Using spatiotemporally distributed patterns of intracranial encephalographic activity to classify previously encountered versus novel stimuli

Alexander Gonzalez, Alan Gordon, and Karen LaRocque

1. Introduction

Intracranial electroencephalography (iEEG) is a neuroscience technique in which electrical recordings are made from electrodes making direct contact with the human brain. While patients with intractable epilepsy are being recorded with pre-operative iEEG, they can opt to participate in cognitive research tasks, thus giving cognitive neuroscientists access to unique and spatiotemporally rich datasets.

An extensive literature from functional magnetic resonance imaging (fMRI) and scalp electroencephalography (sEEG) studies of "old/new recognition" suggests that a distinct set of subregions within the parietal lobe play a role in memory retrieval. Specifically, higher responses to studied relative to unstudied items have been reported in the angular gyrus (AnG) and the inferior parietal sulcus (IPS). A separate set of parietal regions, namely the temporo-parietal junction (TPJ) and superior parietal lobule (SPL), have been posited to play a role in attention rather than in memory retrieval.

Our research aims to advance this literature on memory and attention in the parietal lobe in several ways. First, we iEEG data has better spatial precision than scalp EEG data, and much better temporal precision than fMRI. Second, we employ machine learning algorithms to examine whether the ability of various components of these iEEG signals, such as frequency, power, and different time points, can allow us to perform above-chance classification of old *versus* new retrieval episodes in different parietal regions. Finally, we attempt to pool data from various electrodes in multiple ways in order to see if this improves classification performance rather than the more conventional (and less sophisticated) method of treating data from each electrode as independent observations.

2. Methods

2.1 Data Acquisition

In conjunction with the research labs of Anthony Wagner in the Stanford Psychology Department and Josef Parvizi in the Stanford Medical School, we have made iEEG recordings of one subject performing a recognition memory task.

Subjects are first presented with a "study" list of common words. Then, subjects are presented with a "test" list of words, half of which are repeated from the study list, and half of which are novel. For each test item, the subject is asked to indicate, with a button press, whether she remembers encountering the item in the previous task (an "old" judgment) or whether she does not remember encountering the item (a "new" judgment).

During the memory test, we collected data from a lateral parietal electrode grid in each subject. Each grid has 64 electrodes, which can be labeled as recording signal from AnG, IPS, TPJ, SPL, or a non-parietal region (figure 1).

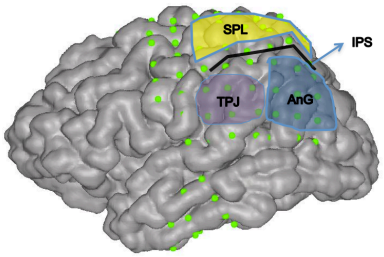

Figure 1: electrode placement

Our goal is to use the recorded electrical activity to classify trials accompanied by *correct* memory detection ("hits") with trials *correctly* identified as novel ("correct rejections"; "CRs"). We have 137 training examples (71 hits, 66 correct rejections)

2.2 Feature Space

For classification analyses that were run separately for each electrode, each training example is a 1000 ms voltage trace (measured in μV; referenced to a common average) that corresponds to the period of time 0 to 1000 ms after stimulus onset during which the subject is making a recognition memory decision. We parsed this trace into features in two different ways. First, we divided the trace into overlapping 100-ms time bins (20 bins) and calculated the mean amplitude for each time bin. These features were used in classification analyses using amplitude. Second, the continuous time course was decomposed into 42 frequency bands ranging from 0.1 to 209 Hz, with a sample compression rate of 7. The power of each band was calculated using a Hilbert transform and then epoched into the events of interest. As in the amplitude case, we divided each 1s event into 100ms bins were the 42 bands were furthered compressed into 6 bands of

interest, by taking the mean power of the band in the bin. These features were used in classification analyses using power. For any classification analyses that relied on multiple (*m*) electrodes, the features from these electrodes were concatenated into one vector yielding *m**20 features for amplitude classification and *m**20*6 features for power classification (or *m**20*42 if all frequency bands are being included).

2.3 Classification Performance Metric

As a metric of classification performance metric we used the signal detection sensitivity index d' = Z(true positive rate) - Z(false negative rate), where Z is the standard normal cumulative distribution function. We chose this metric because, unlike percentage correct, it is invariant to the response bias of a classifier.

3. Results

3.1 Amplitude- vs. frequency-based classification.

L2-regularized leave-one-out logistic regression classification of hits vs. correct rejections was significantly greater when frequency components were used, (max d' = 3.06) relative to amplitude components (max d' = 1.24). This result held even when amplitude was sampled with greater density in the temporal domain, to ensure that the number of features was constant across amplitude and frequency classification. Since frequency information codes information about memory status significantly better then amplitude information, the remaining analyses focused only on frequency data.

3.2 Feature and parameter selection

To assess how feature selection and regularization magnitude affected our classification performance, we sampled a wide space of feature selection and regularization parameters (fig. 2a). Filter feature selection was employed; mutual information between power in each feature across training patterns and the labels of each training pattern was computed. The extent of feature selection was parameterized by pF, where pF was the proportion of features with greatest mutual information scores included in the classification. pF was sampled at 10 points from 10^{-3} to 1, equidistant on a logarithmic scale.

L2 regularized logistic regression was accomplished by maximizing the likelihood function

$$l(\beta) = \sum_i [Y_i \log p(X_i) + (1 - Y_i) \log\{1 - p(X_i)\}] - \lambda\|\beta\|$$

with respect to β, where p is the logistic function

$$p(X_i) = \exp(X_i\beta)/\{1 + \exp(X_i\beta)\}$$

The extent of regularization is parameterized by λ, where greater values of λ result in greater penalization of β vectors with large L2 norms. We sampled λ at 10 points from 10^{-4} to 10^{4} equidistant on a logarithmic scale.

We found that feature selection significantly improved classification at a wide range of pF values (performance was optimal at a pF range of .01 - .46, see fig. 2b). Additionally, λ had a significant effect on classification performance; very low values of λ resulted in poor classification performance, while performance was optimal at a range of .006 - 2.07 (fig. 2c).

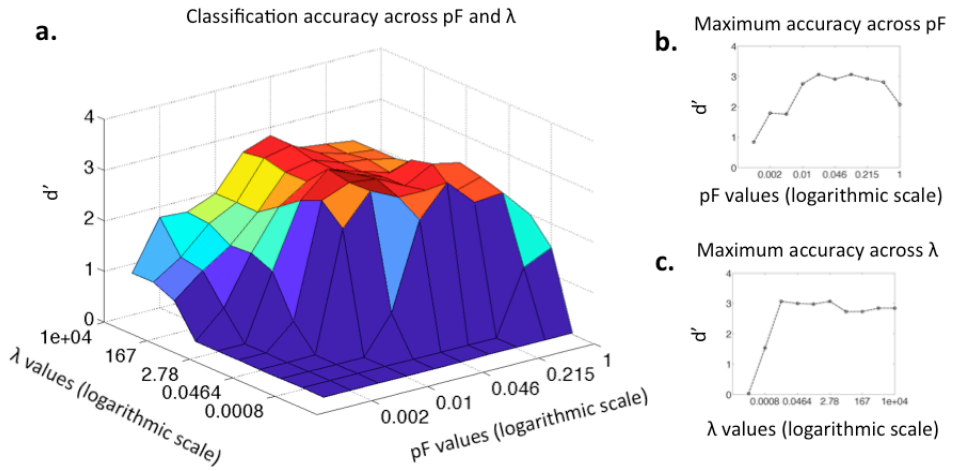

Figure 2: a) Classification performance across pF and λ values. b) Maximum performance across λ values for each given pF. c) Maximum performance across pF values for each given λ.

3.3 Feature importance

To visualize how features differentially drive classification, we created a map of classifier regression coefficients β for each feature. By our convention, positive values drive classification output towards 'hits' and negative values drive classification output towards 'correct rejections.'

A map of beta values across features (fig. 3) identifies two prominent clusters where greater activity predicts hits over correct rejections; a transient (600-800ms) cluster of features in low frequency IPS / AG, and a sustained (600-1000ms) cluster of features in gamma/high gamma anterior SPL.

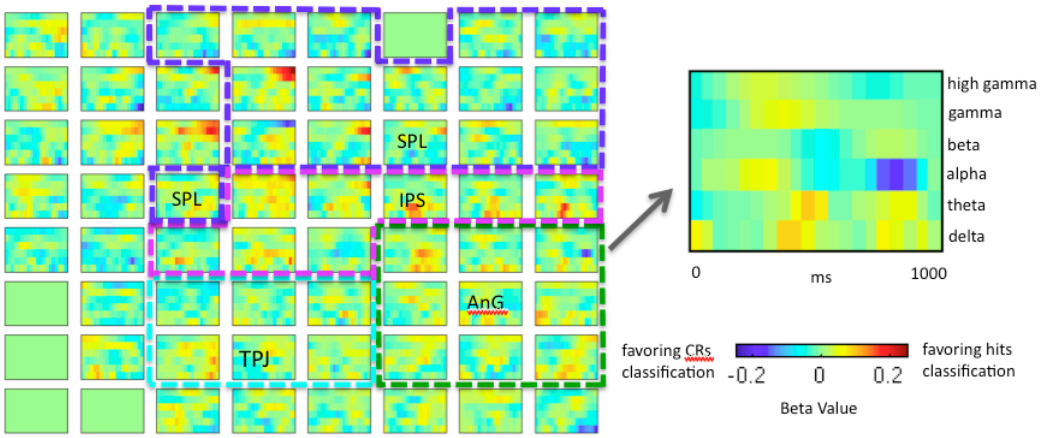

Figure 3: Beta map illustrating features that most strongly drive hits vs. CR classification.

3.4 Classification by Region of Interest

Logistic Regression classification was run again using individual frequency bands and subsets of electrodes that were chosen based on their anatomical location. The cost parameter λ was optimized (using leave one cross validation) for each classification and performance was measured using d' as shown in Fig 4. The SPL electrodes, in conjunction with the gamma band power yielded the best classification of old items versus new items. Gamma band in particular was able to classify better across all areas. The same classification analysis was run using a SVM with a Gaussian kernel (γ=1/#features, libsvm), yielding similar results to Logistic Regression.

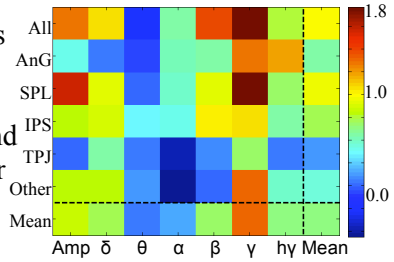

Fig. 4 Classification Performance Map

3.5 Multi-Task Classification

We implemented a small-scale multitask model in which a single classifier utilized training examples that were acquired from two different electrodes. One classification was performed for each pair of adjacent electrodes within the portion of the grid that covered our lateral parietal area of interest. Performance was subdivided based on whether the electrodes came from the same or different parietal subregions. Mean d' scores for classifiers utilizing training examples from the same parietal subregion (M = .20) was significantly greater than scores for data from different parietal subregions (M = -.03), $t(61) = 2.14$, $p < .05$. Moreover, mean d' scores of multitask classifiers utilizing training examples from the same parietal subregion was significantly greater than chance, $t(46) = 3.468$, $p < .005$, and significantly greater than mean d' performance of single electrode classifiers using the same training examples (i.e., classifiers that only use training examples from a single electrode; M = -.043), $t(139) = 2.71$, $p < .01$. Conversely, mean d' scores for multitask classifiers utilizing training examples from different parietal subregions was not significantly different from chance or from mean d' performance of single electrode classifiers using the same training examples (M = -.079), both $p > .50$. Thus, our multitask implementation suggest that single electrode classification can be improved by pooling training examples from electrodes within the same parietal subregion.

3.6 Reaction Time Regression

Epsilon-Support Vector Regression (Gaussian kernel, γ=1/#features, libsvm) was used to independently predict the subject's reaction time for hits and correct rejections. The regression was run across ROIs and different power bands. The cost parameter was optimized for each combination of ROI/band using leave one out cross validation. Performance was measured by the correlation coefficient between predicted reaction time and actual reaction time (MSE was also calculated). Figure 5a and 5b show the correlation regression maps for old items and new items respectively. Strong correlations for high gamma for both old and new items indicate that this frequency band is the best indicator of reaction time. Regions that predicted RT were SPL, IPS and the "other" category. The category "other" seen in these plots corresponds to electrodes not in the predefined ROI's, including some of motor cortex. Not surprisingly, these electrodes predicted reaction time very well, as indicated by their high correlation.

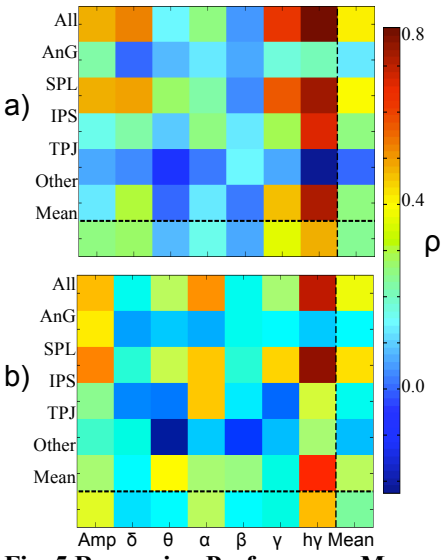

Fig. 5 Regression Performance Map

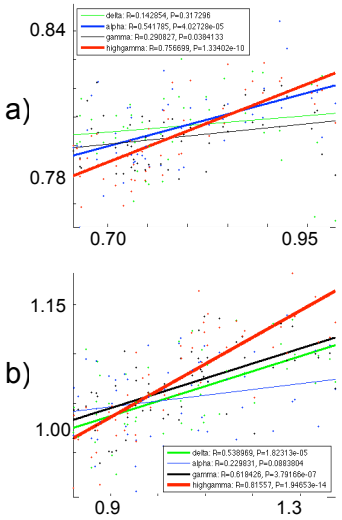

Fig. 6 SVR Correlation plots using all channels

We did not have a priori expectations that prediction accuracy would differ for old vs. new items, however we found that certain frequency bands and regions were able to more accurately predict reaction time for one class of trials relative to the other class. Figure 6 shows the scatter plots of actual reaction times (x axis) and predicted reaction times (y axis) for old items (6a) and new items (6b), using all channels. The thickness of the least squares fitted line indicates the strength of the correlation. Fisher's R to Z transforms indicate significantly greater correlations for new items in the Gamma and Delta bands ($p<0.05$), and marginally greater correlation ($p<0.07$) indicated greater correlations for old items in the alpha band.

3.7 Supervised and Unsupervised Classification of Anatomy based on Response Profile

We used supervised and unsupervised learning methods to determine the extent to which the response profile of an electrode could predict the anatomical placement of that electrode (i.e., whether that electrode recorded data from AnG, SPL, IPS, or TPJ). In these analyses the aggregate data from each individual electrode were used as training examples. Features were created either by collapsing frequency band * time-bin data over hit trials and correct rejection trials separately and then taking the difference of these trial types as features ('difference features) or by collapsing data over all trials regardless of mnemonic status ('collapsed features'). For supervised learning analyses, labels were anatomical location and cost parameters were determined based on the average cost parameter obtained from optimizing over all comparison types.

First, we performed logistic regression for each pair of anatomical regions. Results are displayed in Figure 7 (a: difference features, b: collapsed features). D' for the pair-wise classifications performed over difference features and over collapsed features (M=1.55, 2.23; SD = .77, .85. respectively) were significantly greater than chance, $t(5) = 4.92, 6.39$, respectively, both $p < .005$. There was no significant difference between classification performance based on the two feature types, $t(5) = 1.53$, $p = .19$. In order to assess whether to two classification models were relying on features corresponding to the same time-point and frequency, we correlated the weights of the two models. For all pair-wise classifications other than the AnG-SPL classification the correlation between model weights was significantly greater than chance (all $p < .05$), suggesting that the differences in old-new activity that differentiate between anatomical regions are found in the same frequency bands and time-bins in which trial invariant differences in activity differentiate between anatomical regions. Examination of the d' scores suggests that AnG electrode responses are more distinctive than those in other regions.

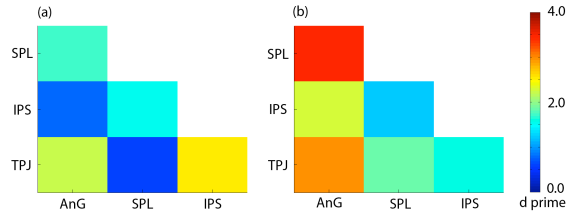

Figure 7. Pair-wise Classification of Region

Second, we used multiclass SVM to examine 4-way classification of anatomical region and computed a d' score for each region based on the rates of correct and incorrect labeling of that region. The d' scores obtained from classification performed over difference features and collapsed features (M = 1.35, 2.26, SD = .50, .84, respectively) were both significant above chance, $t(5) = 5.37, 5.39$, respectively, both $p < .05$, and did not differ from each other $t(5) = 1.47$, $p = .24$. Confusion matrices from this analysis are displayed in Figure 8 (a: difference features, b: collapsed features) and suggest once again that AnG is less confusable with other parietal regions than SPL, IPS, or TPJ.

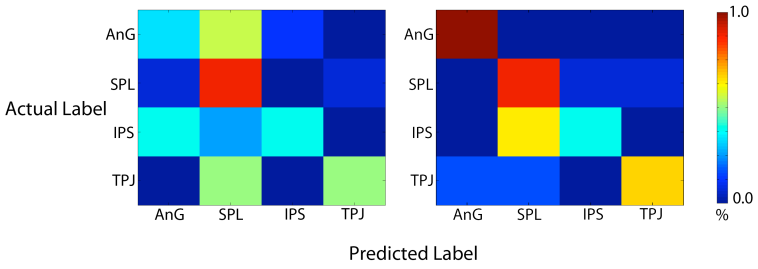

Figure 8: Multiclass Confusion Matrices

Finally, we performed 4-means clustering on the training examples to assess the extent to which unsupervised clustering mirrored anatomical clustering. Cluster assignments of each electrode are visualized in Figure 9 (a: difference features, b: collapsed features). We computed two metrics to assess correspondence between unsupervised clustering and anatomical clustering: the extent to which electrodes from a given region are placed in the same cluster (number of electrodes within a region placed in the modal cluster assigned to that region / total number of electrodes in that region) and the extent to which a cluster is populated by electrodes from the same region (the number of electrodes in a cluster that come from the predominant region in that cluster / total number of electrodes in that region). Clusters resulting from both difference features and collapsed features showed greater clustering on both metrics than would be expected by chance based on a null distribution generated by permuted data, all $p < .05$. A closer inspection of the extent to which electrodes from each region are placed in the same cluster revealed significant clustering for electrodes from SPL and TPJ using difference features, both $p < .05$, but no significant clustering of electrodes from AnG or IPS using difference features, $p = .14, .53$, respectively. Conversely, we observed significantly clustering for electrodes from AnG and TPJ using collapsed features, both $p < .05$, marginally significantly clustering for electrodes from IPS, $p < .10$, and no significant clustering for electrodes from SPL, $p = .48$.

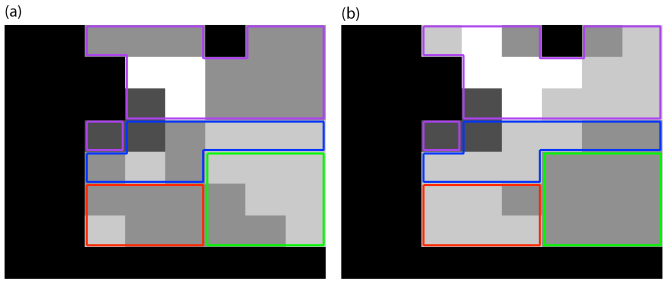

Figure 9: Unsupervised Clusters

4. Conclusions

This study demonstrated that the human parietal lobes code much information relevant to the true mnemonic state of an item. Memory state can be classified based on iEEG data with a d' of up to 3 (corresponding to an accuracy over 90%). Our analyses demonstrate that filter feature selection can increase discrimination performance, and that the features that most drive mnemonic classification include a transient low-frequency response in IPS/AnG and a high-frequency sustained response in anterior SPL.

Analyses of data from individual regions and frequency bands indicate that data from SPL is best able to discriminate mnemonic states and predict reaction time. Moreover, data from higher frequencies (gamma and high gamma) were more important in these classifications and regressions. These results indicate specificity and the importance of gamma power in parietal cortex for recognition memory processes.

Analyses in which we attempt to predict region based on the pattern of data from an electrode suggested that differential response patterns are produced by different parietal regions, and that this may be more true of certain subregions (e.g., AnG) than of others. The responses differentiate on both general stimulus-locked response and on difference in responses to old or new items, and examination of weights suggested that regions may maximally differentiate between old and new items in the same frequency bands and time bins in which they maximally differentiate from each other.

In sum, we sucessfully employed machine learning algorithms to predict of memory status, response latency, and anatomical location based on the response profile of electrodes placed on the lateral surface of the parietal lobe. These analyses will inform future exploration of parietal lobe function in human recognition memory.