

4.1. EEG Data Collection

Five adults between the ages of 20 and 30 years participated in this experiment. All adults had normal or corrected-to-normal vision, and none had a history of neurological abnormalities.

Our stimulus set consisted of 60 randomly ordered presentations of the face and non-face images shown in figure 1. Each trial consisted of stimulus presentation (300ms) and a post-stimulus recording period (1000ms). The inter-trial interval, during which a black fixation cross was presented on a gray background, varied randomly between 1500-2000ms. Participants were required to determine whether each stimulus was a face or a non-face, and responded using two buttons on a button box. Participants were instructed to make their responses as quickly and accurately as possible, and they had 1500ms from stimulus onset to do so.

Electrophysiological Recording and Processing. While participants were performing the above task, continuous EEG was recorded using a 128-channel Geodesic Sensor Net (Electrical Geodesics, Inc.), referenced online to vertex (Cz). The electrical signal was amplified with 0.1 to 100Hz band-pass filtering, digitized at a 250 Hz sampling rate. Data were preprocessed offline using NetStation 4.2 analysis software (Electrical Geodesics, Inc.). The continuous EEG signal was segmented into 900ms epochs, starting 100ms prior to stimulus onset. Data were filtered with a 30Hz low-pass elliptical filter and baseline-corrected to the mean of the 100ms period before stimulus onset. NetStation's automated artifact detection tools combed the data for eye blinks, eye movements, and bad channels. Segments were excluded from further analysis if they contained an eye blink

Table 1. Across subject AUCs for L1 logistic regression and group lasso, in percentage points, for different time ranges.

| Test cases | L1 Logistic Regression | | | | | Group Lasso | | | | |
|------------|------------------------|--------|--------------|---------|------------|--------------|--------|---------|---------|------------|
| | full | 50-100 | 100-200 | 200-300 | ≥ 400 | full | 50-100 | 100-200 | 200-300 | ≥ 400 |
| subject 1 | 82.40 | 70.04 | 80.14 | 78.48 | 57.91 | 88.11 | 66.45 | 79.43 | 68.37 | 55.77 |
| subject 2 | 98.26 | 69.10 | 94.10 | 86.11 | 50.00 | 92.53 | 73.26 | 93.92 | 77.60 | 50.00 |
| subject 3 | 86.88 | 77.16 | 86.11 | 82.72 | 50.00 | 73.46 | 54.78 | 85.49 | 70.06 | 50.00 |
| subject 4 | 84.78 | 74.90 | 82.02 | 80.83 | 67.00 | 87.94 | 66.01 | 82.81 | 79.45 | 64.23 |
| subject 5 | 80.74 | 64.64 | 91.79 | 59.64 | 58.93 | 91.43 | 60.71 | 87.89 | 59.29 | 63.93 |
| Average | 86.61 | 71.17 | 86.83 | 77.56 | 56.77 | 86.69 | 64.24 | 85.91 | 70.95 | 56.77 |

(threshold $\pm 70\mu V$) or eye movement (threshold $\pm 50\mu V$). In the remaining segments, individual channels were marked bad if the difference between the maximum and minimum amplitudes across the entire segment exceeded $80\mu V$. If more than 10% of the 128 channels were marked bad in a segment, the whole segment was excluded from further analysis. If fewer than 10% of the channels were marked bad in a segment, they were replaced using spherical spline interpolation. For the subsequent data analysis we retained signal ranges from 0ms (stimulus onset) up to 700ms.

4.2. EEG Data Analysis

In our experiments we used L1-norm logistic regression and group lasso described in Sec. 3, to design classifiers and spatio-temporal dictionaries for the purpose of predicting the class of visual stimulus to which our subjects were exposed. We used the methods of [15] and [17] to train the two classifiers.

In the data pre-processing stage we filtered the EEG signals by an FIR band-pass filter 8-14 Hz to retain a portion of the alpha range of brain signals. We then normalize each channel independently and windsorize 5% of the peak. Each processed signal forms a $129 \times 700 = 90,300$ dimensional feature vector $\mathbf{X} = [X_{i,t}]$, $i = 1, \dots, 129$, $t = 1, \dots, 700$. For the group lasso setting we formed spatial groups of size five for each channel and its four nearest spatial neighbors. Temporal groups were set to fixed size of 10 samples, set at every 5th sample (i.e., 1-10, 5-15, etc.)

As a baseline classifier, we also use a k-nearest neighbor (kNN) predictor which transfers labels from the training instances using majority voting to the query instance. For kNN we used the following heuristic to remove the outliers. On training data we computed the median of each class as its corresponding center. We classified the training points using those centers with k equal to 90% of the cluster size. We then removed those samples which were not classified correctly and found the centers again. This procedure was repeated several times. The resulting centers were able to classify more than 80% of the training data correctly, which we deemed sufficient to prevent overfitting. We used L1 norm as the measure of distance between points.

To assess the ability of classifiers to identify established signal patterns typically used in visual perception, such as the P100, N170, P250, we contrast the three models constructed over the full range of temporal data (0-700ms) with those constructed with data limited to specific temporal ranges. We chose the ranges of [50ms-100ms], [100ms-200ms], [200ms-300ms], and over 400ms.

We focused on the typically more challenging across-subject classification task where we test the ability of classifiers to generalize across different subjects. To accomplish this we train our classifiers on all but one subject and test the model’s performance on the EEG sequences of the remaining subject. We report the AUC (area under ROC curve) scores for all classifiers as a measure of their classification effectiveness.

4.3. Experimental Results

Table 1 summarizes predictive performance of the two classifiers across the subjects, as well as the average performance. Results for the baseline kNN approach were substantially worse across all temporal ranges. Average AUC scores were 51.63%, 54.79%, 56.16%, 63.51%, 58.65%, starting with the full all the way to ≥ 400 range. Without the aforementioned outlier removal technique in Sec. 4.2 the AUC scores downgrade to 48.29%, 47.26%, 51.46%, 54.8%, 52.70%, respectively.

The compound AUC measure suggests that both L1 and the group lasso can give effective prediction for across-subject predictions but varies significantly across subjects. Responses of some subjects, such as 2 & 3, can be effectively predicted using isolated spatio-temporal features of the L1 regression. On the other hand, responses of subjects 1, 4, and 5 are substantially better predicted using the group model, which signifies the importance of collective activity in select spatio-temporal regions. On average, simultaneous use of features across the full temporal range of 0-700ms leads to best performance, but this is also observed if the temporal range is restricted to 100ms-200ms with the L1 prediction, closely followed by the same restricted range with group lasso. This is clearly in agreement with the known role of the N170 potential in detection of human faces. No other

isolated temporal range has led to similar level of classification performance.

We next examined the spatio-temporal localization of features selected by our sparse predictors, for the across-subject setting. This is illustrated in Fig. 2. Results are shown for average models (over the five subjects) and, as an exemplar, for one specific subject (Subject 1). The two left most columns represent spatial locations and (relative) intensities of non-zero features, accumulated over the full temporal range. Namely, if $\beta_{i,t}$ is the weight of the (processed/filtered) response $X_{i,t}$ of channel i at time t , the weights displayed correspond to $\beta_i = \sum_t \beta_{i,t}$. The first of those columns indicates spatial locations of features present when subjects are exposed to 'face' stimuli, while the next column shows the same for 'vase' stimuli. The spatial locations show consistency both across subjects and across the two models (L1 and group lasso). Group models induce selection of larger spatial groups of electrodes, based on our definition of spatial neighborhoods. In general and as expected, L1 models reduce this spread. It is interesting that the sensors selected by our classifiers mostly lay in the occipital and temporal areas of the brain, the visual regions that have been implicated in object recognition.

The right-most column in Figure 2 displays the temporal distribution of weights, accumulated over all electrodes: the cumulative weight is $\beta_t = \sum_i \beta_{i,t}$, computed separately for all positive (face/blue) weights and negative (vase/red) weights. Again, we observe significant amount of consistency across five subjects. Moreover, the temporal placement of the selected weights shows interesting correlation with known indicators such as the N170 and P250, with the weight around N170 showing very strong peaks. We also observe a considerable concentration of 'face' weights around 400ms, which can be associated with N400. The third concentration around 600ms can be seen as a continuation of the N400. The temporal position of the 'vase' features exhibits very strong peaks slightly before the 'face' signals, around 120-130ms.

In the case of the L1 model, which enforces no temporal smoothness (grouping) of selected features, we also observe strong peaks in some subjects in the range of 50ms-100ms. Not unexpectedly, the temporal signatures show more variability than the corresponding temporal signatures of group lasso model. It should finally be noted that in the case of L1 model and subject 5 we observe temporal signatures that deviate from those in all other cases. This, together with the relatively low AUC scores for L1 on subject 5 and the high scores for the same subject with restricted temporal range (100-200), suggests the existence of spurious (outlier) features at fine temporal scales. However, group lasso successfully eliminated these outliers by requiring collective excitation over larger temporal windows.

It is important to stress that under either of the two mod-

eling approaches our data-driven models recovered specific stimulus signatures that are very closely associated with, but not identical to, established indicators. This is significant from two perspectives. First, it validates the analysis methodology presented here by showing that one can pick out accepted spatio-temporal correlates in a completely agnostic fashion. This is thus a potentially general methodology for use in identifying EEG markers for tasks that hitherto have not been associated with any distinct ERP components. Second, and more specifically, to the extent our results differ from the discrete ERP face correlates, they suggest that additional temporal epochs, besides just the 170 or 250 ms points, might carry information regarding face perception. Future investigations can probe whether this additional information is associated with aspects of face perception beyond just labeling a pattern as a face.

5. CONCLUSION AND DISCUSSION

We conducted L1 logistic regression and group lasso on a novel face perception experiment. The results show promising accuracy. Additionally, the learned weights by regression techniques assert the conventional markers of face perception besides pointing to other information rich zones of the EEG trace. An interesting zone is the period beyond 600ms emphasized by group lasso since it is mostly connected to face recognition [4] (although recognition of individual faces is not the subject of this study). Group lasso and L1 logistic regression display similar performance in terms of AUC, but complementary strength across different subjects. Group lasso often gives us more information regarding the spatio-temporal salient features due to extended temporal filtering. In general, the methodology presented in this paper is potentially a general one for identifying the spatio-temporal correlates of face-perception as well as other perceptual/cognitive tasks.

6. REFERENCES

- [1] J. V. Haxby, E. a. Hoffman, and M. I. Gobbini, "Human neural systems for face recognition and social communication," *Biological Psychiatry*, vol. 51, no. 1, pp. 59–67, Jan 02.
- [2] M. C. Moulson, B. Balas, C. Nelson, P. Sinha, and C. Scences, "EEG Correlates of Categorical and Graded Face Perception," *Journal of Vision*, vol. 8, no. 6, p. 533, 2008.
- [3] S. Bentin, T. Allison, A. Puce, E. Perez, and G. McCarthy, "Electrophysiological studies of face perception in humans," *Journal of Cognitive Neuroscience*, vol. 8, pp. 551–565, 1996.
- [4] M. Eimer, "Event-related brain potentials distinguish processing stages involved in face perception and recognition," *Clinical Neurophysiology*, vol. 111, no. 4, pp. 694–705, 2000.

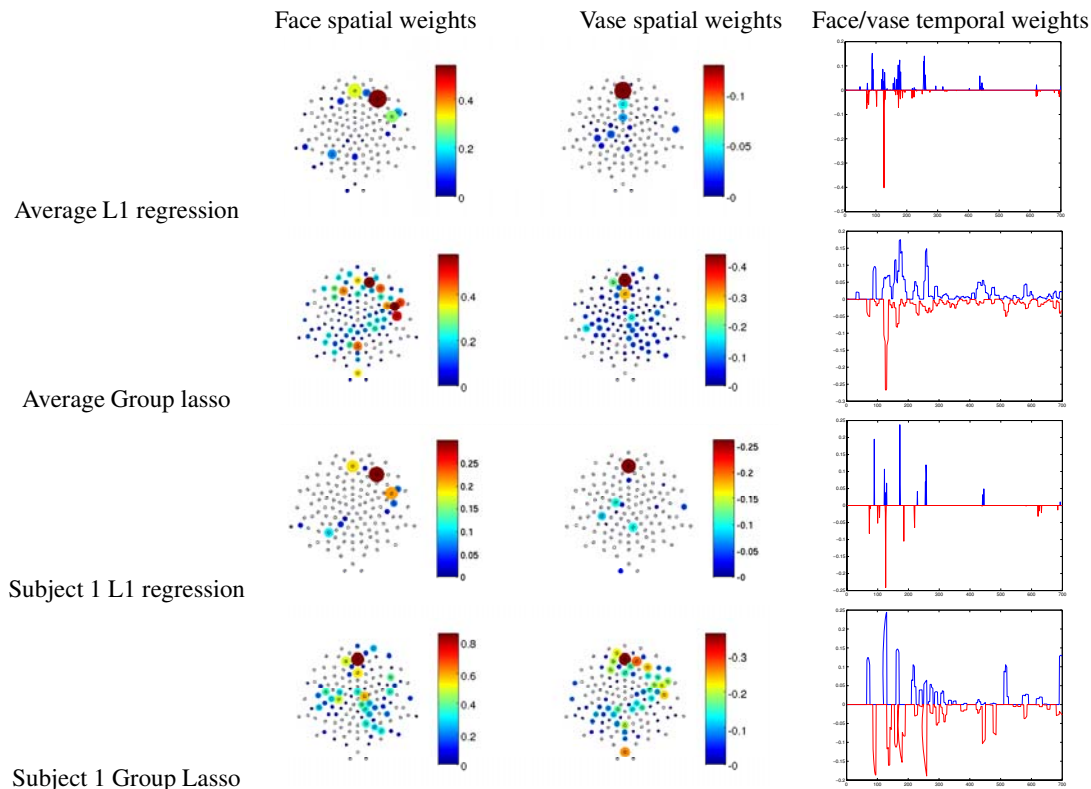


Fig. 2. Average and Subject 1 spatial and temporal distribution of positive (face/blue) and negative (vase/red) weights estimated by L1 logistic regression and group lasso models. Top of each sensor map corresponds to the occipital brain region.

- [5] J.-D. Haynes and G. Rees, “Decoding mental states from brain activity in humans.” *Nature Reviews. Neuroscience*, vol. 7, no. 7, pp. 523–34, 2006.
- [6] A. Yazdani, T. Ebrahimi, and U. Hoffmann, “Classification of EEG Signals Using Dempster Shafer Theory and a K-Nearest Neighbor Classifier,” *Signal Processing*, pp. 327–330, 2009.
- [7] A. Alkan, E. Koklukaya, and A. Subasi, “Automatic seizure detection in eeg using logistic regression and artificial neural network,” *Journal of Neuroscience Methods*, vol. 148, no. 2, pp. 167 – 176, 2005.
- [8] G. Garcia, T. Ebrahimi, and J.-M. Vesin, “Support vector EEG classification in the Fourier and time-frequency correlation domains,” in *First International IEEE EMBS Conference on Neural Engineering*, 2003, pp. 591–594.
- [9] U. Hoffmann, J.-m. Vesin, T. Ebrahimi, and K. Diserens, “An efficient P300-based brain-computer interface for disabled subjects,” *Signal Processing*, no. March, 2007.
- [10] M. van Greven, C. Hesse, O. Jensen, and T. Heskes, “Interpreting single trial data using groupwise regularisation,” *NeuroImage*, vol. 46, no. 3, pp. 665 – 676, 2009.
- [11] S. Makeig, A. Bell, T. Jung, and T. e. a. Sejnowski, “Independent component analysis of electroencephalographic data,” *Advances in Neural Information Processing Systems*, p. 145151, 1996.
- [12] S. Makeig, “Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones.” *Electroencephalography and Clinical Neurophysiology*, vol. 86, no. 4, pp. 283–93, April 1993.
- [13] R. Tomioka, S. Lemm, and M. Kawanabe, “Optimizing Spatial Filters for Robust EEG Single-Trial Analysis,” *IEEE Signal Processing Magazine*, pp. 41–56, 2008.
- [14] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [15] J. Friedman, T. Hastie, and R. Tibshirani, “Regularization paths for generalized linear models via coordinate descent,” *Department of Statistics, Stanford University, Tech. Rep.*, pp. 1–22, 2008.
- [16] M. Yuan and Y. Lin, “Model selection and estimation in regression with grouped variables,” *Journal-Royal Statistical Society Series B Statistical Methodology*, vol. 68, no. 1, p. 49, 2006.
- [17] V. Roth and B. Fischer, “The group-lasso for generalized linear models: uniqueness of solutions and efficient algorithms,” in *Proceedings of the 25th International Conference on Machine learning*. ACM, 2008, pp. 848–855.