

# Epilepsy Classification, EEG Analysis, and EEG-fMRI fusion

Kathryn Heal, Kaitlin Navarro, Margalit Wollner, and Eddie Yan  
Jérôme Gilles, Wesley Kerr, Pamela K. Douglas, and Travis Meyer

August 9, 2013

## Introduction

This study addresses three distinct challenges within the field of medical imaging.

We first approach the topic of epilepsy diagnosis from the supervised and unsupervised perspectives. Fludeoxyglucose Positron Emission Tomography (FDG-PET) scans are used to aid in the diagnosis of epileptic patients. Two challenges which are encountered during the interpretation of these scans are the need for expert analysis, and the level of difficulty in detecting the subtle changes in brain metabolism and structure. We aim to create a computer-aided classifier that incorporates Bilateral Temporal Lobe Epilepsy (BTLE) patients, leading to more robust and accurate classification and diagnosis. To this end, we explore various supervised and unsupervised algorithms. By observing the behavior of our classifiers we obtain novel pathological insight.

We then offer novel methods to visualize neural patterns. Electroencephalography (EEG) is widely used to obtain information about neural activity in a temporal context. In the field of neuroscience, various types of spectrograms resulting from continuous wavelet transforms are currently used for analyzing spectral patterns. In this study, we consider whether there exists a set of spectral bands which are superior to those currently used by neuroscientists to filter EEG signals before the data is interpreted. We offer a new set of adaptive spectral bands, and a corresponding adaptation of the Empirical Wavelet Transform (EWT), which could have useful applications to neuroscience. In addition to this, we introduce a new visualization tool which could offer a more precise representation of neural rhythms with respect to both time and frequency.

We finally pursue the fusion of information obtained from two different imaging modalities. As EEG and Functional Magnetic Resonance Imaging (fMRI) are two methods of recording neural activity, EEG-fMRI fusion is an area of active research. Since they are vastly different measures of activity, mere juxtaposition of the measurements is an insufficient comparison. The first challenge of comparison is that the measurements exist in two different spaces; the second is using a comparative metric that takes into account both spatial and temporal shifts in activity. We use the method of Standardized Low Resolution Brain Electromagnetic Tomography (sLORETA) to first transform the EEG into 3-d space. We then use Temporal Kernel Canonical Correlation Analysis (tkCCA) to measure the correlation between the EEG and fMRI.

# Chapter 1

## Epilepsy Classification

### 1.1 Temporal Lobe Epilepsy

Epilepsy is one of the most common neurological disorders, affecting approximately 1-2% of the general population [18]. It encompasses many disorders that are best characterized by recurrent and unprovoked seizures. Specific seizure types can be split into groups according to their features, but the disease as a whole is split into two main types: generalized and partial. Generalized epilepsies are characterized by seizures that begin in both hemispheres of the brain and do not have an identifiable seizure onset zone. In contrast, partial epileptic seizures originate in a localized section of the brain and can then spread to other regions. Reports show that Temporal lobe epilepsy (TLE) is the most common seizure subtype and comprises 50 to 60% of all focal epilepsies [2].

Positron Emission Tomography (PET) brain scans produce a 3-dimensional image of metabolic processes across the brain. The diagnosis of the disease relies on expert analysis of these scans. Areas of focal, asymmetric hypo metabolism indicate the location of a potential seizure onset zone. However, metabolic abnormalities may not be observable until 10 to 15 years after the onset of seizures due to the progressive nature of the disease.

In order to avoid misdiagnosis, it is crucial to distinguish between those with epilepsy and those with seizure-like disorders such as (Persons with) Non-Epileptic Seizures (NES). Non-epileptic seizures are psychiatric events that mimic many of the physical symptoms of epilepsy without having any of its electrographic features [17]. One third of "medication refractory" epileptic patients actually have NES [16]. Unfortunately, many patients with NES are diagnosed with temporal lobe epilepsy and are erroneously treated with Anti Epileptic Drugs (AED) [27], [8]. These drugs are accompanied with potentially fatal side effects and cause the patient to incur unnecessary treatment expenses (\$100,000/year). When medication fails to effectively control seizures, resective surgery often becomes the main treatment option [17]. More than this, it has been shown that surgery early on in the disease will be more effective in seizure control [21]. Finally, due to the high intrahemispheric hippocampal connectivity, unilateral disease may progress toward bilateral disease. Bilateral TLE patients are no longer eligible for surgery [17], as resection of the hippocampus would severely damage long-term episodic memory. This would leave the patient with permanent anterograde amnesia, unable to create new memories, similar to the famous case of HM [4].

The relationship between BTLE and other temporal lobe epilepsies remains unclear. It is possible that BTLE is independent of both Left Temporal Lobe Epilepsy (LTLE) and Right Temporal Lobe Epilepsy (RTLE), exhibiting its own unique patterns and characteristics. It has also been

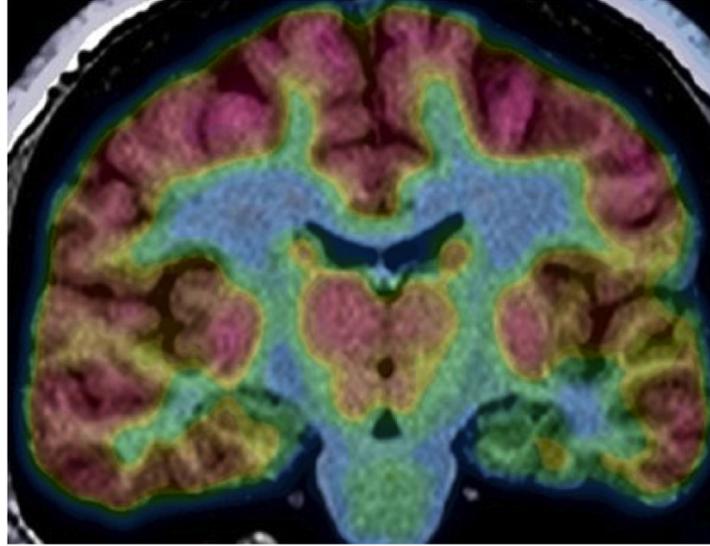


Figure 1.1: PET scan  
Keihaninejad et al. PLoS One 2012

suggested that Temporal Lobe Epilepsy (TLE) is itself a bilateral disease, but abnormalities may exhibit themselves more prominently in one lobe over the other [17]. If BTLE does share characteristics with both LTLE and RTLE, then the unilateral cases can provide important information for the classification of the bilateral case [22].

## 1.2 Project Goal

From previous work, we have a classifier that simultaneously diagnoses and lateralizes NES, LTLE, and RTLE with 76% accuracy [17]. Our goal is to incorporate BTLE patients into the existing classifier. We wish to supplement, rather than replace, expert analysis with computer aided classification. This can be used as an additional tool to either aid in detecting subtle metabolic abnormalities in PET scans or as a preliminary screen to later confirm findings made by neuro-radiologists. Either case would advance diagnostic capabilities and contribute to earlier and more effective treatment of patients.

Adding in data from BTLE patients increases our classification complexity as we have relatively few BTLE patient data to work with compared to other seizure groups ( $n_{BTLE} = 32$  vs  $n_{NES} = 32$ ). If the information from patients with LTLE and RTLE can help characterize the patients with BTLE, this limited data may be sufficient to build an accurate diagnostic tool.

## 1.3 Data

All of the 123 patients that were included in our analysis were admitted to the University of California, Los Angeles (UCLA) Seizure Disorder Centers video-EEG Epilepsy Monitoring Unit (EMU) between 2005 and 2012. Each patients diagnosis was based on a consensus panel review of their clinical history, physical and neurological exam, neuropsychiatric testing, video-EEG, interictal

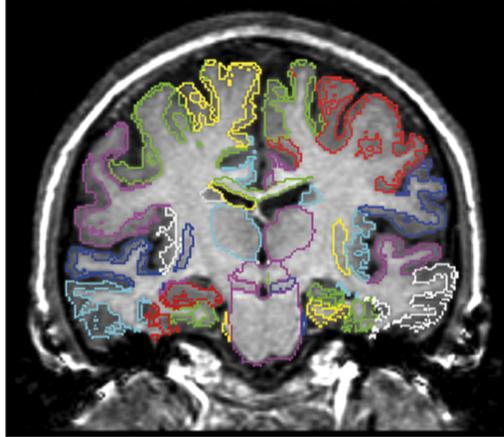


Figure 1.2: =  
Regions of Interest within the brain  
Engel et al. JAMA 2012

FDG-PET, ictal FDG-PET, structural and diffusion Magnetic Resonance Imaging (MRI) and/or CT scan. This multimodal assessment is the gold standard for epilepsy diagnosis and localization of the epileptic focus [19], [6]. The patients included in this analysis were chosen because they had an FDG-PET after 2005 and had no history of penetrative neurotrauma, including neurosurgery. These patients were diagnosed either with non-epileptic seizures, left temporal lobe epilepsy, right temporal lobe epilepsy, bilateral temporal lobe epilepsy, or unspecified temporal lobe epilepsy.

Patient Type	Number of Patients
NES	32
LTLE	39
RTLE	34
BTLE	14
UTLE	5

PET and MRI images were acquired according to the best clinical practices at the time of acquisition. PET/CT studies were acquired using a Siemens Biograph scanner. After a minimum fasting period of 6 hours, patients received 0.14 mCi/kg of  $^{18}\text{F}$ -FDG-PET intravenously. During the ensuing 40 minute uptake period with concomitant EEG monitoring to confirm interictal status, the patients waited in a quiet, dimly lit room with their eyes open. PET images were reconstructed with an iterative algorithm (OSEM: 2 iterations, 8 subsets). CT images were reconstructed using filtered back projection at 3.4 mm axial intervals to match the slice separation of the PET data, and used for attenuation correction. PET images were determined to be interictal by clinical findings and concurrent scalp EEG.

## 1.4 Data Processing

The number of voxels in the PET image is orders of magnitude larger than the number of samples, therefore we have insufficient data to effectively characterize the diagnostic effect of each

voxel. Therefore, we averaged the metabolism of voxels within 47 anatomically defined regions of interest as defined by NeuroQ. This serves as a preliminary biologically-motivated feature selection method that will likely improve our performance.

## 1.5 Methods

### 1.5.1 Cross Validation

The performance of each of our classifiers was evaluated using Cyclical Leave-One-Out Cross Validation (CL1OCV). In CL1OCV, a single data entry is removed and considered the validation data (leave-one-out) while the remaining data is considered the training data. The process cycles through the data such that each data point is treated as the validation once.

We compared the performance of each classifier that of the naïve classifier using a 95% Confidence Interval (CI). The naïve classifier classifies each patient as the most frequent class of the data.

We did not train our classifiers on unspecified patients—classifiers that incorporate unspecified patients incorporate them only as part of the test set.

### 1.5.2 The Permutation Test

In the case where there is bias in our results, specifically where feature selection was based on cross-validation accuracy, we determined the significance of our results using the permutation test. The permutation test randomly permutes the labels of the training data and reruns feature selection on the scrambled data. By comparing the performance of feature selection on the scrambled data with the performance of feature selection on the unperturbed data, we determined if the features selected were meaningful.

### 1.5.3 Clustering Methods

#### Unsupervised Clustering

Unsupervised clustering methods attempt to separate the data into classes without knowledge of the true labels of the data during the clustering step. We attempted classification based on a majority-vote method (knowledge of the true labels is incorporated at this step) with both  $k$ -means and spectral clustering. Both clustering techniques while unsupervised, differ in their methods of separating data.

**$k$ -means Clustering**  $k$ -means determines the clusters by assigning each point to the cluster with the closest floating centroid (center of the cluster). The location of the centroid is recomputed iteratively based on the data points assigned to its cluster. Because the centroids are initialized randomly to observed points in the data, the  $k$ -means algorithm is not deterministic and the identity of the clusters may vary wildly between successive executions on the same dataset. In practice,  $k$ -means is usually executed repeatedly and the result that best clusters the training data is selected. The best-case cluster is determined by some criteria such as minimum total intra-cluster distance. We repeated the  $k$ -means clustering process 20 times each time the algorithm is used—this is important as  $k$ -means is also used as a step in the spectral clustering process.

As part of a classifier,  $k$ -means was used to cluster the training data with the intention of producing clusters that meaningfully correspond to the training labels in some way. (As the labels

of the data are not explicitly associated with the distance function used by  $k$ -means, this far from guaranteed.) The test point is classified by deriving which cluster of the test data it belongs to and assigning it a label determined by the majority of the cluster (majority vote). In the case of  $k$ -means, this is trivially done by applying the rules of the algorithm (assign the test point to the cluster with the nearest centroid).

**Spectral Clustering** Spectral clustering is defined by two main algorithms, normalized spectral clustering and unnormalized spectral clustering. These algorithms can be thought of as relaxed versions of the RatioCut and Ncut problems, respectively [29].

Spectral clustering plays the same role in an unsupervised classifier that  $k$ -means does. Spectral clustering, however, determines clusters using the eigenvectors of the graph laplacian  $L$  derived from similarity (in the form of a similarity matrix  $W$ , see [29] for details) whereas  $k$ -means essentially uses the inverse of a similarity metric (distance). The choice of similarity is a crucial decision, and is a challenging problem in and of itself beyond the scope of this report. We use Gaussian similarity, defined as:

$$w = e^{-\frac{(x_1-x_2)^2}{2\sigma^2}}.$$

For classification, we used a majority vote process similar to that used with  $k$ -means, but deriving the cluster of the test point without explicitly clustering it with the training data is slightly more involved. For our purposes, we consider the RatioCut and Ncut objective functions the normalized and unnormalized spectral clustering cases. We assume, that because the types of spectral clustering used were relaxed form of the cut problems, that they provide a good approximation of spectral clustering. To define the RatioCut and Ncut objective functions, we first consider the adjacency matrix  $W$  where  $w_{ij}$  gives the weight of the edge connecting vertices  $i$  and  $j$ . From this definition of the adjacency matrix, we adopt the notation used by Luxburg *et al.*:

$$W(A, B) := \sum_{i \in A, j \in B} w_{ij}.$$

The definition of a cut that produces partitions of the graph  $A_1, \dots, A_k$  then follows as

$$\text{cut}(A_1, \dots, A_k) := \frac{1}{2} \sum_{i=1}^k W(A_i, \bar{A}_i).$$

From the definition of a cut, RatioCut and Ncut are defined:

$$\text{RatioCut} := \sum_{i=1}^k \frac{\text{cut}(A_i, \bar{A}_i)}{|A_i|}$$

$$\text{Ncut} := \sum_{i=1}^k \frac{\text{cut}(A_i, \bar{A}_i)}{\text{vol}(A_i)}.$$

where  $|A|$  is the number of vertices in  $A$  and  $\text{vol}(A)$  is the sum of the edge weights in  $A$  [29]. Note that to minimize the value of the objective functions, the clusters found should have low similarity with each other. Typically, these objective functions are used to determine the graph cuts given a number of desired partitions. In the case of classification, because the graph cuts are known *a priori*, we can use the RatioCut and Ncut values to determine the appropriate cluster. That is, we tentatively add the test point to each of the clusters and compute the corresponding objective function and ultimately choose the cluster that minimized the objective function.

**Spectral Learning** “Spectral learning” as defined by Kamvar *et al.* attempts to the ground truth of data in spectral clustering by enforcing a similarity of 1 when two vertices occupy the same class and 0 otherwise. We used this method alongside unsupervised spectral clustering.

**Choice of  $k$**  Because the choice of  $k$  is made *a priori* and is not limited to 2 in the case of  $k$ -means and spectral clustering, it is possible to use these clustering methods in two ways: the  $k = 2$  case, and the  $k > 2$  case. The  $k > 2$  case simply corresponds to majority vote system described above, while the  $k = 2$  case corresponds to a decomposition problem equivalent to that used by most supervised clustering algorithms. A  $k$  value not equal to the number of classes can be considered as an attempt to capture subpopulations within each class.

#### 1.5.4 Neural Networks

A neural network is an information processing structure arranged as interconnected nodes in successive layers. Each layer of the neural network have outputs connected to the following layer (feedforward) and the previous layer (backpropagation). The output of a given layer is determined by the weights associated with each of the nodes in the layer—these weights are readjusted by the backpropagation of error through the network. We used the neural network implementation in the MATLAB® neural network toolbox with `trainlm` as the training function. The only additional parameters specified are the number of hidden nodes per layer and the number of hidden layers. We used both a one versus all and a modified one versus all model in order to adapt our multi-class problem to neural networks.

##### One vs. all and modified one vs. all

The one vs. all model corresponds to the following labeling of classes:

NES		1000
LTLE		0100
RTLE		0010
BTLE		0001

The modified one vs. all model operates under the assumption that BTLE is the combination of LTLE and RTLE:

NES		100
LTLE		010
RTLE		001
BTLE		011

In the modified one vs. all case, we classified an instance of test data as BTLE if the initial classification (based on the position of the maximum element of the output vector) was LTLE or RTLE *and* the ratio between the LTLE and RTLE components of the output vector is within a certain heuristically determined threshold  $[3/4, 4/3]$ .

## Optimizing the number of nodes and layers

We used a grid search method to find a suitable number of hidden nodes and layers, but we tend to favor arrangements with fewer nodes per layer due to our limited computational time. We considered a search space of up to 6 hidden layers and 30 nodes per layer. For the sake of computational expense, we use an architecture with the same number of nodes for every hidden layer. Note that the performance of a given configuration is variable in the grid search as MATLAB’s neural network implementation is not deterministic with respect to the training step.

### 1.5.5 Support Vector Machines

We implemented a soft margin support vector machine (Cortes & Vapnick 1995) using the LibSVM package for MATLAB® (Chang *et al.* 2011) [3] where the objective function of the hyperplane is:

$$\frac{1}{2}\mathbf{w}^2 + CF \left( \sum_{i=1}^l \xi_i \right)$$

subject to the constraints

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \\ \xi_i \geq 0$$

Where each instance of data  $x_i$  has a label  $y_i \in \{-1, 1\}$ ,  $C$  is a constant,  $F$  is a monotonic convex function, and the sum of  $\xi_i$  is the sum of training errors [5].

In accordance with the guide developed by (Hsu *et al.* 2003), we used a linear kernel instead of the default radial basis function, as it is suggested that a non-linear remapping of high-dimensional data does not seem to improve performance [13]. All other LibSVM parameters were defaults.

### 1.5.6 Feature Selection

Feature selection is motivated by several problems associated with data with a large number of dimensions (in this case averaged and normalized brain radioactivity for each Region(s) of Interest (ROI). Namely, we wish to avoid the “curse of dimensionality” and eliminate features that are irrelevant to the classification problem.

We distinguish between two classes of feature selection techniques: “supervised” and “unsupervised.” In general, we refer to feature selection techniques that are tied closely to classification performance as “supervised” and techniques that do not consider performance directly as “unsupervised.” Unsupervised techniques are generally based on statistical parameters, such as Mahalanobis distance, covariance (Principal Component Analysis (PCA)) [26], variance (PCA) [26], and joint probability density (Independent Component Analysis (ICA)) [14].

#### “Unsupervised” Feature Selection

ICA is a linear projection designed to separate mixed signals into independent components that were not directly observable in the data. Instead of diagonalizing the covariance matrix, the independent components are constrained to be maximally statistically independent. To accomplish this, we used the iterative fastICA algorithm in MATLAB (Mathworks, MA) [14]. In our case, this means separating the overall radioactivity/metabolism in individual regions of interest into what

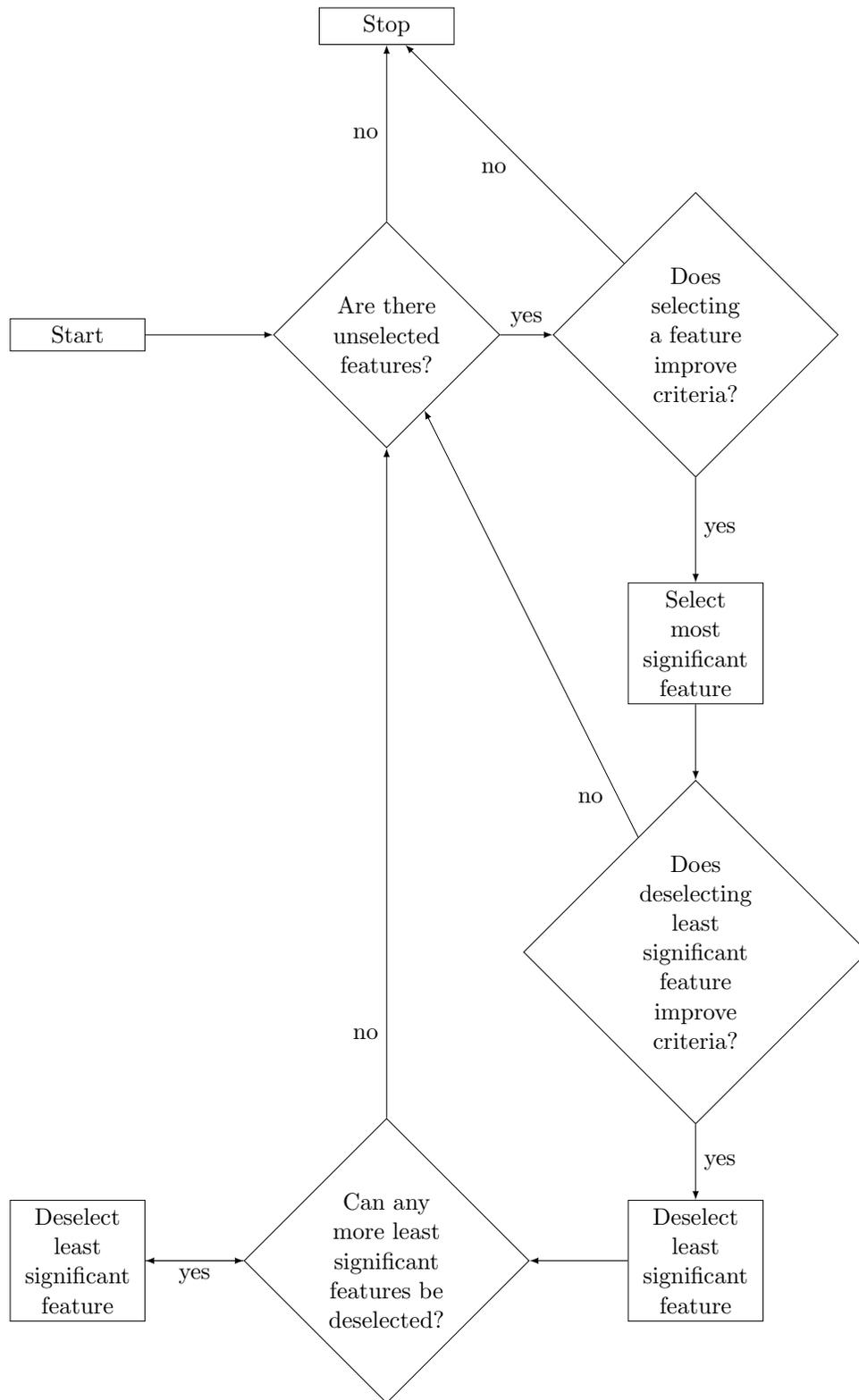
we think represent metabolic networks across the brain, which we can then examine for patterns that may differ between seizure groups. The number of independent components we select allows us to control the size of the brain networks we will focus on: fewer ICs means larger networks, while more ICs means smaller networks [14].

PCA works to reduce the dimensionality of our data projecting it onto a lower dimension, high variance, subspace. This method assumes that noise has low variance compared to meaningful features. A linear transformation is applied to the data that diagonalizes the covariance matrix. The eigenvalues are then ordered such that the first few principal components will contain as much of the original data information as possible [25].

### **“Supervised” Feature Selection**

We explored the use of several “supervised” feature selection algorithms: (Sequential Forward Selection (SFS), Sequential Backward Selection (SBS), Sequential Floating Forward Selection (SFFS), Sequential Floating Backward Selection (SFBS)). SFS and SFFS share a *bottom-up* approach where features are sequentially added while SBS and SFBS share a *top-down* approach where features are selectively removed. That is, each algorithm attempts to find a *subset* of features that give the best accuracy in classification. [24] Pudil et al. introduced the concept of *floating* to sequential feature selection so that features may be added when previously removed in the case of SFBS and vice versa in the case of SFFS. [24]. The algorithm stops when no more features can be added without lowering the value of the objective function (bottom-up) or when no more features can be removed without lowering the value of the objective function (top-down).

Figure 1.3: Simplified Flowchart for SFFS



## 1.6 Results

### 1.6.1 Clustering

We present the results of  $k$ -means and spectral clustering in the case of a trinary classifier (excludes BTLE).

#### Unsupervised clustering

Classifier	$k$	Accuracy
$k$ -means	3	35.238%
$k$ -means	5	20.952%
$k$ -means	10	29.524%
Unnormalized spectral clustering	2	38.095%
Normalized spectral clustering	2	26.667%
Unnormalized spectral clustering	3	39.048%
Unnormalized spectral clustering	5	37.143%
Unnormalized spectral clustering	10	35.238%
Normalized spectral clustering	3	39.048%
Normalized spectral clustering	5	33.333%
Normalized spectral clustering	10	39.048%
Naïve	N/A	<b>37.143%</b>
Naïve 95% CI upper bound	N/A	<b>46.385%</b>

No unsupervised method was able to reliably outperform the naïve classifier.

#### Supervised clustering/spectral learning

Classifier	k	Accuracy
Unnormalized spectral clustering	2	39.048%
Normalized spectral clustering	2	39.048%
Naïve	N/A	<b>37.143%</b>
Naïve 95% CI upper bound	N/A	<b>46.385%</b>

No supervised spectral clustering method was able to reliably outperform the naïve classifier.

### 1.6.2 Neural Networks

We employed a grid search in both the one versus all case and the modified one versus all case.

Figure 1.4: One vs. All

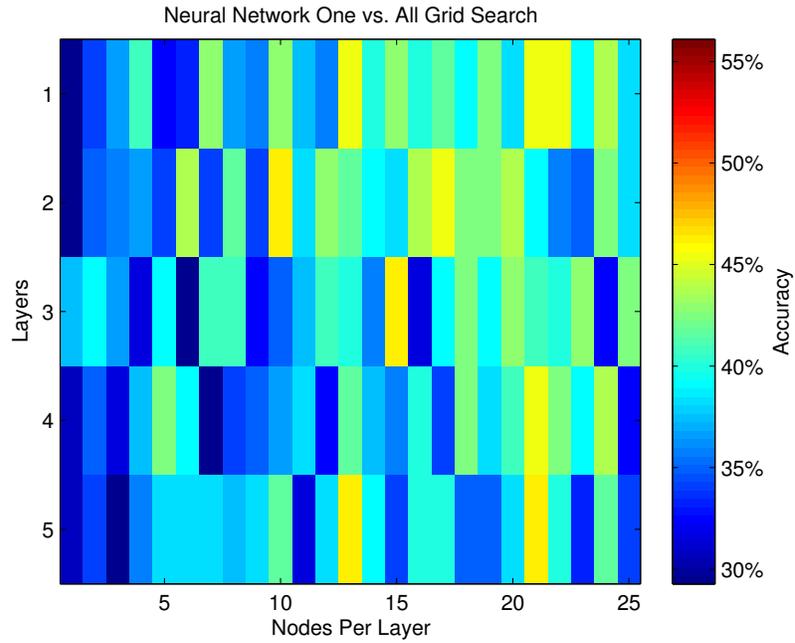
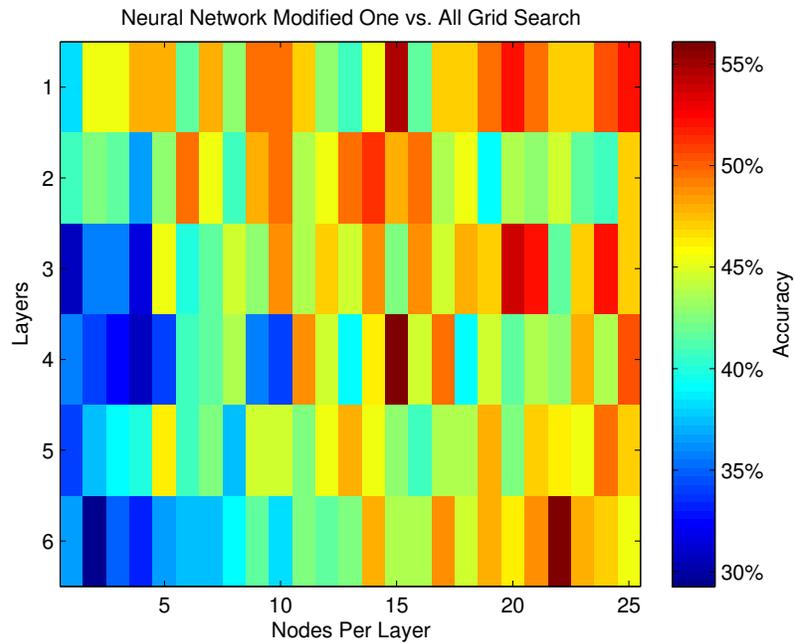


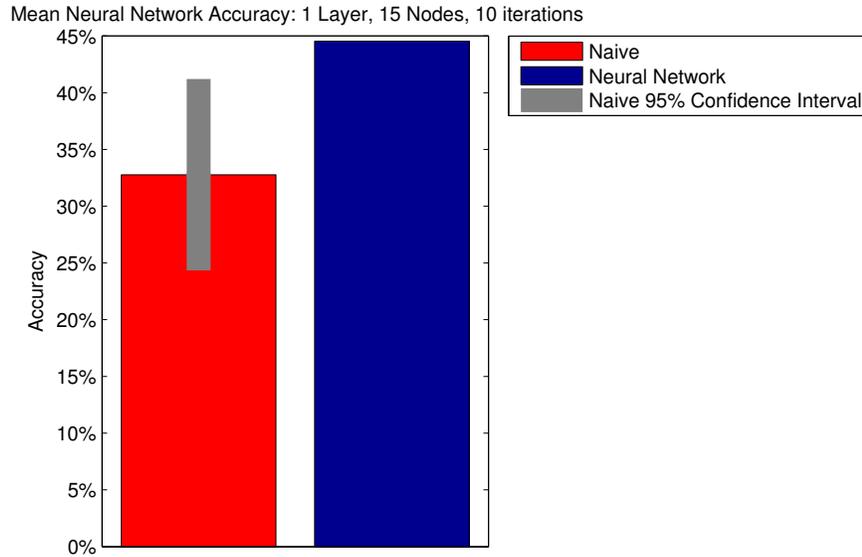
Figure 1.5: Modified One vs. All



The results of both grid searches seem to favor an increased number of nodes per layer, followed by an increased number of layers. This trend is somewhat more pronounced in the modified one

vs. all case. We look in detail at one of the hotspots of the grid search from the modified one vs. all case—1 layer and 15 nodes:

Figure 1.6: 1 Layer, 15 Nodes, 10 iterations

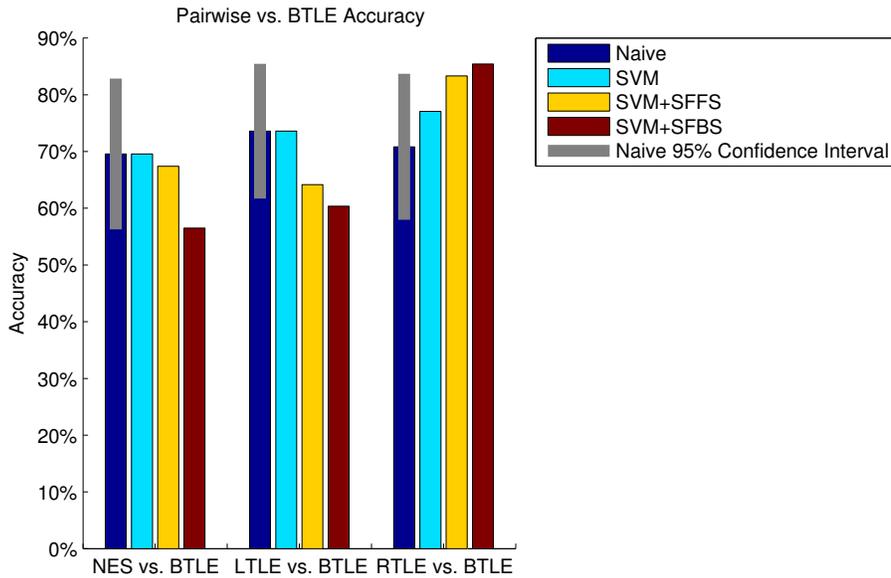


The relatively higher accuracy seen for this configuration of nodes and layers now seems to be an outlier when average over multiple iterations. Neural networks manage to outperform the Naïve classifier, but not by an impressive margin.

### 1.6.3 Support Vector Machines (SVM)

As multi-class support vector machines are built on binary classifiers, we first consider the case of using a support vector machine to distinguish between BTLE and NES, LTLE, and RTLE. Note that these pairs represent only a *subset* of all the classifiers needed to build a single multi-class classifier with these four classes.

Figure 1.7: SVM: Pairwise binary classifiers

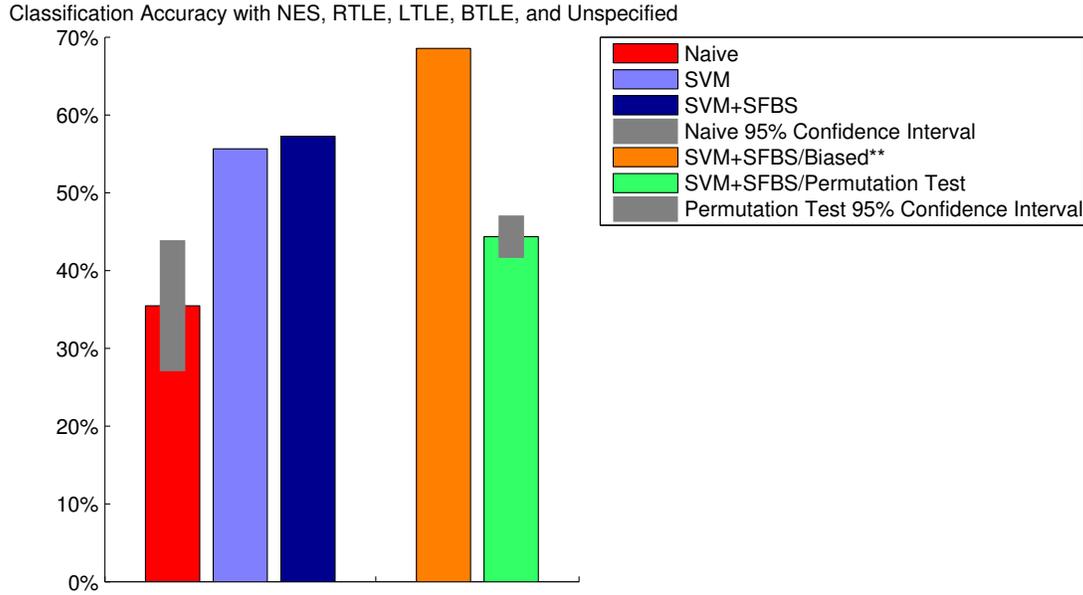


\*+SFBS,SFBS indicate that the corresponding feature selection algorithm was used.

Our classifier was not able to outperform the naïve classifier in the case of LTLE versus BTLE and NES versus BTLE. In the RTLE versus BTLE case, feature selection considerably improved performance and was able to achieve an accuracy of about 85%.

Next, we present the quaternary classifier, which attempts to distinguish between NES, LTLE, RTLE, and BTLE. This classifier also incorporated the unspecified patients. The unspecified patients do not affect the training of the classifier, and they only affect the feature selection of the classifier in the *biased* case.

Figure 1.8: Quaternary Classifier



\*+SFBS indicates feature selection was used. \*\*Biased indicates feature selection trained on cross-validation accuracy.

**The biased case** Normally, feature selection is performed during the training phase of the algorithm, and feature selection is used to maximize some criteria such as training accuracy or a nested cross-validation accuracy. In the biased case, in order to consider the features most relevant to classification, we instead selected features based on the CLIOCV accuracy. To show that these selected features are meaningful in terms of the actual classes of the data, we perform the permutation test.

**Features Selected** We find that SFBS was the most successful feature algorithm we implemented, likely because it is the least susceptible to overfitting (both because of its top-down approach and because of its floating property). Using SFBS, we present a ranked list of the size most significant features with respect to classification performance in the quaternary case.

Rank	Feature
1	right pos Medial Temporal Cortex
2	right inf lat ant Temporal Cortex
3	right Parietotemporal Cortex
4	right sup lat Temporal Cortex
4	right inf Frontal Cortex
4	left inf lat pos Temporal Cortex

## 1.7 Discussion

**Unsupervised Classifier** Results with unsupervised clustering algorithms indicate that traditional similarity measures do not generalize well to FDG-PET data. We suspect that this is the result of Euclidean distance not generalizing well to our data, due to its high-dimensionality and the significance of each dimension not being known *a priori*. If unsupervised clustering algorithms could be adapted to the epilepsy problem, it would be interesting to further investigate the values of  $k$ , as  $k$  can be chosen in an attempt to identify new and different ways that epilepsy may be expressed within the traditional classes.

**Supervised Classifier** We find that spectral learning is best suited for the case where the data is a *mixture* of labeled and unlabeled data and the behavior of the similarity function is well understood with regard to the labels. In our case, there is a ground truth for almost all the data, so clustering in the case where the number of clusters  $k$  is the same as the number of distinct classes is a trivial problem where classification is reduced to choosing the cluster with the highest average similarity to the test point. This is roughly equivalent to the  $k$ -means algorithm, as essentially all of the steps of the spectral clustering algorithm are lost. On the other hand, choosing a number of clusters  $k$  greater than the number of classes does not improve this scenario as it is not meaningful to attempt to cut a connected component where all of the edges have weight 1. Results with spectral learning agree with our doubts regarding the applicability of this supervised technique with our data.

**Neural Network Grid Search** From the overall trends shown in the grid search, it appears that the representation of BTLE as a combination of both LTLE and RTLE gives better overall classification performance—suggesting that this may be a better characterization of the disease. We suggest that given the availability of more computational time, that the grid search method be modified to consider the *mean* performance of each configuration due to the variability seen in our investigation of a configuration that seemed to perform well initially. Though neural networks gave considerably better results than both supervised and unsupervised clustering algorithms, it was not computationally feasible for us to consider more exotic neural network architectures or feature selection that we implemented with LibSVM.

**Quaternary Classifier** These results seem to indicate that SVM were not able to extract meaningful trends in the data in the case of RTLE and NES. That SFFS and SFBS subsequently worsened the accuracy of the classifier suggests that feature selection overfit the data. It is interesting to note, in the RTLE vs. BTLE classifier, however, that SVM found meaningful trends initially and that feature selection managed to select even more meaningful features. It may be possible that this is due to some similarity in the underlying pathology of BTLE to both NES and LTLE but not RTLE. This would be expected if BTLE could be likened to a form with LTLE with a higher level of symmetry.

**“Unsupervised” versus “Supervised” Feature Selection** Overall, we found that unsupervised feature selection failed to extract tractable patterns in our data. We hypothesize that the ineffectiveness of “unsupervised” techniques is the result of classification performance being dependent on *subtle* trends expressed by *networks* of brain regions, and not necessarily by the most

*separable* components or networks. Additionally, “unsupervised” factor analysis introduces the problems of choosing the number of and components to use—which are also nontrivial.

**SFBS Features** We caution that the rank order of features found by SFBS not be weighted too heavily, as the *floating* nature of the algorithm emphasizes finding groups of features (which correspond to brain *networks* in our case) rather finding features strictly by their individual significance.

## 1.8 Conclusion

We find that unsupervised classification techniques, in their current form, are inadequate for the machine learning diagnosis of epilepsy. Overall, we find the combination of support vector machines and supervised feature selection to be the most promising direction for machine learning diagnosis. Perhaps with the availability of a larger and more comprehensive dataset, the generality of support vector machines in the diagnosis of temporal lobe epilepsy can be further characterized.

## Chapter 2

# Empirical Wavelet Transform

### 2.1 Introduction/Motivation

When a subject is presented with a visual stimulus (in our experiments, a flash of a symbol in a darkened environment), different parts of their brain emit waves of certain frequencies. A system of 256 extracranial EEG sensors, or channels, pick up these signals and record them as one-dimensional signals in the time domain. Each experiment is temporally partitioned into 144 successive epochs, each lasting for ten seconds. The flash is presented at second two of each epoch. This process is shown in Figure 2.1.

Many times analysts will examine the EEG signals using signal processing techniques, searching for patterns in the data. Neuroscientists primarily use the spectral bands. Many times it is useful to apply a band pass filter on the original time-domain signal to extract relevant information about the aspects of each wave which correspond to each of the five historically defined spectral bands.

We will refer to the  $\delta, \theta, \alpha, \beta, \gamma$  arrangement, shown in Table 2.1, as the “traditional” set of spectral bands. After discussing our novel boundary search methods purely within the context of signal processing, we will provide our results and discuss their potential relevance to the field of neuroscience.

One pattern which is of particular interest within the context of this experiment is that of the alpha power over time, especially in regions near the visual cortex. We expect the alpha power to drop immediately following the presentation of the flash, since the brain is alerted by the stimulus.

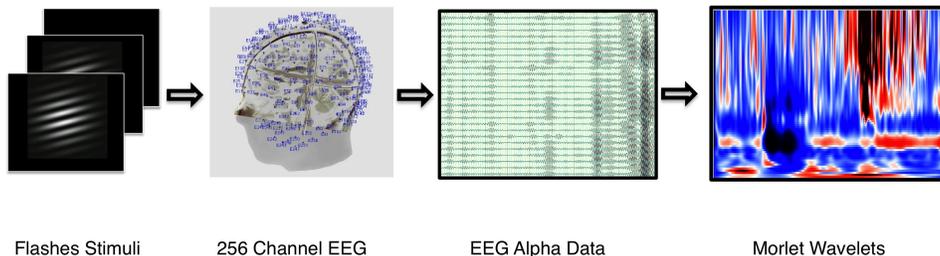


Figure 2.1: EEG Process

Band Name	Frequency Range (Hz)
Delta	approx. 0.1 – 4
Theta	4 – 8
Alpha	8 – 13
Beta	13 – 30
Gamma	30 – variable

Table 2.1: Traditional (Fixed) Spectral Band Definitions

After some time, the region of the brain which is activated by the visual stimulus will return to its resting state. Consequently, the alpha power will return to its baseline amplitude. In our experiment we jitter the flash times to avoid anticipatory brain activations.

## 2.2 Previous Work

Wavelet transforms are commonly used to decompose EEG signals. Typically Morlet wavelets are preferred in the continuous wavelet transform, and the spectrogram is the visualization method of choice. Figure 2.2 is the spectrogram which results from a continuous wavelet transform, whose Morlet wavelets are determined by a continuous sweeping of the dilation coefficient.

In general, classical wavelets provide some sort of prescribed partitioning strategy. We use empirical wavelets instead because they allow us the flexibility to specifically select the set of boundaries  $\omega_n$ .

Figure 2.3 depicts the empirical wavelet construction process. First, a set of boundaries  $\omega_n$  are defined. They can be chosen according to either a prescribed scheme (rigid wavelets) or an adaptive scheme, which we will discuss in depth in this paper. The interval between two successive boundary points will roughly define the support of each empirical wavelet. This set of empirical wavelets will serve as band pass filters. As we would expect, the band pass filters will be directly influenced by the choice of spectral band boundaries  $\omega_n$ .

In [11] the EWT offers an alternative to prescribed wavelet transforms. Part of the EWT method uses boundary search methods to choose wavelet supports which are adaptable to the particular signal at hand. Given a target number  $N$  of spectral bands, one boundary search method included in [11] finds the boundaries of the  $N+1$  largest local maxima of the frequency spectrum of the signal. Between each two successive local maxima, the algorithm finds the global minimum, and defines the resulting value as the boundary  $\omega_n$ . The second boundary search method featured in [11] again finds the  $N+1$  largest local maxima of the frequency spectrum, but then defines the boundary  $\omega_n$  as the midway point between the successive local maxima.

What motivates these methods is the idea that when looking at the frequency spectrum of the signal, each mode corresponds to a complete “mountain”. We would consider Figure 2.4(a) to be a desirable choice of boundaries. Figure 2.4(b), on the other hand, would be an undesirable choice

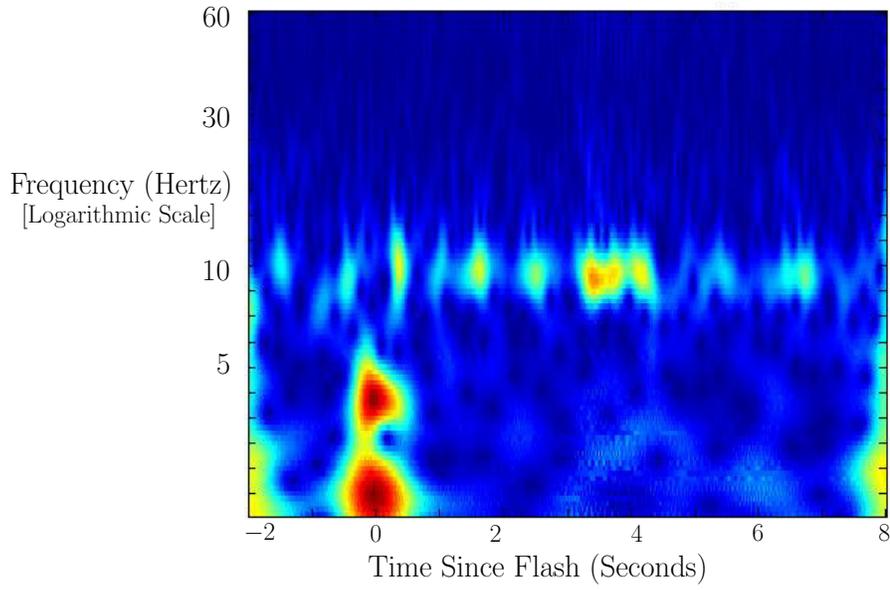


Figure 2.2: Spectrogram, CWT Using Morlet Wavelets

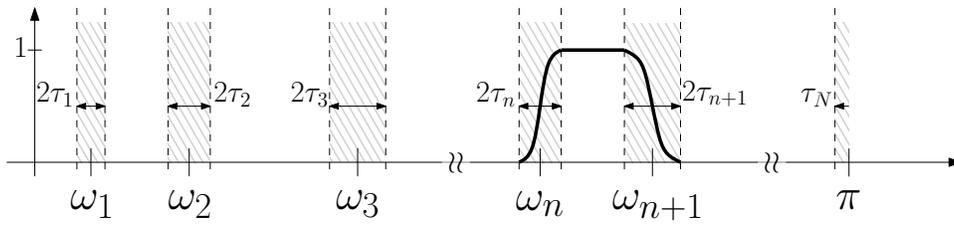


Figure 2.3: Empirical Wavelet Construction

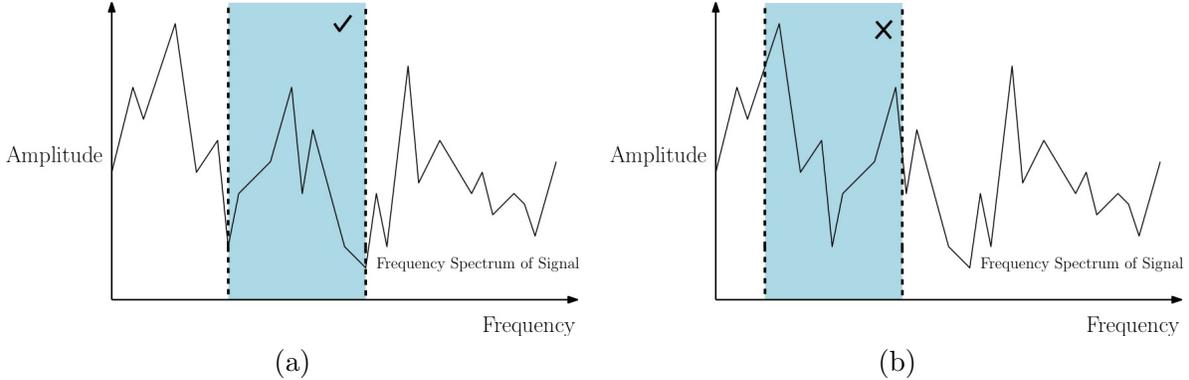


Figure 2.4: (a) Desirable and (b) Undesirable Boundary Choices

of boundaries.

In our experiments, we run our boundary detection algorithm on the average of all channels' signals. This yields a set of boundaries which is particular to an individual subject. We use these boundaries to build a single set of empirical wavelets for that individual, and use the EWT to filter the individual channels, one by one, using that set.

However, the methods proposed in [11] run into issues with EEG data, because the frequency signals of EEG tend to exhibit a  $\frac{1}{\omega}$  trend, and because the EEG data is characteristically jagged and irregular. This causes the previous methods to become caught in local maximum “traps”, seemingly insignificant pockets in the downhill slope of the signal as the time domain increases from zero. It also causes the first of the two algorithms above to blindly find whatever is the global minimum without checking whether the neighboring region represents a minimum of the function as a whole. To combat this phenomenon, we employ two strategies: trend removal and regularization on the frequency spectrum. The methods presented in [11] also inspired us to develop new adaptive methods which were less susceptible to getting trapped in local extrema along the downhill slope of the frequency spectrum.

### 2.3 Regularizing the Frequency Spectrum

Before we search for boundaries, we supply three customizable regularization techniques to smooth the typically jagged nature of the EEG frequency spectrum at hand. A Gaussian filter, an averaging filter, or replacement by the function's morphological closing operator may be used in place of the signal's original spectrum. The Gaussian filter takes as a parameter the frame size and sigma value, the averaging filter has a variable frame size, and the closing operator's structural element can be specified by the user. In our experiments we have favored frame size 10, sigma value 1.5, and structural element between 5 and 15. Figure 2.5 illustrates the effect of such regularization techniques on the frequency spectrum of the original signal. The smoother the frequency spectrum, the less likely the boundary search method might be to get caught in relatively insignificant local extrema. By regularizing the function, we emphasize the most important global trends in the signal, which encourages the algorithm to choose modes which retain the most overall information.

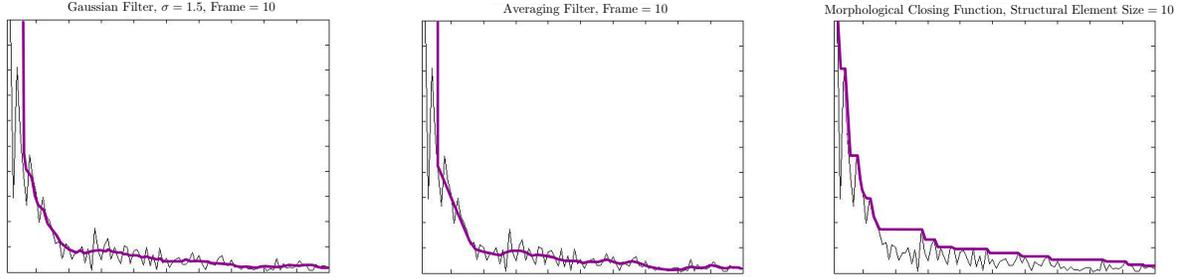


Figure 2.5: Frequency Spectrum Regularization Techniques

## 2.4 Adaptive Boundary Search Methods

We use a variety of boundary search methods to identify points that best partition the frequency domain so that the principal mountains remain intact. We use the pre-existing search methods provided in [11]; we also propose two new adaptive search methods, which use information about the spectrum to better inform boundary decisions.

### 2.4.1 Epsilon-Neighborhood Method

This method attempts to maintain the integrity of the traditional band arrangement, while making slight adjustments to adapt to the individual signal (Figure 2.6). That is, for the  $n$ th boundary point (in the figure, the white dotted line), which separates the  $(n - 1)$ th and  $n$ th traditional bands, we consider a neighborhood of radius  $\epsilon_n$  around that point. The length of this interval,  $\epsilon_n$ , is defined as follows: compare the lengths of the  $(n - 1)$ th and  $n$ th traditional band intervals. Take the interval of minimal length, and halve that interval. Set  $\epsilon_n$  equal to this length. This method ensures that no two  $\epsilon_n$  search regions will overlap. Within each neighborhood, we search for the two local maxima of the frequency spectrum and then further restrict our search regions. We find the global minimum within each search region, and finally define the  $n$ th adaptive boundary point to be the returned minimum. Here, the selected boundaries are represented by the black dotted line.

### 2.4.2 Closure Method

The closure method (Figure 2.7) is a departure from the traditional spectral bands altogether. It uses the morphological closing function to emphasize the most critical peaks of the frequency spectrum and then selects the top  $(N - 1)$  of those peaks. We then search for the global minimum of the original spectrum on a subinterval of the frequency domain which is defined between every two successive critical peaks retained by the previous step.

Figure 2.8 shows the boundaries returned by the Closure Boundary Search algorithm on the spectrum of the averaged channels. It is plotted here with the frequency spectrum of the 118th channel. We are especially interested in channel 118 because that electrode is located near the center of the visual cortex. Thus our visual stimuli would have the biggest perceived effect near this channel. Because we apply the algorithm to the average over all channels, we do not expect

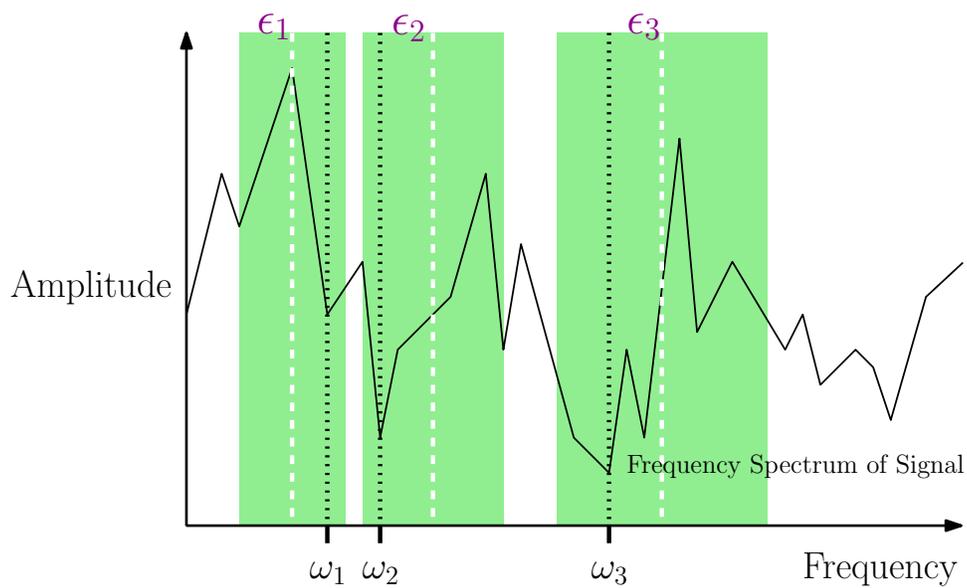


Figure 2.6: The Epsilon-Neighborhood Boundary Search Method

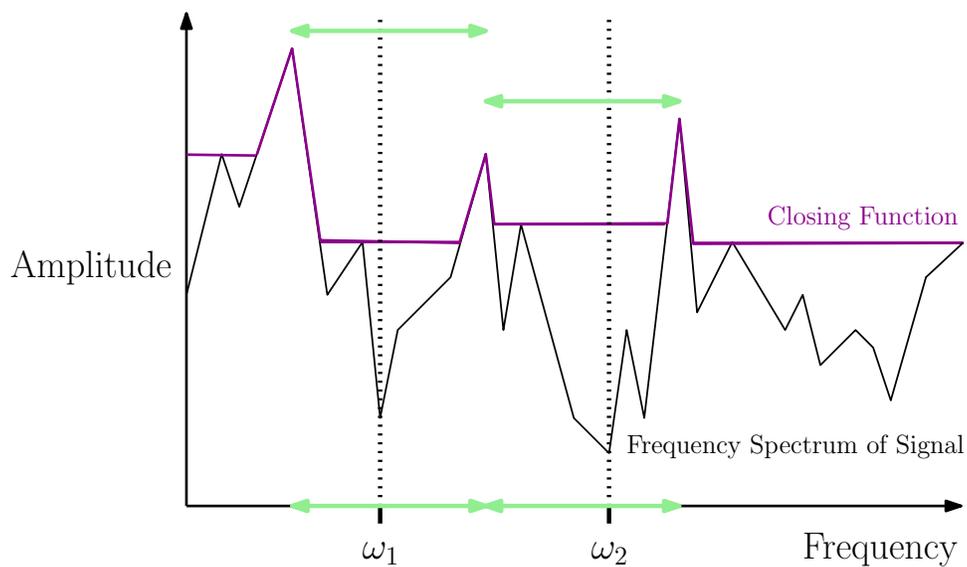


Figure 2.7: The Closure Boundary Search Method

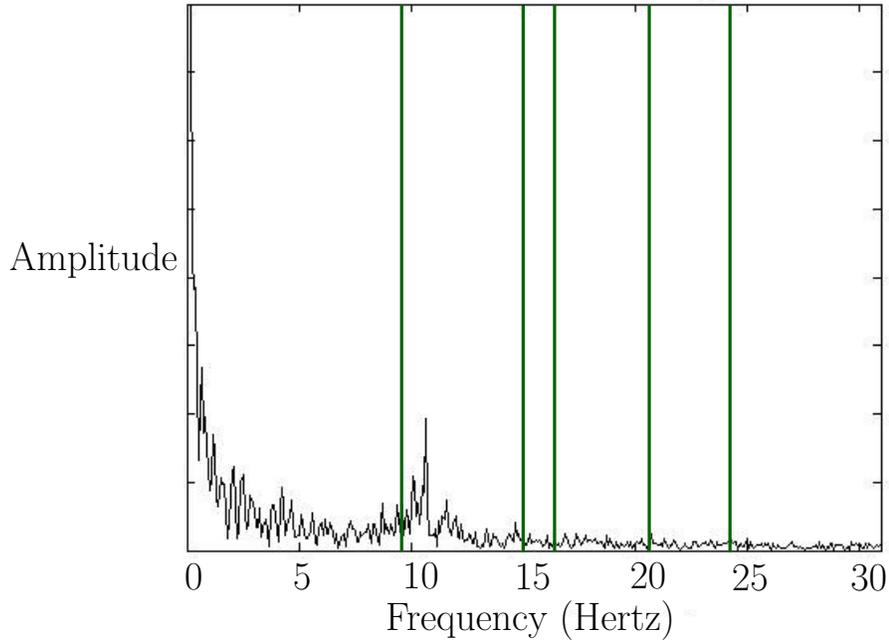


Figure 2.8: Spectral Band Boundaries Chosen Using Closure Method

the vertical lines in the figure to necessarily correspond to local minima of this particular channel’s frequency spectrum.

## 2.5 Decomposition into Spectral Modes

In neuroscience, it is useful to decompose a signal in the time domain into components which correspond to an established set of spectral bands. We use an Empirical Wavelet Transform rather than a Fourier transform. Upon first inspection, using the adaptive boundary search methods introduced in [11], we noticed few if any discernible patterns within the decomposed signals. When the wavelet transform is applied, according to the adaptive boundaries found by the Closure Method, the decomposed signals tend to exhibit “beat” patterns. This could perhaps allow for the analysis of the envelope in the search for patterns in different kinds of neural activity.

Figure 2.9 depicts the modes determined by the EWT; the mode functions are the summands of

$$f(t) = \sum_j a_j(t) \cos(\phi_j(t)).$$

Further, the maxima of the decomposed signals’ envelopes corresponded to amplitude peaks (and subsequent neural events of interest) in the time-frequency plane.

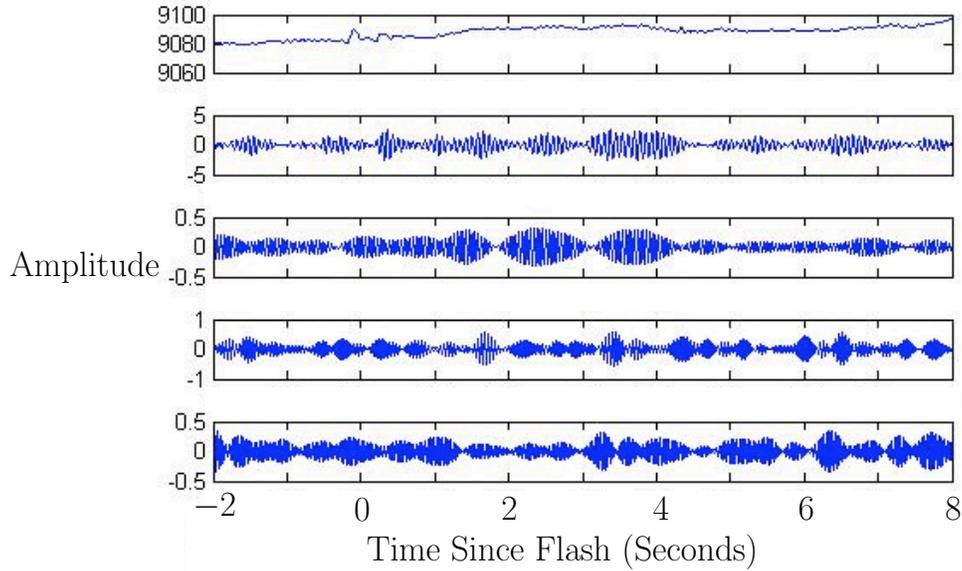


Figure 2.9: Components of Original Signal, Using Adapted Spectral Bands

## 2.6 New Time Frequency Representation

The CWT spectrogram is limited to an imprecise, splotchy representation of the important spectral rhythms that we wish to analyze. Our signal is a composition of Amplitude-Modulated, Frequency-Modulated (AMFM) mode functions. This makes the Hilbert transform a natural tool to extract the amplitude  $a_j$  and instantaneous frequency  $\frac{d\phi_j}{dt}$  of each mode over time. In order to obtain a more precise time-frequency interpretation of neural behavior, we apply the Hilbert transform to the decomposed modes of one channel at a time. The resulting time-frequency plane, which we will refer to as the “Hilbert plane”, offers a more precise pointwise interpretation of the relationship between time, frequency and amplitude of the signal. We use this tool to an alternative to the original spectrogram.

Figure 2.10 shows the Hilbert plane of channel 118 when the Empirical Wavelets are built based on the traditional spectral bands. Zooming in on the spectral patterns, we see the same alpha band activity as before, in a more precise presentation than the spectrogram provided. The horizontal axis is time.

Figure 2.11 is the Hilbert plane of channel 118 when the Empirical Wavelets are based on the spectral bands chosen by the adaptive Closure Method described above, using a structural element of size ten. Again, the Hilbert plane gives a more precise representation of the alpha band activity. Moreover, it gives information about two new spectral patterns which were not accessible at all in the spectrogram.

By varying the parameters of our adaptive boundary search method, we can also capture the

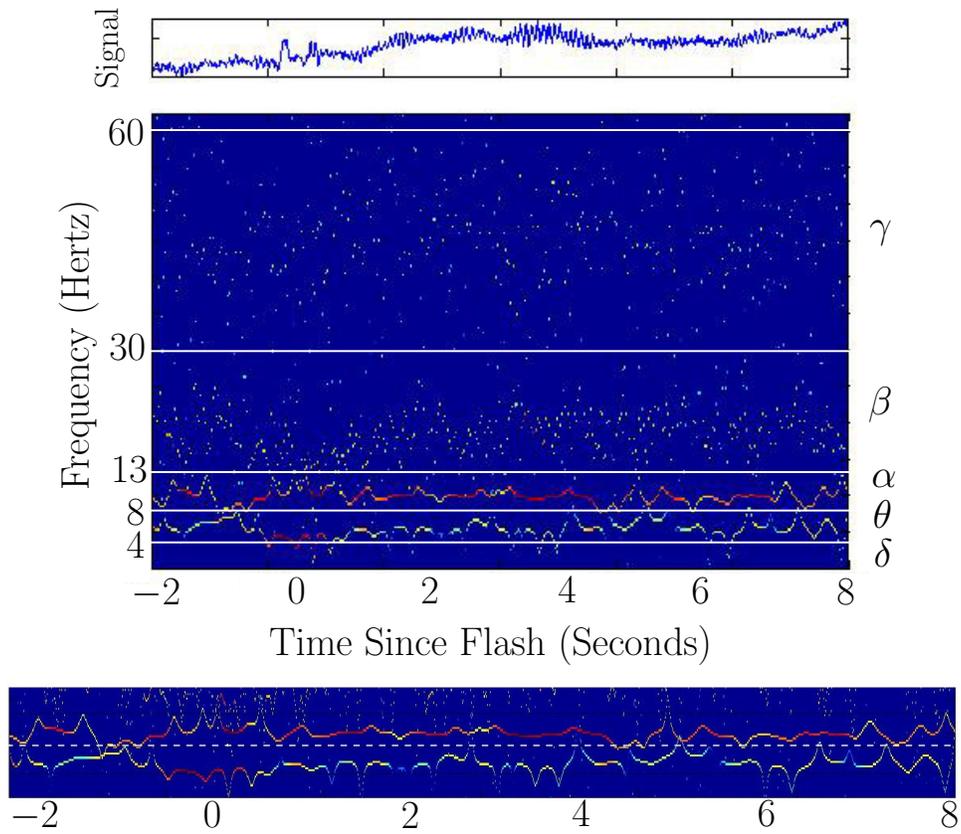


Figure 2.10: Hilbert Plane with Traditional Spectral Bands

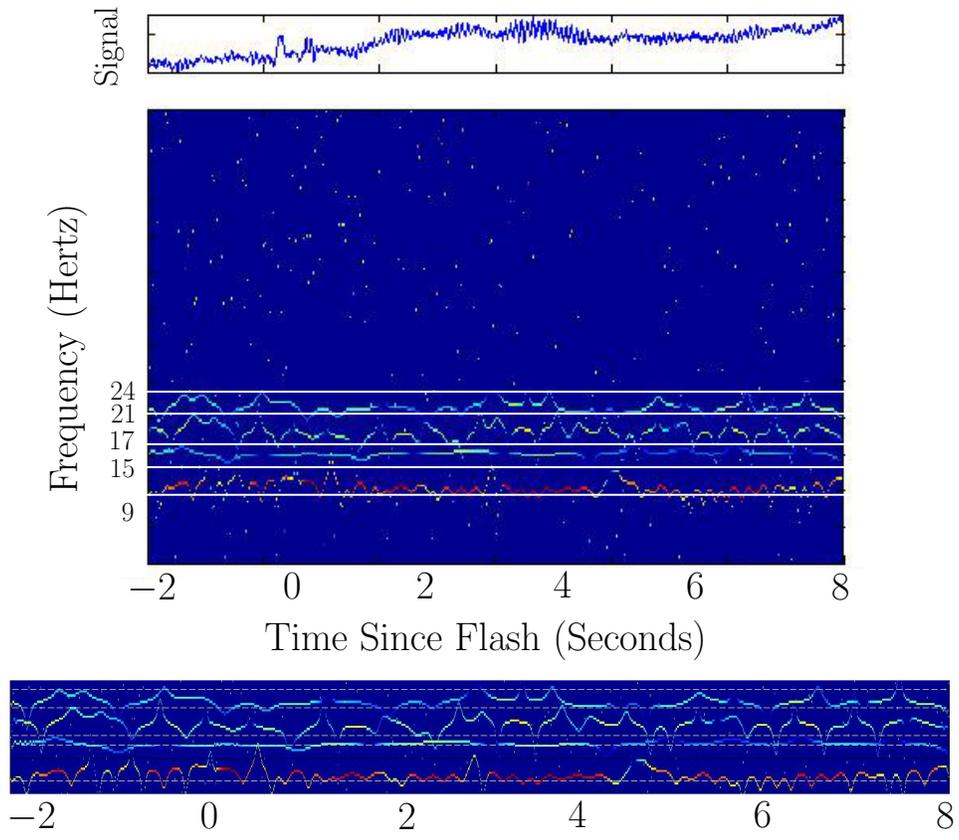


Figure 2.11: Hilbert Plane with Adapted Spectral Bands  
Structural Element Size is 10

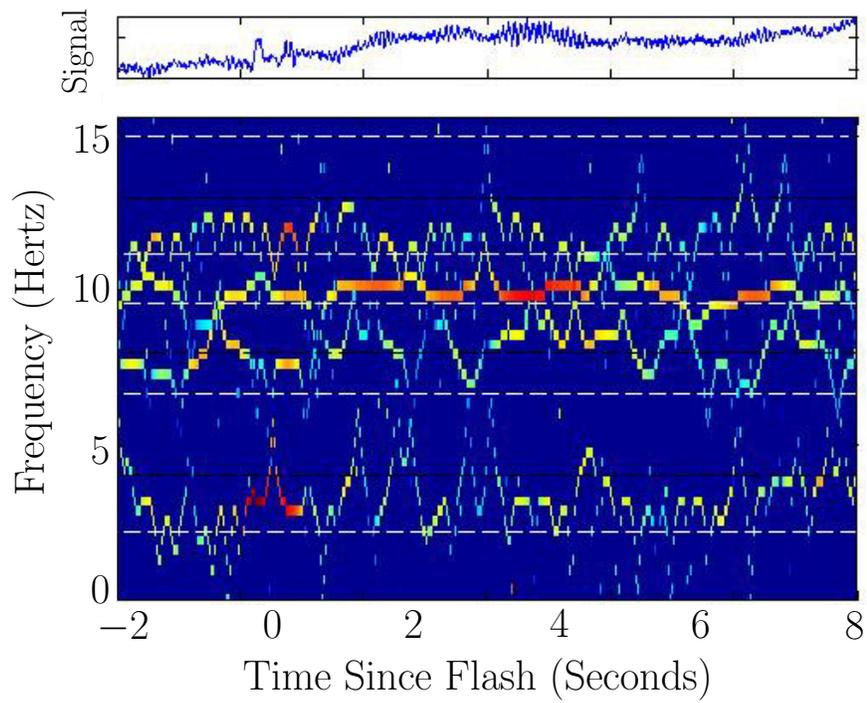


Figure 2.12: Hilbert Plane with Adapted Spectral Bands  
Structural Element Size is 5

same lower frequency patterns in the Hilbert plane that were featured in the spectrogram around time zero. Figure 2.12 illustrates the effect of decreasing the structural element size to five. Aided perhaps by of the natural  $\frac{1}{\omega}$  trend of the EEG frequency signal, we find that the weaker the regularization strength, the closer to zero the boundaries chosen by the adaptive boundary search methods are clumped. Therefore, for a smaller structural element size, the lower-frequency patterns should be better captured by the Hilbert plane. This sentiment is indeed reflected in the figure; we can now see not only the initial drop and subsequent return of the previously-termed “alpha” power, but we can also see the sharp increase in lower-frequency power around time zero. This pleasing result can be compared to the spectrogram in Figure 2.2. In our future work, we will further discuss varying parameters, especially optimizing the number of spectral bands to be defined. It appears that with an optimal choice of parameters, the adaptive spectral band Hilbert plane could prove to offer a more precise interpretation of neural activity without a significant loss of information.

## 2.7 Relevance to Neuroscience

In the Hilbert plane, as we might expect, we see that the alpha band phenomena (a sharp energy decrease immediately following each flash and eventual return to baseline value) appears more pronouncedly in channels which are closer to the visual cortex, such as 118, than those farther from the visual region, such as the 189th.

## 2.8 Future Work

In the future, we would like to use these adaptive spectral bands to provide relevant information for cross-correlation methods, to move towards EEG-fMRI fusion.

Rather than averaging over all 256 channels to generate the signal used in our boundary search, we would like to try averaging only those channels which correspond to the region specifically affected by visual stimuli. This could include the vicinity of the visual cortex, for example. We posit that this may result in the selection of adaptive spectral bands which are predisposed to better specifically capture the spectral rhythms of some region of interest.

Alternatively, instead of averaging at all we could choose one channel located near the center of the visual cortex as a representative and choose the boundaries based on that channel.

The allowance for a greater number of spectral bands (and consequently for additional modes) will allow us to see more distinct spectral patterns. However, for computational efficiency as well as ease of analysis, we would like to focus on only the most important patterns. To approach these desires we may pursue the development of algorithms which detect the optimal number of modes for a given signal.

Inspired by the work of [10], we are drawn to the traditionally defined  $\gamma$ -band in particular. We would like to know whether potential areas of interest may reside within the  $\gamma$ -band, regions which had been previously overlooked by the traditional spectral bands.

On the subject of parameter optimization, when using the closure boundary search method we might also look into optimizing the size of the structural element given a particular averaged signal.

We also see potential for finding distinguishing characteristics of various diseases based on the positions of boundaries chosen using the closure boundary search method.

## Chapter 3

# EEG-fMRI Fusion

### 3.1 Introduction

fMRI is a four dimensional brain scan (x,y,z by time), measuring relative changes in oxygenated (diamagnetic) to deoxygenated (paramagnetic) hemoglobin. As neuronal firing increases, more oxygen rich blood flows to the active brain region, generating the Blood Oxygen Level Dependent (BOLD) signal. This is an indirect measure of neural activity and neuronal firing is coupled to the BOLD signal by what is known as the Hemodynamic Response Function (HRF), and the speed of this reaction is limited by the vasal properties of blood flow.

Another complication with this measure of neural activity is that the relationship between the exact location of oxygenation and the origin of neural electrical activity is uncertain [7]. The location of oxygenation can not be assumed to be the exact origin of neural activity. Thus metrics must be devised to precisely relate the BOLD signal to neural activity. One such studied and measured metric is the aforementioned HRF, a highly informative model of temporal correlation between neuronal local field potentials and the BOLD signal. In practice, the HRF is convolved with a time series indicating when a task took place. The result of this convolution is then compared with the time series in each brain voxel independently – generating a p-value for each voxel in a mass univariate t-test [20].

The goal for this section of the project is to develop a novel process that finds the coupling between EEG and fMRI data - which is currently unknown. Although neuronal firing is coupled to the BOLD with the HRF, EEG is measured at the scalp, and it is unclear exactly what aspects of neuronal firing that EEG is capturing. The chosen algorithm to detect – or deconvolve – this unknown relationship is known as tkCCA. tkCCA is a metric for finding projections of the data that maximally correlates the two data sets, while allowing shifting of the data sets with respect to each other [1]. This allows for a multivariate correlation between two data sets whose correlation might be higher if there is a different alignment of the data. This is ideal for EEG-fMRI fusion as the correlation between the two is likely to be highest with a temporal shift [1]. The graph of the correlations at different time shifts is referred to as the canonical correlogram (CC) and is compared to the HRF.

A benefit of tkCCA is that the two datasets require only one dimension in common, in the Bießmann et al case this was the temporal dimension [1]. For this study we will also take the temporal dimension to be the common dimension. However, in addition the second dimension will be shared, the spatial dimension. This will be done by transforming the EEG data from the

time-frequency plane into 3-d space by time using a method called sLORETA. sLORETA is one of the more recent solutions to the EEG inverse problem. It has the benefits of returning a smoothed solution, so there are no isolated pockets of inactivity surrounded by activity, and it has good accuracy for placing deep sources [23]. We then use an implemented version of our own algorithm to project EEG sources into voxel space [9].

The motivation for transforming the EEG data into fMRI voxel space before computing correlation with tkCCA will be discussed later in this report. Also, the tkCCA of this experimental data without first transforming the EEG data into voxel space is being explored by another project working on this data set.

## 3.2 Methods

The data we are using for this portion of the project is from the same study as the data we use in Empirical Wavelet Analysis portion of this report. We are using the same EEG data. We are also using fMRI data collected with a three Tesla Siemens fMRI scanner. The experimental parameters of randomly presented flashes, separated by at least eight seconds, are the same for the data collected in the fMRI as they are of the EEG data collected. While the fMRI was performed, simultaneous EEG was collected using a 256 Magnetic Resonance (MR) safe electrode cap. In the future we plan to perform this method of analysis using this artifact cleaned EEG data.

The EEG data is decomposed into the traditional spectral bands. This will allow for more meaningful and comparable results.

For both the EEG and fMRI data the recordings are segmented into ten second segments with the stimulus at two seconds. Then sets of twelve epochs (segments around a single stimulus) are averaged together and our process of sLORETA and tkCCA is run on this preprocessed data set.

### 3.2.1 Standardized Low Resolution Brain Electromagnetic Tomography

As previously stated, sLORETA is a solution to the EEG inverse problem. It localizes the point sources from which neural activity originates using the electric potentials measured by the EEG cap electrodes. This method has high accuracy for localizing deep sources onto deeper cortex instead of places these as sources on outer cortex. This is a problem with many inverse problem solutions as the EEG electrodes most effectively measure neural activity originating from the outer cortex as this is closer to the recording apparatus and the signal is stronger. sLORETA has the advantage of parsing out the lower powered signals originating from deeper cortical sources and placing them in the accurate locations, instead of mistaking them as originating from outer cortical sources [23].

A second important benefit of sLORETA is that it returns a smoothed result. While other solutions to the EEG inverse problem can return sources where there are regions of activity surrounding small areas of inactivity, sLORETA returns a result with smoothed regions of activity. This is based on the idea that if an area of the brain is active then this activity will not surround inactivity but that the activity should continue through the entirety of an area [23].

Data analysis was performed with Brainstorm (Tadel et al. 2011), which is documented and freely available for download online under the GNU general public license (<http://neuroimage.usc.edu/brainstorm>). The Brainstorm sLORETA tool is used to compute the solution to the inverse problem [28]. However first the BEM surfaces must be generated [12, 15] and the head model must be computed, a step also performed using the Brainstorm tools. The parameters for the inverse solution by the

sLORETA method are unconstrained source orientations, whitening using PCA, a signal-to-noise ratio of 3, the full noise covariance matrix regularized with parameter 0.1, and depth weighting of order 0.5 and maximal depth 10.

The inverse solution to the EEG signal decomposed using the traditional spectral bands is computed individually and exported as 4-d matrices. Figure 3.1 shows the results of the inverse method for one epoch of alpha frequency data, as visualized on the cortical surface [28].

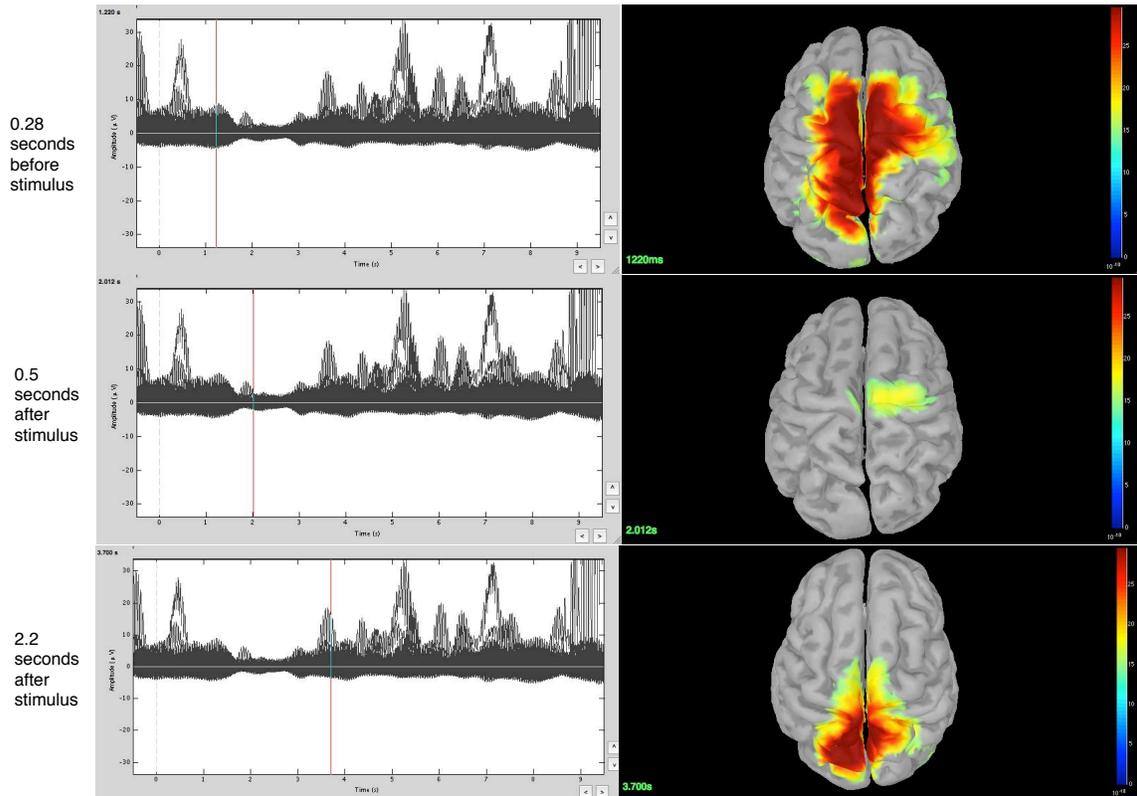


Figure 3.1: The left column shows the power of the alpha frequency band of collected EEG data for one averaged epoch, the data is visualized as time by power with each line corresponding to a channel (a single electrode recording data). The movement of the red line in the three otherwise identical graphs marks the current time, which corresponds to the figure in the right column on the same row. These figures are the solutions to the inverse problem using sLORETA. They are projections of the power measured by the electrodes onto the cortex. This graph shows the expected alpha activity before stimulus, the nearly complete drop-off of activity shortly after the stimulus, and the re-emergence of activity further post-stimulus. This image was created using Brainstorm [28].

### 3.2.2 Temporal Kernel Canonical Correlation Analysis

Our selected method of analyzing correlation is tkCCA. tkCCA has the benefits of determining the multivariate correlation between two data sets, allowing for relative shifting of the data sets as a method of improving correlation, and utilizing a kernel trick to decrease computational complexity.

tkCCA is based on the method of Canonical Correlation Analysis (CCA) which is a metric for finding the maximally correlated features of two data sets.

For two multivariate  $x \in \mathbb{R}^M$  and  $y \in \mathbb{R}^N$ , CCA estimates two normalized linear filters  $w_x \in \mathbb{R}^M$  and  $w_y \in \mathbb{R}^N$ , called *canonical variates*, such that the correlation between the projections  $w_x^T x$  and  $w_y^T y$  is maximized:

$$\arg \max_{w_x, w_y} w_x^T C_{xy} w_y$$

$$\text{s.t. } w_x^T C_{xx} w_x = 1 \text{ and } w_y^T C_{yy} w_y = 1.$$

where  $C_{xx} \in \mathbb{R}^{M \times M}$  and  $C_{yy} \in \mathbb{R}^{N \times N}$  denote the respective auto-covariance matrices and  $C_{xy} \in \mathbb{R}^{M \times N}$  the crosscovariance matrix between  $x$  and  $y$  [1].

Kernel Canonical Correlation Analysis (kCCA) decreases the computational complexity by utilizing a kernel trick and working with:

$$K_X = X^T X \text{ and } K_Y = Y^T Y$$

instead of the covariance matrix, where  $X$  and  $Y$  are the data matrices of the two data sets [1]. This also allows for working with data matrices of variables instead of two multivariate variables.

The filters in the input space of each variable are then given as a linear expansion fo the data points

$$w_x = X\alpha \text{ and } w_y = Y\beta$$

where the vectors  $\alpha \in \mathbb{R}^L$  and  $\beta \in \mathbb{R}^L$  are the solutions to the generalized eigenvalue problem in kernel space [1]:

$$\begin{bmatrix} 0 & K_X K_Y \\ K_Y K_X & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \rho \begin{bmatrix} K_X^2 & 0 \\ 0 & K_Y^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

In order to prevent overfitting, regularization parameters ( $\kappa_X, \kappa_Y$ ) are added to the generalized eigenvalue problem, which becomes [1]:

$$\begin{bmatrix} 0 & K_X K_Y \\ K_Y K_X & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \rho \begin{bmatrix} K_X^2 + \kappa_X K_X & 0 \\ 0 & K_Y^2 + \kappa_Y K_Y \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

The key benefit of tkCCA is the use of a time shift. This is done simply by using  $\tilde{X}$  in place of  $X$  where:

$$\tilde{X} = \begin{bmatrix} X_{\tau_1} \\ X_{\tau_2} \\ \vdots \\ X_{\tau_T} \end{bmatrix} \in \mathbb{R}^{MT \times L}$$

where each  $X_{\tau_i}$  is a copy of  $X$  shifted by  $\tau_i$  with respect to  $Y$  [1]. This also requires the use of:

$$\tilde{w}_x = \begin{bmatrix} w_x(\tau_1) \\ w_x(\tau_2) \\ \vdots \\ w_x(\tau_T) \end{bmatrix}$$

where  $w_x(\tau) = X_{\tau} \alpha_{\tau}$  [1].

The regularization parameters were optimized by running tkCCA on 25 sets of parameters and finding the optimal parameters for each run.

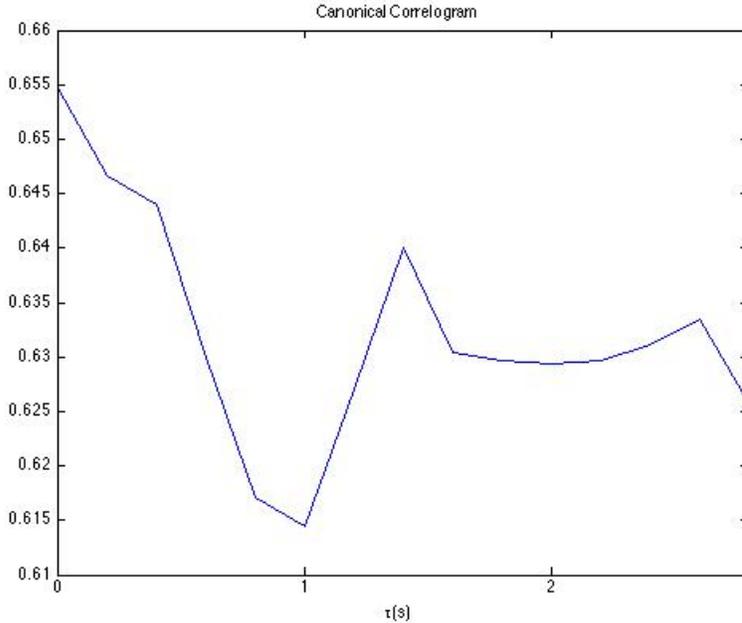


Figure 3.2: This is an example of a CC. This CC is from running tkCCA to find the correlation of the EEG spectral band alpha and the fMRI data with time shifts of 0.2 from time zero to 2.8. The  $x$ -axis is the temporal shift and the  $y$ -axis is the correlation of the two data sets. The  $y$ -axis ranges from zero to one with zero being absolutely no correlation of the two data sets at that time and one being complete correlation at that time, meaning the data sets are identical.

### 3.3 Results

The visualizable results of tkCCA are the CC, which is a plot of the correlation scores of the data sets at each time shift  $\tau$ , and the regions of highest correlation in each of the data sets. Figure 3.2 is the CC for the EEG spectral band alpha and the fMRI data with time shifts from zero to 2.8 seconds, shifted in increments of 0.2 seconds. Figure 3.3 shows the corresponding regions of correlation of the EEG spectral alpha band.

The purpose of the tkCCA is to look at the CC's for all spectral bands for time shifts of up to eight seconds. This is the amount of time shift necessary for capturing the pattern of the HRF and shifting any more than ten seconds will serve only to realign the data with a new epoch. Because of this realignment, shifts modulus ten should be identical. The CC's for all of the spectral bands for time shifts from zero to eight seconds in shifts of 0.2 seconds are shown in Figure 3.4.

Ideally these results would follow a similar pattern to the HRF, starting with low correlation, beginning to increase at about two seconds, and peaking somewhere between four and eight seconds. The wild divergence of these results from the expected result could be caused by any number of things. A main challenge the current examples posed is that the data set used for testing this method is small, one subject with 144 trials of flashes, each of these trials grouped into sets of twelve and averaged. The result could improve once the other subjects' data is added. Another potential cause of the unexpected result is the unexpected solution to the inverse problem, as illustrated in Figure 3.1. This figure places the alpha activity in the parietal lobe when the activity is expected to be in the occipital lobe as that is the area of the brain corresponding to visual processing. Modification of the parameters used in computing the inverse of the EEG by sLORETA could easily change this result.

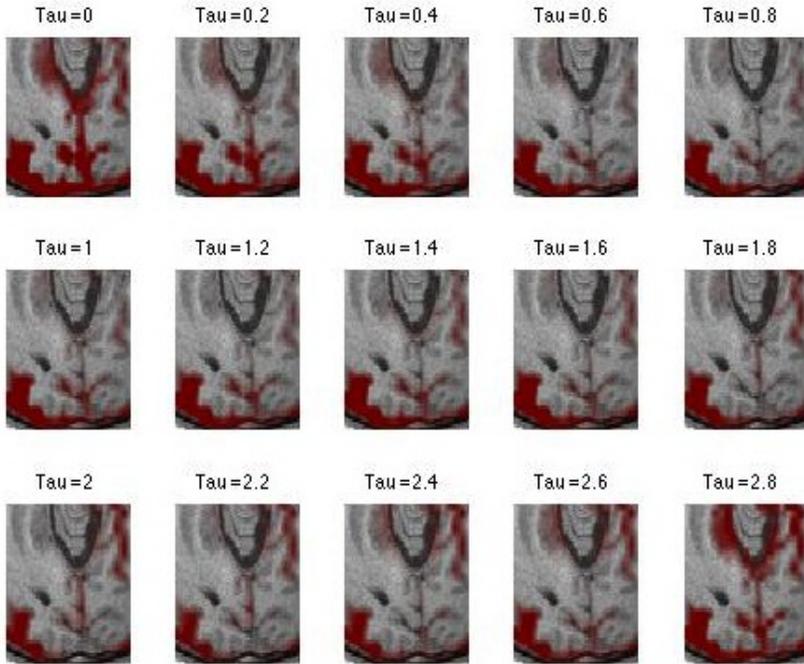


Figure 3.3: This figure shows the regions of correlation of the EEG spectral band alpha with time shifts from 0 to 2.8, shifted in increments of 0.2. The red regions are those of high correlation with the fMRI activation. The time shifts whose corresponding images have large regions of correlation correspond to the time shifts which high overall correlation, as seen in Figure 3.2.

### 3.4 Future Research

These methods have significant room for improvement as well as a large variety of application potential. The same processing of the EEG and fMRI data can be done after altering the parameters of the sLORETA inverse solution. This would show what methods have the best result.

Ideally these methods would be performed on simultaneously collected EEG and fMRI data. This data has been collected and will be available for use after additional cleaning to remove noise and artifacts the scanner causes. Using simultaneously collected data would allow for correlation not only in response to the stimulus. As both modalities capture all brain activity, this activity would be equivalent, unlike in non-simultaneously collected data where the brain will show different activity during the same experiment performed at different times, just as each epoch of data is not identical. This could lead to a higher correlation between the EEG and fMRI data.

Another future project would be to run this method of analysis using the new spectral bands found using the Empirical Wavelet Transform, as described earlier in this report, instead of the traditional spectral bands. This could provide any unknown amount of new insight to EEG-fMRI coupling.

A final future project, the one motivating the transformation of EEG data into voxel-space, is to use a spatial  $\tau$  shift instead of a temporal  $\tau$  shift. This would provide information about spatial overlap of EEG and fMRI. This would be highly informative as there is a prevalent assumption of direct overlap when that could easily not be the case. Using a spatial shift would show exactly which relative shifting gives the highest correlation between the two data sets.

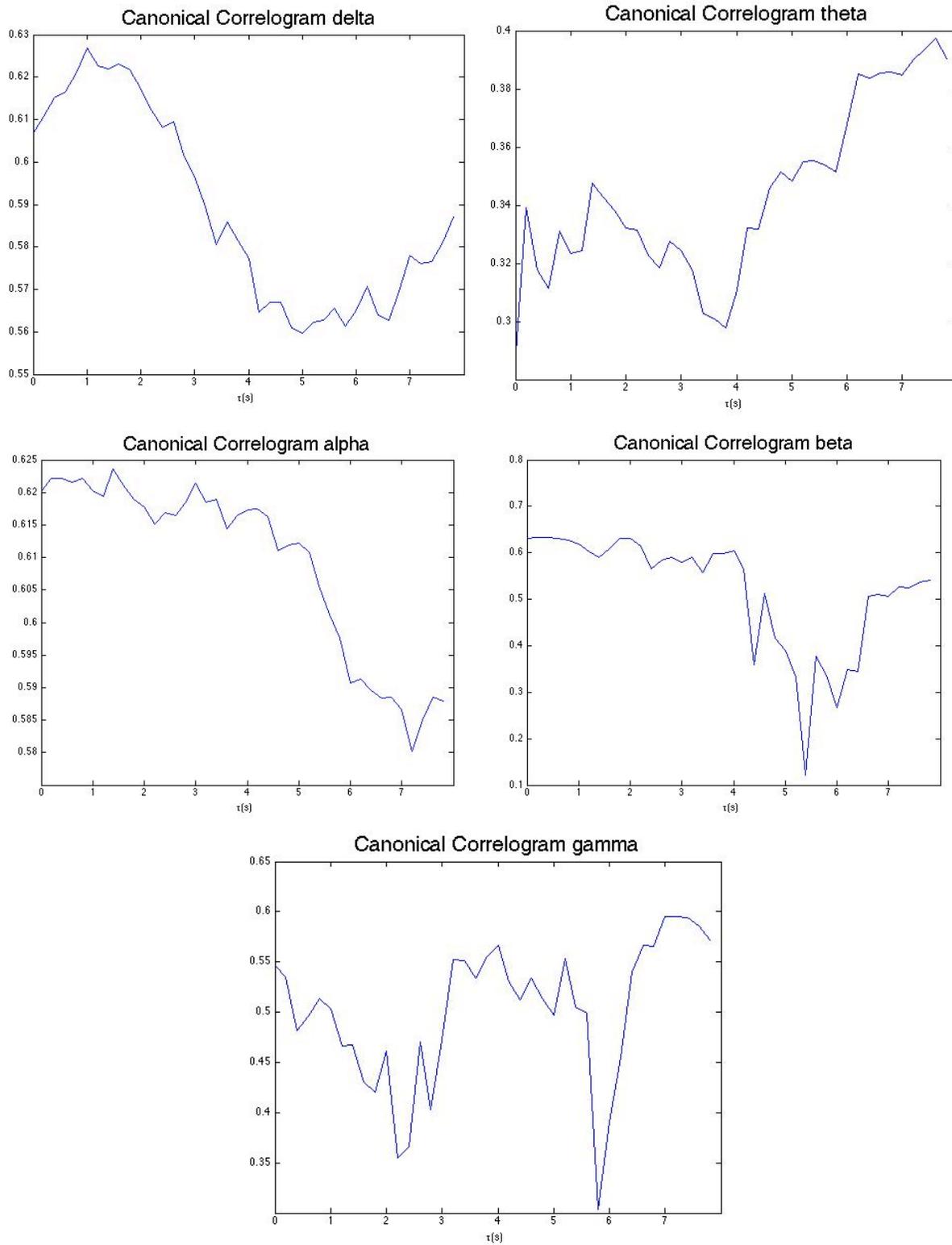


Figure 3.4: The CC's for all of the spectral bands. Note that the range of the  $y$ -axis changes for each graph.

# Bibliography

- [1] Felix Bießmann, Frank C. Meinecke, Arthur Gretton, Alexander Rauch, Gregor Rainer, Nikos K. Logothetis, and Klaus-Robert Müller. Temporal kernel cca and its application in multimodal neuronal data analysis. *Machine Learning*, 79(1-2):5–27, 2010.
- [2] Bernard S. Chang and Daniel H. Lowenstein. Epilepsy. *New England Journal of Medicine*, 349(13):1257–1266, 2003. PMID: 14507951.
- [3] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.
- [4] Suzanne Corkin. What’s new with the amnesic patient hm? *Nature Reviews Neuroscience*, 3(2):153–160, 2002.
- [5] Corinna Cortes and Vladimir Vapnik. Support-vector networks. In *Machine Learning*, pages 273–297, 1995.
- [6] Dona E Cragar, David TR Berry, Toufic A Fakhoury, Jean E Cibula, and Frederick A Schmitt. A review of diagnostic techniques in the differential diagnosis of epileptic and nonepileptic seizures. *Neuropsychology Review*, 12(1):31–64, 2002.
- [7] Jean Daunizeau, Helmut Laufs, and Karl J. Friston. Eeg-fmri information fusion: biophysics and data analysis. In C. Mulert and L. Lemieux, editors, *Eeg - Fmri: Physiological Basis, Technique, and Applications*, pages 511–526. Springer Berlin Heidelberg, 2010.
- [8] Philip Dickinson and Karl J Looper. Psychogenic nonepileptic seizures: a current overview. *Epilepsia*, 53(10):1679–1689, 2012.
- [9] PK Douglas, D Moyer, and MS Cohen. Colocalizing EEG and fMRI in space. In *19th Annual Meeting*. Human Brain Mapping, 2013.
- [10] Andrew D. Engell, Scott Huettel, and Gregory McCarthy. The fMRI {BOLD} signal tracks electrophysiological spectral perturbations, not event-related potentials. *NeuroImage*, 59(3):2600 – 2606, 2012.
- [11] Jérôme Gilles. Empirical Wavelet Transform. *IEEE Transactions on Signal Processing*, 2013:3999–4010, 2013.
- [12] A. Gramfort, T. Papadopoulo, E. Olivi, and M. Clerc. Openmeeg: opensource software for quasistatic bioelectromagnetics. *BioMedical Engineering OnLine*, 9, 2010.

- [13] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. A practical guide to support vector classification, 2003.
- [14] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13:411–430, 2000.
- [15] Kybic J, Clerc M, Abboud T, Faugeras O, Keriven R, and Papadopoulos T. A common formalism for the integral formulations of the forward eeg problem. *IEEE Transactions on Medical Imaging*, 24:12–28, 2005.
- [16] Wesley T Kerr, Ariana Anderson, Edward P Lau, Andrew Y Cho, Hongjing Xia, Jennifer Bramen, Pamela K Douglas, Eric S Braun, John M Stern, and Mark S Cohen. Automated diagnosis of epilepsy using eeg power spectrum. *Epilepsia*, 53(11):e189–e192, 2012.
- [17] Wesley T. Kerr, Stefan T. Nguyen, Andrew Y. Cho, Edward P Lau, Daniel H. Silverman, Pamela K. Douglas, Navya M. Reddy, Ariana Anderson, Jennifer Bramen, and Noriko Salamon. Computer-aided diagnosis and localization of lateralized temporal lobe epilepsy using interictal FDG-PET. *Frontiers in Neurology*, 2013.
- [18] R Kobau, MPH, Y.H. Luo, M Zack, S Helmers, and D Thurman. Epilepsy in adults and access to care- united states, 2010. *Morbidity and Mortality Weekly Report*, 61(45):909–913, 2012.
- [19] W Curt LaFrance and Orrin Devinsky. The treatment of nonepileptic seizures: historical perspectives and future directions. *Epilepsia*, 45(s2):15–21, 2004.
- [20] Cesare Magri, Ulrich Schridde, Yusuke Murayama, Stefano Panzeri, and Nikos K. Logothetis. The amplitude and timing of the bold signal reflects the relationship between local field potential power at different frequencies. *The Journal of Neuroscience*, 32(4):1395–1407, 2012.
- [21] RS McLachlin, S Wiebe, JT Langfitt, J Stern, S Dewar, et al. Early surgical therapy for drug-resistant temporal lobe epilepsy. *JAMA*, 307:922–30, 2012.
- [22] Victoria L Morgan, Baxter P Rogers, Hasan H Sonmezturk, John C Gore, and Bassel Abou-Khalil. Cross hippocampal influence in mesial temporal lobe epilepsy measured with high temporal resolution functional magnetic resonance imaging. *Epilepsia*, 52(9):1741–1749, 2011.
- [23] RD Pascual-Marqui. Standardized low-resolution brain electromagnetic tomography (sLORETA): Technical details. *Methods and Findings in Experimental and Clinical Pharmacology*, 24(D):5–12, 2002. 12th Meeting of the International-Pharmaco-EEG-Group, Barcelona, Spain, NOV 21-24, 2002.
- [24] Pavel Pudil, Jana Novovičová, and Josef Kittler. Floating search methods in feature selection. *Pattern recognition letters*, 15(11):1119–1125, 1994.
- [25] John Shlens. A tutorial on principal component analysis derivation, discussion and singular value decomposition. Technical report, University of California San Diego, March 2003.
- [26] Jonathon Shlens. A tutorial on principal component analysis. In *Systems Neurobiology Laboratory*, Salk Institute for Biological Studies, 2005.

- [27] Tanvir U Syed, W Curt LaFrance, Emine S Kahriman, Saba N Hasan, Vijayalakshmi Rajasekaran, Deepak Gulati, Samip Borad, Asim Shahid, Guadalupe Fernandez-Baca, Naiara Garcia, et al. Can semiology predict psychogenic nonepileptic seizures? a prospective study. *Annals of neurology*, 69(6):997–1004, 2011.
- [28] F Tadel, S Baillet, JC Mosher, D Pantazis, and RM Leahy. Brainstorm: A user-friendly application for meg/eeg analysis. *Computational Intelligence and Neuroscience*, 2011, 2011.
- [29] Ulrike von Luxburg. A tutorial on spectral clustering. *CoRR*, abs/0711.0189, 2007.