USING STATISTICAL CLASSIFICATION ALGORITHMS TO DECODE COVERT SPEECH STATES WITH FUNCTIONAL MAGNETIC RESONANCE IMAGING

by

Devin McCorry
A Thesis
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Master of Arts
Psychology

Committee:

Jim Thompson ________________ Director

____

____

Deborah A Boehm-Davis ________________ Department Chairperson

________________ Dean, College of Humanities and Social Sciences

Date: July 9, 2010 ________________

Summer Semester 2010
George Mason University
Fairfax, VA

Using Statistical Classification Algorithms to Decode Covert Speech States with Functional Magnetic Resonance Imaging

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Arts at George Mason University

By

Devin M. McCorry
Bachelor of Science
The University of Michigan, 2006

Director: James C. Thompson, Professor
Department of Psychology

Summer Semester 2010
George Mason University
Fairfax, VA

ACKNOWLEDGEMENTS

I would like to thank all those who have supported me towards the completion of this thesis. To my advisor, Dr. James Thompson, I am deeply indebted—I could not have done it without your guidance and support. I am also immensely grateful for the advice and encouragement from the other members of my committee, Drs. Patrick McKnight and Jane Flinn. I further wish to express my thanks for comments from Dr. Adam Winsler which were of great help.

Also, I could not have finished this thesis without the assistance and support of my friends and colleagues. In particular, I wish to thank Ashley Safford, Wendy Baccus, Elizabeth Hussey, and Shira Levy.

Lastly, I owe my deepest gratitude to my family for all of their invaluable help and encouragement, especially to my mother and sister Chelsea.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

USING STATISTICAL CLASSIFICATION ALGORITHMS TO DECODE COVERT SPEECH STATES WITH FUNCTIONAL MAGNETIC RESONANCE IMAGING

Devin M. McCorry, M.A.

George Mason University, 2010

Thesis Director: Dr. James C. Thompson

An effective covert speech brain-computer interface (BCI) would be a system that decodes a subject's internal speech-related neural activity, translating it into text or synthesized speech. Multiple clinical populations stand to benefit from such a BCI, most notable of which are patients with locked-in syndrome—a disorder marked by near-complete motor paralysis, despite relatively unimpaired cognition. Determining which regions of the brain contain speech information that could be decoded in a covert speech BCI is an important and necessary step towards actual BCI implementation. In the present study I investigated whether cortical areas involved in *motor speech production* and *speech perception* contain such information. I used functional magnetic resonance imaging (fMRI) to scan subjects (2 males and 3 females, mean age = 23.6) while they completed 6 runs of a speech task, in which they were prompted to speak one of two syllables: /pah/ or /tah/. In one scanning session, subjects spoke these syllables *overtly* (aloud); in another session, they *covertly* "spoke" the syllables in their mind without moving. Results from two independent tasks—a task involving motor articulation, and a task involving perception of speech—were used to select regions of interest (ROIs) for each subject. These ROIs were used to subset the activation observed during the speech tasks. I then used multi-voxel pattern analysis (MVPA) to train statistical algorithms to classify which of the two syllables was spoken (overtly or covertly)—based solely on the subset of fMRI data during the speech tasks. The MVPA was conducted using regressed parameter estimates of the syllables from each of the 6 runs. Classification accuracy was significantly above chance in both the speech production and speech perception ROIs for both overt and covert speech ($p < .05$). Accuracy was significantly higher for overt versus covert speech ($p < .05$), and a nonsignificant trend suggested higher accuracy in the motor ROI than in the perception ROI. These findings are significant in that they indicate that neural patterns of activity during covert and overt speech may be similar enough to apply overt speech models to methods of decoding inner speech. Importantly, speech motor and perception regions may encode sufficient detail about a person's internal speech states to decode in a future implementation of a covert speech BCI. Furthermore, the results of this study demonstrate the utility in using MVPA to map out regions to use in future BCIs based on decoding cognitive states.

1. Introduction

1.1 Overview

The purpose of this paper is to describe a study in which I aimed to accurately classify covert speech-states of subjects using functional magnetic resonance imaging (fMRI).

I used fMRI to scan subjects while they completed speech tasks in which they were prompted to speak one of two syllables. In one scanning session, subjects spoke these syllables aloud; in another session, they covertly "spoke" the syllables in their mind without moving. Results from two other tasks—a task involving motor articulation, and a task involving perception of speech—were used to select regions of interest (ROIs) for each subject. The activation observed during the speech tasks was subset using these ROIs. I then employed multivariate pattern analyses to train statistical algorithms to accurately classify which of the two syllables was spoken (overtly or covertly)—based solely on the subset of fMRI data during the speech tasks (see Figure 1).

Functional MRI was used for two primary reasons: first, it is a well-established and non-invasive neuroimaging technique for measuring neural activity; second, fMRI has a high spatial resolution, which was necessary for the statistical classifiers to discriminate between the spatial patterns of activity associated with each of the two syllables.

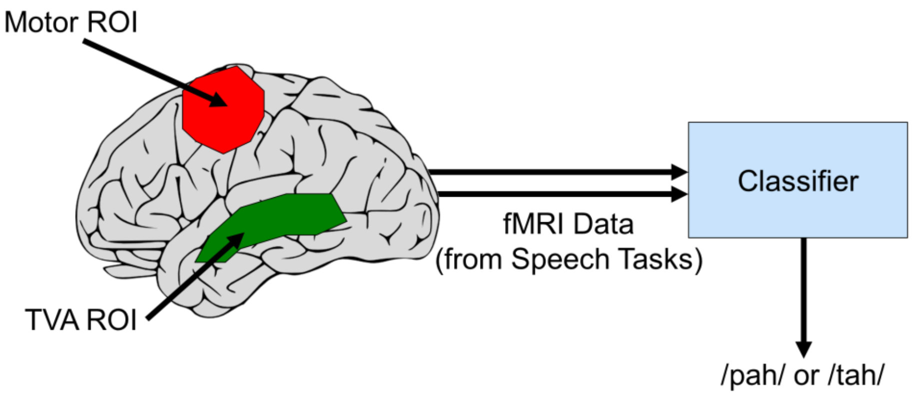


Figure 1. Overview of the approach used in the present study.

1.2 Motivations and Purpose

Before enumerating the specific aims of this study, I will discuss the larger motivations for this research.

The primary motivation of this study is for the eventual development of a covert speech brain-computer interface (BCI). A BCI is a system which decodes signals read from the brain into information or commands a computer can then translate into actions. For the purposes of the present study, a covert speech BCI would be a system that decodes a subject's internal speech-related neural activity, so that it can either be translated into text on a computer screen or synthesized speech. A prime clinical population standing to benefit from such a system is that of patients with locked-in syndrome—a disorder marked by near-complete motor paralysis, despite relatively unimpaired cognition.

However, the present study did not aim to implement a BCI. Rather, I sought to develop a model system for a future applied BCI system: *I hoped to determine what regions of the brain contain information that could then be decoded with more invasive techniques in actual patients*. Accordingly, by demonstrating that simple speech states can be discriminated using neuroimaging data from localized regions of human cortex, this would afford one some confidence that these localized neural regions may contain sufficient information of a person's internal speech states, and argues for the future investigation of focusing on these regions in the design of future BCIs for the decoding of speech.

It is hoped that this study demonstrates directions for future studies sharing the goal of creating a BCI for those who cannot speak—despite normal cortical functioning, such as the locked-in patient.

1.3 Specific Aims of Study

As a means by which to address the above-stated goals, this study addressed the following *three specific aims*:

1.3.1 Specific Aim One

Train a statistical classifier to accurately discriminate between two syllables *overtly*-spoken by a subject in an fMRI scanner, using only acquired fMRI data in brain regions implicated in the production or perception of speech.

1.3.2 Specific Aim Two

Train a statistical classifier to accurately discriminate between two syllables *covertly*-spoken by a subject in an fMRI scanner, using only acquired fMRI data in brain regions implicated in the production or perception of speech.

1.3.3 Specific Aim Three

Compare the classification of *overtly*- and *covertly*-spoken syllables in brain regions implicated in the production or perception of speech.

1.4 Structure of This Paper

The ensuing sections of this paper will first provide relevant background, and then detail the specific methods and results of the current study—all in the context of addressing these aims. I will conclude with a discussion of an interpretation of the results, as well as comments on their significance towards the larger goal of developing a BCI for covert speech.

2. Background

It is necessary to first review background material which will provide the basis of the rationale for the methods which will be proposed to address the specific aims of the study. Background will include a discussion on motivations to develop a speech brain-computer interface, overt and covert speech, differing approaches to decoding speech states, and what unit of speech to decode. Following these will be a description of fMRI methods relevant to studies of speech and how statistical classifiers can be used to decode cognitive states from acquired fMRI data.

2.1 Motivations to Develop a Covert Speech BCI

As has been stated, a primary goal of this study is to provide a foundation for the future development of a brain-computer interface (BCI) to decode covert speech. The primary motivation is to provide a means by which individuals with *locked-in syndrome* can communicate. Locked-in syndrome is a severe, debilitating condition, in which patients display near-complete paralysis; it is defined by the presentation of quadriplegia and an inability to articulate speech, but consciousness is spared (Smith & Delargy 2005). Some patients are able to blink or move their eyes—but only vertically; yet others display a complete absence of voluntary motor control (Smith & Delargy 2005). Although this condition can present itself in the late-stages of amyotrophic lateral sclerosis (ALS), it most commonly results following a lesion to the hindbrain—typically the ventral pons (Fenton & Alpert 2008).

Use of a covert speech BCI to restore communication in locked-in patients is, in theory, feasible. Evidence comes from the crude methods by which some locked-in individuals have been able to communicate with others. As some patients retain voluntary control over vertical gaze and blinking, they are able to use “blinking alphabets” to communicate, albeit at a very slow rate (Fenton & Alpert 2008). The use of a blinking alphabet has perhaps been made most well-known by Jean Dominique Bauby, a locked-in patient who wrote an autobiography, *The Diving Bell and the Butterfly*, which he “dictated” solely by blinks of his left eyelid (Bauby 1998).

Importantly, as these individuals are able to use primitive signals to convey speech intent, they likely have residual speech-related motor functioning—thus a BCI might be able to decode the internal speech intentions of those with locked-in syndrome. Furthermore, a BCI has the potential to decode and convey covert speech much faster than through the use of blinks and vertical eye movement.

A covert speech BCI may also be able to help other patient populations. Recent work by Monti et al. (2010) suggests that a subset of patients with disorders of consciousness demonstrate brain activity indicative of both awareness and cognition—despite their behavior being consistent with a diagnosis of *persistent vegetative state* (PVS). These patients demonstrated brain activity similar to controls in motor and spatial imagery tasks—and were able to communicate “yes” and “no” responses to questions using gross-level changes in fMRI patterns of activity through motor imagery, such as imagining playing tennis (Owen & Coleman 2008). Patients with *amyotrophic lateral sclerosis* (ALS) would also benefit from a covert speech BCI; similar to locked-in syndrome, ALS leaves patients with profound upper and lower paralysis, often sparing the oculomotor muscles. Most importantly, cognitive function is often spared, making it feasible to decode covert speech.

In conclusion, studying methods to decode speech-states using neuroimaging methods will aid in the future implementation of covert speech BCIs. Importantly, these BCIs may help to restore speech to clinical populations of individuals who suffer from profound paralysis, but retain normal cognitive functioning.

2.2 Covert versus Overt Speech

In order to discuss methods by which to decode overt and covert speech, these types of speech should first be more clearly defined so as to differentiate the two. *Overt speech* is simply the vocalized speech with which we are all well familiar; it is spoken aloud in social conversation or when occasionally talking out loud to ourselves (private speech). Importantly, overt speech is characterized by voluntary motor action of the speech articulators. While there are many articulators of speech (Scully 1987), the bilablial (lips) and alveolar (tongue) articulators are most pertinent to this study—partly because they use distinct muscle groups, and contrasts between these articulators have been extensively studied (e.g., Pulvermüller et al. 2006; D'Ausilio et al. 2009; Ackermann et al. 2004).

In contrast, c*overt speech,* also referred to as *inner speech*, is characterized by internal "speech thought"—lacking any intentional use of the motor articulators of speech. I say that inner speech lacks "intentional" motor articulation as there is much evidence that the distinction between overt and covert speech—at the level of both cortical activity and peripheral motor activity—is not as clear as differences between the gross behavioral manifestations of the two speech types. The findings of a study by Livesay et al. (1996) demonstrated the presence of subthreshold motor activity in the lips during covert speech using electromyography (EMG). Similarly, McGuigan and Dollins (1989) demonstrated EMG activity in the lips and tongue during covert speech behavior. Although the observed EMG amplitudes were too low to induce overt articulation, the same patterns of activity were seen when identical phonemes were produced overtly.

It should also be mentioned, with the primary motivation of this study being to provide a foundation for the future development of a covert speech BCI for locked-in syndrome, that the covert speech of the locked-in patient will not produce *any* peripheral motor activity, due to the nature of the paralysis of speech. However, there is sufficient evidence to believe that similar cortical mechanisms may be involved in the covert speech of locked-in individuals and controls. Some evidence for this comes from the above-mentioned study by Monti et al. (2010) in which some individuals with a presentation of symptoms similar to locked-in syndrome displayed similar neural activation patterns to controls during motor and spatial imagery tasks. Similarities between the cortical activity during covert and overt speech will be discussed further in the next section, where I will explain the theoretical underpinnings to the approaches used by the present study to decode overt and covert speech.

Finally, for the purposes of this paper, I will define these two speech types with the following distinction: *overt speech* is deliberately vocalized speech, and thus there is voluntary action of the speech articulators; in contrast, *covert speech* is inner-thought resembling vocalized speech in syntax and structure, but lacking in voluntary intention to engage motor speech action, with no more than subthreshold EMG activity of articulators.

2.3 Two Approaches: Motor and Auditory

In this section I will discuss the theories underlying the two approaches I used to decode overt and covert speech from cortical activity. One approach is to decode speech states from motor areas. The other approach is to use activity in speech-perception areas. I will also discuss combining the two approaches.

2.3.1 Motor Approach – Production of Speech

The first approach I introduce concerns the use of *decoding inner speech states from activity in motor-related regions of cortex*. This approach relies on the theory that covert speech uses the same motor mechanisms of overt speech—there is just not enough motor output to produce movement.

In an fMRI study investigating differences in functional activation between overt and covert speech, Shuster and Lemieux (2005) found greater levels of activation in motor regions of cortex during overt speech. Similarly, Dogil et al. (2002) found greater activation in motor regions during overt versus covert articulation. However, they found that both overt and covert articulation were associated with significantly higher fMRI signal compared to baseline in a "motor speech network"—primary motor cortex, supplementary motor area, and cerebellum. Using a different approach, Aziz-Zadeh et al. (2005) used repetitive transcranial magnetic stimulation (rTMS) pulses over motor regions to induce speech arrest during both overt and covert speech. They assessed the level of speech arrest by measuring differences in response latencies in a speech-related task. Evidence from this study argues for the necessity of motor regions used during overt speech in the production of covert speech, adding credibility to the proposed motor approach of speech decoding.

Additional evidence comes from the motor theory of speech perception (MTSP), which postulates the same neural mechanisms are involved in both speech perception and speech production (Liberman et al. 1967; Liberman & Mattingly 1985; Liberman & Whalen 2000). If this is the case, then it is probable that covert speech—viewed as mental stimulation of speech—uses the same neural mechanisms of perception and production. Somewhat recently, Pulvermüller et al. (2006) demonstrated a clear link between the phonological mechanisms for the production and perception of syllables. Evidence for this link was strengthened by D'Ausilio et al. (2009).

The argument that covert speech uses similar neural mechanisms to overt speech can also be made by viewing covert speech as a type of *motor imagery*—imagining or simulating a motor action within the mind. Jeannerod and Frak (1999) shed light on the possible covert aspects of motor activity. They found evidence that mental simulation of motor action relies on neural mechanisms of motor execution. While they investigated non-speech motor actions, the same is likely the case for motor imagery of speech, especially given results from other studies as discussed above.

2.3.2 Auditory Approach – Perception of Speech

The second approach I introduce concerns the use of *decoding inner speech states from activity in speech-perception regions of cortex*. This approach relies on the theory that covert speech production is a form of "auditory verbal imagery" that elicits activity in auditory areas activated during the perception of speech.

Visual imagery—internally imagining visual scenes—has been shown to activate visual cortical areas in a topographic manner (Kosslyn et al. 1993). Similarly, covert speech could be viewed as a form of *auditory verbal imagery*. A study by Sherghill et al. (2001) provided evidence that inner speech is associated with some activation in lateral temporal regions associated with speech perception (Belin et al. 2000).

Formisano et al. (2008) have shown that it is possible to accurately decode which syllables subjects listened to using fMRI MVPA techniques (see below). If covert speech does indeed activate speech perception regions in a similar manner to the perception of actual speech, then it may be possible to decode using this approach.

2.3.3 Combining Approaches

These approaches need not be mutually exclusive—it may be possible to decode speech from both motor *and* perceptual areas. In this case, decoding accuracy might improve by using both motor and perceptual regions. As stated earlier, despite similarities, there are also differences in activity in these regions between overt and covert speech. Therefore there may be differential success in decoding each speech type from motor, perceptual, or both regions. I hypothesized that higher accuracies could be achieved in decoding from these areas during overt *versus* covert speech. However, I anticipated above-chance accuracy for both overt *and* covert speech. Addressing these questions is the main goal of Specific Aim Three, as stated in the introduction of this paper.

I will later discuss the methods I used corresponding to an implementation of each approach, as well as the results that were obtained.

2.4 Decoding Syllables as Units of Speech

The aims of the study involved the decoding of syllables as simple speech-states. This is a reasonable goal with which to start in working towards the longer-term goal of decoding full covert speech. While syllables, phonemes, and words are each linguistic units of speech (Scully 1987), conflicting evidence does not clearly suggest which of these is the fundamental unit in speech *production* (Wilshire & Nespoulous 2003; Cholin et al. 2003; Levelt 1992; and Levelt & Wheeldon 1994). However, for the purposes of the present study, the choice of syllables as simple units of speech to decode is not difficult to justify. First, syllables are well-instantiated in terms of their associated motor articulators. Additionally, attempting to decode syllables is a *more tractable* problem as compared to phonemes and especially words, which are a more abstract representation of speech. Furthermore, in using syllables, it was possible to use a paradigm for the motor approach (see above) similar to that used in a previous study (Pulvermüller et al. 2006).

2.5 Functional Magnetic Resonance Imaging

Functional magnetic resonance imaging (fMRI) was used for two primary reasons: first, it is a well-established and non-invasive neuroimaging technique for measuring neural activity; second, fMRI has a high spatial resolution, which was necessary for the statistical classifiers to discriminate between the spatial patterns of activity associated with each of the two syllables.

2.5.1 Basic fMRI Background

As the aims of the study involve the use of functional magnetic resonance imaging (fMRI), some general background and issues with its use in speech-related studies is appropriate. Functional MRI is a neuroimaging technique often used to examine the activity of the human brain during functional tasks. It is quite important to note, however, that fMRI is *not* a method to *directly* measure the activity of neurons in the brain, such as the firing of action potentials. Instead, fMRI measures what is termed the blood-oxygenation-level dependent (BOLD) contrast, an index of the ratio of deoxygenated hemoglobin to oxygenated hemoglobin in vasculature. This can be measured with MRI as deoxygenated hemoglobin has a much higher magnetic susceptibility than oxygenated hemoglobin (Huettel et al. 2009). As oxygen is necessary for and used up during much of neuronal activity, the BOLD hemodynamic response is thought to correlate with neuronal activity. Although it is not an entirely resolved issue, it has been shown that observed BOLD activity may correlate more with local field potentials (LFPs) than action potentials (Logothetis et al. 2001).

Images are acquired by the fMRI scanner by taking individual *slices* in quick succession which form a three-dimensional (3D) *volume*. Multiple 3D volumes are taken over time in a functional scan (typically around two seconds between the acquisitions of each 3D volume) and are referred to as four-dimensional (4D) volumes. Volumes are composed of a 3D matrix of image constituents with identical dimensions called *voxels*. The voxel is then the smallest unit on which analyses can be performed, and their dimensions determine the spatial resolution. Typical voxel dimensions are 1 mm^3 for structural volumes, and closer to 4 mm^3 in functional scans. Increases in magnetic field strengths allow for the acquisition of higher-resolution images, but at the cost of increasing the impact of undesired noise and artifact (Huettel et al. 2009).

The shape of the BOLD hemodynamic response, after a stimulus leads to activation of a voxel, is characterized by a quick rise in amplitude, a peak (followed by a plateau if the stimulus is longer lasting), and a fall in amplitude to a level below baseline, which leads to an undershoot that more slowly recovers. The timescale of the BOLD response to reach its peak is on the order of 5-8 seconds (Hall et al. 1999). Since this is orders of magnitude slower than the timescale of actual neuronal firing, this limits what types of questions can be addressed with fMRI. This "sluggishness" of the hemodynamic response and frequency at which we acquire volumes (typically not less than 1.5 s) determines the temporal resolution (Huettel et al. 2009). Therefore, when studying speech and speech perception with fMRI, we can investigate BOLD activation associated with speaking or listening to specific words, types of words, syllables, phonemes, or even types of sentences, but we cannot easily determine the BOLD response seen for each word in a sentence uttered by or presented to the subject.

2.5.2 Scanning During Speech Tasks

In this study I attempted to decode both overt and covert speech with fMRI (Specific Aims One and Two). To do this, speech tasks were completed by subjects in the fMRI scanner. However, studying speech—both overt and covert—with fMRI presents particular difficulties. Measurement of activity associated with both classes of speech must find ways of dealing with the significant between-subject variability in the functional topography when speaking. The measurement of inner speech is likely to be problematic, primarily in that it is, by definition, covert, and there are no methods to guarantee that the subject is producing the desired covert words or syllables. Furthermore, the neural systems involved with production of inner speech are not fully understood and it is reasonable to expect weaker signal in covert versus overt speech, as the phonetic representations remain cognitive and fewer motor-related areas show activation (Shuster & Lemeiux 2005).

As stated above, overt speech requires movements of the jaws, larynx, tongue, and lips to produce and articulate the finely-tuned sounds. Unfortunately, this movement can cause significant distortion in the images acquired by the scanner (Dogil et al. 2002). This undesired effect may be moderated to an extent if subjects are asked to use only the minimal motion required.

2.5.3 Issues with the Presentation of Auditory Stimuli

To study speech perception, auditory stimuli must be generated and presented to the subject during scanning. Several factors must be taken into consideration. First, high-fidelity magnetic speakers cannot be placed in the scanner for obvious reasons. One often-employed method is the use of pneumatic headphones (Gaab et al. 2007), but sound quality is lost to at least some level, and this may be problematic when the stimuli are verbal utterances, which require high-fidelity to accurately interpret.

The loud sounds of the scanner (110-140 dB) present another problem as they can easily drown out the desired auditory stimuli (Hall et al. 1999). One way to deal with the scanner noise is to present the stimuli only between acquisitions. However, this limits the duration of sounds one can present if using a normal time between acquisitions (TR), such as 2 s: due to the lag in the hemodynamic response (much greater than 2 s), it is difficult or not possible to separate signal due to scanner noise from signal due to the stimulus.

The method of sparse temporal sampling, as demonstrated by Hall et al. (1999), addresses both of these issues by acquiring volumes only near the maximum and minimum of the hemodynamic response to maximize contrast while reducing inter-volume scanner noise. Therefore, the TR needs to be large enough such that a new volume is only acquired after the majority of BOLD activation from the previous stimulus is no longer present *and* the volume is acquired near the peak of the "current" stimulus. As mentioned above, the peak of the hemodynamic response has a latency of 5-8 s after stimulus onset; the TR used in a sparse temporal sampling sequence should exceed this latency (Hall et al. 1999).

From the discussion in this section, it can be seen why fMRI can be used to learn what regions of the brain contain information we can decode to determine speech content. However, a practical BCI could not be based on the use of fMRI to acquire estimates of neural activity—costs associated with scanner usage are quite high, and the temporal resolution is too low to be of real-time use for speech.

2.5.4 Temporal Voice Area

As a part of this study, I investigated the auditory approach to decoding covert speech (see above). To do this, areas of the brain involved in speech perception first had to be localized.

This sparse temporal sampling method was used by Belin et al. (2000; 2002) to locate areas of human cortex responding selectively to vocal versus non-voice sounds. They succeeded in functionally localizing a "Temporal Voice Area" (TVA) present bilaterally, on the upper bank of the superior temporal sulcus (STS). The localization of this region is important if we are interested in decoding inner speech, as perception of one's own covertly-spoken voice may activate this area similarly to an overt vocal stimulus.

Therefore, activity in the TVA, in addition to activity in speech motor areas, may be useful as features used to decode covert speech.

2.6 Multivariate Pattern Analysis

The Specific Aims of the present study involve the use of training statistical classifiers to discriminate between the speech states of subjects in an fMRI scanner. To accomplish this, I used multi-voxel pattern analysis (MVPA) as a “decoding” technique. This section will discuss this technique in some detail.

2.6.1 MVPA Approach versus Univariate Methods: Decoding Cognitive States

The "traditional" objective of an analysis of data collected using fMRI (or any other neuroimaging modality for that matter) is to be able to make a statement of the form "when subjects are doing task *T* or presented with stimulus *S*, a statistically significant BOLD response is observed in voxels *V*." However, the goal of *decoding* methods is the opposite; we hope to be able to *predict* accurately that a subject is engaged in task *T* or viewing a stimulus *S*. Stated differently, such methods aim to decode behavioral or cognitive states of individuals using only the observed BOLD activity (Mitchell et al. 2004). Decoding is accomplished either through classification or reconstruction of cognitive states. Classification methods are limited to a finite number of discrete states which can be decoded, whereas reconstructive approaches are not limited in this way. Predicting covertly-spoken syllables lends itself nicely to the classification technique, especially when the number of syllables is small.

The MVPA approach enables such classification by doing multivariate analysis on the *patterns* of fMRI activation *across many voxels*. Single voxels contain information from many thousands of neurons, and of overlapping populations coding different states of interest. For example, in visual cortex, populations of individual cells code the orientation of visually-observed objects (Hubel & Wiesel 1962). Single voxels contain neurons encoding many different orientations—this makes classification of orientation impossible if looking at signal changes from individual voxels. However, as demonstrated by Kamitani and Tong (2005), the orientation of visual stimuli can be accurately classified by measuring differential patterns of activity across many voxels in visual cortex.

Decoding cognitive information directly from neuroimaging data has been a growing trend, and has been demonstrated with significant success by many in different domains (e.g., Haxby et al. 2001; Cox & Savoy 2003; Kamitani & Tong 2005; Kriegskorte et al. 2007; and Formisano et al. 2008). Both whole-brain (Mitchell et al. 2008) and in regions-of-interest (ROI) methods have been used (Etzel et al. 2009). In the whole-brain case, one uses the voxels from the entire brain (or cortex) to train and test a classifier to decode classes of interest. In contrast, ROI-based approaches train and test classifiers using a subset of voxels. Here, voxels used are in ROIs defined *prior* to the training and testing of classifiers. This is important so as to not introduce statistical non-independence into the process (Kriegskorte et al. 2009). Common a priori methods of ROI selection, which allow one to stay clear of non-independence issues, include using structurally-defined ROIs as well as well as functionally-defined ROIs. These two methods are often combined, in which voxels are selected from a functional-localizer, but only significant voxels within an anatomical region are selected—an approach used by the present study.

2.6.2 Steps in Performing a Classification Analysis

As reviewed by Pereira et al. (2009), to perform a classification-based analysis one must define the classes, features to be classified, examples, and the choice of the classifier itself. The classes are discrete labels for the data, and they are usually task conditions or categories of stimuli. Features are the predictors used to classify the data—often voxels in fMRI. Choosing which features to use is a critical step; with fMRI data, feature selection usually involves deciding how many and which voxels are relevant and likely to contain information that varies depending upon class. Voxels can be selected by defining regions of interest (ROIs) that are determined anatomically (e.g., voxels in the anterior cingulate cortex) or by using functional localizers to locate clusters of voxels which show activation during a specific task. It is important to mention, however, that *independent* data must be used to define the features (Kriegeskorte et al. 2009). This means that features should not be chosen using the same functional data we wish to classify, as this may introduce statistical bias in the results.

It may be desirable to reduce the number of features, as it reduces the dimensionality of the analysis (and thus computational power needed) and can make the resultant classification easier to present and interpret (having more than three features is very difficult to graph effectively). However, dimensionality reduction should be performed only with sound theoretical reasons for doing so (Cox et al. 2003; Pereira et al. 2009). Some simple methods of feature reduction include principal components analysis (PCA), its cousin, independent components analysis (ICA), and mean voxel-activations of ROIs can be used in place of all individual voxel values (Pereira et al. 2009).

The data which are to be classified are divided into *examples*, each of which corresponds to a single sample in the feature-space. For classifiers with parameters which must be trained, examples are the inputs to the classifier, and are divided into a *training set* and a *test set*. Examples in both sets need to be independently drawn from an "example distribution" with no overlap (Pereira et al. 2009). Training data must have an equal number of examples with each class. The training of the classifier is done by feeding it examples from the training set and the corresponding known class labels and the classifier parameters are optimally adjusted. The testing is performed by feeding the trained classifier test examples, and comparing the output label with the correct class for the example. Accuracy can be measured simply as the percentage of correct classifications. Any accuracy significantly above chance-level (50% for a two-class dataset) suggests that the classifier demonstrates at least some level of efficacy in decoding classes (e.g., cognitive states) from data in the feature space (e.g., fMRI signal in a defined ROI).

If one is fortunate enough to have a very large example set, then separate training and test sets can be partitioned which are sufficiently large. However, this is rarely the case in fMRI paradigms, where one may often have fewer than ten examples (often corresponding to runs). Luckily, there are methods which allow for effective use of small data sets through methods of *cross-validation* such as “leave-one-out” and *k*-fold cross validation (Pereira et al. 2009). These are methods by which a classifier is trained on a large proportion of the examples, and tested on the remaining examples, and additional classifiers are similarly trained and tested on all remaining combinations of examples (of the same proportions). The overall classifier accuracy is then calculated as the mean accuracy of each individually-trained and tested classifier.

2.6.3 Linear Support Vector Machines and Correlation-Based Classification

The pattern classification discipline is in no shortage of classifier algorithms which can be used to decode cognitive states from fMRI, each with their advantages and disadvantages, and are reviewed in depth elsewhere (Bishop 2006; Pereira et al. 2009; Theodoridis & Koutroumbas 2006). Two methods which are applicable in the classification of syllables of speech are support vector machines and a correlation-based approach.

Linear support vector machines (SVM) are linear maximum-margin classifiers, well suited for high-dimensional datasets (Ben-Hur et al. 2008) which make them appropriate for classification of fMRI data (Cox and Savoy 2003). Mathematically, a linear SVM is a linear discriminant function of the form $y(\mathbf{x}) = \mathbf{w}\mathbf{x} + b$, where $\mathbf{x}$ is an example in feature space, b is a “bias” parameter, y is the class output by the SVM, and $\mathbf{w}$ is a vector of weights which are the classifier parameters that must be trained, such as to maximize the distance between the decision boundary, where $y(\mathbf{x}) = 0$, and the closest data points (Bishop 2006).

In addition, a correlation-based classification algorithm (as in Haxby et al. 2001; Kanwisher et al. 1997) is a simple method which does not require the training of any classifier parameters. It allows one to demonstrate that different patterns of feature-values may be separable by class. One does this by comparing within-class correlations of data to between-class correlations (Haxby et al. 2001). If the within-class correlations are significantly greater than zero and greater than the between-class correlations, than this is evidence that the classes have differing patterns in the feature space. Although not *strictly* a method of decoding, this approach could be used to classify "test" data by determining which class test examples correlate more closely with.

3. Methods

I will now present in detail the methods of data collection and analysis used to address the above-mentioned aims of the study. The majority of methods were within-subject, and will be described first. Group-level methods of analysis will then be covered.

3.1 Overview

Here I will discuss in detail the fMRI task paradigm used in the present study, as well as subject-level analyses performed on all acquired data. To address the aims of this study, I used fMRI in conjunction with the multi-voxel pattern analysis (MVPA) technique to test the motor and auditory approaches of decoding speech. Specifically, it was investigated whether patterns of fMRI activity in cortical areas involved in *speech production* and *speech perception* contain sufficient information to correctly classify which of two syllables a subject overtly (Specific Aim One) or covertly (Specific Aim Two) produced.

I used fMRI to scan subjects as they completed speech tasks in which they were prompted to speak either the syllable /pah/ or /tah/. In one scanning session, subjects spoke these syllables *overtly* (aloud); in another session, they *covertly* “spoke” the syllables in their mind without moving. Results from two independent tasks—a task involving motor articulation, and a task involving speech perception—were used to select regions of interest (ROIs) for each subject. The ROIs were used to subset the fMRI data acquired by the speech tasks. I then used a correlational approach to first demonstrate the patterns of activation associated with speaking each syllable were unique. With the same data, MVPA was used to train statistical algorithms to accurately classify which of the two syllables was spoken. This process was within-subject, and separate classifiers were trained and tested for each ROI—and for both the overt and covert speech conditions. The syllables /pah/ and /tah/ were chosen as they map to different motor articulators: utterances of /pah/ require articulation of the lips; utterances of /tah/ require articulation of the tongue. As these associated lip and tongue movements have somatotopically-distinct representations in motor cortex (Lotze et al. 2000; Hesselmann et al. 2004), they are good syllables to use in fMRI tasks (Pulvermüller et al. 2006).

Methods by which the accuracies of the classifiers were determined will also be discussed—along with assessing their statistical significance. Lastly, a control MVPA comparison will be described, designed to rule out a possible confound.

3.2 Experimental Paradigm

This section will describe in detail each of the tasks performed by each subject during scanning with fMRI. Two tasks were used as functional localizers, to identify the subject-specific regions corresponding to speech production and speech perception—the motor localizer task and the temporal voice area localizer task, respectively. The third task was the speech task, with both an overt and a covert variant. Data from the localizer tasks was used to determine what subset of the data from the speech tasks was to be used in the classification analyses.

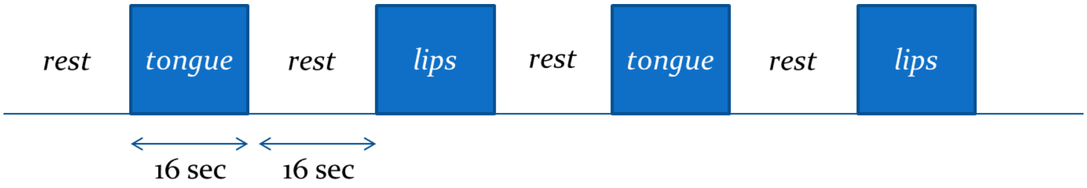


Figure 2. Paradigm for the motor localizer task. There were 16 second blocks, alternating with rest of equal duration. During the blocks subjects had to make slow, minimal movements of either their tongue or lips, as prompted.

3.2.1 Motor Localizer Task

A simple task adapted from Pulvermüller et al. (2006) was used to localize lip and tongue motor areas for each subject. The paradigm consisted of two on-off-on-off blocks, alternating between slow minimal tongue movements and slow minimal lip movements (Figure 2). Attempting to localize motor regions, the subjects were asked to minimize any tactile contact with the tongue and the roof/floor of the mouth, as well as to try and avoid any lip closure, so as to prevent undesired task-related somatosensory activity.

Likewise, it was important for subjects to minimize undesired movement. When the subjects practiced the task (see below), the importance of not moving the jaw while moving lips was emphasized —they were instructed to hold their jaw while they practiced so they could learn to better move their lips without their jaw, if it was not already natural to them.

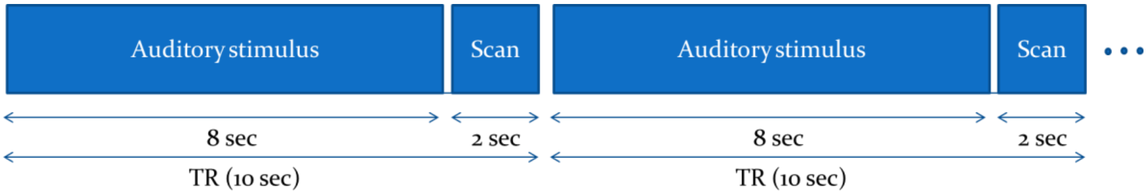


Figure 3. Paradigm for the Temporal Voice Area localizer task.

3.2.2 Temporal Voice Area Localizer Task

I used a paradigm and stimuli adapted from Belin et al. (2000) to localize the temporal voice area (TVA) of each subject. Auditory stimuli, 8 s in duration, were presented in a pseudorandom order. Half of the stimuli consisted of vocal (speech and non-speech) sounds, with the rest being non-vocal sounds. Periods absent of auditory stimulation were also present, to allow for a baseline measurement. This task made use of the sparse temporal sampling technique to deal with issues of auditory tasks (as described above), and volumes were acquired only once every 10 seconds (Figure 3).

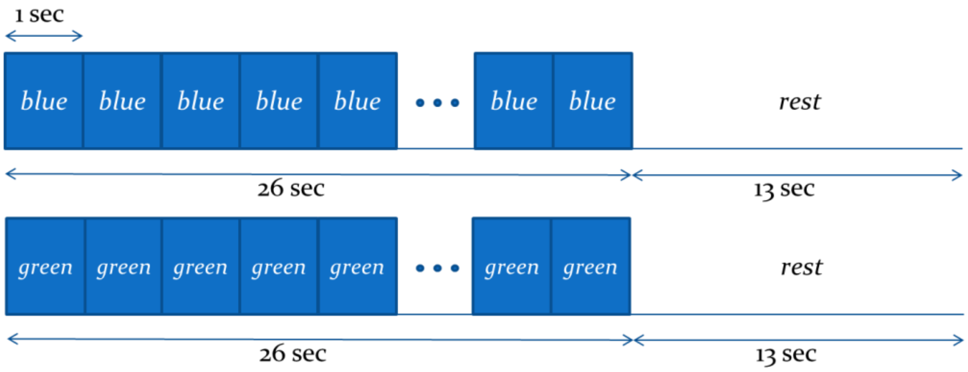


Figure 4. Paradigm for the speech tasks. One pair of blocks is shown.

3.2.3 Speech Task – Two Variants: Overt and Covert

In this task, subjects had to speak one of two syllables—/pah/ or /tah/—as prompted by a visual stimulus; a blue square indicated one syllable, a green square indicated the other. The task was organized in blocks of 26 repetitions of the syllable indicated by blue, followed by an interstimulus interval (ISI) of 13 seconds, and a block of 26 repetitions of the syllable indicated by green, which were again followed by an ISI (see Figure 4). There were four pairs of blocks in each of six runs. To ensure there was no effect from the color of the prompting visual stimulus, the syllables indicated by the blue and green stimuli alternated between runs. The correct color-syllable mappings were explicitly indicated to subjects in visually-presented instructions prior to the start of each run.

Importantly, there were two variants of this task—*overt* and *covert*. In the overt task, subjects spoke the syllables aloud. They were, however, asked to try to limiting movement somewhat, so as to minimize motion-related artifact. In the covert task, subjects "spoke" the syllables in their mind without moving. Overt and covert runs were completed in different scanning sessions (see below).

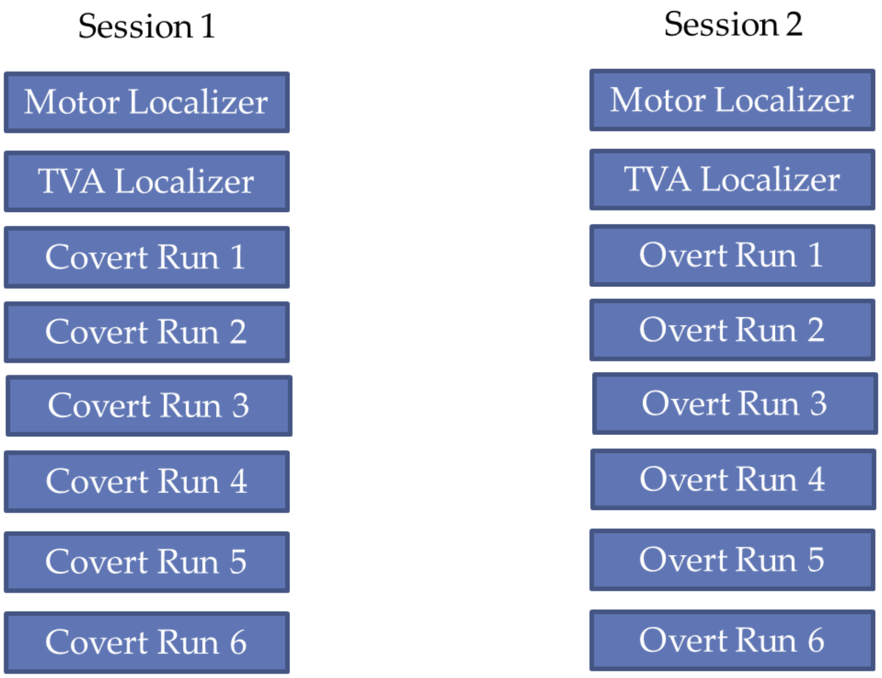


Figure 5. Paradigm overview: order of tasks in each of the two scanning sessions. As I attempted to counter-balance across subjects which speech tasks would be in each session, about half of the subjects did the covert speech task in the first session as indicated in the figure, and the remaining subjects did the overt speech task in the first session.

3.3 Defining Terminology (to be used in remainder of paper)

In order to reduce confusion, as there are two scanning sessions for each subject, in which the same tasks, or variants of the same task, are run in each, I now introduce terminology to be used for the remainder of this paper. *Motor localizer* (or *motor localizer task/runs*) will refer to the motor localizer task runs from either (or both) of the sessions. Similarly, *TVA localizer* (or *TVA localizer task/runs*) will refer to the motor localizer task runs from either (or both) of the sessions. *Localizers* (or *localizer tasks/runs*) will refer, collectively, to the motor localizer and TVA localizer runs (from both sessions). Only where it is made explicit will there be a distinction between the localizer tasks/runs from the two sessions. Furthermore, *overt task/runs* will refer to the overt-variant of the speech task, of which there are six in either the first or second session (as this was counter-balanced across subjects). Similarly, *covert task/runs* will refer to the covert-variant of the speech task, of which there are six in the other session. Speech tasks (and their runs) of both the overt and covert variation will be referred to, simply, as *speech tasks/runs*.

Finally, it is essential to note that unless explicitly stated, all methods, analyses, and results described will refer to those as they apply to a single subject; except for comparisons between-subjects—which will be made explicit—methods and analyses are within-subject. Having defined this terminology, it is hoped that subsequent sections of this paper will be understood with greater clarity.

3.4 Subjects

Six healthy, right-handed, and English-speaking subjects (three females, three males) were recruited to complete two scanning sessions each. However, through debriefing it was discovered that one subject completed the one session incorrectly—the subject admitted making mistakes with the stimulus-syllable associations, and he performed some of the covert tasks overtly. The data from this subject was therefore not used. Only the remaining five subjects will be referred to in the remainder of this document—three females, two males; age ranging from 22 to 26 years.

As seen in Figure 5, subjects completed the overt speech task in one session and the covert speech task in the other; I attempted to counter-balance the order of the sessions across individuals. The primary constraint on subject size was budgeting, as usage of the scanner is expensive. However, a modest sample size was sufficient for the purposes of this study, in that most analyses are within-subject, and, as previously mentioned, this study sought primarily to *demonstrate* the possibility of a classification method. As such, the sample size should allow the demonstration of the feasibility of the present approach. All subjects had normal or corrected-to-normal vision. If a subject had corrected-to-normal vision, they were asked to wear contacts in the scanner. Also for safety reasons, females were asked to take a simple pregnancy test immediately before each scanning session. All subjects received informed consent documentation which they signed prior to their participation in the study. The George Mason University Human Subjects Review Board approved all subject-related tasks in the scanner, as well as the consent documentation they signed.

All subjects received a short training session before each of the scanning sessions in order to familiarize them with the task in an attempt to both minimize confusion on the part of the subjects and to increase the quality of the data during the scanning sessions. Training consisted of the subjects sitting in front of a computer where they were presented with identical stimuli to what they later saw in the scanner. They were first given instruction on how to engage in the task, and then received shortened practice rounds. In giving the instructions to the subjects prior to the scanning session, I emphasized the importance of staying still, remaining fixated at the center of the screen (except for the TVA localizer task), and that for the tasks that required movement (motor localizer and the overt speech tasks), movement should be minimal.

During actual scanning, subjects were reminded of the instructions between runs, and, as a further reminder, written instructions were visually presented in the scanner as scanning on each run began.

3.5 Image Acquisition

A 3 Tesla Siemens Magnetom Allegra "head-only" scanner, with a "bird cage" head coil, was used to acquire images both functional and structural images in each of two scanning sessions. A scanning session consisted of eight functional runs and a single structural run. The functional runs corresponded to the above-defined tasks: one run of the motor localizer task, one run of the TVA localizer task, and six runs of the speech task (overt for one session, and covert for the other). The pulse sequences used during the scanning of each run are detailed below.

As discussed earlier, movement during scanning creates substantial artifact in the acquired images. Accordingly, subjects' head movements were restricted with foam padding. The subjects had a "panic button" in their right hand to press in case they would become uncomfortable or needed to be let out of the scanner. The scanner is located on campus at the Krasnow Institute of George Mason University.

3.5.1 Functional Runs

For each of the eight functional runs, a gradient-echo, echoplanar pulse sequence was used to measure the blood-oxygen-level dependent (BOLD) signal during the functional tasks. The same slice-prescription was used for all functional runs: 33 axial slices were obtained per volume (64 x 64 matrix; 3.75 x 3.75 mm in-plane resolution; 4 mm slice thickness with a 1 mm gap).

For the motor localizer run, 78 volumes were collected over 2 min 40 sec (TR/TE = 2000/30 ms). As was described above, the TVA localizer run used the sparse temporal sampling method, and only 63 volumes were collected over a 10 min 40 sec duration in time (TR/TE = 10,000/30 ms). Following the localizer runs, six runs of the speech task were completed. In each speech run, 175 volumes were collected over a 5 min 54 sec duration (TR/TE = 2000/30 ms). Importantly, for these runs, I alternated which stimulus colors (blue and green) were associated with the prompt of the syllables so as to counter-balance and to ensure stimulus color would not confound results.

3.5.2 Presentation of Stimuli and Acquisition of Stimulus-Event Timings

Stimuli were back-projected onto a screen directly in the line of vision of the subjects as they were in the scanner in a supine position. The computer used to project the stimuli, which will hereafter be referred to as the *stimulus computer*, was a Dell Precision M4300 PC laptop (Intel® Core™ 2 Duo CPU, T9300 @ 2.50 GHz; 772 MHz, 2 GB RAM; Microsoft Windows XP Professional, Version 2002 with Service Pack 2).

On the stimulus computer, I used the program Presentation (Version 13.0, Build 01.23.2009), software developed by Neurobehavioral Systems, Inc. (Albany, CA; www.neurobs.com). I used Presentation to program the stimuli and their timings, which were completed—and tested with mock runs—well in advance of when I began scanning subjects. During the functional scans, Presentation recorded to log files the stimulus timings *and* the timings of each volume acquired by the scanner. This information was important, as it was necessary during analyses to know which volumes corresponded to which tasks.

3.5.3 Anatomicals

Following functional scans, a high-resolution T1-weighted structural scan using a three-dimensional magnetization-prepared rapid-acquisition gradient echo (MPRAGE) pulse sequence (160 1mm thick sagittal slices, 256 x 256 matrix, TR/TE = 2300/3.37 ms, flip-angle = 7 degrees, isotropic 0.94 mm voxels). Each structural scan took 8 minutes and 37 seconds to complete. The collection of these three-dimensional structural volumes was necessary during image analyses to localize functional activity and to allow for the spatial normalization of all acquired images.

3.6 Univariate Analysis Procedures

3.6.1 Overview

This section will describe the methods used to analyze the acquired data so as to address the aforementioned aims of the study. All primary analyses conducted were within subjects. First, the fMRI volumes were preprocessed. This was done in such a way as to align all 3D and 4D volumes used in later analyses in the same “native space.” The reasons why I chose to do this are detailed below. Next, I conducted standard univariate analyses on each run with a general linear model. Functional contrasts resulting from these univariate analyses were then used to aid in the creation of regions of interest (ROIs). Results from these univariate analyses and the ROIs were needed for multivariate pattern classification, as described in a later section.

3.6.2 Data Preparation

To put the data in the necessary format that was needed for downstream analyses, I performed several initial data extractions and conversions. Immediately following each scanning session, acquired image data was initially saved to compact discs (CDs) in the Digital Imaging and Communications in Medicine (DICOM) format (http://dicom.nema.org). All DICOM files were then copied to what I will refer to as the *analysis computer*, a Dell Precision WorkSation T7400 (Intel® Xeon® CPU, X5482 @ 3.20 GHz; 3.19 GHz, 8 GB RAM). Proprietary software, DicomWorks (http://dicom.online.fr), was used to extract the image volumes for each run, and DCM2NII (http://www.cabiatl.com/mricro/mricron/dcm2nii.html) was then used to convert the extracted volumes to the NIfTI format (http://nifti.nimh.nih.gov/). These data manipulations were done in Microsoft Windows XP (Professional x64 Edition, Version 2003 with Service Pack 2; www.microsoft.com), and I did all subsequent image analyses (on a dual boot of the analysis computer) using the Fedora 11 distribution of GNU Linux (Kernel version 2.6.30.8-64.fc11.x86_64; http://fedoraproject.org).

For each run, I first removed unnecessary "leading" (and for some runs, "trailing") volumes. This was done as it requires some time for the full effects of the MR gradient coils to "ramp up" and reach a steady-state at the beginning of each run. Also, as mentioned above, instructions were displayed to the subject at the beginning of each run, and volumes acquired during this period were irrelevant to the experiment. Accordingly, for the motor localizer runs, 78 volumes were acquired during scanning, and the first 9 volumes were removed, leaving 69 volumes to use in the analysis; for the TVA localizer runs, 63 volumes were acquired, and the first 2 were removed, leaving 61 volumes for analysis; and for the speech runs, 175 volumes were acquired, and the first 9 and last 2 were removed, leaving 164 volumes for analysis. This step was performed with the FSL software suite (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl).

In addition to the data preparation for the image data, I took preparatory steps to extract the necessary information from the recorded stimulus-event timings recorded by Presentation during image acquisition. The event timings are essential in that they were used to create the model to which the image data was later fit (see below). Using a built-in tool from Presentation in conjunction with Microsoft Excel, I determined the temporal onset and duration of relevant stimuli in each run. This step was completed on the *stimulus computer*.

3.6.3 Within-Subject Coregistration Methods – Preserving the "Native Space"

In this study, I made within-subject comparisons across imaging data. However, individuals move their heads both within and between runs—introducing motion artifact. Furthermore, I needed to compare data across the two sessions for each subject. To address this, I used motion correction and image coregistration methods. This was done *before* fitting the data to a general linear model to extract parameter estimates using FEAT (see below). The following strategy was used:

- Motion-corrected to the "mean of runs" for all covert runs—the mean covert volume becomes the *covert space*
- Motion-corrected to "mean of runs" for overt runs
- Ran FEAT (see below) on covert runs, registering to MNI template—gives us a *covert space*-to-MNI transform
- Ran FEAT (see below) on overt runs, registering to *covert space*

- Ran “low-level” FEAT (see below) on localizer runs—two motor localizer and two TVA localizer runs—registering to *covert space*
- Ran “high-level” analysis with FEAT (see below) for localizer runs; since there were two runs for each localizer, the parameter estimates were averaged together to gain more accurate parameter estimates and *contrasts* between conditions: contrasts are z-score comparisons across condition, with the denominator of the calculation being the residual variance across all of the runs
- Transformed overt run parameter estimates—one for each condition, /pah/ and /tah/—as well as the filtered functional image data—4D volume which has only been preprocessed—into *covert space* using the *overt space*-to-*covert space* transform computed when running FEAT on the overt runs
- Transformed contrast images (*tongue vs. lips* and *vocal vs. nonvocal*) to *covert space* using the transforms computed when running FEAT on each localizer

The importance of ensuring that all overt and covert speech runs are in the same *native* space cannot be understated. Given the sensitivity of the classification algorithms (as will be discussed below) to variations in the pattern of activity of small groups of voxels, even minor deviations in the native space between runs could lead to sufficient noise as to significantly affect the classification results.

3.6.4 Parameter Estimation with FEAT

I conducted standard within-subject functional data analyses using FEAT (FMRI Expert Analysis Tool), which is distributed with FSL. FEAT attempts to fit image data to a general linear model (GLM) to extract *parameter estimates* (the regressors) for each voxel. The model, or *design matrix*, is defined by the task conditions (e.g., lip movement and tongue movement in the motor localizer) and any confounds for which one wishes to account (such as movement). The calculation is essentially a linear regression in which the relative contributions of each condition to the total BOLD signal at each voxel are determined. As is common practice, I convolved a gamma-based convolution with each regressor to compensate for the shape of the hemodynamic response curve (Huettel et al. 2009).

As a part of FEAT, and prior to the computation of parameter estimates and contrasts, standard preprocessing was done: Skull-stripping was performed to remove any non-brain voxels using FSL's BET (Smith 2002); a high-pass filter was applied to remove low-frequency noise (e.g., respiratory artifact), using a threshold period of 116 s; and spatial smoothing was performed using a Gaussian kernel with a full-width, half-maximum length of 5 mm.

3.6.5 Regions of Interest (ROIs)

This section will describe the methods by which the regions of interest (ROIs), used for feature selection, were determined. ROI selection was the critical next step that used results from the analysis of the functional localizer runs to choose which voxels to use as features in training and testing the classifiers with the data from the speech runs.

I used the Harvard-Oxford MNI anatomical atlas, distributed with FSL, to create structural masks. The following masks were created:

- A "whole-brain" mask—consisting of all cortical regions
- Two temporal voice area (TVA) masks—one for each hemisphere—composed of the posterior division of the middle temporal gyrus, the posterior division of superior temporal gyrus, and the planum temporale
- Motor masks—again, one per hemisphere—composed of the precentral gyrus, the postcentral gyrus, and the inferior frontal gyrus, pars opercularis (cortex posterior to the ascending ramus of the lateral fissure)

I was able to easily transform the structural masks from the atlas' MNI space to each subject's native space using the appropriate transformation matrix computed during coregistration (see above). Each of the contrasts—*tongue vs. lips* and *vocal vs. nonvocal* (as computed above with FEAT on the localizer runs for each subject) were subset by each of the four native anatomical masks. Then, the 50 voxels with the highest z-scores in each of these defined the four main ROIs for each the subject—*Motor LH, Motor RH, TVA LH, TVA RH*. The *Whole Brain* ROI included all cortical voxels of the subject, and was not subset by the contrasts.

3.6.6 Univariate Tests of Difference between Syllables

Although not directly addressing the Specific Aims of the present study, I conducted univariate tests of difference. Paired single-sample *t*-tests were run using the mean difference between /pah/ and /tah/ parameter estimates (as averaged over voxels in each ROI) for each run. Significance was computed using the null hypothesis of a mean difference of zero. The results from these tests were used to compare with those of the multivariate analyses as described below.

3.7 Multivariate Pattern Classification

This section will address the main components of Specific Aims One and Two—the training of statistical classifiers to discriminate between the two syllables. Again, this process was done *within-subjects*—the distinct ROIs calculated for each subject (as detailed above) were used to extract features from each subject's covert and overt speech runs, and then separate classifiers were trained and tested within subject. Thus, separate classifier accuracies were determined for each subject.

The steps for the classification analysis were decided using a step-by-step approach as reviewed by Pereira et al. (2009). Accordingly, I will now define the classes, features, data instances, and the choice of the classifiers.

3.7.1 Classes

The classes are the two conditions from the scanning paradigm we wish to discriminate: the covert or overt utterances of the syllables /pah/ and /tah/.

3.7.2 Feature Selection

Parameter estimates from the speech runs, along with the ROIs from the localizers, were imported into Matlab R2007b (Mathworks, Natick, MA) for the classification analyses. The ROIs were used to subset the parameter estimate data from the speech task runs; extracted voxels were used as features in the classification analyses.

3.7.3 Data Used as Instances (Data Points to Classify)

The data instances I attempted to classify were the parameter estimates (PEs)—one for /pah/, one for /tah/—from the speech runs. Therefore, as there were 6 speech runs, there were 12 data instances, 6 of each class. The sparseness of data increased the importance of the chosen feature selection method (Pereira et al. 2009).

3.7.4 Correlation Classifier

I first attempted to demonstrate that the data corresponding to speech conditions /pah/ and /tah/ are separate classes in the extracted features. This was done by correlating within-class parameter estimates of the speech task runs and between-class parameter estimates from even and odd runs (as in Haxby et al. 2001). The data can be "successfully" classified with this method if within-class correlations are significantly greater than zero (use NHST) and greater than between-class correlations. Should this method be unsuccessful, it does not imply that /pah/ and /tah/ cannot be classified with the extracted data. Rather, a more sophisticated classifier may be needed, as separation between classes may not be detectable through simple correlation comparisons. For this reason, I attempted to classify the data with a linear support vector machine.

3.7.5 Linear SVM Classifier

As the number of features is large, and there is only a sparse sampling of data (only six runs), a linear SVM is a good candidate classifier (Pereira et al. 2009). The same classes (/pah/ and /tah/) and features (voxels masked by the localizer ROIs) were used. Instances (samples in the feature-space) were the parameter estimates derived from the speech task runs. I used LIBSVM (http://www.csie.ntu.edu.tw/~cjlin/libsvm), a standard Matlab-compatible SVM package frequently used for fMRI pattern analysis (Cox & Savoy 2003), to train and test linear SVM classifiers.

3.7.6 Cross-validation Accuracy Testing

In order to compute a classifier accuracy measurement, I trained and tested SVM classifiers on examples using the leave-one-out cross-validation method (reviewed in Pereira et al. 2009), leaving out a different run on each training iteration. Accuracy of each classifier is defined by the percentage of correct classifications for the test data set.

3.7.7 Assessing Confidence in Accuracy with Permutation Testing

To determine the statistical significance of a classifier's accuracy a randomization test known as permutation testing (Golland & Fischl 2003) was used. For each dataset (ROI x subject), I iterated over all unique permutations of the class labels. For each iteration, I computed the cross-validation accuracy of classifiers *trained with the permuted labels,* but *tested on the true labels*. This yielded a distribution of accuracies across all possible class-training labels. The likelihood that the classifier trained with the *correct* labels had cross-validation accuracy significantly above chance-level was computed as the proportion of class permutations yielding accuracies less than that of the correctly-trained classifier.

3.7.8 Assessing Confidence in Accuracy with a Binomial Test

In addition to permutation testing, I used a binomial test to calculate a second measure of significance. This method is also often used to assess classifier accuracy (e.g., Harrison & Tong 2009); while it may be a more liberal test of significance than permutation testing, the number of instances used in this study is small enough to warrant caution in interpreting the results of permutations testing—the fewer the instances, the fewer the possible permutations of class labels, yielding a sparser "permutation distribution."

Using a binomial distribution, I tested the null hypothesis that classifier accuracy was at chance—50%. A one-tailed test of significance gave the likelihood of the computed cross-validation accuracies being greater than chance.

3.8 Multivariate Control Analysis

3.8.1 Train Classifiers Using Stimulus Color Labels

To provide further evidence that the present approach demonstrated syllables can be decoded using this method, I introduced a control comparison. I used the same linear SVM classifier methods as above, but I *replaced the class labels* to indicate the *color of the stimuli* used *instead of the syllables* prompted by the stimuli. I then used the same cross-validation accuracy assessment, using permutation and binomial testing to obtain likelihood estimates for the computed accuracies. This was a valid control comparison in that color and syllable were orthogonal factors—ensured by the counter-balancing of color-syllable pairings in the design of the speech tasks (see above).

3.9 Group-Level Analyses

In addition to the within-subjects analysis described in the previous sections, I conducted the following group-level analyses.

3.9.1 Group-Level Univariate Tests of Difference between Syllables

Similar to the within-subjects univariate tests described above (see section 3.6.6), paired single-sample *t*-tests were run using each subject's mean difference between /pah/ and /tah/ parameter estimates (as averaged over voxels in each ROI). Significance was computed using the null hypothesis of a mean difference of zero. The results from these tests were used to compare with those of the multivariate group-level analyses as described below.

3.9.2 Group-Level Correlation Analyses

I examined group-level results from the correlation classifier results. Correlations were averaged over subjects for each ROI. Significance was determined with independent one-sample *t*-tests; the null hypothesis *t*-distribution had a mean correlation of zero.

3.9.3 Group-Level SVM Analyses

As with the correlations, mean SVM cross-validation accuracies were computed by *averaging over subjects for each of the ROIs*. Significance of mean accuracies was computed with independent one-sample *t*-tests comparing to population distribution with a mean of .50 (chance-level accuracy). These were one-tailed tests, as the hypothesis that observed accuracies were greater than chance was being tested. This was done separately for both the overt and covert conditions.

As per the Three Specific Aims, classifiers were trained separately on different brain regions for both overt and covert speech data. I therefore conducted a repeated measures ANOVA testing for effects due to three factors: *condition* (*Overt vs. Covert*); *region* (*Motor vs. TVA*); and *hemisphere* (*LH vs. RH*). Additionally, the data were examined for any interaction effects between the factors.

An additional ANOVA was run as a control to compare the mean accuracy with the *Whole-Brain* ROI as compared to the mean accuracies of the other ROIs combined.

The level of significance of these computed main and interaction effects allowed for conclusions regarding the comparative performance of the classifiers using fMRI data from motor versus perceptual regions and during overt versus covert speech tasks. These results will be presented next, followed by a discussion of the findings as they relate to both the Specific Aims and the more general purpose of the study.

4. Results

In this section, I will present the results of this study. Detailed discussion of the results—their significance—will be included in the following section.

4.1 Univariate Results

4.1.1 Regions of Interest

As stated earlier, I computed univariate statistics on each subject's data prior to doing any multivariate pattern analysis. The primary purpose of this step was to select regions of interest (ROIs) to be used by the classifiers. All fMRI volumes were motion-corrected and co-registered to the same "native space." Parameter estimates (PEs) for each run were obtained by fitting the data to a general linear model defined by the design matrix of the run. PEs from the localizer runs were used to create functional contrast images: *tongue vs. lips* from the motor localizer; *vocal vs. nonvocal* from the TVA localizer.

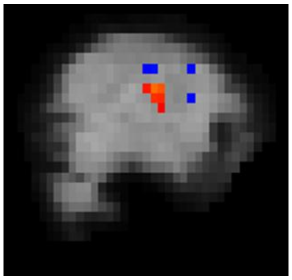


Figure 6. The *tongue vs. lips* contrast for subject S5. Red indicates voxels responding more to tongue than lips; blue indicates voxels responding more to lips than tongue. Only values within the motor mask are shown.

Figure 6 shows the *tongue vs. lips* contrast computed for one subject. A clear separation can be seen between voxels showing greater activation during tongue versus lip movement (in red) and those showing greater activation during lip versus tongue movement (in blue). This demonstrates the distinct somatotopic representations for areas more sensitive to either tongue or lip movement.

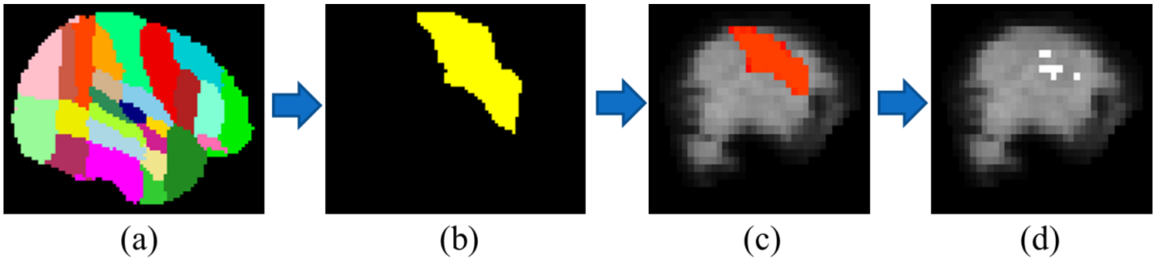


Figure 7. The ROI-selection process for one of the subjects (S5). (a) The Harvard-Oxford anatomical atlas in MNI standardized space; (b) selected motor anatomical regions from the atlas as a mask; (c) the mask as converted into the subject's native space; and (d) the resulting *Motor LH* ROI consisting of the 50 voxels from the *tongue vs. lips* contrast with the greatest contrast (the 25 voxels with the highest values and the 25 with the lowest) within the spatial extent of the mask.

Figure 7 displays the process by which the *Motor RH* ROI was created for one subject. The motor anatomical regions were selected from The Harvard-Oxford anatomical atlas in MNI standardized space (Figure 7a) and were used to select the motor anatomical regions as a mask (Figure 7b). The mask was converted into the subject’s native space (Figure 7c), and the 50 voxels from the *tongue vs. lips* contrast with the greatest contrast (the 25 voxels with the highest values and the 25 with the lowest) within the spatial extent of the mask defined the ROI (Figure 7d). This same method was applied to create all ROIs for each subject, with one exception—the *Whole brain* ROIs for each subject consisted of all cortical voxels, and so they varied in size (see Table 1). All other ROIs—*Motor LH*, *Motor RH*, *TVA LH*, and *TVA RH*—consisted of 50 voxels each, selected as just described.

Table 1. Number of voxels in the *Whole Brain* ROI for each subject

Subject	S1	S2	S3	S4	S5
Whole Brain ROI Voxels	15137	17241	17801	24207	14298

4.1.2 Univariate Tests of Difference between Syllables

To compare with the results of the multivariate analyses (see below), I conducted univariate tests of difference. First, paired single-sample *t*-tests were run using the mean difference between /pah/ and /tah/ parameter estimates (as averaged over voxels in each ROI) for each run—a within-subjects comparison. Significance was computed using the null hypothesis of a mean difference of zero. The results from these tests are displayed in Table 2. Group-level univariate tests were then conducted similarly: Paired single-sample *t*-tests were run using each subject's mean difference between /pah/ and /tah/ parameter estimates (as averaged over voxels in each ROI). Again, significance was computed using the null hypothesis of a mean difference of zero. The results from the group-level tests are displayed in Table 3.

While it can be seen that some of the within-subjects comparisons were statistically significant, all comparisons made at the group-level were nonsignificant (p > .05) for each ROI and for both conditions. These results will be further mentioned below in comparison to the multivariate results.

Table 2. Univariate comparisons of difference between parameter estimates averaged over all voxels in each ROI. Values are results from paired *t*-tests computed per subject by ROI and speech condition. *p<.05; **p<.01; ***p<.001

Condition	Subject	Whole brain	Motor (LH)	Motor (RH)	TVA (LH)	TVA (RH)
Overt	S1	19.243 ***	9.124 ***	27.035 ***	11.261 ***	7.415 ***
	S2	9.741 ***	5.805 **	4.035 *	8.162 ***	3.204 *
	S3	2.214	1.042	1.161	-0.147	-0.186
	S4	-3.907 *	-1.828	-2.100	-4.370 **	-6.961 ***
	S5	8.651 ***	1.911	5.624 **	3.330 *	-2.624 *
Covert	S1	1.067	1.044	1.147	1.393	1.591
	S2	3.704 *	6.733 **	3.415 *	0.165	1.283
	S3	0.258	-0.463	-0.515	-0.287	-0.441
	S4	-2.548	0.769	1.808	-1.428	-1.427
	S5	3.337 *	7.552 ***	3.699 *	1.893	2.277

Table 3. Univariate group-level comparisons of mean differences between the average difference in parameter estimates across class for each subject. Values are results from paired *t*-tests computed for each ROI from both overt and covert speech runs.

Condition	ROI	t	p
Overt	Whole brain	1.534	0.200
	Motor (LH)	1.563	0.193
	Motor (RH)	2.021	0.113
	TVA (LH)	0.997	0.375
	TVA (RH)	-0.249	0.816
Covert	Whole brain	1.506	0.207
	Motor (LH)	1.731	0.159
	Motor (RH)	1.709	0.163
	TVA (LH)	0.971	0.386
	TVA (RH)	1.233	0.285

4.2 Multivariate Results

The ROIs were used to select features from the data instances—the parameter estimates (PEs) from the speech runs. There were 12 instances (6 of class /pah/ and 6 of class /tah/) for each speech condition (overt and covert). The results from the multi-voxel pattern analysis (MVPA) will now be presented.

4.2.1 Correlation Classification

I first sought to demonstrate that the data corresponding to speech conditions /pah/ and /tah/ are separate classes in the extracted features. This was done by correlating within-class parameter estimates of the speech task runs and between-class parameter estimates from even and odd runs. Correlations were computed using Pearson's *r*. Meaning over subjects, I used single-sample *t*-tests to compute the significance of the mean correlations for each ROI. These results are displayed in Table 4. There is a clear pattern of more significance in within-class correlations than between-class. There is also an apparent difference in the strength of the correlations across speech condition—correlations for the overt condition are generally larger than those for the covert condition. This appears to be most noticeable for the within-class correlations for the syllable /tah/.

Table 4. Results from the correlation classification: within-class and between-class correlations averaged over subject. Significance computed with *t*-tests: *p<.05; **p<.01; ***p<.001

		Within-Class Correlations		Between-Class Correlations	
Condition	ROI	odd /pah/ even /pah/	odd /tah/ even /tah/	odd /pah/ even /tah/	odd /tah/ even /tah/
Overt	Whole brain	0.619 *	0.717 **	0.328 *	0.287 **
	Motor (LH)	0.677 *	0.818 ***	0.343 *	0.534 **
	Motor (RH)	0.652 *	0.670 **	0.445 ***	0.152
	TVA (LH)	0.721 **	0.785 ***	0.122	0.269
	TVA (RH)	0.624 **	0.756 ***	0.271 *	0.373 *
Covert	Whole brain	0.473 **	0.218	0.130	0.045
	Motor (LH)	0.533 *	0.561 *	0.399 *	0.315
	Motor (RH)	0.669 **	0.489 **	0.198	0.217
	TVA (LH)	0.426 **	-0.100	0.107	0.040
	TVA (RH)	0.524 *	0.169	0.195	-0.005

4.2.2 Linear SVM Results

I then trained and tested linear SVM classifiers on the same parameter estimates for the two syllables from each run. For each ROI, a leave-one-out cross-validation (CV) method of calculating the accuracy of the classifier was used. I performed permutation testing to compute a likelihood estimate of CV accuracy. In addition, binomial testing was used to obtain an alternate measure of likelihood, as the permutations testing is limited in the tails of the distribution with there being so few instances—12 in the current case, yielding less than 300 unique permutations. The CV results for the overt and covert conditions can be seen in Table 5. Comparing these results with those from the within-subject univariate results (Table 2), it is clear that while the univariate tests were able to detect differences between the conditions /pah/ and /tah/ when meaning over the voxels (features) in an ROI, a greater number of SVM accuracies were significant. This demonstrates the SVM approach was more *sensitive* to differences in syllable production.

Table 5. All cross-validation accuracies (as a percentage) computed per subject by ROI and speech condition.

_p_-values computed with permutation testing: *_p_<.05; **_p_<.01; ***_p_<.001

and _p_-values computed with a binomial test: +_p_<.05; ++_p_<.01; +++_p_<.001

Condition	Subject	Whole brain	Motor (LH)	Motor (RH)	TVA (LH)	TVA (RH)
Overt	S1	100.000 **,+++	91.667 *,++	100.000 **,+++	100.000 *,+++	100.000 *,+++
	S2	100.000 **,+++	100.000 **,+++	100.000 **,+++	100.000 *,+++	100.000 **,+++
	S3	66.667	58.333	83.333 *,+	58.333	75.000
	S4	91.667 ++	91.667 *,++	100.000 **,+++	91.667 *,++	100.000 **,+++
	S5	100.000 **,+++	91.667 *,++	100.000 **,+++	91.667 *,++	83.333 +
Covert	S1	50.000	66.667	66.667	66.667	58.333
	S2	83.333 +	100.000 **,+++	75.000	66.667	58.333
	S3	58.333	50.000	58.333	50.000	50.000
	S4	83.333 +	66.667	91.667 *,++	66.667	75.000
	S5	83.333 *,+	91.667 *,++	100.000 **,+++	83.333 *,+	75.000

As with the correlation results presented above, I computed the mean across subjects and used a single-sample *t*-test to compute whether they were significantly above chance

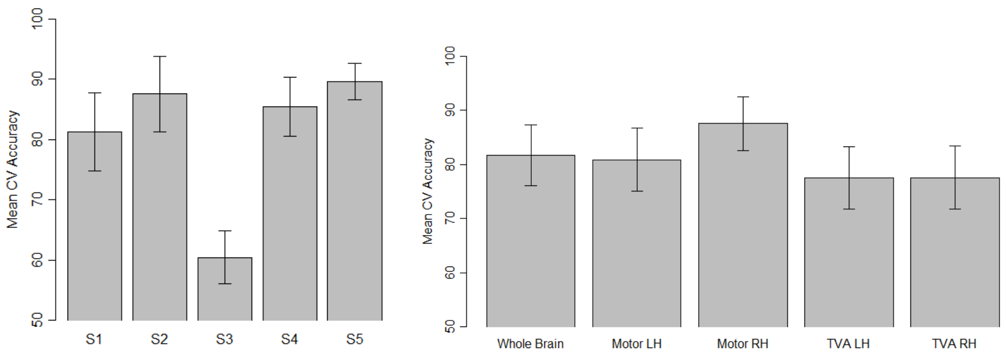
(Table 6). These results indicate that mean cross-validation accuracy was statistically significant using all ROIs (see Figure 8b), and across both speech conditions—contrasting sharping with the univariate group-level tests (Table 3), which were all nonsignificant (p > .05).

The computed *t*-statistics suggest higher accuracy for the overt condition (at least p < .01 for each ROI) than for the covert condition (p < .05 for each ROI). This was examined more formally in an analysis of variance (ANOVA), the results of which I will now discuss.

Table 6. Mean cross-validation accuracies (as a percentage) for each ROI from both overt and covert speech runs. Significance computed with *t*-tests: *p<.05; **p<.01; ***p<.001

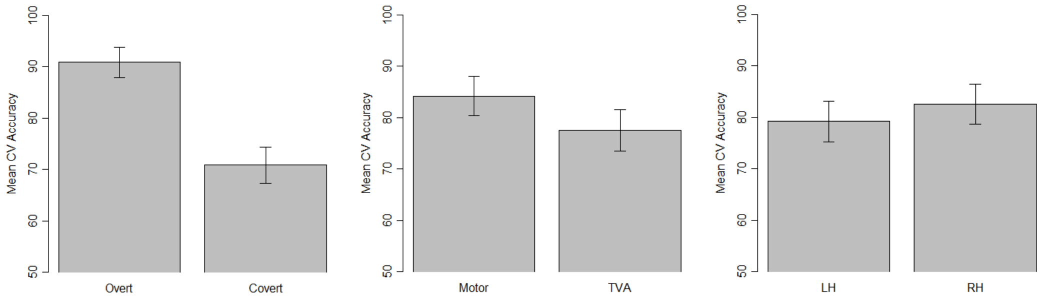
Condition	ROI	mean acc.
Overt	Whole brain	91.667 **
	Motor (LH)	86.667 **
	Motor (RH)	96.667 ***
	TVA (LH)	88.333 **
	TVA (RH)	91.667 **
Covert	Whole brain	71.667 *
	Motor (LH)	75.000 *
	Motor (RH)	78.333 *
	TVA (LH)	66.667 *
	TVA (RH)	63.333 *

To further investigate the SVM CV-accuracy results, I conducted a repeated measures ANOVA testing for effects due to three factors: *condition* (*Overt vs. Covert*); *region* (*Motor vs. TVA*); and *hemisphere* (*LH vs. RH*). The results indicated a main effect of *condition*, *F*(1,4) = 17.194, *p* = .0143; a main effect of *region* approached significance, *F*(1,4) = 7.111, *p* = .056; and a nonsignificant main effect of *hemisphere*, *F*(1,4) = .681, *p* = .456. The differences across the levels of each of these three factors can be seen graphically in Figure 9.



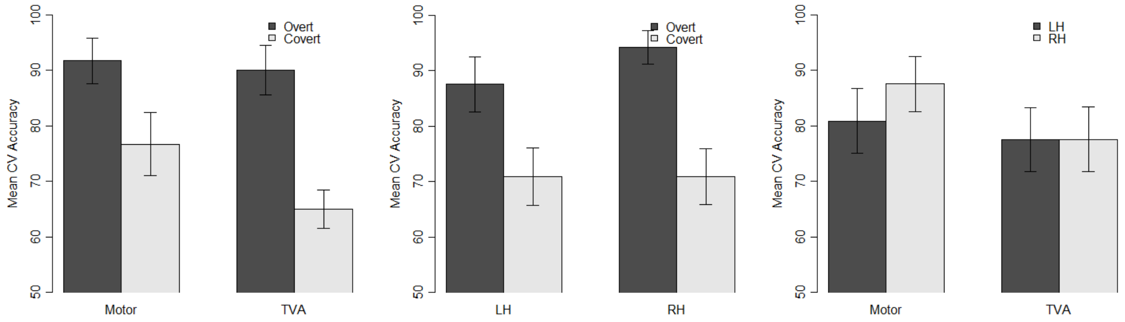
(a) (b)

Figure 8. Comparisons of mean cross-validation accuracies over subjects in (a), and each ROI in (b). ***Note*****: the *Whole Brain* ROI is not included in the means computed for each subject in (a).**



(a) (b) (c)

Figure 9. Comparisons of mean cross-validation accuracies for the factors (a) *condition*, (d) *region*, and (e) *hemisphere*.



(a) (b) (c)

Figure 10. Mean cross-validation accuracies of the factors (a) *region* and (b) *hemisphere*, grouped by *condition*, and (c) *region*, grouped by *hemisphere*.

There were no significant interactions for *condition* x *hemisphere* or *region* x *hemisphere*—*F*(1,4) = 1.882, *p* = .242 and *F*(1,4) = 2.667, *p* = .178, respectively. However, there was an interaction between *condition* and *region* which approached significance, *F*(1,4) = 6.000, *p* = .0705. These interactions can be seen graphically in Figure 10. Furthermore, the three-way interaction of *condition* x *region* x *hemisphere* was not significant, *F*(1,4) = 0.000, *p* = 1.000.

Also, in a separate ANOVA, it was found that there was no significant difference between the *Whole Brain* ROI and the other (*Motor* and *TVA*) ROIs, *F*(1,4) = .211, *p* = .670.

4.2.3 SVM Control Comparison Results

To serve as a control basis for comparison, I also trained and tested linear SVM classifiers using the same features and instances, but using classes orthogonal to the speech task—the color of the stimuli. Mean cross-validation accuracies across all subjects and ROIs were uniformly at chance-level (50% correct). Furthermore, identical permutation testing found—with the exception of one classifier—accuracies to be nonsignificant ($p > .05$).

5. Discussion

5.1 Summary

The objective of this study was to determine whether *speech production* and *speech perception* regions of the brain contain information that could be decoded in a covert speech brain-computer interface (BCI). I investigated this by scanning subjects with fMRI during overt and covert speech tasks in which they produced the syllables /pah/ and /tah/. Separate scans localized motor and perceptual regions that were used to subset data from the speech tasks. I then trained and tested multi-voxel pattern analysis (MVPA) classifiers in an attempt to decode which syllable was spoken using subset fMRI data from the speech tasks.

Specific Aim One was successfully addressed, demonstrated by results revealing that significantly accurate decoding of overt syllables was possible in all five subjects in at least one region; covert syllable decoding accuracy was significant in three of the five subjects, addressing Specific Aim Two. The significance of these accuracies was shown in both permutation and binomial tests. Permutation tests determine the likelihood of classifier accuracy by testing the null hypothesis that the class labels contain no information relevant to classification. An empirical probability distribution is computed by calculating the accuracies for the set of classifiers trained on all possible permutations of the class labels—/pah/ and /tah/ for the present study. The accuracy computed with the correct labels is then compared to this distribution, and an exact *p*-value is computed. However, the reliability of this method may be questionable when the number of permutations is small, and the distribution may be under-sampled (Golland & Fischl 2003)—particularly at tails of the distribution. As this study trained classifiers with a relatively small number of instances, the number of unique class permutations was also small (less than 300). I therefore used binomial tests of significance to provide an alternative measure. In this case, each individual classification was viewed as either a success or a failure, with the null hypothesis of equal chance of achieving either. Both of these methods indicated similar significance estimates for each of the classifiers trained at the within-subject level.

Group-level analyses indicated significant accuracy in all investigated ROIs as determined by using *t*-tests comparing accuracy to chance. In addressing Specific Aim Three, classification of overt syllables was found to be significantly more accurate than covert syllables, as demonstrated by differences in an analysis of variance (ANOVA). Interestingly, the ANOVA also suggests that there may be greater accuracy when using the motor versus perception regions—especially in the covert case.

These findings are significant in that they indicate that neural patterns of activity during covert and overt speech may be similar enough to approach the decoding of inner speech with methods proven to be successful in decoding overt speech. Importantly, speech motor and perception regions may encode sufficient detail about a person’s internal speech states to decode in a future implementation of a covert speech BCI. Furthermore, these results demonstrate the utility in using MVPA to map out regions to use in future BCIs based on decoding speech states.

5.2 Specific Points

5.2.1 Covert Speech Brain-Computer Interface

This study used multivariate pattern analysis methods to analyze fMRI data of subjects speaking two syllables overtly and covertly. Classifiers were trained using data from speech production and perception regions. Successful classification was possible in both overt and covert speech, and with data from each region.

Although accuracy was shown to be significantly above chance in each region and for both the overt and covert conditions, accuracy was still significantly lower in decoding covert speech. Furthermore, the classification process is computationally intensive, and not suitable as a real-time system. However, the purpose of this study was to demonstrate that motor and perception regions contain information that can be decoded in future BCI systems.

To translate the findings of the present study towards the development of an actual BCI, certain issues must be considered. First, decoding is accomplished either through *classification* or *reconstruction* of cognitive states. Classification methods use patterns of imaging data to predict cognitive states (classes)—such as production of different syllables, as per the present study. Importantly, the states must be both finite, discrete (Kay & Gallant 2009), and defined *a priori*. However, the cognitive states one wishes to decode may not be finite and discrete; in these cases, classification can only be applied through the artificial reduction of the complexity of the states.

An alternative to classification is to decode cognitive states through the reconstruction of the complex features of the cognitive state of interest (Kay et al. 2008). As described by Kay and Gallant (2009), an early study by Thirion et al. (2006) demonstrated this method—albeit with limited success—by reconstructing seen images from fMRI data. More recently, Miyawaki et al. (2008) were able to quite accurately reconstruct contrast-defined images seen by subjects in the scanner from patterns of BOLD data in early visual cortex (V1 – V3). Demonstrating the distinction between classification and reconstruction methods of decoding, the reconstructed "images" that Miyawaki et al. created were 10 x 10 matrices of gray-scale pixels. It is important to note that accurate reconstruction requires a good theoretical understanding and model of how neural signals (e.g., as indirectly-measured by fMRI) represent the cognitive states that are being decoded (Kay & Gallant 2009). In contrast, classification methods are less reliant on such models—but, selecting regions of the brain to be used as features can, and arguably *should* rely on theory (e.g., the motor and auditory models of covert speech underlying the ROI-selection in the present study).

Predicting covertly-spoken syllables lends itself nicely to the classification technique, especially when the number of syllables is small, as in the present study. However, this method may not scale well, and it may be awkward for a patient to use this method in an actual covert speech BCI. Another, more “natural” approach, is to extract continuous speech information associated with covert speech utterances—which would require reconstruction. Therefore, the reconstruction approach to decoding may be preferable in a future speech BCI.

Second, to be of beneficial clinical use, the decoding (classification or reconstruction) of covert speech with a BCI will need to be faster than the process employed in this study. The temporal resolution of fMRI is limited to a few seconds, due to the hemodynamic lag of the BOLD response. EEG is faster, but has poor spatial resolution (see below). Invasive techniques—in which recording devices, such as electrodes, are implanted into cortex—are likely to yield the best methods of recording neural information for a speech BCI to decode.

5.2.2 Overt versus Covert Speech

The present study demonstrated that statistical classifiers were able to much more accurately predict syllables spoken overtly than covertly. This was largely suspected, as covert speech has been shown to have weaker fMRI responses than overt speech (Shuster & Lemieux 2005). One possible explanation for why this signal is weaker is that since overt speech involves actual muscle movement, covert speech could simply be due to a weaker signal being sent to muscles. Evidence that the magnitude of the signal in motor cortex depends on the force of an action comes from studies in which the fMRI signal in motor areas correlates with the force of squeezing a ball (Thickbroom et al. 1998). Also, subthreshold EMG in muscles during covert speech is consistent with this theory—covert speech could be simply a “low-force” motor action—not enough force to produce overt articulation. However, it is unclear if “imagined” force can increase motor signal (like a more intense motor imagery). To test this, participants could be instructed to (overtly) speak a prompted phrase very softly and very loudly in an fMRI scanner; participants would then repeat the task, but with covert utterances. The results of Thickbroom et al. (1998) suggest that a greater motor signal in motor cortex would be observed during loud versus soft overt speech; if a similar pattern were to be observed during covert speech, it would provide evidence that “imagined” force of motor imagery modulates the associated motor activity in cortex.

Furthermore, that significant accuracies were also attained in decoding covert syllables indicates covert and overt speech are not orthogonal in their neural representations. This is important, as it suggests methods can be used to decode covert speech with a BCI using models derived from overt speech.

The present findings of significant above-chance accuracies could be made more persuasive by using methods to behaviorally validate subjects’ performance during the covert task—to ensure proper task completion. As mentioned earlier, subthreshold EMG can be measured during covert speech in the lips and tongue (Livesay et al. 1996; McGuigan & Dollins 1989). Therefore, recording EMG activity associated with the tongue or lips during the covert task could provide a measure of subject-compliance. However, use of EMG in the fMRI scanner, while possible, remains technically challenging (Sörös et al. 2010; Laufs et al. 2008).

Alternatively, as the utterance of covert syllables is a form of mental imagery and priming effects due to imagery have been demonstrated (Michelon & Zacks 2003), it may be possible to validate covert speech by *priming* for reaction to stimuli presented immediately after the covert tasks that are compatible or incompatible with the "uttered" syllable. For example, the following method may reasonably-address this issue for the speech task of the present study: As the syllables /pah/ and /tah/ lack meaningful semantic associations, the simple visual stimuli of either "PAH" or "TAH" could be presented, following trials of covert utterances; subjects would be instructed to press a button as soon as they see the stimulus. The reaction time should be faster for compatible stimuli. The use of this or a similar metric would allow for greater confidence that subjects are performing the task correctly—increasing the credibility of the findings.

5.2.3 Motor versus Auditory Approach

The results of the present study demonstrate that it is possible to decode speech states accurately using patterns of fMRI activity in motor areas involved in speech production *and* in lateral temporal areas implicated in auditory processing of speech. There is some suggestion that a motor approach to decoding speech is better, particularly in decoding covert syllables, although the difference only approached significance. Nevertheless, due to the small sample size of the study, this is an important finding as it suggests that future investigations for regions that contain information to decode convert speech might favor the motor approach.

Alternative explanations for the trend towards greater significance of decoding accuracy using the motor versus auditory approach should be noted. For example, noise in the scanner might contribute to lower decoding accuracy in the TVA ROIs; the sparse temporal scanning method—used in the TVA localizer task to reduce scanner interference with the auditory stimuli—was not employed in the speech tasks. Therefore, the scanner noise may have degraded auditory vocal stimuli (in the overt task) and auditory vocal imagery (in the covert task) sufficiently such that the signal could not be processed as well by the TVA. If this was the case, the accuracy differences could be explained in one of two ways: the TVA simply was unable in many cases to discriminate between sounds amidst the noise, while discrimination in the motor region was not affected; relatedly, due to the noise in the auditory signal, motor circuits active during speech perception (Pulvermüller et al. 2006) may be up-regulated in their sensitivity as a mechanism of accommodation.

Another alternative concerns the possibility that auditory regions, such as the TVA, are more sensitive to *vowel sound* differences in syllables. Both /pah/ and /tah/ have the same vowel—/a/. Perhaps the TVA would show greater discrimination between the syllables in which the articulated consonant is held constant, but the vowel is varied, as in /pah/ and /pih/. Evidence supporting this argument comes from a study by Formisano et al. (2008). Their study demonstrated that MVPA methods could accurately decode which of three vowels—/a/, /i/, and /u/—a subject heard. Vowels were presented as overt auditory stimuli from three speakers; decoding used voxel-based features in lateral temporal regions—overlapping the TVA ROIs used in the present study. The alternative that patterns of activity in auditory regions are more sensitive to the decoding of vowels—versus leading consonant articulation in simple CV (consonant-vowel) syllables—should be investigated in future research. The approach of the present study, or that used by Formisano et al. (2008), could be employed to compare the decoding of CV syllables in which both consonants and vowels vary as independent factors: e.g., /pah/, /tah/, /pih/, and /tih/. Furthermore, this should be examined for both overt and covert speech—to determine which approach is more likely to yield success in a covert speech BCI, as per the motivation of this study.

Finally, the nonsignificant trend of the present results suggesting greater decoding accuracy motor regions versus auditory regions might simply be an artifact of the small sample size (N = 5). Future studies employing the methods of the current study must use a greater number of subjects to address this alternative.

However, I re-emphasize that the present study demonstrated significantly above-chance decoding of syllables. Future research *is* needed to conclusively determine the relative differences of activation patterns between motor and auditory regions during covert speech—but both approaches are worth investigating in future speech BCI research, and the present study suggests that a motor-based approach to a covert speech BCI may be preferable.

5.2.4 Use of fMRI

This study demonstrated the successful use of fMRI as a means by which to collect data during speech and auditory tasks, and then apply pattern analysis techniques to classify task conditions (see below). Functional MRI was used for two primary reasons: First, it is a well-established and non-invasive neuroimaging technique for measuring neural activity. Invasive techniques require surgery, and carry with them the risk of infection due to the surgery, in addition to immune system responses to the implants after the surgery. Therefore, non-invasive techniques, particularly when trying to develop methods to use in BCIs, are quite valuable.

Secondly, fMRI has a high spatial resolution, necessary for the statistical classifiers to discriminate between the spatial patterns of activity associated with each of the two syllables. In the present study, voxels were of the dimensions 3.75 x 3.75 x 5 mm^3—dependent upon the 3T field strength of the scanner. Use of high-field (7T) fMRI has been used recently (Hoffman et al. 2009), but although higher field strength does allow for smaller voxels (greater spatial resolution), more noise is introduced: The fMRI signal itself increases quadratically with field strength—but as thermal noise increases linearly with field strength, the raw signal-to-noise ratio increases only linearly. Furthermore, physiological noise (due to variation in respiratory and cardiac activity) increases quadratically with field strength. Therefore, the *functional* signal-to-noise ratio (SNR)—*a measure of great significance in fMRI studies*—will likely level-off with increasing field strength (Huettel 2009). However, methods do exist to obtain higher spatial resolution at lower field strengths. Most notably, the use of parallel imaging with multichannel receiver coils can greatly increase spatial resolution—a 16-channel coil at 3T can yield voxel dimensions of 1.1 x 1.1 x 1.5 mm^3 (de Zwart et al. 2006). Increasing spatial resolution by using high-field fMRI or parallel imaging should prove useful in future studies with similar aims.

The manner in which one would *translate* decoding methods demonstrated with fMRI *to a real BCI*, such as an invasive one, is dependent upon the relationship between the fMRI BOLD signal and neuronal activity. BOLD allows measurement of the change in blood-oxygenated level correlated with neuronal activity, and is primarily a local detail signal—it is more associated with blood-oxygenated level changes in the capillaries than in larger blood vessels (Huettel 2009).   However, it is still not fully clear what neuronal processes lead to the BOLD signal (e.g., action potentials), but there is much evidence now that local field potentials most strongly correlate with the BOLD signal (Logothetis et al.

2001). This is important to consider when creating an invasive BCI using a model developed with fMRI; specifically, it argues that electrodes should be implanted to measure local field potentials and not action potential spikes, if the decoding behavior of the invasive BCI is to be similar to the noninvasive fMRI model.

5.2.5 Multi-Voxel Pattern Analysis

I used the multi-voxel pattern analysis (MVPA) approach to analyze fMRI data from the speech tasks to predict which of two syllables a subject was overtly or covertly saying in the scanner.

MVPA is a multivariate fMRI analysis technique that takes advantage of the pattern of responses across voxels, rather than merely examining changes in signal over pooled-groups of voxels. The advantages of this approach are well-exemplified in one of the key early studies in the literature: Kamitani and Tong (2005) used MVPA in a study in which subjects were shown visual stimuli of eight orientations. They used a similar linear support vector machine (SVM) algorithm to that used in the present study to classify the orientation of the stimuli from patterns of voxels in early visual cortices. They compared classification results across retinotopically-defined ROIs (V1 – V4, and MT+), and found the highest decoding accuracy in V1 and V2. As a single voxel in early visual cortex contains many thousands of neurons, not all responding to the same stimulus orientation, such decoding is not possible at the level of a single voxel—and for the same reason, not by computing a mean signal over a large group of voxels. Instead, by looking at differential patterns of activation across many voxels, Kamitani and Tong (2005) demonstrated the value of MVPA in decoding information represented at the sub-voxel level.

In the present study, I used MVPA to demonstrate that information about covertly spoken syllables can be decoded by using the differential patterns of activation across many voxels. The voxels used as features were selected by motor and auditory ROIs—defined by univariate analyses on functional localizer runs. These features were then used to train and test classifiers on classes corresponding to the two syllables, /pah/ and /tah/.

It was possible to differentiate between the two syllables spoken overtly or covertly using a simple correlation-based method. However, as the predictive power of this method is limited, linear SVMs were used to train and test classifiers to predict which syllables were spoken, using the features. I used a leave-one-out cross-validation method to estimate the accuracy of each classifier. Permutation testing and binomial tests were employed to determine the significance of the cross-validation accuracies—there were few differences in the results of each of these tests, suggesting that either method may be suitable.

Individual cross-validation accuracies were largely significant, with group-level analyses indicating greater accuracy in the classification of overt speech, along with a suggested trend of greater accuracy using features from the motor region as compared to the auditory region in the classification of covert speech.

5.3 General Discussion

5.3.1 Temporal Resolution

In the present study, I performed classification analyses using parameter estimates derived from each of the six speech runs (per condition). This was done as an initial step to demonstrate the validity of the motor and auditory approaches, as each instance being classified was essentially a summary measure of four blocks of a *run*, with a temporal resolution of minutes. Other MVPA studies have also used instances derived in a similar manner. For example, Serences and Boynton (2007) used fMRI activations averaged over runs to decode which of two directions of attended visual motion subjects were shown in the scanner—eight instances per each of two classes (directions of motion).

An obvious next step is to use instances at the level of *blocks* from each run, and then single time-point acquisitions—3D volumes from *single TRs*. The advantage of increasing the temporal resolution of the classification analysis is obvious—it moves closer to demonstrating that real-time data can be used for successful decoding in a BCI. The primary disadvantage is that the signal-to-noise ratio will drop steeply, as fMRI data is inherently noisy. Nonetheless, MVPA performs well in situations with noisy data, as long as the differential patterns of activity are still present. Kamitani and Tong (2006) used instances of feature values averaged over 16-second blocks to decode the direction in which visual stimuli were moving—22 instances per each of eight classes (directions of motion). As another example, Formisano et al. (2008) used 60 instances derived from individual *trials* to decode which vowel-sounds were heard by subjects.

Taking the above points into consideration, applying the methods of the present study to higher-resolution data is therefore strongly suggested.

5.3.2 Alternative Methods of Data Collection

Another approach would be to use electroencephalography (EEG) instead of fMRI to record patterns of neural activity. The advantage of EEG is that it has a much higher temporal resolution—on the order of milliseconds; the disadvantage is that the spatial resolution is quite poor. The spatial resolution is low, as EEG measures changes in electric potential on the scalp—and there is significant degradation of signal from the cortex as it passes through the skull and scalp. In addition, there is the infamous "inverse problem"—there are an infinite number of possible dipole configurations within the brain that will produce the same pattern of scalp potentials as measured by EEG. This can be addressed to some extent using convergent methods, such as cortical source density (Liu & He 2008), in which MRI techniques are used to reduce the number of possible solutions to the inverse problem. Still, EEG uses an array of scalp electrodes—32, 64, or 128 are often used—and apart from the inverse problem, as compared to fMRI, much larger populations of neurons are being measured. However, with non-invasive techniques, there remains a tradeoff between spatial and temporal density: high temporal resolution will be necessary for any real-time system, but the signal-to-noise ratio of relevant features drops off steeply as the spatial resolution declines. BCIs using EEG tend to rely on particular observed frequencies. For example, Pineda et al. (2003) trained subjects to alter mu-frequency activity differences between their hemispheres to control left or right movement in a three-dimensional video game. While this method demonstrates the ability to use EEG in a binary-choice BCI, the long-term goal of a covert speech BCI would need to be able to quickly decode (through classification or reconstruction as discussed above) much richer data.

5.3.3 Alternative Methods of Feature Selection

While the ROI-based method of feature selection used in the present study has shown it to be possible to decode covertly spoken syllables from motor and voice-related areas of cortex, other regions may have been neglected—other cortical areas may also contain relevant covert speech information. For example, there may be distinct areas involved in motor planning versus motor execution—including for speech. Desmurget et al. (2009) applied intracranial electrical stimulation to premotor and parietal regions of patients undergoing open-brain surgery. Interestingly, they found that stimulation of premotor regions caused overt mouth movement—yet the patients denied having moved. When the left inferior parietal region was stimulated, patients had the awareness of an intention to talk or move their lips—but there was no EMG activity recorded in the lips. This study demonstrated a double dissociation between speech movement intention/awareness and actual speech-related movement. If covert speech is a form of motor imagery, then it might include an aspect of motor planning as well. Whereas the results of the present thesis indicate that covert speech states can be decoded from motor areas, if this is the case, it may also be possible to decode covert speech *intention* from parietal areas.

In addition, using a multivariate feature selection method, such as Recursive Feature Elimination (De Martino et al. 2008) might shed light on which voxels contain the most pattern information. Recursive Feature Elimination (RFE) starts by including all voxels as features to train a classifier. A subset of voxels is removed, and the difference in classification accuracy is used as a measure of the predictive power of the removed voxels. As the name suggests, this is a recursive iterative process; while computationally intensive, it may be possible to use this method to empirically determine the voxels most-discriminative to the classification syllables or other speech-states. A similar technique is the multivariate searchlight technique (Kriegskorte et al. 2006) which assumes discriminative pattern information is contained in a small group of neighboring voxels within a "searchlight" distance. In this method, an entire volume is searched with this "searchlight," running the classification analyses at each step, and weighting the predictive accuracy of each neighborhood of voxels; those areas with the highest relative predictive accuracy are selected as the "best" features. To accurately distinguish between greater numbers of covert speech states—as will be necessary in a classification-based covert speech BCI—the problem of low contrast-to-noise will become more apparent, and these multivariate feature selection methods can be quite useful in low contrast-to-noise situations (De Martino et al. 2008).

5.3.4 Individual Differences in Speech – Subject S3

Classifier accuracies—averaged across condition, hemisphere, and region—did not vary significantly between subjects, with the notable exception of subject S3. One possible explanation of this observed difference in S3 is the following: The mapping from patterns of cortical activity to differing motor articulations may not be distinct enough to discriminate in some individuals—at least at the spatial and temporal resolution of the data in the present study.

It is also possible that S3 may have performed the tasks differently than the other subjects: Some subjects needed more training than others to be able to move their lips without also significantly moving their jaws. Thus, it is possible that the motor localizer task was not optimal, and voxels were included in the motor ROIs for some subjects that were too-dependent upon jaw movement. Use of an alternate paradigm, such as the pursing of the lips (Lotze et al. 2000), could be used to compare motor localizer results to address this issue. As for the auditory localizer, it is likely that variation in attentiveness and wakefulness across subjects would affect the results. Furthermore, if subjects attended more or less to differing features of the auditory stimuli, this could likewise be significant.

Individuals may also perform covert speech tasks differently—employing different types of *imagery*: For some, "covert speech" may be more like auditory verbal imaging; for others, it may be more similar to motor imagery (imagining doing the task). Some may be engaging in something more similar to subvocal speech (which would yield greater motor cortex activity). Individuals may differ in their ability to perform mental imagery tasks, and there is evidence that a subject's subjective measure of the "vividness" of imagery correlates with the observed fMRI signal in imagery tasks (Cui et al. 2007).

Given the small sample size, it is not possible to state how significant individual differences would be in a larger sample. However, it is unlikely that the exact same methods will yield significant decoding accuracies in all individuals—therefore, the above issues are important to consider in the design of future speech BCIs.

5.3.5 Generalizing Covert Speech Decoding of Controls to that of Locked-in Patients

As mentioned earlier, covert speech has been shown to elicit subthreshold EMG activity in speech articulators (Livesay et al. 1996). As the present study used normal, healthy individuals as subjects, it may not be fully valid to generalize the results regarding covert speech decoding in controls to that in patient populations—the covert speech of locked-in patients (or others with similar motor paralysis) is truly "covert" in the sense that motor articulation is impossible. There are also some reasons to question whether the covert speech mechanisms of locked-in patients are similar to those of controls. For example, it might be the case that actual peripheral articulators are required for the normal functioning of motor cortex. In this case, generalizations of the present study's findings to the covert speech of locked-in patients would be tenuous. Evidence to counter this model comes from the demonstration that motor imagery (e.g., imagining oneself playing tennis) in patients with a presentation similar to locked-in syndrome produces similar activation to that of controls (Owen & Coleman 2008). Alternatively, a signal may be properly generated by the motor cortex, but not propagate further than the corticospinal tracts. This is certainly an issue that should be addressed in future research, perhaps in using both control subjects and patients in a similar speech or motor imagery-decoding study. Furthermore, receiving feedback from patients and ensuring that they are performing the tasks properly is a significant hurdle that may be difficult to overcome. It may be possible to—at least partly—address this issue using methods of covert behavior validation similar to those mentioned above in the case of controls during covert speech tasks.

5.3.6 Advantage of the MVPA Technique

I end this general discussion by drawing attention to the benefit of the demonstrated application of MVPA of the present study in aiding the design of future BCIs: Previous studies demonstrated greater BOLD amplitude of overt compared to covert speech in the areas used as the motor and auditory ROIs in the present study (e.g., Shuster and Lemieux 2005). However, the current results do not simply indicate that motor and middle temporal regions are involved in covert speech—more importantly, the patterns of activity in these regions may contain *sufficient information* to encode the contents of covertly-spoken syllables. Taking this together with findings that motor regions may also be *necessary* to produce covert speech (Aziz-Zadeh et al. 2005), a strong case can be made for the role of the regions investigated in the present study in covert speech. Importantly, this argues that a BCI could accurately decode some covert speech states from these regions. Future studies should thus investigate whether more complex covert speech states can be decoded from patterns in these regions.

5.4 Concluding Remarks

In conclusion, the findings of this study indicate that neural patterns of activity during overt and covert speech may be similar enough to apply overt speech models to methods of decoding inner speech—although motor regions may be more similar than auditory regions. Importantly, speech motor and perception regions may encode sufficient detail about a person's internal speech states to decode in a future implementation of a covert speech BCI. Therefore, future investigations should focus on these regions in the design of such BCIs. Furthermore, the results demonstrate the utility in using MVPA to map out regions to use in future BCIs based on decoding cognitive states.

It is hoped that this present thesis study has convincingly demonstrated directions for future studies sharing the goal of creating a BCI for those who cannot speak—despite normal cortical functioning, such as the locked-in patient.

REFERENCES

REFERENCES

Ackermann, H., Mathiak, K., & Ivry, R. B. (2004). Temporal Organization of "Internal Speech" As a Basis for Cerebellar Modulation of Cognitive Functions. *Behavioral and Cognitive Neuroscience Reviews, 3*(1), 14-22.

Aziz-Zadeh, L., Cattaneo, L., Rochat, M., & Rizzolatti, G. (2005). Covert Speech Arrest Induced by rTMS over Both Motor and Nonmotor Left Hemisphere Frontal Sites. *Journal of Cognitive Neuroscience*, *17*(6), 928-938.

Bauby, J. (1998). *The diving bell and the butterfly*. Random House, Inc.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, *13*(1), 17-26.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309-312.

Ben-Hur, A., Ong, C. S., Sonnenburg, S., Schölkopf, B., & Rätsch, G. (2008). Support Vector Machines and Kernels for Computational Biology. *PLoS Computational Biology*, *4*(10), e1000173.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Cholin, J., Schiller, N. O., & Levelt, W. J. M. (2004). The preparation of syllables in speech production. *Journal of Memory and Language*, *50*(1), 47-61.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*, *19*(2), 261–270.

Cui, X., Jeter, C. B., Yang, D., Montague, P. R., & Eagleman, D. M. (2007). Vividness of mental imagery: Individual variability can be measured objectively. *Vision Research*, *47*(4), 474-478.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The Motor Somatotopy of Speech Perception. *Current Biology, 19*(5), 381-385.

De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., & Formisano, E. (2008). Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *NeuroImage*, *43*(1), 44-58.

Desmurget, M., Reilly, K. T., Richard, N., Szathmari, A., Mottolese, C., & Sirigu, A. (2009). Movement Intention After Parietal Cortex Stimulation in Humans. *Science*, *324*(5928), 811-813.

Dogil, G., Ackermann, H., Grodd, W., Haider, H., Kamp, H., Mayer, J., Riecker, A., et al. (2002). The speaking brain: a tutorial introduction to fMRI experiments in the production of speech, prosody and syntax. *Journal of Neurolinguistics*, *15*(1), 59-90.

Etzel, J. A., Gazzola, V., & Keysers, C. (2009). An introduction to anatomical ROI-based fMRI classification analysis. *Brain Research*, *1282*, 114-125.

Fenton, A., & Alpert, S. (2008). Extending Our View on Using BCIs for Locked-in Syndrome. *Neuroethics, 1*(2), 119-132.

Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). "Who" Is Saying "What"? Brain-Based Decoding of Human Voice and Speech. *Science*, *322*(5903), 970-973.

Gaab, N., Gabrieli, J. D., & Glover, G. H. (2007). Assessing the influence of scanner background noise on auditory processing. II. An fMRI study comparing auditory processing in the absence and presence of recorded scanner noise using a sparse design. *Human Brain Mapping*, *28*(8), 721-732.

Golland, P., & Fischl, B. (2003). Permutation Tests for Classification: Towards Statistical Significance in Image-Based Studies. In *Information Processing in Medical Imaging* (pp. 330-341).

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., Gurney, E. M., et al. (1999). "Sparse" temporal sampling in auditory fMRI. *Human Brain Mapping*, *7*(3), 213-223.

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632-635.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, *293*(5539), 2425-2430.

Hesselmann, V., Sorger, B., Lasek, K., Guntinas-Lichius, O., Krug, B., Sturm, V., Goebel, R., et al. (2004). Discriminating the Cortical Representation Sites of Tongue and Lip Movement by Functional MRI. *Brain Topography*, *16*(3), 159-167.

Hoffmann, M. B., Stadler, J., Kanowski, M., & Speck, O. (2009). Retinotopic mapping of the human visual cortex at a magnetic field strength of 7 T. *Clinical Neurophysiology*, *120*(1), 108-116.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), 106-154.2.

Huettel, S. A., Song, A. W., & McCarthy, G. (2009). *Functional magnetic resonance imaging*. Sinauer Associates.

Jeannerod, M., & Frak, V. (1999). Mental imaging of motor activity in humans. *Current Opinion in Neurobiology*, *9*(6), 735-739.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience, 8*(5), 679-685.

Kamitani, Y., & Tong, F. (2006). Decoding Seen and Attended Motion Directions from Activity in the Human Visual Cortex. *Current Biology*, *16*(11), 1096-1102.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *Journal of Neuroscience*, *17*(11), 4302-4311.

Kay, K. N., & Gallant, J. L. (2009). I can see what you see. *Nature Neuroscience*, *12*(3), 245.

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, *452*(7185), 352-355.

Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Maljkovic, V., Weise, S. B., Chabris, C. F., Hamilton, S. E., et al. (1993). Visual Mental Imagery Activates Topographically Organized Visual Cortex: PET Investigations. *Journal of Cognitive Neuroscience*, *5*(3), 263-287.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540.

Kriegeskorte, N., Formisano, E., Sorger, B., & Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences*, *104*(51), 20600-20605.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, *103*(10), 3863-3868.

Laufs, H., Daunizeau, J., Carmichael, D., & Kleinschmidt, A. (2008). Recent advances in recording electrophysiological data simultaneously with magnetic resonance imaging. *NeuroImage*, *40*(2), 515-528.

Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*(1-3), 239-269.

Levelt, W. J. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*(1-3), 1-22.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461.

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, *4*(5), 187-196.

Liu, Z., & He, B. (2008). fMRI-EEG integrated cortical source imaging by use of time-variant spatial constraints. *NeuroImage*, *39*(3), 1198-1214.

Livesay, J., Liebke, A., Samaras, M., & Stanley, A. (1996). Covert speech behavior during a silent language recitation task. *Perceptual and Motor Skills*, *83*(3 Pt 2), 1355-1362.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*(6843), 150-157.

Lotze, M., Seggewies, G., Erb, M., Grodd, W., & Birbaumer, N. (2000). The representation of articulation in the primary sensorimotor cortex. *NeuroReport*, *11*(13), 2985-2989.

McGuigan, F., & Dollins, A. (1989). Patterns of covert speech behavior and phonetic coding. *Integrative Psychological and Behavioral Science*, *24*(1), 19-26.

Michelon, P., & Zacks, J. (2003). What is primed in priming from imagery? *Psychological Research*, *67*(2), 71-79.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting Human Brain Activity Associated with the Meanings of Nouns. *Science*, *320*(5880), 1191-1195.

Mitchell, T. M., Hutchinson, R., Niculescu, R. S., Pereira, F., Wang, X., Just, M., & Newman, S. (2004). Learning to Decode Cognitive States from Brain Images. *Machine Learning*, *57*(1), 145-175.

Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M., Morito, Y., Tanabe, H., Sadato, N., et al. (2008). Visual Image Reconstruction from Human Brain Activity using a Combination of Multiscale Local Image Decoders. *Neuron*, *60*(5), 915-929.

Monti, M. M., Vanhaudenhuyse, A., Coleman, M. R., Boly, M., Pickard, J. D., Tshibanda, L., Owen, A. M., et al. (2010). Willful modulation of brain activity in disorders of consciousness. *New England Journal of Medicine*, *362*(7), 579.

Owen, A. M., & Coleman, M. R. (2008). Functional neuroimaging of the vegetative state. *Nature Reviews Neuroscience*, *9*(3), 235-243.

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage*, *45*(1S1), 199–209.

Pineda, J. A., Silverman, D. S., Vankov, A., & Hestenes, J. (2003). Learning to control brain rhythms: making a brain-computer interface possible. *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, *11*(2), 181-184.

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences, 103*(20), 7865-7870.

Scully, C. (1987). Linguistic units and units of speech production. *Speech Communication*, *6*(2), 77-142.

Serences, J. T., & Boynton, G. M. (2007). Feature-Based Attentional Modulations in the Absence of Direct Visual Stimulation. *Neuron*, *55*(2), 301-312.

Shergill, S. S., Bullmore, E. T., Brammer, M. J., Williams, S. C. R., Murray, R. M., & McGuire, P. K. (2001). A Functional Study of Auditory Verbal Imagery. *Psychological Medicine*, *31*(02), 241-253.

Shuster, L. I., & Lemieux, S. K. (2005). An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain and Language*, *93*(1), 20-31.

Smith, E., & Delargy, M. (2005). Locked-in syndrome. *BMJ, 330*(7488), 406-409.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping, 17*(3), 143-155.

Sörös, P., Macintosh, B. J., Tam, F., & Graham, S. J. (2010). fMRI-Compatible Registration of Jaw Movements Using a Fiber-Optic Bend Sensor. *Frontiers in Human Neuroscience*, *4*, 24.

Theodoridis, S., & Koutroumbas, K. (2006). *Pattern recognition*. Academic Press.

Thickbroom, G. W., Phillips, B. A., Morris, I., Byrnes, M. L., & Mastaglia, F. L. (1998). Isometric force-related activity in sensorimotor cortex measured with functional MRI. *Experimental Brain Research*, *121*(1), 59-64.

Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J. B., Lebihan, D., & Dehaene, S. (2006). Inverse retinotopy: inferring the visual content of images from brain activation patterns. *NeuroImage, 33*(4), 1104–1116.

Wilshire, C. E., & Nespoulous, J. (2003). Syllables as units in speech production: Data from aphasia. *Brain and Language*, *84*(3), 424-447.

de Zwart, J. A., Gelderen, P. V., Golay, X., Ikonomidou, V. N., & Duyn, J. H. (2006). Accelerated parallel imaging for functional imaging of the human brain. *NMR in Biomedicine*, *19*(3), 342-351.

CURRICULUM VITAE

Devin M. McCorry graduated from Centreville High School, Clifton, Virginia, in 2002. He received his Bachelor of Science in Computer Science from the University of Michigan, Ann Arbor, in 2006. He was employed as a software engineer for two years at Science Applications International Corporation in Chantilly, Virginia. He received his Master of Arts in Psychology from George Mason University in 2010.