UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

AUTOMATICALLY IDENTIFYING STIMULI FROM FIRING PATTERNS IN THE AUDITORY CORTEX OF RAT

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

KIMBERLY C HOUCK
Norman, Oklahoma
2012

AUTOMATICALLY IDENTIFYING STIMULI FROM FIRING PATTERNS IN THE AUDITORY CORTEX OF RAT

A THESIS APPROVED FOR THE
SCHOOL OF COMPUTER SCIENCE

BY

Dr. Andrew H. Fagg

Dr. Robort R. Rennaker

Dr. Amy McGovern

Acknowledgments

I would like to acknowledge Dr. Andrew Sloan and Dr. Robert Rennaker at the University of Texas at Dallas for providing the data set used in this work. I would especially like to thank Drew for his help in understanding the data and the domain background.

I would also like to thank my advisor, Dr. Andrew H. Fagg, for his patience and guidance during my time at the University of Oklahoma, and for teaching me about research in computer science and how to write about it. I also appreciate my other committee members, Dr. Amy McGovern and Dr. Robert Rennaker, for their guidance in preparing this document.

Additionally, I'd like to thank my fellow students in the Symbiotic Computing Laboratory for their advice, and especially Tom Palmer and Joshua Southerland for their involvement in the early stages of this project and the construction of the initial codebase to work with the data.

Finally, this research would not have been possible without funding from NIH grant number NIDCD R01DC008982.

Table of Contents

List of Figures

Abstract

Sensory information is encoded in the mammalian brain through the activity of many individual neurons. While neural activity can be observed, the finer details of this encoding is not always readily apparent. Machine learning techniques can be used to extract information from vast and complex inputs, and might provide a means to help understand such natural systems. In the case of auditory processing, the auditory cortex is the initial stage in the neocortex to process aural inputs. In this brain region, individual neurons have been shown to respond preferentially to particular ranges of auditory frequencies. With the ability to record simultaneously from many sites within the auditory cortex, one can ask how such a population of neurons encodes both simple and complex sounds. Additionally, is the auditory cortex a pure sensory area, or does it play a larger role in processing information in a task dependent way?

In this thesis, I approach these questions by using Support Vector Machine techniques to construct classifiers that can distinguish different categories of information captured in the neural population. The state of the neural population in response to a stimulus, a pure tone, multi-frequency noise signal, or speech sound, is captured as a set of features, each of which describes the binned action potential count of a single multi-neuron within a given time range, or the power of the Local Field Potential (LFP) signal at a particular range of frequencies. By using classifiers with these features, I show that the population response encodes the type of aural stimulus. Additionally, my results show that the population response is sensitive to whether the aural stimulus is used in the task. For the case of stimuli within the task context, my results show that under certain conditions, the population response better captures the rat's response to the stimulus than the actual stimulus type. Finally, I show that while the recorded action potentials encode stimulus type, the LFP signal appears to

better capture this information.

Chapter 1

Introduction

The Primary Auditory Cortex (A1) is the initial stage in the neocortex to process aural inputs. This area has been shown to encode information such as sound magnitude and frequency, and is tonotopically organized, with individual neurons responding preferentially to particular ranges of auditory frequencies (Kandel et al., 1991; Malmierca, 2003). The rat auditory cortex has been shown to contain cells that respond with large burst of action potentials to auditory input induced activity in lower brain areas. Other cells respond to these inputs with fewer action potentials, but more do so more selectively (Malmierca, 2003).

Sloan (2009) describes the temporal structure of responses in the rat auditory cortex to single frequency stimuli, both in contexts in which the stimuli are behaviorally relevant and outside of such contexts. The responses are characterized by a large initial excitatory, "phasic" response peaking at about 10-15 ms after the onset of the stimulus. For frequencies close to the preferred frequency of a given neuron, this initial response is followed by a less active, but lasting, "tonic" excitatory response for the remainder of the stimulus' duration. However, for frequencies to which a neuron is not tuned, the phasic response period does not last beyond the initial (approximately 30 ms) following the stimulus onset, with a period of suppression of neural activity occurring for the remainder of the stimulus duration (Sloan, 2009).

Hubel et al. (1959) describes neurons in the cat auditory cortex that to responded to novel stimuli when those stimuli appeared to capture the animals' attention, but did not respond otherwise. This phenomenon, indicating that some responses to stimuli might be modulated by attention, has been further

investigated by others since. For example, Fritz et al. (2007) study the A1 response in ferret to multi-frequency waveform and single frequency tone stimuli, observing changes in neural firing patterns during auditory related tasks. Sloan (2009) also shows contextual differences in the rat auditory cortex. This work suggests that the auditory cortex is not a pure sensory area, capturing information without regard to task context. Rather, it suggests that the auditory cortex instead plays a larger role in processing information in a task dependent way, encoding aural inputs differently based on the context. Specifically, in task contexts in which stimuli affect expected behavior, the relevant stimuli elicit more focused neural activity compared to the same stimuli outside of such a context. However, the details and the meaning of these changes in encoding have not been examined.

Individual neurons in the neocortex encode information in terms of the frequency of events, referred to as *action potentials* or "spikes." There are multiple methods to extract discrete metrics of neural activity from series of spike times from a population of neurons, one being to quantify the spike rate and its variation over time (Victor, 2005). A series of binned action potential counts, determined by the number of action potentials that fall in a bin centered about that time step, is known as a Peri-Stimulus Time Histogram (PSTH). The PSTH is a fixed-length representation of the neural activity seen by an electrode channel over a fixed period of time. The representations of multiple channels can be concatenated together to create a representation of that set of channels.

The timing of individual action potentials is only present from one or a few neurons closest to each electrode. These action potentials are high frequency features that can be extracted from the overall voltage signal, called the Local Field Potential (LFP). The LFP captures a measure of the average activity of many neurons in the vicinity of an electrode, and contains information about their overall activity at lower frequencies than the action potentials themselves (Liu and Newsome, 2006). Information has been successfully decoded from LFPs recorded from the monkey motor cortex (Mehring et al., 2003; Bansal et al., 2011; Lindberg et al., 2011), so it is expected that information about the stimulus that generated the neural response can also be decoded from the full LFP signal.

Previous work comparing spike rate and LFP information from the primary motor cortex of monkey has shown that for large numbers of electrodes the spike rates of individual neurons provide more information for the continuous decoding of hand position. However, LFP generally contains more information if fewer electrodes are used (Mehring et al., 2003; Bansal et al., 2011). A similar tradeoff between the spatial and temporal representations is expected in the case of stimulus classification as well. Since the LFP signal is influenced by the activity of neurons further from an electrode, in addition to those neurons from which individual action potentials can be detected, it allows for information to be captured from a greater number of neurons (Lindberg et al., 2011). However, the precision of the exact temporal information regarding the activity of individual neurons is lost in the LFP signal. One can ask whether the spike rate representation, with its higher temporal resolution, or the LFP signal, containing information from a greater number of neurons, captures more useful information about stimuli encoded in the auditory cortex. This question is meaningful, since the LFP can present a detectable signal to an electrode for longer periods of time than individual action potentials, which at best can be recorded for a year or so (Lindberg et al., 2011).

With the ability to record these neural signals simultaneously from many sites within the auditory cortex, one can ask how a population of neurons encodes both simple and complex sounds. Neural activity can be difficult to interpret, however, due the complexity of the information encoding, with the complexity increasing as additional channels add new information from additional neurons. However, one can make use of machine learning classifiers to decode the information contained in the neural signal. Using such a classifier, one can determine whether a neural population encodes a given binary, or n-ary concept, such as the stimulus type, or whether the encoding varies with stimulus context. By comparing the performance of the classifiers at predicting attributes of neural responses to different stimuli, or at predicting different types attributes of the same data, one can make inferences about the relative information content or the neural signal.

One such classifier, described by Foffani and Moxon (2004) for classifying the discrete action potentials in response to a stimulus, uses a "prototype" PSTH for a group of responses to the same stimulus type, simply using the average

of those the PSTHs of those responses. These prototypes can then be used as part of a simple classifier to predict the stimulus type of a novel histogram. This is done by treating each bin in the histogram as a dimension, calculating the Euclidean distance between the novel example's PSTH vector and multiple prototypes, and assigning the novel example to the nearest prototype's group (Foffani and Moxon, 2004). While these classifiers can differentiate between neural responses to different stimuli, their mathematical properties are not well suited to high dimensional data. Specifically, the Euclidean distance metric loses its significance as dimensionality increases (Aggarwal et al., 2001; Beyer et al., 1999). Support Vector Machines (Vapnik, 1995), might better address these complexity and dimensionality concerns, since SVMs allow for non-linear separation of the data, along with the avoidance of some of the shortcomings of Euclidean distance. The SVM classifier is also more flexible, as it can be used with action potential histograms like PSTH classifier, or use other inputs, including representations of the LFP.

In Chapter 2, I describe the encoding of aural information in the auditory cortex, and the use of SVMs and related classifiers to decode neural signals. In Chapter 3, I discuss the specific approaches for inferring the information content contained in the neural activity of rat A1 in response to auditory stimuli. In Chapter 4, I describe the performance of these classifiers. I use the classifiers to predict the stimulus type that generated neural responses to both simple, single frequency tone or multi-frequency noise signals, or more complex natural speech sounds. I then discuss what the relative performance of the classifiers reveals about the neural encoding.

Chapter 2

Prior Work

2.1 Information Encoding in Neurons

Neurons are individual cells that are the fundamental unit of computation and communication in the nervous system. These cells have three main components: the cell body, the dendrites and the axon. There is a small difference in electrical potential across the cell membrane, with the internal electric charge of the cell is lower than that outside of the cell. Input into the cell initiates local chemical reactions, normally at the dendrites, that cause changes in this electrical potential. Once the change in potential propagates through the dendrites and cell body, it reaches the interface between the cell body and the axon, called the *axon hillock*. The cell membrane of this region, and of the axon itself, contain active processes that respond once the local potential reaches a critical threshold that is higher than the resting potential. This response results in a rapid and large increase in the electrical potential, followed by a rapid decrease. The increase in potential causes neighboring cellular regions to also rise above threshold. As a result, an *action potential*, or *spike*, travels from the axon hillock along the length of the axon. When this electrical signal reaches the end of the axon, it causes the release of a chemical signal across the synapse to the neighboring cells' dendrites.

The form and peak voltage of individual action potentials is consistent regardless of the input (Adrian, 1959), and, as such, the amplitudes of individual spikes encode no information. Instead, the encoding of information is captured by the frequency of the action potentials and the cells to which they propagate (Kandel et al., 1991). The generation of an action potential by an individual

cell also has an effect on the potential of the extracellular medium. While these individual action potentials are discernible from neurons located near the electrode, the magnitude of the individual signals drops rapidly with increasing distance. Nevertheless, many neurons acting in concert can still exhibit measurable effects. These local field potentials (LFPs) capture information about neural activity on a spatial scale greater than that of individual neurons and seem to be correlated to the information encoded by these larger groups of neurons (Liu and Newsome, 2006).

The primary auditory cortex (A1) serves as a gateway for auditory information to enter into the neocortex from subcortical auditory areas. Frequency information, along with other attributes of the sound, are known to be encoded in A1 (Kandel et al., 1991). This encoding of aural inputs has been shown to be affected by factors such as attention and stimulus novelty. The initial evidence for an attentional effect on neural activity was observed in the auditory cortex of the awake cat (Hubel et al., 1959). They found that, while some neurons did not respond reliably to simple recorded sounds, there was a response to complex or novel sounds that appeared to draw the animals' attention.

Fritz et al. (2007) trained ferrets on various auditory related tasks involving the detection of and discrimination between combinations of single and multi-frequency stimuli. The discrimination task used a series of reference stimuli, followed by a target stimulus. These stimuli consisted of a multi-frequency part followed by a single frequency signal, with this single frequency portion being shifted for the target stimuli. The neural responses to the stimuli inside the task context differ compared to responses to the same stimuli outside of the task context, producing a more focused response that peaks sooner after the onset of a stimulus within the task context.

Sloan (2009) reports differences in the responses of neurons in the rat auditory cortex to the same single frequency stimuli, depending on whether the stimulus was behaviorally relevant to the behavior context within which it was presented. Specifically, during engagement in a frequency change identification task, a decrease in both spontaneous neural activity and neural activity after an initial response to repeating identical tones is shown. Additionally, Sloan (2009) shows an increase in activity in response to the differing frequency tones that the animals were trained to identify, compared to the same stimuli outside

of the task context. This leads to larger differences being present between the repeat and novel tones in the task context. While the variation in the intensity of the neural response between repeat and novel stimuli, and between the passive and task contexts, makes it appear that the amount of information encoded is changing, this question warrants further investigation in order to understand it more clearly.

Whether recorded neural signals are in the form of discrete spikes or LFP signals, a means is needed to understand the content of these signals and their relation to the task in which the animal is engaged. Such a method should be able to determine whether information about the stimulus is encoded in the population and whether this encoding is affected by the context of the stimulus. One way to approach this problem is to first construct a model to predict some attribute of the stimulus or task given the associated neural activity. A set of such models can then be constructed, each of which predicting different aspects of the task or using forms of the available neural information. The performance of these models can then be compared to determine the relative extent to which the stimulus attributes are encoded in the population data. I discuss several such classification approaches in the coming sections.

2.2 Prototype Based Clustering

One way to neural activity is to capture the spike rate as a function of time. One way to approximate this function is to bin action potentials into fixed-length windows. Over a fixed number of time bins, this spike rate representation is a fixed-length vector. Such vectors for a population of neurons, when concatenated together, become a feature vector, which we refer to as $\mathbf{x}$. Each feature vector is assigned a label, y, denoting some class, such as the stimulus associated with the neural response, or an attribute of that stimulus. Given a set of $< \mathbf{x}, y >$ tuples, various supervised learning based classification methods can be used to construct models for predicting the label of a novel vector.

The Peri-Stimulus Time Histogram (PSTH) based classifiers discussed by Foffani and Moxon (2004) is one such method. Foffani and Moxon (2004) classify

novel examples by measuring the Euclidean distance to a class prototype, $\bar{\mathbf{x}}_\mathbf{Y}$, for each class, Y. Mathematically, this distance is calculated as follows:

$$d_{i,Y} = |x_i - \bar{x}_Y|,$$

where $d_{i,Y}$ is the distance between new example $\mathbf{x_i}$ and the class prototype for class Y. With this distance defined, the feature space between two prototypes is divided by a hyperplane that is equidistant to each prototype at all points on the plane. The class label for the new example is then determined by which side of the hyperplane it falls upon, represented by:

$$\hat{y}_i = \arg\min_Y d_{i,Y}.$$

In order to construct a model, the class prototypes must be calculated. The first step is to create prototype vectors for each each label value in the training set, by taking the mean of the PSTH histogram vectors $\mathbf{x_i}$ in the training set with that label. Formally, Foffani and Moxon (2004) define each prototype as follows:

$$\bar{x}_Y = \frac{1}{|Y|} \sum_{\mathbf{x_i} \in X_Y} x_i. \tag{2.1}$$

where $\bar{\mathbf{x}}_\mathbf{Y}$ is the value of the prototype for class Y and X_Y is the set of training examples for class Y.

Foffani and Moxon (2004) apply this classifier to responses in the primary somatosensory cortex to whisker stimulation, using the classifier to identify the whisker being deflected in an anesthetized animal. They report an accuracy of 37% in classifying between the neural responses to the mechanical deflection of 20 different whiskers, using 40 ms of neural data. The authors show that the PSTH for a novel example will be most similar to the prototype histogram corresponding to the whisker being deflected in the test example with a probability better than chance.

Foffani and Moxon (2004) also compare the performance of the PSTH based classifier to the performance of a classifier using Linear Discriminant Analysis (LDA). LDA, also known as Fisher's Linear Discriminant, separates the classes based on their prototypes. However, while this use of a center point for each class is similar to Foffani and Moxon (2004)'s approach, instead of using Euclidean distance to determine the class, LDA utilizes the variances of each class to divide the classes to maximize the separation between them (Bishop, 2006). Foffani

and Moxon (2004) found that the performance of the PSTH classifier did not differ from that of LDA when the dimensionality was small, either when principle component analysis was used to greatly reduce the dimensionality (50% or less of the variance accounted for), or larger bin sizes were used. However, PSTH performs best for the highest dimensionality cases, when most of the original dimensionality is preserved and small bin sizes are used.

While the PSTH classifier does have the advantage of being relatively simple, the pure Euclidean nature of the distance metric can cause problems for high dimensional feature spaces, such as multi-channel spike rate estimates using high bin size resolution. Beyer et al. (1999) have shown that as dimensionality increases, the difference in the L_2 distances between pairs of points in a uniform and random, though possibly correlated, distribution of points in $\mathbb{R}^d$ become small in relation to the magnitudes of the distances. Further aggravating this lack of contrast is the fact that due to the definition of a Euclidean metric, all dimensions are treated as equally important to determining the class label. As a result, noisy or irrelevant dimensions make the same contribution to the distance as dimensions whose variation is meaningful. Because of this, other, more robust classifiers might make a better choice for classifying neural data.

2.3 Support Vector Machines

Like the PSTH classifier, Support Vector Machine (SVM) classifiers also divide the feature space into classes based on the locations of the training examples in that feature space (Vapnik, 1998). A key difference of the SVM, however, is that a hyperplane is selected so as to best separate the training samples of the two classes and to maximize the distance between the hyperplane and the training points. The former property allows one to address differences in the covariance between the classes. The latter property attempts to enforce the selection of as conservative a boundary as possible. Specifically, small variations in positions of the training set points should not cross the hyperplane. The class y of a novel vector $\mathbf{x}$ is determined by the sign of the following equation:

$$f(\mathbf{x}) = \mathbf{w}^T \Phi(\mathbf{x}) + b, \tag{2.2}$$

where $\Phi(\mathbf{x})$ is some expansion of the feature vector $\mathbf{x}$, $\mathbf{w}$ is a vector of weights, and b is an offset constant, and the class label of $\mathbf{x}$; $y = \{-1, 1\}$, is determined

by the sign of $f(\mathbf{x})$. Additionally, the absolute value of $f(\mathbf{x})$ in Eq. 2.2 can be required to be ≥ 1. Given a training set of tuples, $< x_i, y_i > \in B_N$, where B_N is the set of N training examples, one can solve for w and b by minimizing (Vapnik, 1998; Schölkopf and Smola, 2002):

$$E = \frac{1}{2}\mathbf{w}^T\mathbf{w}, \tag{2.3}$$

subject to the constraint:

$$y_i(\mathbf{w}^T\Phi(\mathbf{x_i}) + b) \geq 1. \tag{2.4}$$

The minimization in Equation 2.3 is intended to keep the weights, $\mathbf{w}$, small, while selecting values that properly divide the training examples into classes. Equation 2.4 ensures that the values selected for $\mathbf{w}$ and b divide the training examples with a non-trivial margin.

While in the ideal case, the two classes can be cleanly divided, in reality, dividing the training data perfectly may not be possible. Additionally, even if the data are separable, perfect separation may result in a model that, while fitting every outlier in the training data, does not generalize well in the case of noisy training data. To address these two cases, the above formulation can be modified to allow for some misclassification of the training examples in exchange for a more general solution. For this non-separable case, the formulation becomes:

$$\hat{E} = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_i^N \xi_i, \tag{2.5}$$

subject to the constraint, and where the slack variable, ξ_i, is an error term and is non-negative:

$$y_i(\mathbf{w}^T\Phi(\mathbf{x_i}) + b) \geq 1 - \xi_i. \tag{2.6}$$

The introduction of ξ_i allows for some errors in the classification of the training examples. However, due to the $C\sum_i^N \xi_i$ term in Equation 2.5, the total magnitude of these errors is controlled, with the parameter C determining how much weight this error is given compared to the minimization of $\mathbf{w}$.

The Lagrangian Multiplier approach for solving such a constrained optimization problem yields the following cost function:

$$L = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^N \xi_i - \sum_{i=1}^N \alpha_i(y_i(\mathbf{w}^T\Phi(\mathbf{x_i}) + b) - 1 + \xi_i) - \sum_{i=1}^N \beta_i\xi_i. \tag{2.7}$$

The first term in Equation 2.7 represents the overall magnitude of the weights, $\mathbf{w}$, and the second controls the amount of error allowed. The third term encodes the constraint in Eq (2.6), and the forth term encodes the constraint that $\xi \geq 0$. One property to note about the values of the Lagrangian multipliers in the third term is the fact that they control the influence of the individual training vectors $\mathbf{x_i}$. Thus an individual $\mathbf{x_i}$ only contributes to the position of the hyperplane when the corresponding α_i is non-zero.

The optimal value of $\mathbf{w}$ lies at a saddle point in L, and can be found by minimizing Eq. 2.7 with respect to $\mathbf{w}, \mathbf{b}$ and ξ_i and maximizing it with respect to α_i and β_i. Minimizing Eq. 2.7 with respect to $\mathbf{w}$ and b ensures that these terms do not become unnecessarily large while partitioning the training examples, avoiding degenerate solutions. Minimizing with respect to ξ_i keeps the error low. Maximizing L with respect to α_i drives α_i to zero when the corresponding training example is far from the boundary between the two classes, eliminating the influence of these training examples since they do not add information about the location of the boundary. Finally, maximizing L with respect to β_i allows the relative penalty of individual error values to be controlled by selecting individual penalties for the error terms ξ_i.

Because the solution to this optimization is at a saddle point, the solution can be found by setting the following partial derivatives are equal to zero:

$$\frac{\partial L}{\partial \alpha_i} = -(y_i(\mathbf{w}^T \Phi(\mathbf{x}) + b) - 1 + \xi_i) = 0,$$

which simplifies to:

$$\mathbf{w}^T \Phi(\mathbf{x}) + b - 1 + \xi_i = \frac{1 - \xi_i}{y_i} = y_i(1 - \xi_i),$$

since $y_i = \frac{1}{y_i}$. This equation is important because it recovers the equality of the constraint in Eq 2.6.

The partial derivatives with respect to β_i and ξ_i are also used to find the solution:

$$\frac{\partial L}{\partial \beta_i} = -\xi_i = 0,$$

and

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \beta_i = 0. \tag{2.8}$$

However, Eq. 2.8 can be written as:

$$C = \alpha_i + \beta_i. \tag{2.9}$$

Since a smaller value of β_i reduces the cost of misclassifying $\mathbf{x_i}$, the relationship between α_i and β_i dictated by Eq. 2.9 ensures that either the error ξ_i is more highly penalized or $\mathbf{x_i}$ is given a larger influence in determining the boundary surface.

Finally, the derivatives for the hyperplane parameters $\mathbf{w}$ and b are:

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{N} \alpha_i y_i = 0, \tag{2.10}$$

and

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w}^T - \sum_{i=1}^{N} [\alpha_i y_i \Phi(\mathbf{x_i})]^{\mathbf{T}} = \mathbf{0}, \tag{2.11}$$

which becomes:

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \Phi(\mathbf{x_i}). \tag{2.12}$$

The $\mathbf{w}$ term in Eq. 2.7 can be replaced with 2.12, allowing the equation to be expressed in terms of the Lagrangian multipliers α, the offset b, the training set labels y and vectors and $\mathbf{x}$:

$$L = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \Phi(\mathbf{x_i})^T \Phi(\mathbf{x_j}) - b\sum_{i=1}^{N} y_i \alpha_i + \sum_{i=1}^{N} \alpha_i - \sum_{i=1}^{N} \alpha_i \xi_i - \sum_{i=1}^{N} \beta_i \xi_i + C\sum_{i=1}^{N} \xi_i. \tag{2.13}$$

By Equation 2.10 the $-b\sum_{i=1}^{N} y_i \alpha_i$ term is zero at the saddle point. Since $C = \alpha_i + \beta_i$, $C\sum_{i=1}^{N} \xi_i$ becomes $\sum_{i=1}^{N} \alpha_i \xi_i - \sum_{i=1}^{N} \beta_i \xi_i$ and the various ξ_i terms cancel out, yielding:

$$L = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \Phi(\mathbf{x_i})^T \Phi(\mathbf{x_j}) + \sum_{i=1}^{N} \alpha_i, \tag{2.14}$$

this can then be maximized with respect to α, subject to the constraints:

$$0 \leq \alpha_i \leq C,$$

and

$$\sum_{i=1}^{N} \alpha_i y_i = 0.$$

Computing the solution to Eq. 2.14 can be difficult or impossible, however, depending on the dimensionality of the expansion $\Phi(\mathbf{x})$. Specifically, the

$\Phi(\mathbf{x_i})^T\Phi(\mathbf{x_j})$ term in Eq. 2.14, can be of very high, or even infinite, dimensionality. While the expansion $\Phi(\mathbf{x})$ provides a means to represent non-linear relationships in the feature space, if the dimensionality is prohibitively large, it can preclude the computation of a solution to the optimization problem. To make the problem more feasible, this term can be transformed into a more manageable form using a *kernel function*(Boser et al., 1992):

$$\Phi(\mathbf{x})^T\Phi(\mathbf{x_i}) = K(\mathbf{x}, \mathbf{x_i}). \tag{2.15}$$

Substituting Eq. 2.15 into Eq. 2.14 yields:

$$L = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(\mathbf{x}, \mathbf{x_i}) + \sum_{i=1}^{N}\alpha_i. \tag{2.16}$$

Once the values of the Lagrange multipliers, α_i, have been computed, those with nonzero values correspond to the *support vectors*, defined as those training vectors that are closest to, and thereby define, the hyperplane. The values for α and b can then be used to predict the label of a novel feature vector using the following equation (Vapnik, 1998):

$$f(\mathbf{x}) = \sum_{i=1}^{N} y_i\alpha_i K(\mathbf{x}, \mathbf{x_i}) + b. \tag{2.17}$$

For a function to be used as a kernel function, $K(\mathbf{x}, \mathbf{x_i})$ it must meet the criteria of the Mercer theorem. Namely that $K(\mathbf{x}, \mathbf{x_i})$ is positive definite, meaning that:

$$\sum_{i,j} c_i c_j K(\mathbf{x_i}, \mathbf{x_j}) \geq 0$$

where $\mathbf{x_i}$ and $\mathbf{x_j}$ are training examples and c_i and c_j are constants in $\mathbb{R}$ (Schölkopf and Smola, 2002). If the above requirement is met, the kernel $K(\mathbf{x_i}, \mathbf{x_j})$ is a Mercer kernel and can be approximated by:

$$\sum_{k=1}^{n} \lambda_k \phi(\mathbf{x})^T\phi(\mathbf{x_i}), \tag{2.18}$$

where n is the number of terms being used for the approximation, and the constant $\lambda_k > 0$ (Schölkopf and Smola, 2002).

There are many functions meeting the Mercer criteria that are commonly used as the $K(\mathbf{x}, \mathbf{x_i})$ similarity metric described above. The simplest such kernel is the linear kernel ($\Phi(\mathbf{x}) = \mathbf{x}$). The kernel function is simply:

$$K(\mathbf{x}, \mathbf{x_i}) = \mathbf{x}^T\mathbf{x_i}.$$

In another common, yet more complex, kernel is the polynomial kernel (Vapnik, 1998), where $\Phi(\mathbf{x})$ consists of all possible combinations containing up to q dimensions of , i.e:

$$\Phi(\mathbf{x}) = \begin{bmatrix} x_1^2 \\ x_1 x_2 \\ x_2 x_1 \\ x_2^2 \end{bmatrix}$$

when $q = 2$ and x_1 and x_2 are elements in $\mathbf{x}$, a two element vector. When $q > 1$ this expression expands to include all possible products of individual elements in $\mathbf{x}$ and $\mathbf{x_i}$ with q determining the maximum number of elements of $\mathbf{x}$ and $\mathbf{x_i}$ that can be combined into a single term:

$$K(\mathbf{x}, \mathbf{x_i}) = (\mathbf{x}^T \mathbf{x_i} + 1)^q.$$

Another common kernel is the Radial Basis Function (RBF), also called the Gaussian kernel. The dimensionality of $\Phi(\mathbf{x})$ is of unlimited dimensionality, as it is equal to the total number of training examples (Schölkopf and Smola, 2002). This kernel can be represented in a simple manner, however, as the dual space form is simply a radial basis function:

$$K(\mathbf{x}, \mathbf{x_i}) = e^{-\gamma ||\mathbf{x} - \mathbf{x_i}||^2}, \tag{2.19}$$

where γ defines the width of the kernel function and is a parameter to be selected. Assuming that all of the training vectors $\mathbf{x_i}$ are unique, the RBF kernel is associated with an expansion $\Phi(\mathbf{x})$ of dimensionality equal to the number of training vectors (Schölkopf and Smola, 2002). Additionally, since the points $\Phi(\mathbf{x_i})$ are linearly independent when all $\mathbf{x_i}$ are unique, Schölkopf and Smola (2002) show that the training examples are separable for all possible assignments of the labels $y = \{-1, 1\}$.

2.4 Fourier Space Features

The kernels discussed above are well suited to relatively small numbers of discrete features, which works well for data such as binned spike counts. Unlike the discrete spike times, LFPs are a continuous voltage as a function of time. Such continuous signals must be temporally aligned before they can be compared in

the time domain. Alternatively, the signals can be represented in the frequency domain, avoiding the potential difficulty of precise alignment. In addition to this simplification, the significance of the frequency components of LFPs makes frequency space features a logical choice for LFP signals (Wolpaw et al., 2002).

The discrete Fourier transform provides a means to derive the frequency domain representation from the LFP. As a continuous, periodic, function can be represented by a Fourier series, a finite timeseries S_t ranging from $t = 0$ to $t = N - 1$ can be represented by a discrete Fourier transform, a summation of terms of the form (Cooley et al., 1969):

$$S_t = \sum_{n=0}^{N-1} A_n e^{2i\pi tn} \tag{2.19}$$

where A_n is a complex number representing the frequency component n (Schoenberg, 1950), and i is $\sqrt{-1}$

The coefficients A_n can be calculated as follows (Cooley et al., 1969):

$$A_n = \frac{1}{N} \sum_{t=0}^{N-1} S_t e^{2i\pi tn/N}, \tag{2.19}$$

where S_t is the value at time t for $t = 0$ to $N - 1$. The equations of Schoenberg (1950) describe the real component of A_n as representing the signal's *cos* component at frequency n and the imaginary component of A_n is the magnitude of the *sin* component of the signal. These coefficients can then become features, either as a single value for each frequency or as separate real and complex components. Since neighboring frequencies should be making similar contributions, the coefficients from neighboring frequencies in the same frequency band average to create *band power* features. These features can then be used as the feature vector for a standard classifier, such as SVMs. Fourier based features have been used by Lindberg et al. (2011) to decode information recorded from the primary motor cortex, using the frequency components of the neural signal to predict the recorded electrical, or electromyogram (EMG), signal in the corresponding muscle groups. The authors use the frequency components with a Wiener cascade model to model the non-linear EMG signal. Lindberg et al. (2011) compare the performance of the model built from the LFP signal to a model using discrete spike information, showing that the LFP based model was nearly as accurate as the spike based model.

2.5 Classifiers for Interpreting Neural Data

The relationship between the class label of a stimulus and the corresponding population activity, whether described as a set of spike counts or as LFP signals, is complex. While it cannot be determined directly how much information about a given stimulus is reflected in the activity of a set of cells, the relative strengths of the information encoded about various aspects of the corresponding stimuli can be inferred. Specifically, by training classifiers to predict different attributes of the data and then comparing their performance, one can infer the relative degree to which the population signal encodes these attributes. Additionally, classifiers can be trained to use different aspects of the neural activity information, yielding insights into the spatial and temporal encoding of the information.

Various classifiers, including those described above, have been used in the analysis of neural data. Funamizu et al. (2009) compare several classifier types, using binned spike counts from multiple channels, comparing the performance of SVM based classifiers, k-nearest neighbors (kNN) classifiers and LDA based classifiers on population data recorded from the rat auditory cortex. They used the classifiers to identify the frequency of single tone stimuli presented to anesthetized animals. The spike counts were binned into 3 ms bins, and the bins with the least contribution to the classifier performance were dropped one at a time, thus identifying the the subset of bins with the largest contribution to classifier performance. Funamizu et al. (2009) show that the best performing SVM outperforms kNN, but they do not show a significant difference between SVM and LDA, with the best subset of bins being selected for each classifier. However, to achieve peak performance LDA utilized a smaller subset of the full feature vector than SVM. This indicates that LDA is more negatively impacted by the inclusion of unnecessary features. The SVM kernels, along the selection of only some training examples as support vectors for the model, allow it to better handle these extraneous dimensions.

While SVMs perform well with high dimensional and complex data, their higher complexity relative to other common classifiers is one of the main disadvantages of SVMs. These disadvantages include the fact that the learned model is difficult to interpret through inspection. Additionally, the high computational requirements of SVMs may not be suitable for large amounts of data and/or

time critical applications (Lotte et al., 2007). However, SVMs have been successfully applied to online classification applications (Olson et al., 2005; Garcia et al., 2003). Specifically, Olson et al. (2005) use SVMs to classify spike rate information as part of a real-time interface. Classifiers were trained using data from the rat motor cortex while the rats pressed a paddle to their left or right in response to a cue. Once trained, the classifiers were used to decode the movement information in real time, using the output of the classifier in place of the actual paddle actuation. The successful use of SVMs in this system shows that SVMs are both a effective and efficient tool for classifying neural signals.

Bansal et al. (2011) compare the information available with spike counts and low frequency LFP in the primary motor cortex and ventral premotor area of monkeys reaching for and grasping moving objects. They used a Kalman filter based approach to predict position and velocity information from the low frequency components of the LFP signal. They show that a single LFP channel, or a small number of such channels, contained could be decoded more accurately than the action potentials recorded from the same number of sites. However, the best spike signals, or larger populations of such signals, outperformed the low-frequency LFP. Mehring et al. (2003) also show that LFP performs better than spike counts for predicting arm kinematics using recordings from the motor cortex of monkeys when small numbers of channels are used.

2.6 Performance Metrics

While classifiers can be employed as a means of inferring the information encoded by the set of cells being recorded, in order to compare the performance of different classifiers, a performance metric is needed. While percentage accuracy is the simplest method for evaluating the performance of a classifier, it does not always correctly represent performance, as the accuracy metric gives equal credit for all correct predictions and cannot identify a classifier that predicts one class extremely well but performs poorly on a less common, but equally or more important case. An extreme example of this is that for data containing a majority of a given stimulus label, a classifier could achieve high accuracy by simply assigning all samples the most common label, never successfully identifying the less common stimulus cases. To prevent this, we select a performance metric

that accounts for such degenerate cases by requiring a classifier to correctly identify all labels in order to perform well. While several such metrics exist, we choose the Peirce Skill Score (PSS) as a performance metric (Peirce, 1884). This metric, while simple, accounts for false positives and is robust to unbalanced class labels (Woodcock, 1976). The definition of PSS can be expressed for the multi-class case as follows (Wilks, 2006):

$$PSS = \frac{\sum_{k=1}^{R} p(r_k, w_k) - \sum_{k=1}^{R} p(r_k)p(w_k)}{1 - \sum_{l=1}^{W} p(w_l)^2},$$

where r is the predicted outcome, w is the observed outcome, R is the total number of possible predicted outcomes and W is the total number of possible observed outcomes. Note that in this general case, it is possible for the number of possible observed outcomes to differ from the number of possible predicted outcomes.

In this two-class case of reference and target stimuli, PSS simplifies to the following equation:

$$PSS = \frac{\text{hits}}{\text{hits} + \text{misses}} - \frac{\text{false alarms}}{\text{false alarms} + \text{correct rejections}}.$$

For target stimuli, hits are defined as correctly identified targets, and misses as incorrectly identified targets. Correct rejections are reference stimuli that were correctly identified as such while false alarms are reference stimuli mis-identified as target stimuli. The *true positive rate* is defined as:

$$\frac{\text{hits}}{\text{hits} + \text{misses}}.$$

Similarly, *false positive rate* is:

$$\frac{\text{false alarms}}{\text{false alarms} + \text{correct rejections}}.$$

Doswell et al. (1990) discuss concerns about the use of PSS in evaluating the prediction of unbalanced labels. Specifically, these concerns include the fact that as the number of actual reference examples gets sufficiently large compared to the total number of examples, the false positive rate trends to zero as long as the total number of false alarms remains small. The concern is that this reduction in the score penalty of false alarms will lead to a biased classifier still appearing to perform well.

Chapter 3

Experimental and Analysis Methods

In this chapter I discuss specific approaches for inferring the information content contained in the neural activity of rat A1 in response to auditory stimuli. Specifically, I discuss the experiments during which the neural signals were recorded, how these signals are represented, and the machine learning classifiers that I use to examine to information content of the signals.

3.1 Auditory Change Detection Task

All rat experiments were performed in the Rennaker Laboratory at the University of Texas at Dallas. Fifteen rats were trained to listen to a series of single frequency tone stimuli and respond when either a tone of a different frequency or a multi-frequency *broadband noise* sound was heard, (Sloan et al., 2009). The rats remain stationary with their nose in a nose-poke for the duration of the reference tones. The animals are then rewarded for responding by removing their nose from the nose-poke upon hearing the target stimulus.

A trial ends either with a pair of target stimuli, or when the animal withdraws without a target stimulus being played (a *false alarm*). Each of the stimuli are 200 ms in duration. Tones are separated by silent periods that are either 100 or 200 ms in duration; this duration was consistent for all sessions for a given rat. The trial is considered a *hit* if the animal withdrew from the nose-poke within 600 ms of the onset of the first target stimulus. In addition to the normal trials, catch trials were included to estimate the false alarm rate. These trials consisted of a finite series of reference tones with no target at the

end. If the animal did not react during the reference sequence, it was considered a *correct rejection*; otherwise the trial was considered a *false alarm*.

A reference tone frequency was selected for each trial from a set of frequencies between 2.31 kHz and 27.7 kHz (Sloan et al., 2009). For the discrete tone variation of the task, target frequencies were generated by shifting the reference frequency by a discrete percentage. The shift could be up or down and was a percentage of the trial's original reference frequency from the following set: {1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 12%, 14%, 16%, 18%, 20%, 36%, and 72%}. Instead of the shifted frequency target tones, in a broadband noise version of the task used a multi-frequency noise signal as a target tone, the animals were able to identify with a much higher accuracy than the discrete tones (Sloan, 2012).

In the passive version of the task, sequences of these same stimuli were also played while the animals were in a different cage and not engaged in the task, but were free to move around as they wished. The neural responses to the stimuli recorded to provide a *passive* context comparison to the responses to the stimuli within the task context.

In all, for the frequency change detection task, 227 sessions were recorded, along with an additional 139 passive sessions using these stimuli. For the broadband noise detection task 33 separate data collection sessions were collected. Additionally, there are 69 passive sessions using the broadband noise detection stimuli.

The discrete tone frequency change and broadband noise stimuli tasks provide a simple case in which to compare neural responses during the task with responses to the same stimuli outside of the task context. These simple stimuli, however, provide little opportunity to study the encoding of sounds with a more complex structure. Additionally, because of concerns about including movement correlates in the voltage signal, the animals' rapid reaction times to the simple stimuli allowed for only very short windows of neural data to be used.

In order to investigate the encoding of more complex stimuli, a running speech discrimination task is also used. This study uses 13 different animals trained to perform a word recognition task. The mechanics of this task are similar to the discrete tone and broadband noise tasks, in that the animals are again trained to react to a target stimulus. However, in this case, the stimuli

are similar words with differing first consonants, as described by Engineer et al. (2008): "Bad," "Gad," "Sad," "Tad," and the target word: "Dad." One major difference between this and the discrete tone and broadband noise studies is that the any of the four reference words could be presented in any order prior to the target word. This makes the sequence of reference stimuli non-homogeneous and unpredictable. In the running speech discrimination task, the animals had to identify the target word specifically, rather than recognizing a change in the stimulus, as in the tone and noise tasks.

Unlike the tone and noise signal task, no equivalent passive versions of the task were performed. As such, these data cannot be compared to the responses to equivalent stimuli presented outside of the task context.

3.2 Recording and Signal Processing

After being trained on the above tasks, the rats were implanted with 15 channel microwire electrode arrays in the right auditory cortex (Rennaker et al., 2005). After implantation, data were recorded from the auditory cortex while the animals performed the task. Similar recordings were also made while a static sequence of the task stimuli were played without the animal being engaged in the task. The signals were recorded at 25 kHz, with a bandpass filter from 5-5000 Hz. Action potentials were identified using a threshold of twice the RMS value of the signal. While an attempt was made to separate these spikes by individual neurons, this was often impossible to do so effectively and the data channels are instead considered as Multi-Unit Clusters (MUCs), rather than individual neurons. The MUCs were manually rated as to whether they were *driven* by a neural signal by confirming that the peak spike rate following a stimulus was at least two standard deviations above the average spike rate in the 50 ms period prior to all trials in that session.

A behavioral session consisted of a series of trials presented to an animal on the same day. The number of stimuli presented during each behavioral session, depended on the motivation of the rats and their behavioral performance. A set of inclusion criterion were used to ensure that a minimum number of trials were available for analysis. First, a behavioral sessions had to include at least 200 individual stimulus presentations. Second, the rats behavior performance, as

measured with the PSS, had to exceed 0.5, since poor behavioral performance on the task was presumed to indicate inadequate task engagement. Additionally, in order to ensure that neural data of sufficient quality is recorded during a session, at least one MUC in that session's recording has to have been identified as being driven. As long as a session met these criteria, all available channels are included in the feature vector, regardless of whether all channels were labeled as driven. While MUCs that appear undriven might not hold very much information, they still may have some information content that the classifier could use.

Once a session is selected for inclusion, the trials within that session, along with the responses to individual stimuli, are then filtered to reduce the number of non-representative examples being used for training and testing the classifier. First, the beginning three tones from each trial are excluded from the training data to ensure that those samples used to build the classifier contain only neural signals during which the animal was engaged in the task. This is due to the potential for significant unrelated activity in the neural signal before the rat is engaged in the task. For the test data, this exclusion is slightly less conservative, as only the first tone from each trial is excluded. The assumption is that, while the rat may still not have been engaged in the task during the second and third reference stimuli, this is unlikely, and once a classifier model has been generated using high quality examples, the lower quality examples may be used for testing.

Additionally, since target stimuli that are unlikely to have actually been correctly identified by the animal provide poor examples for building a classifier to identify target stimuli, such lower quality examples are removed in order to give the classifiers more representative training data. Specifically, when a classifier is being trained to classify the responses to the tone stimuli, only those tones with a frequency shift greater than ten percent are used, since very small frequency shifts are too close to the animals' discrimination thresholds (Sloan et al., 2009). The target stimuli for the broadband noise or word recognition do not need to be filtered in this way since the target stimuli for these tasks are fixed.

Classifiers that predict the animal's response to a stimulus, the target tones are filtered differently. Specifically, training the classifier using target stimuli with very small frequency shifts is no longer a concern, because the classifier is identifying the animals' responses rather than the actual stimulus type. Instead,

target stimuli for which the animal responded correctly, but if the animal's reaction time doesn't coincide with the expected timing of a movement in response to the stimulus, it indicates that the animal's movement was not in response to the stimuli. This is a concern, as such examples could introduce misleading examples into the training data. For this reason, the target stimuli to be included for training response prediction classifiers for the frequency shift task are selected by reaction time. Reaction times that are too fast are assumed to reflect a decision to move before the stimulus was heard. Similarly, reactions that were too slow are also assumed to reflect behavior decisions not made in response to the stimulus. Only stimuli from the the middle 80% of reaction times are used to ensure that the responses were indeed reactions to that tone.

Finally, because there were many more examples of responses to perceived or actual reference stimuli than to the equivalent target stimuli, the training examples were subsampled to include equal numbers of each class. The training sets were balanced in this manner in order to avoid classifiers that simply guess the label to be the majority label.

3.3 Labeling Schemes

The primary analysis method in this study uses two classifier algorithms to classify neural responses to individual tones of the trial sequences using one of two label schemes. The first labeling scheme used for the classifiers is whether the neural response was to a reference stimulus or to a target stimulus, called *true class* prediction. The second labeling scheme was whether the neural response predicted a correct behavioral response from the subject, which may be called *perceived class* prediction.

3.4 Classifier Construction

For spike data from both the tone and speech experiments, neural responses are binned into Peri-Stimulus Time Histograms with 10 ms or 4 ms time bins respectively over a response to a stimulus. For the tone experiments, only the data from tone onset to 50 ms after tone onset is used in order to address concerns about correlated noise from muscle activity when the animal reacted

to the target tone. Subjects had mean nose-poke withdrawal times of 220 ms, meaning that there is no risk of including movement correlate in this initial 50 ms window. For the speech experiments, the reaction times are slightly longer at 246 ms, thus allowing for 128 ms with which to analyze the often more complex responses.

For the speech stimuli, the LFP signals are also used as a classifier inputs. Because the LFP is a continuous signal with strong frequency components to the information encoded, the coefficients generated by a discrete Fourier transform provide a logical means of building meaningful, concise feature vectors. Both real and complex components of the Fourier coefficients are used, with the mean of each being taken over all computed coefficients in the following frequency ranges: 0 Hz (DC offset), 1-4 Hz (Delta band), 4-8 Hz (Theta band), 8-13 Hz (Alpha band), 13-30 Hz (Beta band), 30-40 Hz, 40-50 Hz, 50-60 Hz, 60-70 Hz, 70-80 Hz, 80-90 Hz, 90-100 Hz (Gamma band, divided into 10 Hz steps), and frequencies greater than 100 Hz. All frequency ranges are inclusive of the lower bound but not the upper, so that the frequency bands do not overlap each other or allow gaps between bands.

3.4.1 Classifier Details, Kernel and Parameter Selection

Once the feature vectors are constructed, two classifier algorithms are tested for this analysis: a PSTH-based classifier and a SVM-based classifier. The PSTH based classifier is implemented in MATLAB. For the SVM classifier, I use the SVM-C implementation of LibSVM (Chang and Lin, 2001). Although a number of kernel functions were discussed in the previous chapter, the results described in the following chapter use a radial basis function (RBF) kernel.

For the classifiers using spike bin based feature vectors, I consider possible values of C in the range of 2^{-5} to 2^{15}, and γ in the range of 2^{-15} to 2^{5}. The ideal C and γ values for the LFP Fourier space based feature vectors tended to be higher than those for the spike count based features. As a result I consider values of C in the range of 2^{-3} to 2^{25} and values of γ between 2^{-7} and 2^{15}. A single pair of parameters is selected for all the data sessions using a particular stimulus and of particular label type.

3.4.2 Training Sample Labels and Classifier Training and Evaluation

For a single data session, I train and test using 20-fold cross validation. The examples from a session are partitioned into 20 equal-sized subsets (folds). A total of 20 classifiers are constructed, each trained with 18 of the 20 folds. The 19th fold is used as the *validation data set* for parameter selection, and the 20th fold is used for reporting classifier performance. Across all data sessions of a particular type, we select the parameter pair that maximizes the average *aggregated PSS* across the validation data sets.

Aggregated PSS is used to address the small size of the individual folds. For a single data session, the contingency tables across all 20 models (for either the validation or test data sets) are combined (summed) to construct an aggregated contingency table. The PSS metric is then computed over this aggregated table.

Chapter 4

Experimental Results

4.1 Tone Stimulus Change Detection

I previously discussed two classifier types that have been used to classify neural signals; PSTH (Foffani and Moxon, 2004) and SVMs (Vapnik, 1998). The PSTH based classifier described by Foffani and Moxon (2004) selects a label based on the distance to the mean spike histogram for each example class. SVM classifiers use a more sophisticated method to determine the label, selecting a set of training examples, *support vectors*, that are most indicative of the boundary between classes. For the tone and broadband noise experiments a window size of 50 ms, beginning at stimulus onset, and a bin size of 10 ms is used to selected the portion of the neural signal used for these classifiers. This window is chosen in order to ensure all useful neural information is provided to the classifier, while preventing the introduction of movement correlates.

The SVM classifier is expected to outperform the PSTH based classifier, since the mathematical properties of SVMs better allow them to differentiate between points in high dimension space. The performance of the SVM and PSTH based classifiers is compared for both the discrete tone and broadband noise experiments. The results shown in Figure 4.1 confirm this, with the SVM classifier achieving a PSS of .67 and .19 on the behaving context broadband noise and behaving context discrete tone responses, respectively. The PSTH based classifier achieves a PSS of .60 and .15 on these same responses. These differences are significant with $p < .001$ for both the broadband noise and discrete tone cases.

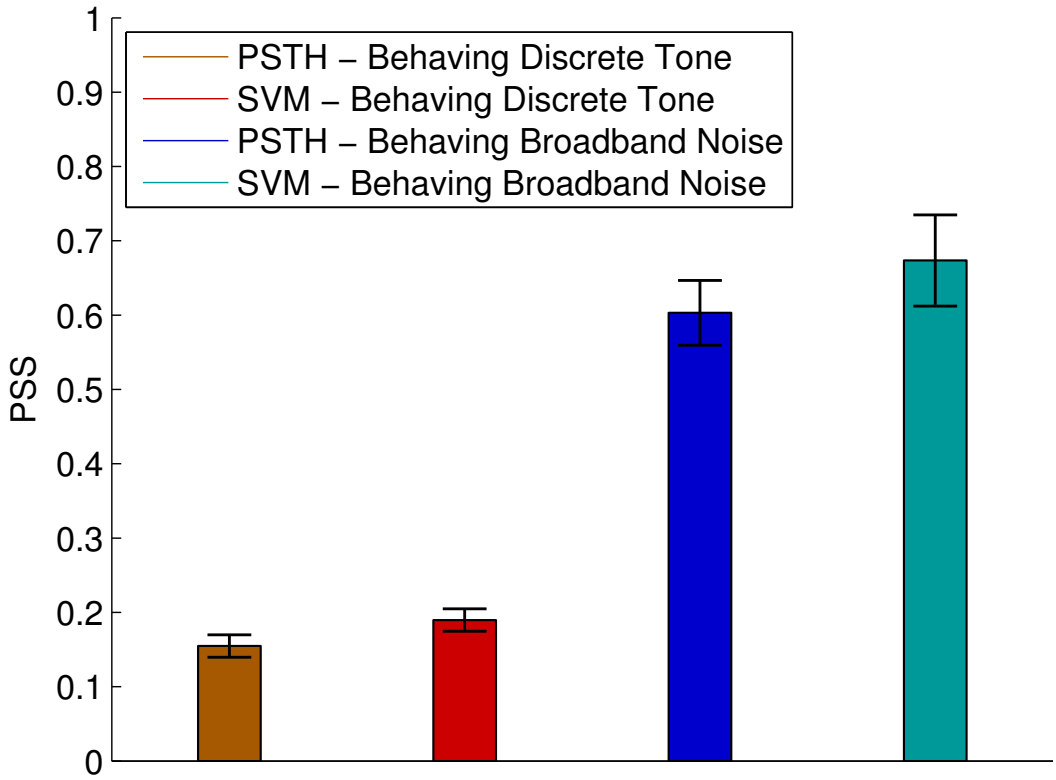
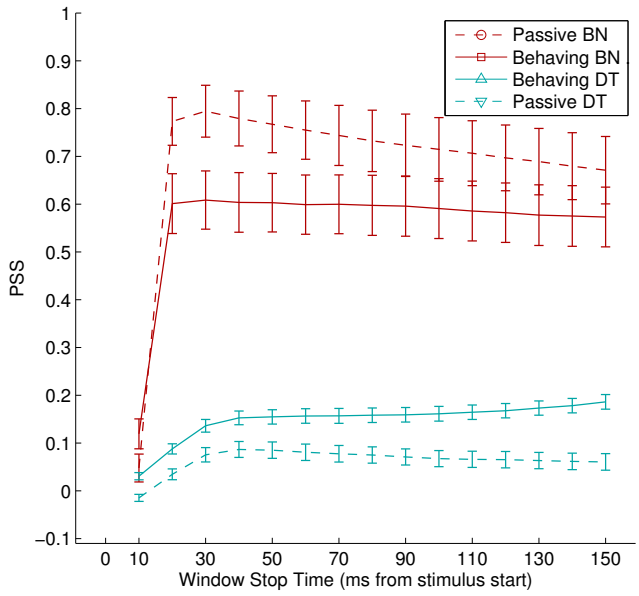


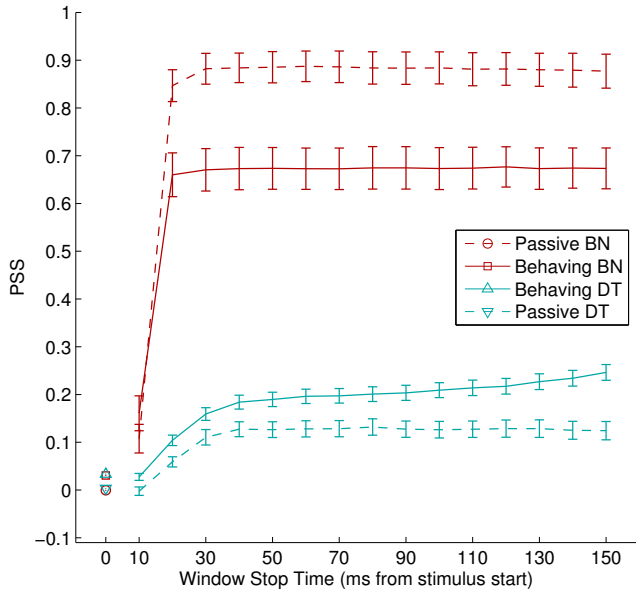
Figure 4.1: Comparison of PSTH based and SVM classifiers, on responses to tone and broadband noise stimuli.

4.1.1 Effect of Time Window

To assess the sensitivity of the classifiers' performance to the choices of window and bin size, and to verify that additional information does not occur in the neural signals outside of the chosen analysis window, different classifiers were constructed using windows of varying sizes. Figure 4.2 shows classifier performance for various window sizes for the broadband noise and discrete tone data sets, using both the PSTH and SVM classifiers. The single points, not connected by lines, in Figure 4.2b represent windows starts 50 ms before the stimulus onset and lasting until stimulus onset. This time window contains no information about the stimulus and is reflected by the fact that performance for these windows is near random ($PSS \approx 0$). Performance is almost as low for windows starting at stimulus onset and only including 10 ms of data. However, as the window grows, performance rises rapidly as the window begins to overlap the stimulus. Performance saturates at 30 ms for the passive and active



(a) PSTH



(b) SVM

Figure 4.2: Mean and standard deviation of classifier performance for both the PSTH (a) and SVM (b) classifiers, as a function of the window size. The points at t = 0 correspond to using time windows starting 50 ms before and ending at stimulus onset. All other windows start at stimulus onset and end as indicated.

discrete tone stimuli and 50 ms for the corresponding broadband noise stimuli, leveling off completely after this point for all cases except the active discrete tone experiments. For the passive cases, the asymptote in dictates that there is no additional information being added after the initial rise in performance at the beginning of the time window. For the active discrete tone case, however, performance does continue to rise slightly as information from a larger period of time is available. The performance increases as the window expands beyond 100 ms from the stimulus onset. This performance increase may be due to correlated movement related modulation that contributes to the information available to the classifier. After the initial increase, the active broadband noise performance levels off almost completely. While there is also a possibility of movement correlates in this case, as well, since performance is already so high ($PSS \approx .9$), there may not be much additional information offered by any correlated noise.

The PSTH classifier, shows a similar pattern for smaller window sizes. However, as the window grows the PSTH classifiers show a slightly different pattern compared to the SVM classifiers. Unlike the SVM classifiers, the performance of the PSTH based classifiers on the passive cases begins to drop after the initial peak. While the SVM based classifier is unaffected by the additional, non-informative features being added as the window is expanded, the performance of the PSTH based classifier is reduced as the dimensionality and amount of uncorrelated noise increases. The Euclidean distance metric used by the PSTH classifier is less able to handle the large number of features. As stated by Beyer et al. (1999), as the number of dimensions increases, the magnitude of the distance between points in that space also increases and assuming that the points are uniformly distributed, the ratio between the nearest and farthest neighbor decreases. However, Beyer et al. (1999) show that although the distance to the nearest neighbor remains small, if the query point is close enough to its nearest neighbor in comparison to other points, the distance metric can remain valid as dimensionality increases. This explains why Euclidean distance based classifiers like the PSTH based classifier can be effective in certain cases. In the case of the PSTH classifier, as the dimensionality increases, test points that are

further away from its class prototype can be similar distances from several prototypes. This similarity thus allows signal noise more of a chance to influence the classification decision.

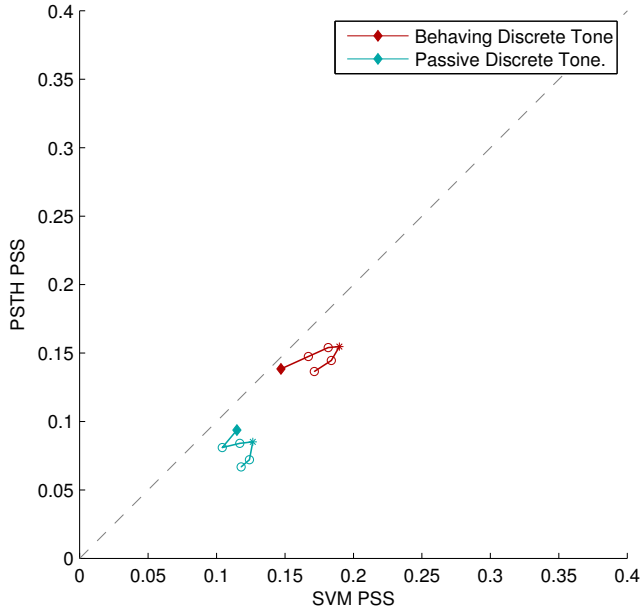
The behaving stimulus context does not show this drop, with the discrete tone case, even showing an improvement in performance beyond 100 ms, as the SVM classifier does with these window sizes. However, while the performance of the PSTH based classifier is harmed by many dimensions with un-correlated noise, the movement correlates in the latter part of the window can provide useful information to the classifier. While additional dimensions that do not contain additional neural information introduce noise that overpowers the distance between classes, adding dimensions with correlated noise increases the contrast in the distances between the classes.

One point to note about Figure 4.2 is that for the discrete tone cases the classifier performs best on the behaving case; $p < .001$ for a paired t-test with the SVM classifier using 50 ms windows. For the broadband noise stimuli this relationship is reversed with the classifier performing better on the passive context stimuli (paired t-test; $p < .001$, SVM classifier, 50 ms window). The result for the broadband noise stimuli was surprising and I will discuss it in more detail later.

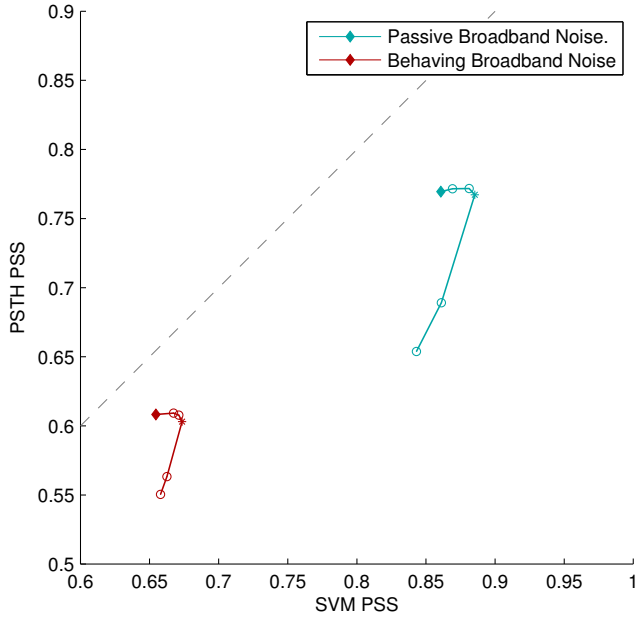
4.1.2 Effect of Bin Size

Another model choice is the spike bin size. Bin size should be chosen so as to provide an accurate estimate of the instantaneous spike rate during the time window within that bin. Larger bin sizes provide a more accurate estimate in that they are less susceptible to noise or small shifts in the timing of spikes, however, information about the temporal variation of the instantaneous spike rate is lost. Smaller bins, on the other hand, provide higher temporal resolution of the spike rate, at the expense of having less information for a given bin with which to make the estimate.

Figure 4.3 shows an overall analysis of bin sizes for the tone and broadband noise cases, testing bin sizes of 1 ms, 2 ms, 5 ms, 10 ms, 25 ms and 50 ms. The PSTH performance is along the vertical axis and the SVM performance along the horizontal axis. A decrease in bin size increases the total number of features. The smallest bin size, denoted by diamonds in the figure, yields a



(a) Discrete Tone



(b) Broadband Noise

Figure 4.3: Classifier sensitivity to bin size, for discrete tone stimuli (a) and broadband noise stimuli (b). The first bin size (1 ms) is denoted by a diamond, with the rest of the sequence connected in order, with the '*' representing 10 ms bins. The bin sizes are 1 ms, 2 ms, 5 ms, 10 ms, 25 ms and 50 ms (with a 50 ms total window).

total of 50 bins per channel. On the other end, the largest bin size yields only one bin per channel, eliminating all temporal information in the window. For the discrete tone cases, the SVM or PSTH classifiers show little variation in performance with bin size. For the broadband noise stimulus cases, however, the PSTH classifier shows a larger difference between the different bin sizes, with 1 ms, 2 ms, 5 ms and 10 ms bins all performing well, but with 25 ms and 50 ms bin sizes showing a large drop in performance. Overall, the 10 ms bin size seems to provide slightly better overall performance compared to smaller or larger bins. Additionally, the performance of the PSTH based classifier seems to depend more on bin size in the case of the broadband noise stimuli, compared to the SVM classifier or the PSTH based classifier in the case of the discrete tone stimuli.

4.1.3 Behaving vs Passive Contexts

The relevance of the stimuli to the task context is expected to affect their encoding in the auditory cortex, and this should be reflected by the performance of classifiers being used to classify the neural responses to stimuli in each context. It has been shown that behavioral relevance affects that encoding on stimuli in A1 (Hubel et al., 1959; Fritz et al., 2007; Sloan, 2009), and it is expected that the information encoded in these responses is also different. Figure 4.2 shows an average performance difference between passive and behaving classifiers. However, it is not clear if this result due to a consistent difference between these classifiers or to variation in a small number of cases. We examine this question through the performance of classifiers constructed on a session-by-session basis. To ensure that biases were not introduced by the presence of additional training data, the training sets were constrained to contain the number of samples for each session's passive/behaving pair. This was necessary because the number of trials, and therefore stimuli, varied between the behaving and passive sessions.

Figure 4.4 shows classifier performance for the passive (vertical axis) and behaving (horizontal axis) responses.For the discrete tone experiments, the classifier performs better for the behaving context stimuli than for the passive, as the points representing most passive-behaving session pairs fall above the diagonal in Figure 4.4. The relationship between the classifiers for the full sets of passive and behaving noise data, like that shown in Figure 4.4, follows this

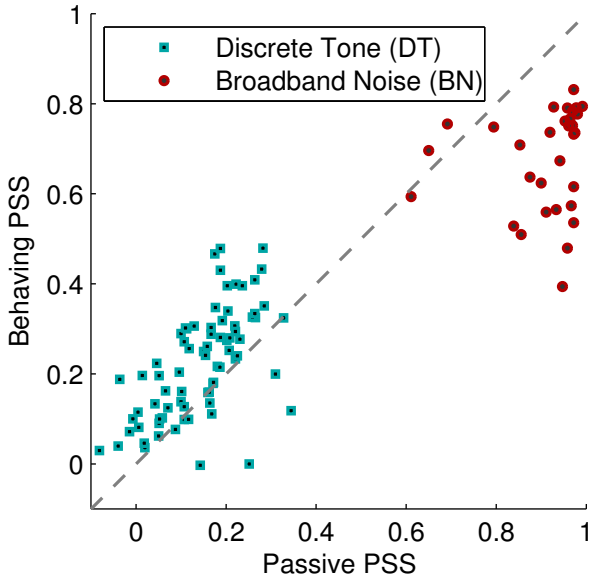


Figure 4.4: Passive versus active performance for both discrete tone (squares) and broadband noise (circles) for sessions with the same rat on the same day. The vertical axis represents the classifier's performance on the behaving context stimuli from that day, and the horizontal axis the passive context stimuli. The dashed diagonal line represents equivalent performance between the behaving and passive cases.

same pattern, with mean behaving classifier performance being overall significantly better than mean passive classifier performance (paired t-test, $p < .001$). Performance for the behaving and passive contexts are correlated ($R = .66$, $p < .001$).

The opposite result is true for the broadband noise stimuli. Passive classifier performance generally outperforms behaving classifier performance, as can be seen by the majority of the points falling below the diagonal in the figure. The relationship between the performance on a matching pair of passive and behaving sessions is less clear, however, as the Pearson's correlation coefficient, $R = .13$ (significance $p = 0.471$). However, the lack of correlation can be explained by a ceiling effect due to the high overall performance and may, in part, also be due to the smaller number of session pairs for the broadband noise experiments.

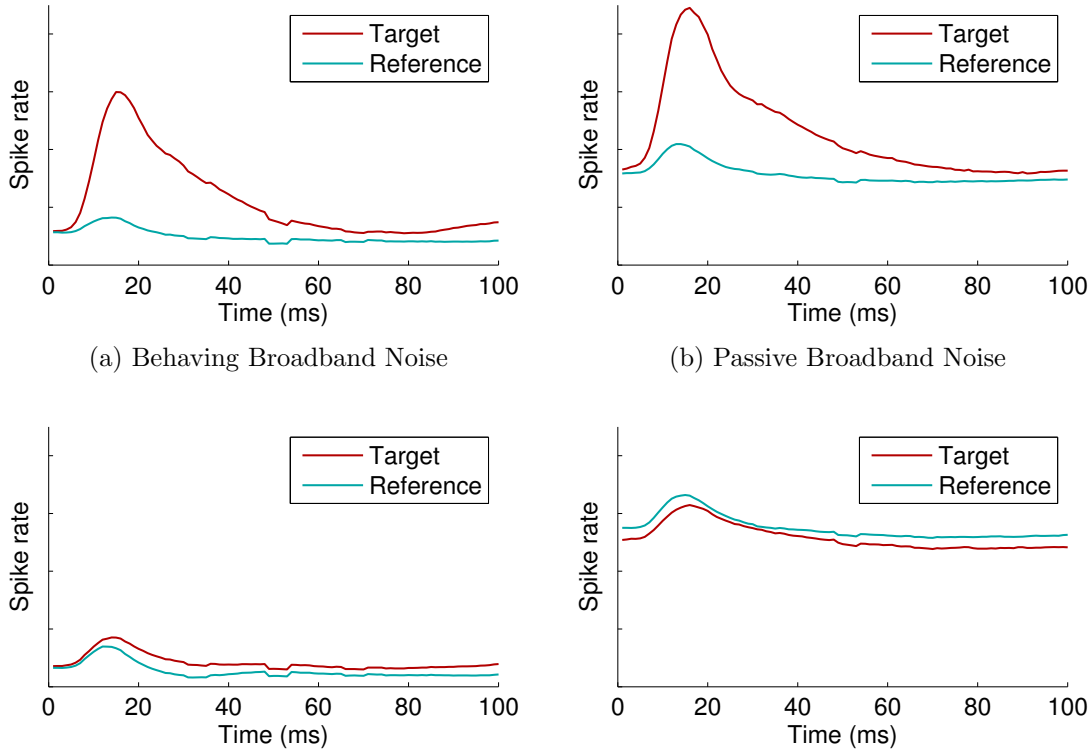
Examining the average responses to the various stimuli, both within and outside of the task context, can help give insight into the the reasons for the

differences in classifier performance. Figure 4.5 shows the average spike rate over all channels and all sessions of each context and stimulus type. The average responses to the broadband noise stimuli are shown in Figures 4.5a and 4.5b for the behaving and passive contexts, respectively. In both cases, the average response to the broadband noise target stimuli is much larger than the average response to the single frequency reference stimuli. This accounts for the high performance in broadband noise cases. Note that in the passive case 4.5b, the background spike rate is substantially higher than the behaving case 4.5b. Nevertheless, the average increase in spike rate of the target over the reference is similar across these two cases.

In contrast, the discrete tone cases (4.5c and 4.5d) exhibit very similar average modulation for the target and reference stimuli. This is likely in part because in this case, both the target and the reference stimuli were single frequency, presumably activating fewer neurons than multi-frequency stimuli due to A1 neurons being tuned to specific frequencies. This greater similarity as compared to the responses to the broadband noise stimuli likely explains the general reduction of performance relative to the broadband noise cases. In addition, the active case represents a larger relative change in modulation as compared to the passive case. This likely explains the fact that the active discrete tone performance is higher than the passive case.

4.1.3.1 Effect of Frequency Shift

In general, the rats perform better at the broadband noise task than at the discrete tone task. The degree of difference between the reference and target stimuli appears to be a factor, since for the discrete tone task, the rats performed better with the larger frequency shifts ($> 10\%$). A critical question is whether a correlate of this performance change with an increase in frequency shift is also observable in auditory cortex. Figure 4.6 shows the aggregated PSS as a function of percent frequency shift for four cases: rat behavior, passive, behaving and *perceived class*. Perceived class is defined as the class of the stimulus that *should* have invoked the observed response from the animal, regardless of what the actual stimulus was. The performance shown is the aggregated performance over the individual test set examples of each frequency shift magnitude. With zero frequency shift, the mean performance of the rat and of the classifiers that



(a) Behaving Broadband Noise

(b) Passive Broadband Noise

(c) Behaving Discrete Tone

(d) Passive Discrete Tone

Figure 4.5: Average spike rate for behaving and passive stimuli for both the frequency shift and broadband noise tasks, smoothed over a 5 ms window.

distinguish the true class of the stimulus is at chance, indicated by a PSS of zero. The classifier that predicts the perceived class shows a positive PSS at 0% frequency shift, as the classifier is successfully identifying stimuli that the rat responded to incorrectly, as if there were a frequency shift.

In addition, the classifier that predicts perceived class has an mean performance of 0.25 over all frequency shifts, higher than the actual true class performance of 0.18. This difference in performance is significant (paired t-test; $p < .001$). With the increasing frequency shifts, the rats' task performance increases, along with the performance of the behaving true class classifiers. A two-way analysis of variance (ANOVA) confirms that the effect of frequency shift ($p < .001$) and class (true and perceived) ($p < .001$) for the behaving discrete tone stimuli are significant. There is also a significant interaction effect ($p < .003$). The differences between the true and perceived class performance are significantly different for frequency shifts of 1% through 6%, and 8%

($p < .006$ for all using a Bonferroni-corrected t-test) (Jensen and Cohen, 2000). While the classifier performance predicting the perceived class of the stimulus shows a moderate increase in performance as frequency shift increases, at higher frequency shifts, the classifier's performance at predicting the perceived class no longer shows a significant difference with of that the true class. This is not surprising, since the rats' performance is also higher at these higher frequency shifts, making the true and perceived class labels equivalent a higher percentage of the time. The fact that the perceived class classifier outperforms the true class classifier when the rat is misidentifying higher numbers of stimuli implies that at least for simple stimuli, when the animal misidentifies a stimulus it is encoded in A1 more similarly to the stimulus that it was misidentifies as. This suggests that when an incorrect behavioral decision is made, the information needed to make the correct decision may not have been encoded in A1.

In addition to the performance at individual frequency shift magnitudes, Figure 4.6 also shows the animals' behavioral performance, and the performance of the classifiers, for the broadband noise stimuli. The broadband noise stimuli can be considered as representing an even larger difference from the reference tones than the difference between a reference tone and even the largest frequency shift magnitude. As described previously, the classifier performs better on the broadband noise stimuli than on the discrete tone stimuli.

4.1.4 Effects of Channel Count

Since each spike rate channel contains information from a unique multi-unit cluster, an increase in the number of driven channels should increase the amount of information available to the classifier. To examine the influence of the number of constituent MUCs on classifier performance, sessions with greater than ten driven MUCs were used for a subsampling analysis. For each number of channels between one and ten the MUCs were sub-sampled to create 20 random subsets of channels. A classifier was then constructed for each subset. Classifier performance for the tone and broadband noise experiments, as a function of the number of available channels, is shown in Figure 4.7. Classifier performance for the broadband noise stimuli increases quickly from one to five MUCs, but then appears to begin approaching an asymptote. This suggests that near maximum performance can be reached with relatively few channels for these stimuli. This

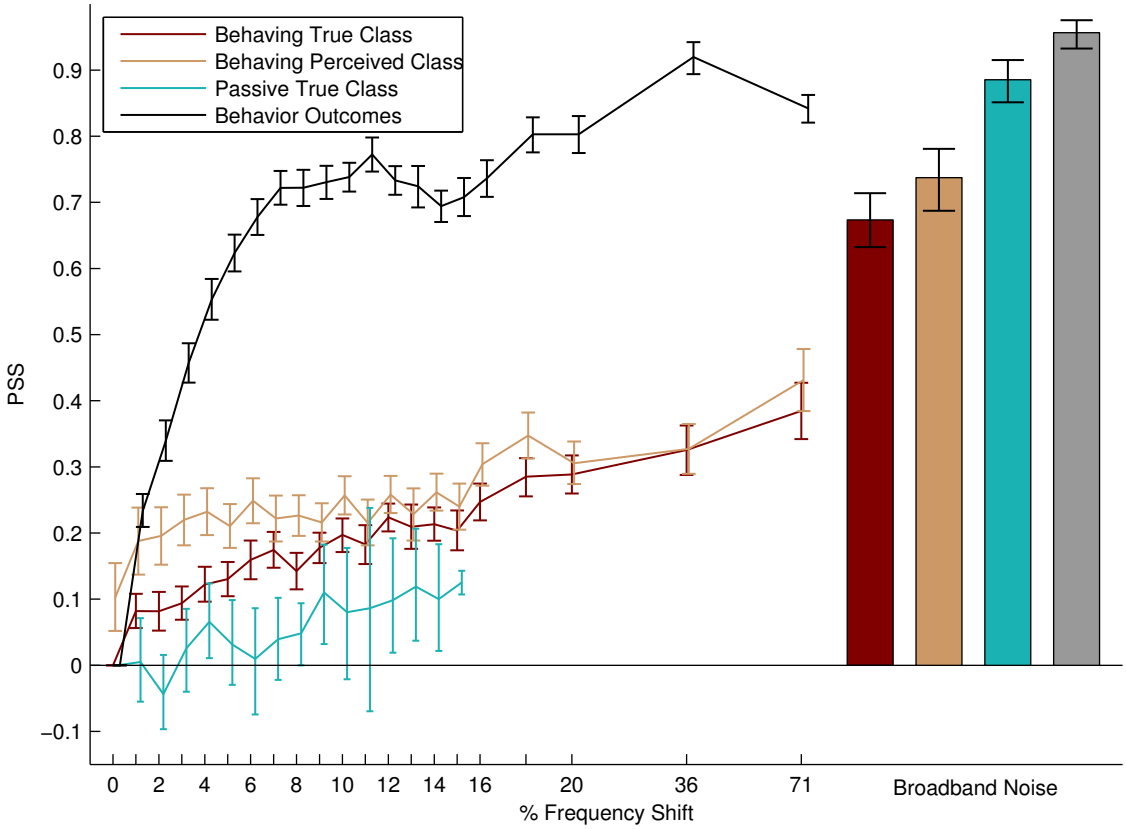


Figure 4.6: Performance for discrete tone stimuli at varying frequency shift magnitudes is shown by the curves on the left, and performance for broadband noise stimuli is shown for by the bars on the right. The rats' performance at the task is represented in black. For behaving task sessions the true class SVM classifier is in red and the perceived class version of the classifier is shown in tan. The classifier performance for the responses to passive context stimuli is in cyan.

implies that the difference between the single frequency reference tones and the multi-frequency noise signal is encoded by the activity of relatively few neurons.

Classifier performance with discrete tone targets increases at a greater rate as more than a handful of channels are added, compared to the broadband noise cases. This implies that, for the more similar stimuli from the frequency shift task, the additional MUCs contain additional information for the classifier. This trend also indicates that if even more channels were available, then performance would continue to increase.

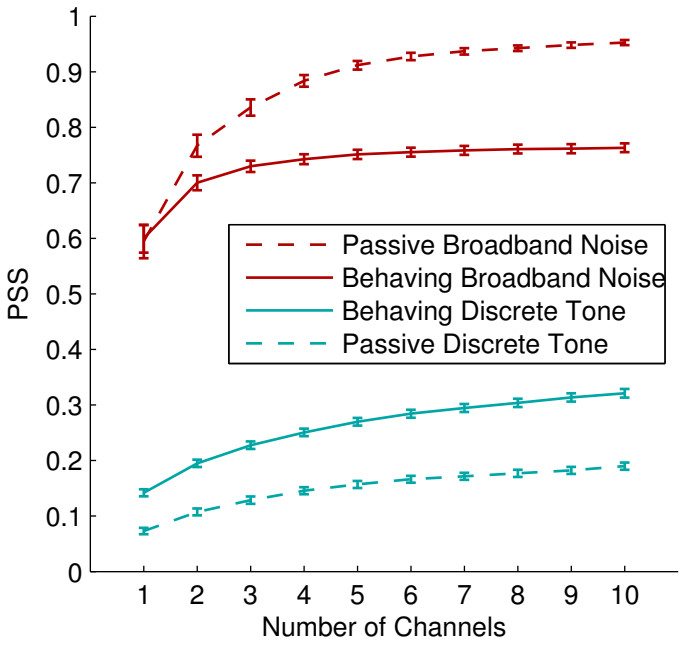


Figure 4.7: PSS as a function of the number of driven channels. Each point represents the mean performance over 20 randomly sampled sub-sets of channels. The 95% confidence intervals are estimated using bootstrap sampling.

4.2 Speech Stimulus Recognition

While the spike rate based approach to generating features from neural signals provides high resolution information about a few neurons, the LFP signal also contains much lower frequency fluctuations due to the activity of a larger group of nearby neurons. The lower frequency, continuous signal can also be used for classification, and while it does not contain the detailed resolution of the higher frequency spikes, it includes the activity of many more neurons. The questions are whether this signal can prove as informative as the spikes themselves, and if so, which frequency bands provide the most information?

The discrete tone and broadband noise stimuli are both simple in temporal structure, in that the stimulus is either temporally unchanging (discrete tone), or of random structure (broadband noise). More complex stimuli, such as natural speech sounds, differ from the tone and broadband noise stimuli in that they have both frequency and temporal complexity. In order to examine the encoding of these more complex stimuli, classifier models were built to predict

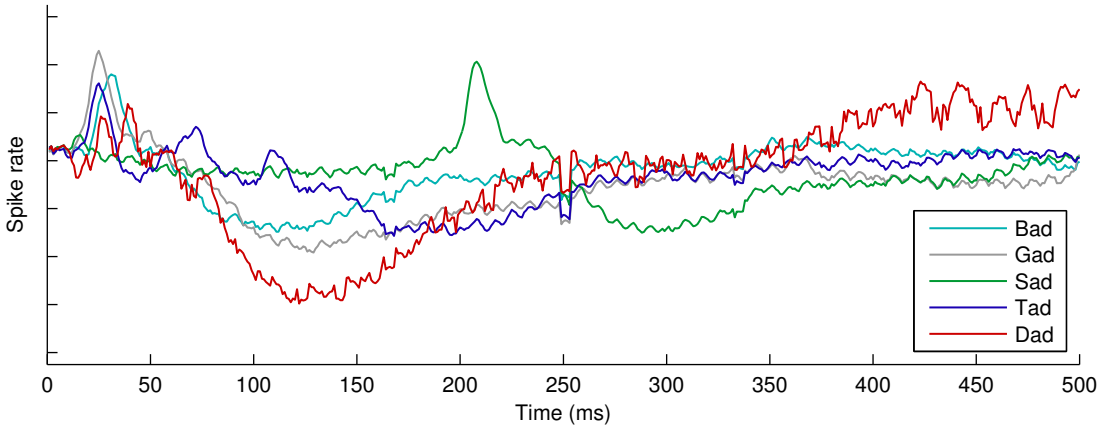


Figure 4.8: 5ms running average spike rate for speech stimuli.

the target/reference label for the speech stimuli. The feature selection for the spike rate classifiers is similar to those for the simpler tone and broadband noise stimuli, however, there are a few differences due to the more complex stimuli. Specifically, since the reaction times to the speech stimuli were slower, and the corresponding neural responses more complex than those of the simpler stimuli, the windows are expanded to include the additional information. Too long of a window, however, still presents concerns about movement correlates. This can be seen in Figure 4.8, which shows the average spiking responses over all animals, sessions and cells. One can see both the similarities and differences between the average response to each target word, with the rise the spike count in response to the target word after the initial 125 ms being of particular note. While it cannot be determined for certain that the increased spike rate for the target word at the end of the window (≈400 ms) is due to movement correlates, the similarity of the ending of all five stimulus words, paired with the dissimilarity of the latter portion of the average response to the target stimulus and those to the reference stimuli makes it a credible concern. Because of this, the window size is set at 128 ms, with a power of two being chosen in order to facilitate the use of discrete Fourier transforms. Additionally, to increase the resolution of the spike rate information, the bin size was reduced to four milliseconds. This bin size also evenly divides the chosen window size.

As shown in Figure 4.9, the classifier achieved a PSS of .30 using 4 ms bins with this 128 ms window. This is significantly better the the classifier's

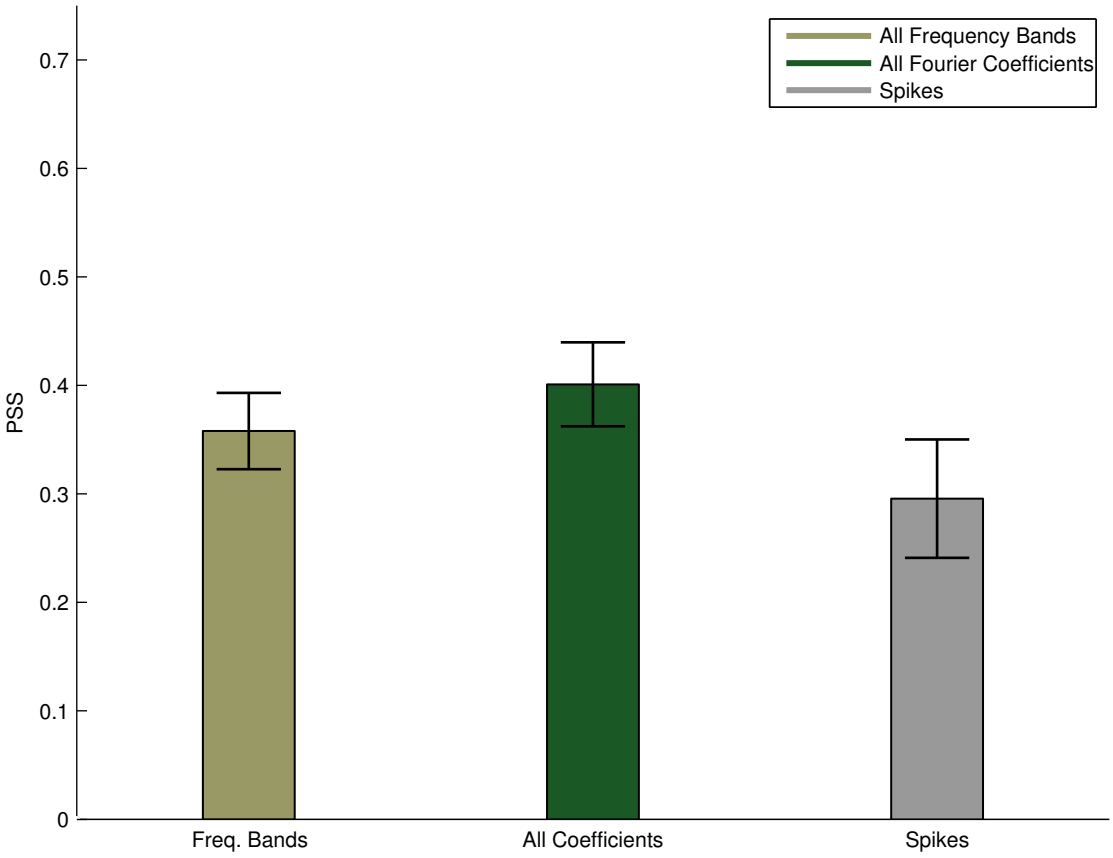


Figure 4.9: Performance for LFP Fourier space SVM classifier for all frequency bands, all available Fourier coefficients, and the spike rate bin based SVM classifier.

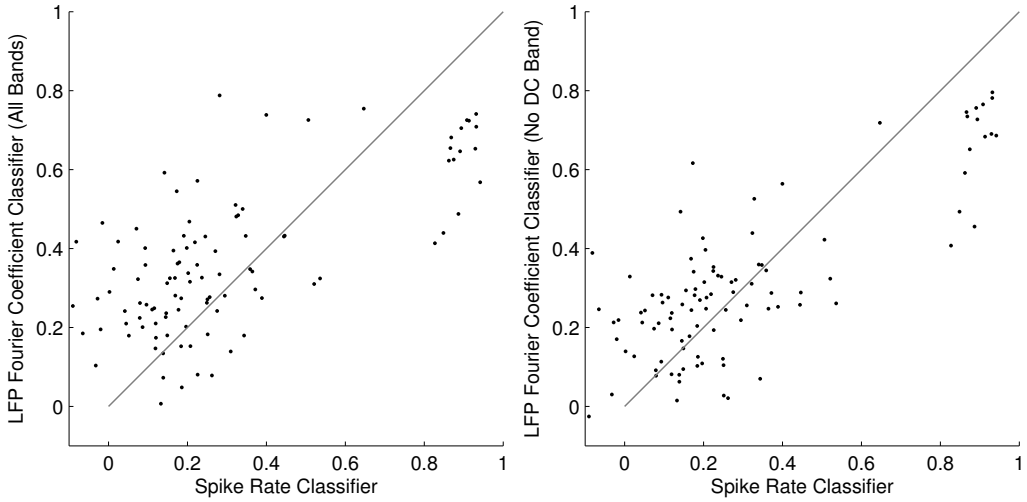
performance on the discrete tone frequency shift responses, but the classifier's performance on the speech stimuli do not outperform performance in the broadband noise case. Variations in the nature of the tasks aside, this performance relationship is not surprising. The average spike rate in response to the speech stimuli shown in Figure 4.8 is more complex compared to the small, but simple response to the tone shift stimuli shown in Figures 4.5c and 4.5d. The much higher peak spike rate in response to the broadband noise stimuli, as compared to the responses to the tone stimuli provides much better contrast, however, and higher classifier performance on the broadband noise stimuli is expected. The question of a perceived class is not addressed for these word recognition experiments, since the rats did very well at identifying the target word, normally misidentifying the target word only a handful of times.

For the speech experiments, a classifier was trained using the LFP signals from the same 128 ms window as the spike rate classifier, shown in Figure 4.9. The LFP band power classifier ($PSS = .37$) outperforms the spike rate based classifier ($p < .02$). All available Fourier coefficients can also be used individually instead of averaging the coefficients from groups of adjacent frequencies into band power features. With this more detailed frequency information, the PSS increases to .41 and significantly outperforms the frequency band-based classifier ($p < .001$). This implies that the additional frequency space resolution available when the coefficients are used individually provides additional information about the stimulus.

Figure 4.10 shows scatter plots of the performance of spike rate and LFP classifiers, with the LFP classifier either including or not including the DC offset. Each point represents a set of stimuli presented to the same animal on the same day. The higher LFP performance in Figure 4.10a shows the performance in individual data sessions for a classifier using all available bands, including the DC offset, and the performance without the DC offset (0Hz) component is shown in Figure 4.10b. The classifier performance for the spike rate and LFP features is correlated ($R = 0.67$, $p < .001$ for the LFP classifier using all bands, $R = 0.76$, $p < .001$ for the LFP classifier using only the non-zero band LFP features $p < .001$). The LFP classifier without the DC offset feature has an overall PSS of 0.31, a significant drop from the classifier using all bands ($p < .001$). This indicates that the DC offset contains information not included in the other bands.

In order to determine which of the other frequency bands provide useful information to the classifier, classifiers were constructed using features from subsets the frequency bands, shown in Figure 4.11. For classifiers trained on only the DC offset of each channel, Figure 4.11 shows a PSS of .26. While this is lower than the performance with all frequency bands, this is still significantly above random (paired t-test; $p < .001$).

The performance using each of the additional frequency band subsets, shown in Figure 4.11, was compared to the performance of an LFP classifier using the features from all the bands. If performance decreases significantly when a frequency band is removed, then that frequency band must have been contributing useful information to the classifier.



(a) All Bands (b) No DC Offset

Figure 4.10: Performance for LFP Fourier space SVM classifier vs spike rate bin based classifier, with and without the LFP's DC offset component.

Without the Theta band (4-8Hz, column 5 in Figure 4.11), $PSS = .32$, the decrease also being significant ($p < .001$). This indicates that the DC offset and Theta band contains useful information not contained in the other bands. Removing a single higher band (Beta, Gamma or frequencies > 100Hz, columns 6-8 in Figure 4.11) does not produce a significant drop in performance ($p < .85$). However, a classifier using only the DC offset and Theta bands (column 3 in Figure 4.11) only achieves a PSS of .34, which is significantly lower than the performance using all bands ($p < .001$).

The 128 ms windows, used for both the classifiers using the LFP signals and the spike rate information from the same experiments, were chosen as a balance between providing a large enough window to include sufficient information to the classifier, and ending the window before correlated movement noise is seen. Figure 4.12 shows a histogram of the reaction times to the target word, which allows a window to be selected that avoids including the resulting movement noise in the feature vector. Figure 4.12a shows the distribution of reaction times over all animals and sessions. This distribution varies greatly across animals. As an example of this, Figures 4.12b and 4.12c, show the distributions of reaction times for the slowest and fastest animals respectively.

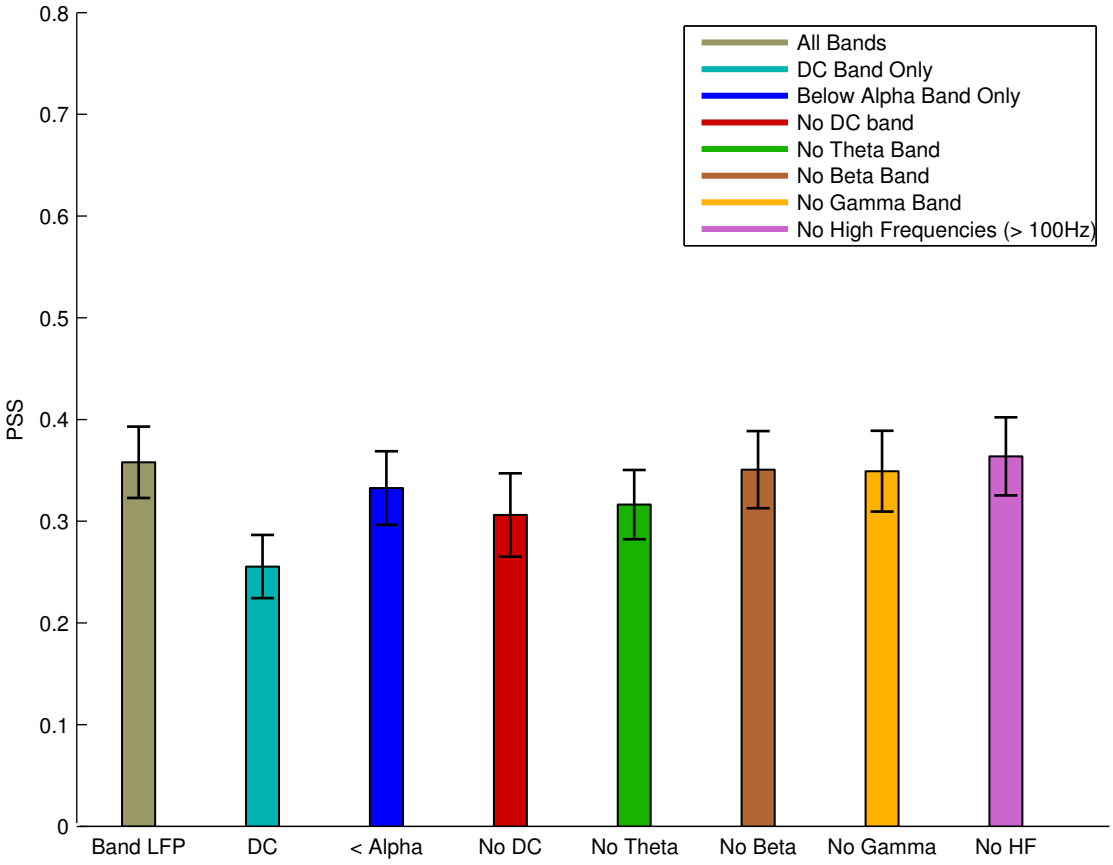
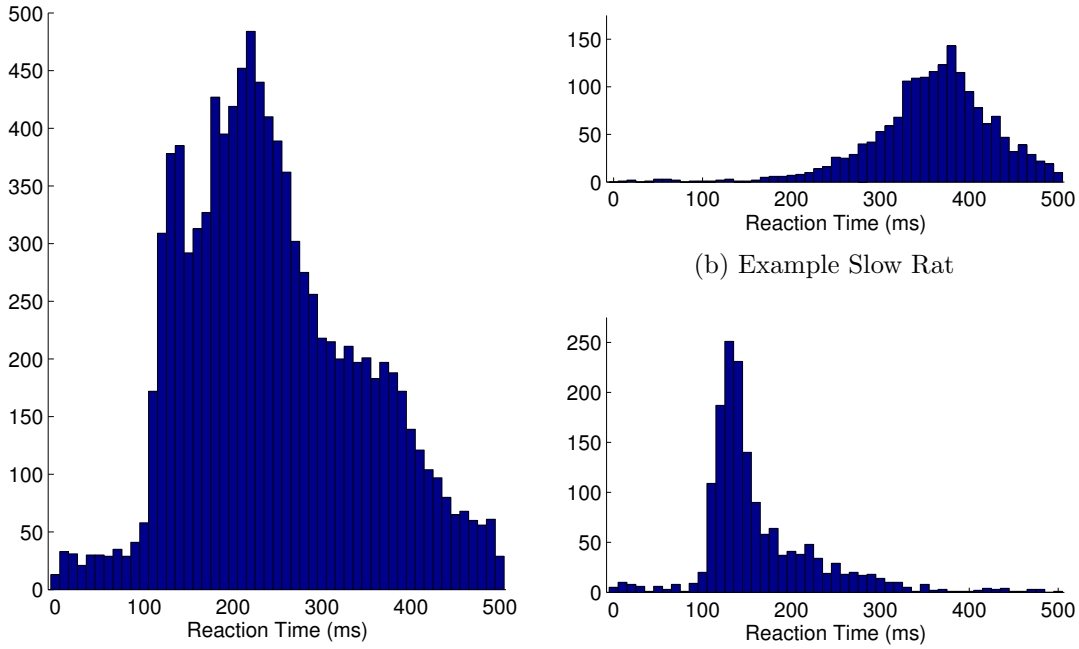


Figure 4.11: Performance for LFP Fourier space SVM classifier for all frequency bands, DC band only, DC, Delta and Theta bands only, and each combination of all bands but one. This performance is also compared to the spike rate bin based SVM classifier.

In addition to managing concerns about movement correlates, the choice of window size is especially difficult when using frequency space features due to the formulation of the discrete Fourier transform. Another significant concern for selecting a window size for the LFP classifier is the fact that not only do shorter windows of the LFP signal contain information about a shorter temporal period, but due to the smaller number of discrete points, the resolution of the frequency space representation is reduced.

Figure 4.13 shows the performance of both the spike rate and LFP based SVM classifiers, at varying window sizes. Results using the LFP based classifier are shown for window sizes of 64 ms, 128 ms, 256 ms and 512 ms. The spike rate classifier is also tested with a window size of 192 ms. According to a 2-way



(a) All Subjects

(b) Example Slow Rat

(c) Example Fast Rat

Figure 4.12: Histogram of reaction times to speech stimuli.

ANOVA, the LFP based classifiers outperform the spike rate based classifiers ($p < .001$). The effect of window size is also significant, with $p < .001$. There is also a significant interaction effect between the two factors, with $p < .04$, indicating that the performance of the LFP based classifier may be growing at a greater rate than that of the spike rate classifier.

Multiple factors may be contributing to the increase in performance as the window size grows. First, for both classifiers, as the window size grows, the amount of neural information available to the classifier also increases. All target word examples where the animal displays a reaction time shorter than 180 ms have been removed from the data set, in order for movement correlates to not be a concern for the smaller windows (64 and 128 ms). However, as the window grows beyond this point, some of information available to the classifier may not in fact be stimulus-related neural data. Finally, for the LFP based classifier, the number of discrete frequencies for which Fourier coefficients are available, and therefore, the resolution in frequency space, increases as window size increases.

To help determine how much these performance differences are due to legitimate advantages of the larger window (more information and higher LFP

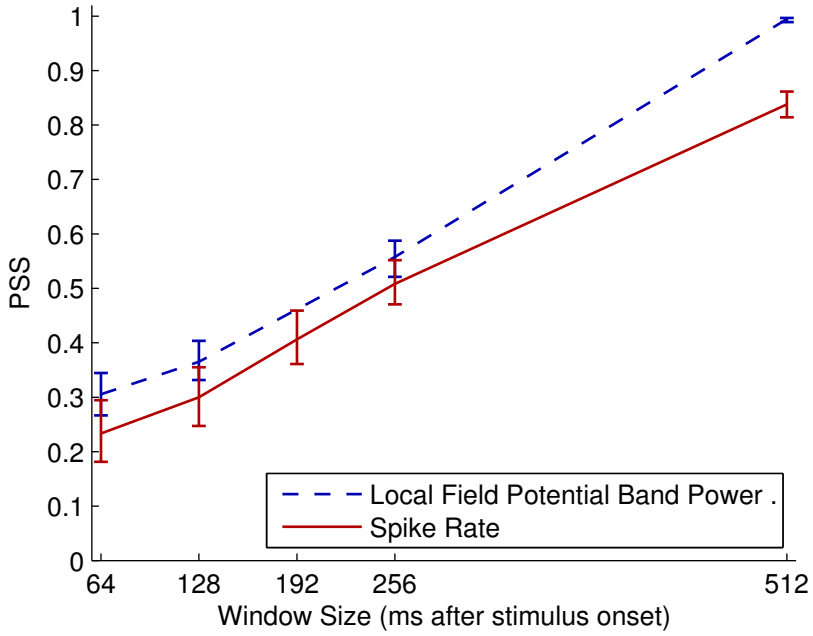


Figure 4.13: Performance for LFP Fourier space SVM classifier classifiers for varying window sizes. Window sizes greater than 128 ms are likely to contain correlated noise due to movement responses to the target stimuli, and thus the classifier performance shown is likely not solely due to decoded neural activity.

frequency resolution) versus how much is due to movement correlates, an additional set of classifiers was constructed to predict only the different reference stimuli. Classifier performance is expected to be lower when the classifier has to discriminate between all four reference words. However, if the performance patterns seen for the reference or target identification classifier are not simply due to movement correlates providing extra information to the classifier, then the same patterns should hold for the prediction of specific reference stimuli, where no correlated movement artifacts should exist.

Figure 4.14 shows the performance of LFP classifiers using all coefficients, LFP classifiers using band power features, and spike rate based classifiers predicting the individual reference words. a two-way ANOVA shows a significant effect for both the classifier type and the window size ($p < .001$ for both), along with a significant interaction effect ($p < .002$). The performance of all three classifiers increases between the 128 ms and 256 ms windows, with $p < .001$ for all three, using Bonferroni correction. The performance of the two LFP based

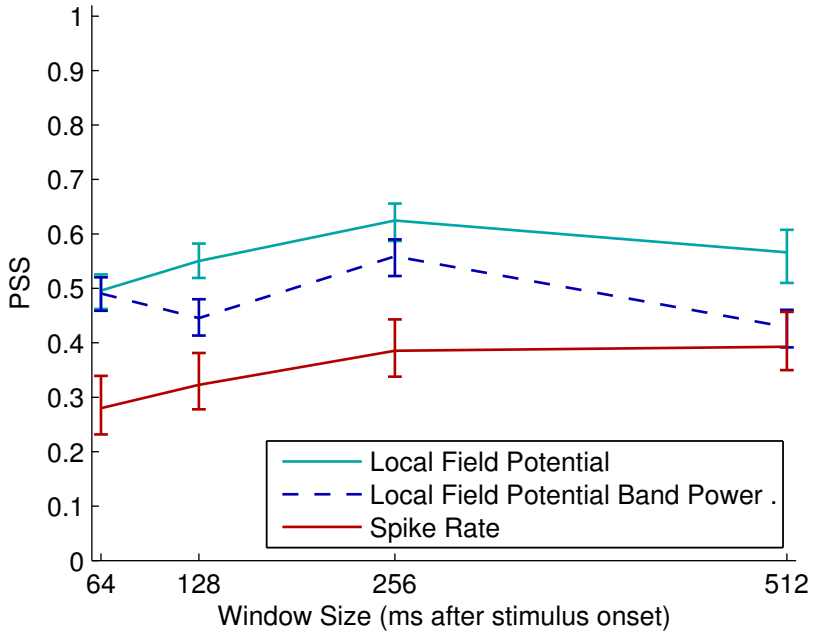


Figure 4.14: Performance for LFP Fourier space and spike rate SVM classifiers for varying window sizes. The solid gray line represents the performance of the classifiers using all of the available Fourier coefficients, while the broken gray line has the coefficients averaged by frequency band. The black dashed line shows the performance of the spike rate based classifiers.

classifiers does not show a significant difference between them for the 64 ms window ($p < .38$), and both LFP classifiers show a significant drop in performance between the 256 ms and 512 ms windows ($p < .001$ for both). There is not a statistically significant difference between the 256 ms and 512 ms windows for the spike rate based classifier, however.

It is unclear why the LFP classifier with the coefficients averaged into band power components shows a drop in performance between the 64 ms and 128 ms windows ($p < .001$, using Bonferroni correction). Looking at the average PSTHs for each word, there may not be much new information between 64 ms and 128 ms, this it is not clear whether this is actually the case. Since the additional frequency space resolution afforded by the larger window is probably limited for the averaged coefficients of the band power classifier, if the larger window contained no new information it would provide no advantages. As such,

increasing the frequency space resolution by increasing the sampling rate might improve classifier performance without changing the size of the analysis window.

Chapter 5

Conclusions

A1 encodes information about aural inputs such as sound location, magnitude and frequency, and is organized tonotopically with individual neurons responding preferentially to particular frequency ranges (Kandel et al., 1991). One major question is whether the actual information encoded about a stimulus by the activity in A1 varies according to the behavioral context of the stimulus. I use machine learning classifiers to examine the information content of the neural responses to the stimuli in order to answer this question about the differences in the information encoded in A1 depending on behavioral context. Specifically, I compare the performance of classifiers predicting information about neural responses to stimuli in different behavioral contexts.

For the case of classifiers identifying whether a single frequency stimulus is of a different frequency than the preceding stimulus performs better when the stimuli were presented within a task context as opposed to outside such a context. While the classifiers can identify the change in frequency for a stimulus presented to the same animals outside such a paradigm, the classifiers do not perform as well as with stimuli from within a behavioral context. This difference in performance is not surprising, since when engaged in a task, there is a greater "focus of attention", which manifests itself in terms of additional inhibition that reduces spontaneous activity in A1 (Sloan, 2009). In turn, this reduces the effects of noise and forces differences in the encoding of different stimuli.

In was unexpected, however, in the case of the broadband noise stimuli, to see better classifier performance for the passive context stimuli. While the cause of this is unclear, there are several reasons why this might be the case. For example, is possible that the rats were startled by the broadband noise stimulus

outside of the task context, thus engaging the animal's attention. Additional stimulus presentation paradigms, such as presenting the broadband noise stimulus after a period of silence, in addition to after the reference tone, might provide more insight into the encoding of the broadband noise stimulus.

For same smaller frequency shift magnitudes ($\leq 10\%$), the classifier is able to predict the rats' response to the stimulus better than the stimulus itself. Since the rats fail to detect many of the smaller frequency shifts, the higher classifier performance for the perceived class indicates that the target stimuli that were not recognized by the rat were encoded more similarly to reference stimuli. This is consistent with the idea that projections from motor decision or execution areas influence the activity in these more sensory areas.

The speech stimuli differ from the pure tone or multi-frequency noise signals in that the speech sounds are more complex. Specifically, the speech stimuli are multi-frequency, like the broadband noise signal, stimulus. Unlike the broadband noise signal, however, the speech stimuli also have temporal and frequency structure. Nevertheless, a direct comparison between the word recognition and tone shift experiments is not possible due to the differences in the tasks. Specifically, for the frequency change detection experiments, it did not matter what the target stimulus was, just that it was different than the preceding reference stimuli. In the speech task, however, the rats identified a stimulus that matched a specific, overlearned auditory pattern.

The LFP signals represent one end of the spatial coverage/temporal resolution spectrum. On the other side of this trade off is the high resolution spike rate information. The spike rate, however, can only be observed for a much smaller number of neurons per electrode. My results show that the performance of the LFP classifiers, with this broader information base, is significantly higher than the spike rate classifier performance for the same stimuli. This indicates that, at least for the number of available channels, with stimuli of this complexity, more information is present in the overall activity of a larger population of neurons in the rat A1 than is present in the spike rate information of just a few neurons.

This result is not surprising, as the relative amount of information encoded in spike rate metrics and LFP signals available from a fixed number of channels from the monkey motor cortex has been shown to vary depending on the number of channels being sampled (Bansal et al., 2011; Mehring et al., 2003).

Specifically, Bansal et al. (2011) show that the LFP signals contain more information on average than spike rate information from a few neurons. They found, however, that larger populations of neurons, or single neurons whose activity is unusually well correlated to the stimulus, in fact, contain more information than the LFP signals for the corresponding channels.

By systematically dropping individual frequency bands, I have shown that lower frequency components of the LFP signal seem to contain the most information about these auditory stimuli. However, while it can be shown that the lower frequency bands contain significant amounts of information, some information is still contained in the higher frequencies.

The increase in classifier performance with larger signal windows indicates that additional information about the stimulus is present in the neural signal after the initial response to the stimulus. This is not surprising, given the more complex average responses to the speech stimuli, as compared to the less complex responses to the lower complexity tone and noise signals. However, it is unlikely that all of the performance increase for the larger window sizes (>128 ms) is due additional stimulus-related information being present. All windows larger than 128ms contain movement correlates in the signal as well and therefore these larger windows contain non-stimulus-related information. The performance of the LFP based classifiers increases with increasing window size at a greater rate than spike rate based classifiers for the same stimuli, possibly indicating that this additional information afforded by the larger window size is better utilized by the LFP based classifier. However, since the windows larger than 128 ms contain movement correlates, whether the classifier is really better utilizing the additional neural information in the larger window, or just the movement correlates. Since the movement correlates are only a concern for the target words, classifiers that differentiate between the four actual stimulus words used as reference words should only have neural information available to them.The fact that the classifiers predicting only the reference words do not show increased performance as the window size grows beyond 256 ms indicates that it may simply be the movement correlates in these larger windows that account for the improvement when identifying the target stimulus.

5.1 Future Work

I have shown that the SVM classifiers outperform the PSTH based classifiers when both classifiers use the same window and bin sizes. However, I have not explored how the window size parameter affects the PSTH based classifier if the bin size is also varied in order to keep the length of the feature vector constant. This might yield higher PSTH based classifier performance for the larger window sizes, since the PSTH based classifier is sensitive to the size of the vector used to represent the neural response. The additional time bins resulting from a longer sample window might actually have a deleterious effect on performance, since they increase the dimensionality of the feature vector. Holding the number of bins constant will still not completely isolate the effect of the additional information available in a longer sample window, due to the decreased temporal resolution allowed by larger bins. Since the effect of the bin size on performance is relatively small, at least for the single frequency stimuli, large effects on performance due to sample window size should still be apparent, however.

Similarly, to isolate the effect of the additional frequency space resolution that a larger sample window gives the discrete Fourier transform, the LFP signals with varying sampling rate could be used. This would allow the comparison of the information content of different frequency space resolution representations of the same time window. This could be done using classifiers that identify the target word on task context data, however, the window size would need to be kept small enough to avoid the introduction of movement correlates. If movement correlates were not a factor, for example, by having the classifier identify individual reference words, then the results would be more indicative of the effect of the frequency resolution on the classification of the pure LFP signal.

Finally, since both the LFP and spike rate metrics provide useful information to the classifier, one can ask whether at least some of the information provided by each is unique. This would mean a classifier using both signals would have even more information available to it. If such a classifier were properly constructed, it is expected that it would outperform either of the single signal type classifiers, as has been done for predicting hand trajectory in monkeys (Mehring et al., 2003).

Bibliography

Adrian, E. D. (1959). *The Mechanism of Nervous Action: Electrical Studies of the Neurone*. The Eldridge Reeves Johnson Foundation Lectures. University of Pennsylvania Press, new edition.

Aggarwal, C., Hinneburg, A., and Keim, D. (2001). On the surprising behavior of distance metrics in high dimensional space. In Van den Bussche, J. and Vianu, V., editors, *Database Theory ? ICDT 2001*, volume 1973 of *Lecture Notes in Computer Science*, pages 420–434. Springer Berlin / Heidelberg.

Bansal, A. K., Vargas-Irwin, C. E., Truccolo, W., and Donoghue, J. P. (2011). Relationships among low-frequency local field potentials, spiking activity, and 3D reach and grasp kinematics in primary motor and ventral premotor cortices. *Journal of Neurophysiology*, 105:1603–1619.

Beyer, K., Goldstein, J., Ramakrishnan, R., and Shaft, U. (1999). When is "nearest neighbor" meaningful? In *International Conference on Database Theory*, pages 217–235.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, COLT '92, pages 144–152, New York, NY, USA. ACM.

Chang, C.-C. and Lin, C.-J. (2001). *LIBSVM: a library for support vector machines*. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Cooley, J., Lewis, P., and Welch, P. (1969). The finite Fourier transform. *IEEE Transactions on Audio and Electroacoustics*, 17(2):77 – 85.

Doswell, C. A., Davies-Jones, R., and Keller, D. L. (1990). On summary measures of skill in rare event forecasting based on contingency tables. *Weather and Forecasting*, 5:576–585.

Engineer, C. T., Perez, C. A., Chen, Y. H., Carraway, R., Reed, A. C., Shetake, J. A., Chang, V. J. K. Q., and Kilgard, M. P. (2008). Cortical activity patterns predict speech discrimination ability. *Nature Neuroscience*, 11(5):603 – 608.

Foffani, G. and Moxon, K. (2004). PSTH-based classification of sensory stimuli using ensembles of single neurons. *Journal of Neuroscience Methods*, 135(1-2):107 – 120.

Fritz, J. B., Elhilali, M., David, S. V., and Shamma, S. A. (2007). Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in A1? *Hearing Research*, 229(1-2):186 – 203.

Funamizu, A., Kanzaki, R., and Takahashi, H. (2009). Different neural activities require different decoders. In *4th International IEEE/EMBS Conference on Neural Engineering, 2009*, pages 287–290. IEEE.

Garcia, G., Ebrahimi, T., and Vesin, J.-M. (2003). Support vector eeg classification in the fourier and time-frequency correlation domains. In *First International IEEE EMBS Conference on Neural Engineering, 2003*, pages 591 – 594.

Hubel, D. H., Henson, C. O., Rupert, A., and Galambos, R. (1959). "Attention" units in the auditory cortex. *Science*, 129(3358):1279–1280.

Jensen, D. D. and Cohen, P. R. (2000). Multiple comparisons in induction algorithms. *Machine Learning*, 38:309–338.

Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (1991). *Princples of Neural Science*. Elsevier, 655 Avenue of the Americas, New York, New York 10010, 3rd edition.

Lindberg, E. W., Miller, L. E., Oby, E. R., and Slutzky, M. W. (2011). Decoding muscle activity with local field potentials. In *Proceedings of the 5th International IEEE EMBS Conference on Neural Engineering, Cancun, Mexico, April 27 - May 1, 2011.*

Liu, J. and Newsome, W. T. (2006). Local field potential in cortical area mt: Stimulus tuning and behavioral correlations. *The Journal of Neuroscience*, 26(30):7779–7790.

Lotte, F., Congedo, M., Lcuyer, A., Lamarche, F., and Arnaldi, B. (2007). A review of classifcation algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 4:R1–R13.

Malmierca, M. S. (2003). The structure and physiology of the rat auditory system: An overview. *International Review of Neurobiology*, 56:147–211.

Mehring, C., Rickert, J., Vaadia, E., de Oliveira, S. C., Aertsen, A., and Rotter, S. (2003). Inference of hand movements from local field potentials in monkey motor cortex. *Nature Neuroscience*, 6:1253–1254.

Olson, B. P., Si, J., Hu, J., and He, J. (2005). Closed-loop cortical control of direction using support vector machines. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(1):72–80.

Peirce, C. (1884). The numerical measure of the success of predictions. *Science*, IV(93):453–454.

Rennaker, R., Ruyle, A., Street, S., and Sloan, A. (2005). An economical multi-channel cortical electrode array for extended periods of recording during behavior. *Journal of Neuroscience Methods*, 142(1):97–105.

Schoenberg, I. J. (1950). The finite fourier series and elementary geometry. *The American Mathematical Monthly*, 57(6):390–404.

Schölkopf, B. and Smola, A. (2002). *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. Adaptive Computation and Machine Learning. The MIT Press.

Sloan, A. (2009). *Spectrotemporal dynamics of neural responses in auditory cortex during frequency discrimination*. Phd dissertation, The University of Oklahoma, United States - Oklahoma.

Sloan, A. (2012). personal communication.

Sloan, A., Dodd, O., and Rennaker, R. (2009). Frequency discrimination in rats measured with tone-step stimuli and discrete pure tones. *Hearing Research*, 251:60–69.

Vapnik, V. N. (1995). *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA.

Vapnik, V. N. (1998). *Statistical Learning Theory*. Adaptive and Learning Systems for Signal Processing, Communications and Control. John Wiley & Sons, Inc.

Victor, J. D. (2005). Spike train metrics. *Current Opinion in Neurobiology*, 15:585–592.

Wilks, D. S. (2006). *Statistical Methods in the Atmospheric Sciences*. Academic Press, 2nd edition.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113:767–791.

Woodcock, F. (1976). The evaluation of yes/no forcasts for scientific and administrative purposes. *Monthly Weather Review*, 104:1209–1214.