Investigation of Speech for Communicative Brain-Computer Interface

BY

EMILY M. MUGLER
B.S., Duke University, 2006

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Bioengineering
in the Graduate College of the
University of Illinois at Chicago, 2013

Chicago, Illinois

Defense Committee:

James L. Patton, Ph.D., Chair and Advisor
Patrick J. Rousche, Ph. D.
John R. Hetling, Ph.D.
Thomas J. Royston, Ph.D.
Marc W. Slutzky, M.D., Ph.D., Northwestern University

This thesis is dedicated to my parents, Karla Mugler, Ph.D. in Counselor Education, and Dale Mugler, Ph.D. in Applied Mathematics, for being my first and forever educators, my bastions of encouragement, and my role models.

Acknowledgements

I would like to thank my thesis committee – Dr. Patrick Rousche, Dr. John Hetling, Dr. Thomas Royston, Dr. Marc Slutzky and Dr. James Patton – and for their support, guidance and assistance. I would also like to thank Dr. Matt Goldrick for his collaboration and counseling on the science of speech production. I further acknowledge and thank fellow lab members Dr. Felix Huang and doctoral candidate Moria Fisher, who not only helped edit this document but also participated in valuable discussions that elicited creative analytical approaches.

Many people provided critical assistance during the electrocorticographic data collection, and I would like to thank them as well: Eric Lindberg, Zachary Wright, Dr. Robert Flint, the nursing and technician staff at the Epilepsy Monitoring Unit at Northwestern Memorial Hospital, and Dean Krusienski at Old Dominion University.

I also would like to acknowledge the financial support I received throughout my doctoral work from National Science Foundation Grants Number 0549489 and Number 0718558 and the Department of Bioengineering at UIC.

Most of all, I would like to thank the individuals who donated their time and effort by volunteering for my studies, for without them, none of this research would be possible.

EMM

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ALS	Amyotrophic Lateral Sclerosis
ASR	Automatic Speech Recognition
BCI	Brain Computer Interface
CLIS	Completely Locked-In Syndrome
ECoG	Electrocorticography
EEG	Electroencephalography
EMG	Electromyography
Hz	Hertz
FFT	Fast Fourier Transform
fMRI	Functional Magnetic Resonance Imaging
ITR	Information Transfer Rate
LDA	Linear Discriminant Analysis
LIS	Locked-in Syndrome
M1	Primary Motor Cortex
MEA	Microelectrode Array
MRT	Modified Rhyme Test
PCA	Principal Component Analysis
QoL	Quality of Life
ROC	Receiver Operating Characteristic
SCP	Slow Cortical Potential
STFT	Short Time Fourier Transform
STG	Superior temporal Gyrus
SSVEP	Steady State Visual Evoked Potential
SVM	Support Vector Machines
wpm	Words per Minute

ABSTRACT

Recent successes in decoding speech from cortical signals provide hope for restoring function to those who have lost the ability to speak normally. Despite these successes, the exact cortical representation and functional dynamics of speech production remain unknown. Prominent theoretical models of speech production in the literature differ in their hypothesized functional organization of speech motor cortex. Using electrocorticography, with its fine spatial and temporal resolution, we can analyze the exact spatial and temporal cortical dynamics related to complex speech mechanisms.

This dissertation addresses various unknowns in the current speech brain-computer interface literature and recommends a methodology for successful speech classification from electrocorticographic electrodes. Addressing the current limitations and barriers to widespread BCI adoption, I here seek to add to the engineering merit of the communicative BCI field with the mechanistic analysis and results of three separate studies. In the first study, I seek to determine what factors contribute to successful phonemic decoding of an ECoG signal. In the second study, I seek to determine cortical representation of phonemic categorization in speech production. In the third study, I leverage classification results to address the structure of cortical correlates of speech production. The result of these studies outlines a set of guidelines for future speech-BCI research that will work towards useful speech-BCI neuroprothetics.

I. INTRODUCTION

1. Communicative Brain Computer Interface

1.1. Defining Communicative Brain-Computer Interface

Brain-computer interface (BCI) technology, in which brain signals are recorded and connected to control computers, electronics, machines or other external devices, was originally developed with the long-term goal of assisting patients in communication (Farwell and Donchin 1988; Kübler et al. 1999). Researchers who had studied human biosignals that demonstrated voluntary changes by the individual then applied these signals to control external devices. An obvious and useful application of this technology, then, was to provide voluntary control of communicative devices for individuals who were no longer able to communicate by traditional means. Individuals with amyotrophic lateral sclerosis (ALS) or other motor neuron diseases who had lost typical innervation of their speech motor faculties therefore became some of the first target populations for BCI technology (Rowland and Shneider 2001; Bach JR. 1994).

Because there are many methods for ascertaining a neurally-controlled signal from the human body, a wide variety of BCIs exists. A BCI, at its most basic level, is comprised of a sensor or electrode that records a neurally-controlled signal and an algorithm that interprets the recorded signal. Sensors can be *electric*, sensing a change in the electric potential of the brain, *magnetic*, sensing the changes in the dipoles or magnetic potentials of the brain, or *optical*, sensing a change in the color spectrum through the skin of blood flow to the brain. They can be *synchronous*, in which an individual must rely on the timing of an external process to control an interface, or *asynchronous*, in which volitional biosignal control directly executes external commands. BCIs can be *invasive*, in which electrodes may directly penetrate neural tissue, or *non-invasive*, in which sensors sit on top of the body and do not penetrate tissue. They can be *closed-loop*, in which an individual's neural feedback is immediately reported to the person, or *open-loop*, in which an individual does not get to observe the execution of the process.

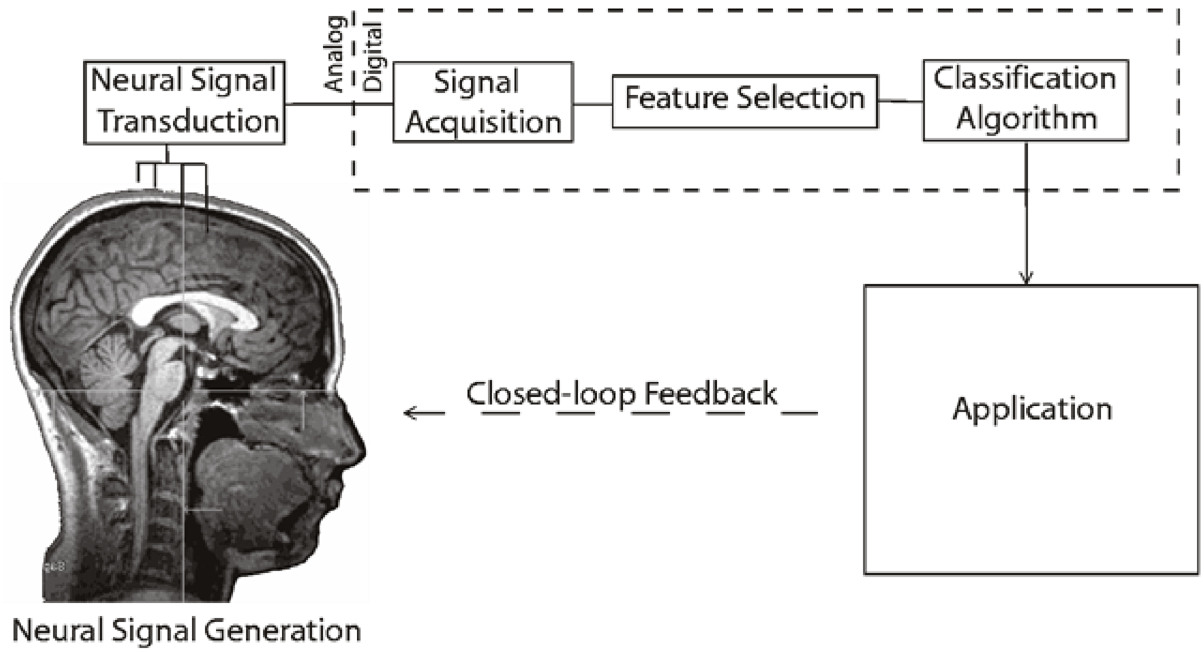


Figure 1. Generalized block-diagram of Brain-Computer Interface

Despite these differences in modality of recording neural signal, here I define *communicative brain-computer interface* as a BCI specifically used for communication purposes. This categorization therefore applies to the specific application of the BCI and includes all aforementioned BCI modalities, which are surveyed in the next section.

1.2.Survey of Communicative BCI in the Literature

Although much of the electrophysiology research of the past 50 years focuses on neural recordings from primates, cats, and other animal models, research into communicative BCI requires human subjects, and with that distinction, specific constraints. The majority of the human research performed to date has been primarily non-invasive, which then dictates the type of neural control system that can be used. Here, I detail BCI systems that have isolated neural control of communication by means of cognitive selection of presented stimuli, which comprises the vast majority of communicative BCI in the literature.

Electroencephalography (EEG), which measures the change in electric potential at the scalp, was employed as the sensor technology in some of the first communicative BCI tasks. These tasks primarily used the “slow cortical potential” (SCP), essentially comprised of a slow drift in the DC voltage recorded at a single electrode. The SCP is thought to be related to abstract, higher-level cognition (Kübler et al. 2001). In SCP-controlled closed-loop BCIs, individuals can achieve better performance after several sessions of training, suggesting some changes in cortical dynamics related to SCP control are elicited from routine use. The presence or absence of the SCP is the principle variable in selection control, so communicative responses are limited to either yes/no paradigms or a system of other binary formulae (Blankertz et al. 2003). Such signals can be used in conjunction with a binary forced-choice basket paradigm to select characters for typing messages (Bensch et al. 2007; Mellinger et al. 2003; Kübler et al. 2001).

Contrastingly, other EEG-recorded synchronous paradigms can be designed to elicit “event-related potentials” or ERPs as a control signal. The most common ERP approach utilizes an easily identifiable neural response to stimuli that are different from the norm. This response to so-called “oddball” stimuli is robust and easily reproducible in comparison with other ERPs. The prominent variant of this technique utilizes the “P300” signal, in which 300 milliseconds following presentation of an oddball stimulus, a robust control signal is elicited (Pritchard 1981; Polich and Kok 1995). Creative applications of this P300 ERP have been developed for communicative applications, most notably the P300 Speller approach (Farwell and Donchin 1988). The P300 speller organization provides a fixed set of characters and flashes them in order to elicit an oddball response for the intended character. This organization has since been widely optimized and adapted to a variety of communicative applications (Emily M Mugler et al. 2010;

E. Mugler et al. 2008; Sellers et al. 2006; Hoffmann et al. 2008; Sellers and Donchin 2006; Nijboer et al. 2008; Furdea et al. 2009).

Another technique that employs EEG electrodes, typically those over the scalp above visual cortex, is the steady-state visual evoked potential (SSVEP). This synchronous paradigm relies on various visual stimuli flashing on a screen at different prime frequencies (Cheng et al. 2002). The selected stimuli can then be deduced from harmonic frequencies present in the EEG signal. In this manner, an individual can effectively select and type letters or commands from presented options in an efficient paradigm (Bin et al. 2011). This signal is so robust that it can require fewer electrodes on the scalp to get a reliable response (Y. Wang et al. 2004). For SSVEP and P300 techniques, visual attention is required, so some residual eye movement must still be present in the BCI user for sufficient control.

EEG paradigms targeting changes in sensorimotor rhythms related to motor imagery can aid potential BCI users lacking voluntary eye movement. For individuals with late-stage ALS, who become functionally "locked-in" and are no longer able to move, even simple eye movement can degrade or altogether vanish. This is termed completely locked-in syndrome (CLIS), and these individuals cannot reliably control BCIs that incorporate even minimal eye movement. For these individuals, who are arguably the intended target population of communicative BCI, BCIs employing motor imagery remain a viable BCI paradigm. In motor imagery BCIs, an individual imagines moving some part of his or her body. When this occurs, there is a decrease in a specific frequency power band (8-13 Hz, termed the mu rhythm) in EEG channels over the motor cortex, which can then be used as a binary control signal to external communicative devices (Kübler, Nijboer, and Mellinger 2005; Birbaumer and Cohen 2007; Wolpaw et al. 2003).

Another non-invasive measure of recording brain activity is functional magnetic resonance imaging (fMRI), in which a person is stationary in a large magnetic field, and with brief pulses to disturb the magnetic field, changes of blood flow to specific parts of the brain over time can be tracked. This changing signal, called the blood-oxygen-level-dependent (BOLD) contrast signal, has a very slow time-course, often with peak reaction time occurring 7 to 8 seconds behind the selected stimulus, but it can reveal volitional changes in attention. There have been few attempts at creating a communication interface using this technology (Sorger et al. 2012), perhaps this BCI technology can only be used in a stationary, hospital setting and is therefore harder to test with locked-in individuals.

Functional Near-Infrared Spectroscopy (fNIRS or NIRS), considered the portable version of the fMRI, measures the BOLD contrast signal, but it does this through the surface of the scalp where hair is not present. Much like pulse oxymetery, changes in the oxygenation reveal where metabolic activity occurs near the surface of the cortex. This enables asynchronous binary command control (Coyle, Ward, and Markham 2007; Jackson et al. 2013; Herff et al. 2013).

Each BCI recording paradigm or technology has its own advantages and disadvantages. I highlight the disadvantages that hinder widespread communicative BCI adoption in the following section.

1.3. Barriers to wide adoption of BCI technology

The need for communication in the target population of CLIS individuals is crucial. For people with ALS, they rate their Quality of Life (QoL) as high if communication can be preserved (Kübler, Nijboer, and Mellinger 2005). This factor drives a need for sufficient communication. Even a slow communication channel, if reliable, can improve quality of life. However, a minimum requirement for reliability, when tested in BCI users, was 70% (Kübler et al. 2001; Kübler and Neumann 2004). For the BCI paradigms that satisfy the requirement of 67% or better correct performance, further issues of longevity of control may impede adoption. Many BCI paradigms require intense focus and concentration on the presented stimuli. These paradigms require vigilance of attention that can be exhausting for BCI users, although it is important to note that not all patients with ALS may consider this vigilance to be tiresome and tedious (Blain-Moraes et al. 2012).

Potential imposition for caregivers is another reason impeding BCI adoption for some patients. For EEG and fNIRS recording devices, the experimental set-up of securing electrodes to the scalp, which must be performed by an aide, can take as long as 15 minutes and can be taxing to the individual. Further, the majority of EEG systems require gel to be inserted through electrodes on to the scalp, which must be cleaned at the conclusion of the experiment, which may add to the imposition of the BCI user or patient caregiver. Most critically, although roughly two-thirds of surveyed ALS patients would not want to receive an implanted BCI device, 41% of surveyed ALS patients would appreciate a wireless, implanted neural interface to reduce burden and overhead to their caretakers (Huggins, Wren, and Gruis 2011; Blain-Moraes et al. 2012). This suggests that an automated, implantable device could reduce the day-to-day overhead for BCI users despite the risk and temporary hardships associated with medical device implantation.

BCI adoption by healthy individuals for control of external devices or silent communication is only beginning to reach the commercial market. This is likely due to the tediousness of applying many of these sensors or electrodes to the body to ascertain a reliable signal, and some research groups are only now developing “dry” electrode systems that do not require this labor-intensive set-up (Popescu et al. 2007; Zander et al. 2011; Grozea, Voinescu, and Fazli 2011)). Communicating silently carries the novelty of communicating “just by thinking,” but also may have practical applications for covert military operations or when privacy is paramount (Denby et al. 2009). Such devices could also supplement normal human communication in environments with high ambient noise (B. J. Betts, Binsted, and Jorgensen 2006). If these silent communication devices can sufficiently augment healthy communication, there could be more widespread adoption among healthy individuals. Furthermore, there seem to be some subsets of roughly 20% of the healthy population that cannot seem to develop control over BCIs, a phenomenon that remains poorly understood (Vidaurre and Blankertz 2010).

Finally, another issue that impedes widespread BCI use is the inefficiency of many devices. Most late-stage ALS patients have at least some communication or volitional switch they can control (L. J. Ball, Beukelman, and Pattee 2004). Others may have small joysticks or sip-and-puff devices that can control external communication tools. Most eye-tracking software used for communication purposes can function at a speed of 10 to 20 words per minute (Ward, Blackwell, and MacKay 2000). If a person maintains reliable control of eye movement, BCI technology – in which typical speeds are an average of one-third as efficient – presents no clear advantages to the user. This efficiency issue ultimately contributes to a lack of communicative BCI adoption.

I thus assert two criteria for BCI technology to be adopted by target populations:

(1) Progress can only be made in typing or communication when successful communication is 67% or higher. This refinement of the Kübler criterion results logically from the fact that in order to make progress when typing, there must be at least 2 correct character selections for each incorrect character to achieve progress.

(2) A BCI must perform at or better than the speed and efficiency of alternative technologies available to the individual.

Because such a variety of BCIs exists, there have been many efforts to quantify the differences between differing systems and algorithms. In the following section I will discuss the most prominent comparative measure of BCI efficiency: the rate of information transfer. This comparative metric will enable speed and efficiency comparisons between communicative modalities that will help identify BCI success.

2. Information Transfer Rate of Communication

2.1. ITR in the communicative BCI literature

Information Transfer Rate, or ITR, is a common metric used to determine the speed and efficiency of communication between two parties in a single channel. First developed by Claude Shannon in the late 1940s and early 1950s in an effort to identify the precise quantity of information lost over a noisy telephone line (Shannon 2001; Shannon 1961), this formula has since been applied to countless scientific and computational fields to determine the mutual information shared by two parties, primarily in the presence of noise. In brain-computer interface literature, this metric has been applied to ascertain the degree of efficiency and accuracy of a BCI system (Wolpaw et al. 2002a; Nykopp 2001; Kronegg, Voloshynovskiy, and Pun 2005). The utility of this measure is in its application to a wide variety of options, and that it can calculate the exact amount of information that is being sent over these differing BCI paradigms.

Typically calculated in bits per second in information technology, this measure can be converted to characters per second by establishing the bits of information inherent in characters of the English language (Shannon 1951). Further, assuming the average length of an English word at 5 letters per word, we can convert to the more common standard of words per minute (wpm), either including or excluding a space between words (MacKenzie and Soukoreff 2003; Bochkarev, Shevlyakova, and Solovyev 2012; Mathematica).

Table 1. Highest Reported Information Transfer Rate for Communicative BCI Modalities

BCI Paradigm	**Highest Reported Information Transfer Rate**		
	Bits / min	**Characters / sec**	**Words / min ***
SSVEP, EEG	123	0.41	4.9
SSVEP, ECoG	113	0.36	4.4
P300 ERP, EEG	37.8	0.12	1.5
SCP, EEG	~15	0.05	0.6
fMRI	~5	0.02	0.2
			***** not including spaces or autocomplete algorithms

Although ITR calculations may differ based upon the quantity of presented options within each BCI paradigm, the quantitative measure remains largely consistent. With this metric, BCI efficiency speeds can be compared to other human-to-human communication modalities, as well as computer and internet speed channels. Each formula can analyze a BCI in full as a system involving an information source (the neural-controlled signal), transmitter (the electrode or sensor), channel (the interpreting algorithm), receiver (the computer), and a destination (the action).

2.2. Survey of human-to-human communication speeds

Human-to-human communication speeds can serve as a target for optimal BCI performance (Figure 2). By surveying speeds of traditional human communication media, the most efficient methods for transmission of communicative signal are revealed, which can then be goals for more efficient communicative BCI. Thus I briefly review a history of human-to-human communication modalities and their average and maximum rates here (Reed and Durlach 1998), with the stated goal of determining best practices for BCI communication[1].

[1] For the purposes of a fair evaluation, language is standardized across modality. In this case, English is being compared.

The oldest form of human communication is speech – more specifically, the combination of speech production and speech comprehension. In this modality, the speaker and the listener must both understand the language or media at hand. If the speaker speaks too quickly or if there is significant ambient noise, and the listener does not understand the speaker, then the mutual information along this channel is lost. Typical rate of human speech in the English language is 150 wpm, though auctioneers are known to speak at 250 wpm and one individual can recite memorized text at over 500 wpm ("Talk About a Fast Show: Londoner Rattles Off Raglan in 25 Seconds" 2011). Thus speech articulation may be the fastest communication modality between humans.

With the advent of the written word over 5000 years ago, human writing and reading became efficient methods for communication in literate peoples. Functional limits to human writing or typing speed as well as reading and comprehension speed exist. A typical person reads at a rate of 250 wpm and cannot read much faster than 500 wpm (Hardcastle and Matthews 1991). A notable exception is skimming or "speed reading," which can purportedly reach speeds of 1000 words per minute. The rate-limiting factor of written communication, then, is not the speed at which an individual can read or comprehend, but the speed of the individual to produce text[2]. Alphanumeric handwriting occurs at an average rate of 20 to 50 wpm (Hardcastle and Matthews 1991). The ensuing section further documents non-handwritten writing technology ITR.

Telegraphy, the transmission of a message via binary electromagnetic commands and therefore more similar to current BCI modalities, was originally developed by Samuel Morse in 1844, with Morse code being the primary encoding language across this medium (Morse 1849). In 2013, two operators with knowledge of Morse code still maintain a higher ITR than the fastest text-messaging writers. Text-messaging, the modern-day equivalent to telegraphy, also involves

[2] Rates vary with pictographic or other non-alphanumeric written languages.

sending short transmissions to another by typing out messages on cellular phones. Texting is often performed by selecting letters with thumbs, either on interactive capacitive displays (93.5 wpm), small keyboards (56 wpm), or phone keys (55 wpm) (Glenday 2013). Many user interface improvements, such as automatic completion of the word that is being typed or optimization of presented stimuli, can be applied to greatly improve overall typing speed. Swype and other graphical path analyzing short-cuts are an example of one such user interface improvement (Kushler and Marsden 2006).

The incorporation to more digits of manual control can yield higher typing speeds. Use of the traditional QWERTY keyboard, introduced in the 1870s, can elicit an average of 40 wpm for a typical person and a top speed of approximately 120 wpm (Ostrach 2005). Varying organizations of the keyboard layout can result in faster typing speeds, such as a maximum rate of 212 wpm using the Dvorak layout (McWhirter 1985). However, the highest typing rates are performed by court stenographers with a maximum speed of 360 wpm, who use a phoneme-based keyboard (Glenday 2013). The stenograph, which uses combinations of alphanumeric characters to simulate phonemes, is therefore capable of producing the highest ITRs of any human-to-human writing modality.

Since the 1980s and 1990s with development of automatic speech recognition algorithms, speech-to-text dictation software has also produced high information transfer rates, often as an alternative to typing or text messaging. Advancements in machine-learning algorithms, such as neural networks and Hidden Markov Models, and the development of huge databases of speech information contribute to high ITR for speech recognition in non-noisy channels. The upper bound for speed in this modality is intrinsically related to the rate of speech dictation, which can reach as high as 200 wpm at maximum speed but is typically 50-70 wpm. This communication modality typically performs poorly in the presence of ambient noise.

Logically, an upper bound for human-to-human communication must exist. One can analyze and compare ITR across all modalities of a specified language to ascertain this upper bound. In this analysis of human-to-human communication rates, illustrated in Figure 2, a trend emerges in which phonemic information entry surpassing its alphanumeric information entry equivalent. The International Phonetic Association has categorized *phonemes*, the smallest discriminable sounds of speech, for all human languages and identified corresponding mutually exclusive vocal tract positions for each given phoneme (Ladefoged 1990; Brown 2013). Phonemes, when combined, produce speech, which has been tied to human evolution for rapid communication (Long and Berke 2010). It is with this in mind that we here investigate speech as a modality to be utilized for communicative BCI control.

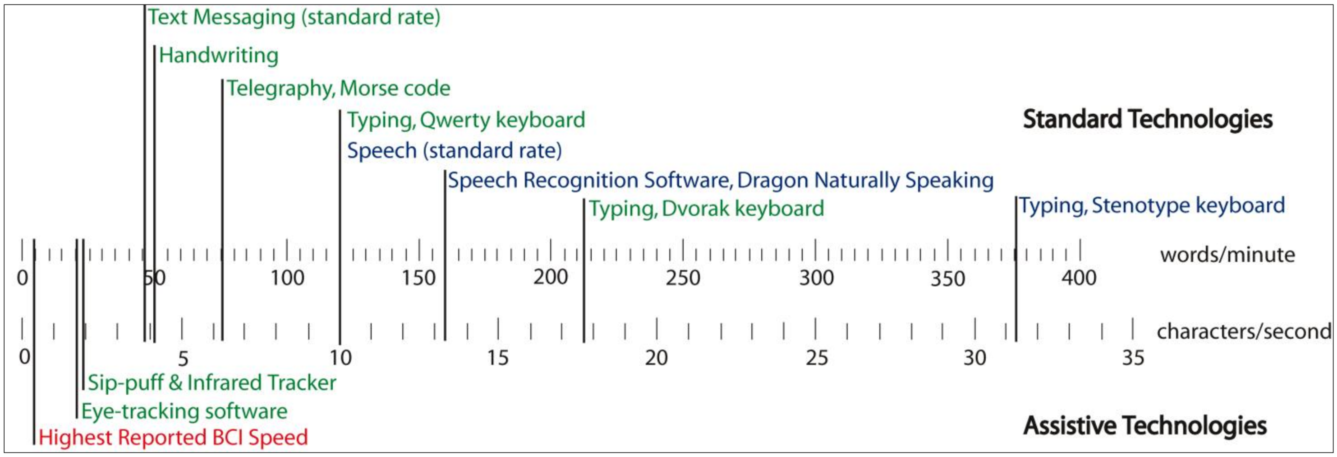


Figure 2. Speed of Media in the English Language. Human-to-human communication speeds reported in words per minute and characters per second, using the assumption as 5 characters per word standard in the literature. Green text indicates alphanumeric input, blue text indicates phonemic input, and the highest reported BCI speed is highlighted in red text. Standard technologies for communication are separated from assistive technologies primarily employed in rehabilitation science.

2.3. Advantages of Speech for BCI

There exists a need for a more reliable, more efficient modality in BCI research. A survey of current BCI users revealed a majority would only be satisfied by over 90% accuracy and by performance of at least 3 times faster than current standard speeds (Huggins, Wren, and Gruis 2011). If speech signal – particularly as it corresponds to phonemic information – could be reliably and successfully ascertained from neural recordings, the potential increase in ITR could yield a BCI that drastically outperforms traditional alphanumeric approaches. Further, if the neural signal for speech motor control could be reliably determined and recorded, it could yield a more intuitive interface for potential BCI users in comparison with more cognitive operations in the Section 1.2. Current users of neuroprosthetic devices that access such motor signals report that use of such prosthetic feels intuitive (Collinger et al. 2013; L. Hochberg and Taylor 2007; L. R. Hochberg et al. 2012; L. R. Hochberg et al. 2006; Scheme, Hudgins, and Parker 2007). Thus there may be untold advantages in recording and interpreting existing neural processes related to speech for use with brain-computer interface.

3. Speech decoding in BCI

3.1. A Comprehensive Survey of the Literature

Speech BCI, a subset of communicative BCI, began in search of a reliable speech-related signal that could be used to control a communicative BCI. Unlike the communicative BCIs described in Section 1.2, which primarily access higher-level or more cognitively-controlled signals for neural control, these BCIs attempt to access the neural signal of speech production.

Though there have been a few non-invasive measures to acquire changes in neural signal due to gross speech activity non-invasively (Herff et al. 2013; DaSalla et al. 2009), these forays have primarily only determined if a person is actively speaking or imagining speaking, and cannot necessarily determine the actual content of the speech. It remains unclear whether any neural signal corresponding to speech content is accessible from non-invasive electrophysiological approaches at the scalp. Therefore it follows that a more direct method of recording cortical speech production signal would be more successful. Most successful approaches for determining content of speech production signal to date have been invasive and are detailed in the remainder of this section.

The cortical speech areas were first discovered by pioneering neurosurgeons Otfrid Foerster and Wilder Penfield during craniotomies, when they stimulated the cortex with current and observed subsequent movements in speech articulator organs (Foerster 1931; W. Penfield and Boldrey 1937). Over the course of a decade of surgical work, Penfield outlined a "somatotopic" motor map of a cortex, in which a large area of motor and sensory areas are dedicated to speech articulation, lip movement, and gustatory activity (Wilder Penfield and Roberts 1959). With the development of fMRI technology in the early 1990s, the findings of Penfield could be further confirmed, showing articulator organ representation throughout the primary motor cortex during speech articulation (Bohland and Guenther 2006). Although the 7 to 8 second time delay related to metabolic activity impeded the ability to identify contextual information related to speech, some studies document the differences in cortical activation between simple motor movement of speech organs and speech articulation (Terumitsu et al. 2006). Further fMRI studies also suggest the left ventral premotor area may have syllable-specific processing in addition to traditional primary motor cortex (M1) areas bilaterally (Guenther, Ghosh, and Tourville 2006). Therefore, sensors that record neural signal related to speech content for BCI applications require a fine spatial resolution to discriminate variations of activity within these areas.

One issue that has impeded speech BCI research is that speech is a complex, dynamic, and mathematically non-stationary process that varies rapidly in time[3]. This uniquely human process (speech) requires measurement with an electrode or sensor capable of recording its fine temporal resolution, or there may be a loss of information that could be useful in interpreting the signal. Although the neural activity associated with speech likely has a dynamic range separate from speech itself, the sensors needed to record such activity must be able to analyze changes on a rapidly-changing time scale. The need for temporal and aforementioned spatial resolution can now be uniquely met by developments in electrocorticographic technology (Schalk 2010).

Electrocorticography (ECoG), sometimes referred to in the literature as intracranial EEG, is the recording of electrical field potentials on the surface of the cortex (Leuthardt et al. 2004). ECoG electrodes can be inserted below the dura (*subdural*) or just above the dura (*epidural*); although the difference in signal quality and robustness may be marginal (Slutzky et al. 2010). ECoG electrode arrays were originally intended for direct current stimulation and corresponding functional mapping, primarily useful in precautionary determination of so-called “eloquent cortex” or functional speech cortex during neurosurgery (G. A. Ojemann 1991; G. A. Ojemann

[3] In contrast, slow cortical potentials and fMRI signals are two neural signals that demonstrate a slow time course.

1979; G. Ojemann et al. 1989). When used to record changes in electrical potential rather than stimulate, ECoG has the spatial resolution necessary to record over a wide area of cortex related to speech production and articulation with standard medical spacing of 1 cm intra-electrode distance. Unlike fMRI, ECoG has the fine temporal resolution to map the brain's responses to rapid processes such as speech. The primary electrophysiological determinant of activity recorded with ECoG, the increase in the high-gamma power band (70-200 Hz), has also been shown to correlate with the BOLD signal of fMRI (Hermes et al. 2012). With the advent of this technology, typically used for recording and detecting seizure activity in epilepsy patients, the ability to record the unique neural dynamics related to speech production become possible.

The first scientific foray with ECoG into speech production investigated differences in one patient during word production and during sign language, indicating differences between language representation and overt speech (Crone et al. 2001). Differences in overt speech were primarily noted in the tongue motor area of M1. Other early studies investigated the difference between receptive and expressive speech areas (Towle et al. 2008), which largely identified two separate areas for producing speech in what is thought of as Broca's area and what is thought of as M1.

The first study investigating discernible content of speech attempted to isolate four separate phonemes and demonstrated that distinct cortical regions could be identified that corresponded with specific phonemic production, loosely suggesting a "phonetotopic" map similar to the somatotopy discovered by Penfield (Blakely et al. 2008). Their results led to the inference that spacing smaller than 1cm may be necessary to discern these separate phoneme-elicited regions with ECoG. This initial phonetotopic map has since been substantiated and further detailed according to speech articulator organs via fine-grid ECoG recordings (Bouchard et al. 2013). This phonetotopy can serve as a guideline for designing viable speech BCI.

Another issue that impedes the progress of speech BCI research is the confusion between speech production, speech reception, and language areas in the cortex. Kellis and colleagues investigated cortical changes during speech in two separate neural areas: facial motor cortex (Brodmann Area 4), which elicits movements in the face when stimulated, and Wernicke's area (superior temporal gyrus (STG), Brodmann Areas 22 and 40), which is thought to be the area where words take on semantic meaning (Kellis et al. 2010b; Kellis et al. 2010a; Wernicke 1874). Discrimination between words that were phonetically similar performed better using information from facial motor cortex electrodes in comparison with Wernicke's area electrodes, demonstrating that recording speech-motor-signal approach may be to discern the content of speech for BCI.

A further issue in speech BCI is determining whether to target cortical activity related to actual *overt* speech as compared with cortical activity related to imagined or *covert* speech. Both closed-loop and open loop approaches have attempted to infer phonemic content of covert speech from neural signal (Pei et al. 2010; Pei et al. 2011; Leuthardt et al. 2011). These studies reveal that covert speech activates a much broader cortical area than overt speech alone, but they also demonstrate that it is possible to discriminate phonemes in real-time with signals from ECoG arrays despite nontraditional cortical areas of control. This covert, inner narrative was further investigated using ECoG arrays over auditory areas, showing a more "top-down" inner voice categorization (Perrone-Bertolotti et al. 2012). Although it is unclear whether covert speech has more to do with imagined speech production or imagined speech reception, one can assume that overt speech may be a closer approximant of intended speech in a CLIS individual as compared to covert speech. Investigations with a patient with LIS prove that attempted overt speech may be sufficient for accessing motor speech signal for partial classification of *intended* speech signal (Guenther et al. 2009; Brumberg et al. 2011). This suggests that overt speech may serve as a basis for communicative BCI for individuals with LIS and CLIS.

The ECoG electrode paradigms detailed here have performed better than any other modality at ascertaining speech signal from the cortex, though other invasive measures have been studied as well with varying success. One implanted device, the neurotrophic or Kennedy electrode, implanted directly into speech motor cortex with a total area of about 1.5 mm and a resolution of 600 microns (Bartels et al. 2008; Kennedy and Bakay 1998), achieved discrimination of three imagined vowels in an individual with locked-in syndrome (LIS) with a success rate of up to 70% (Guenther et al. 2009). Further investigations of phonemes including consonants yielded a best performance of 21% (Brumberg et al. 2011). This was the first study to perform an invasive procedure with the goal of speech-motor-communication through BCI use, and the study highlights the difficulty in terms of unknowns that are involved in imagined speech and neural interfaces. Nevertheless, this study notably confirms abilities of individuals with LIS to produce volitional cortical activity related to intended speech motor output independent of actual motor activity. Despite neuroplastic changes that can occur with loss of sensorimotor activity, Shoham and colleagues have demonstrated that cortical surface activity can remain typical in quadriplegics with intended motor movement at least 5 years post-injury (Shoham et al. 2001). Although the neurotrophic electrode records from deeper cortical structures, it demonstrates potential for success by studies of intended speech motor activity at the cortical surface.

Implanted microelectrode arrays (MEAs), which record field potentials and single-unit activity of neurons in cortex, typically cover a maximum area of 3x3mm, which may be too small of an area to access most speech articulator activity. Although typing interfaces have demonstrated 2-d mouse control via imagined motor movement (L. R. Hochberg et al. 2006), MEA use in humans has been exclusively for motor neuroprosthetic control and has not yet explicitly designed to access speech motor areas in the literature.

An alternative approach to speech-BCI exists in placing electromyographic (EMG) electrodes on the articulatory organs themselves. Although this would presumably provide little assistance to those with CLIS, they could be used by otherwise healthy speakers to improve the fidelity of standard speech in high noise environments (B. Betts and Jorgensen 2005; B. J. Betts, Binsted, and Jorgensen 2006) or even enable silent or whispered speech to be successfully communicated (Metze et al. 2005; Maier-Hein 2005; Jou, Schultz, and Waibel 2005; Chan et al. 2001; Walliczek et al. 2006; Wand et al.; Denby et al. 2009). These processing methods are enabled by direct contact to muscle activity, without the skull and cerebrospinal fluid acting as bandpass filters to impede electrophysiological fidelity as in EEG recording. Although the EMG signal is not typically considered a neural signal, it is controlled by volitional neural processes in healthy individuals and can therefore augment understanding of motor movement and subsequent neural activation involved in speech.

Further analysis of communication interfaces is discussed by Brumberg and colleagues (Brumberg et al. 2010).

3.2 Speech production as a neural process

Speech production is a complex human behavior comprised of rapid transitions between positions of the vocal tract (Simonyan and Horwitz 2011). Speech production involves exact coordination of 68 different muscles[4] (Epstein, Hacopian, and Ladefoged 2002). The corresponding neural process involves increases in activity in several regions of cortex from the original spontaneous generation of the word concept and its articulation. Therefore many factors confound the isolation of speech production for neuroscience investigations, including the phonemic context of speech, the simultaneous production of auditory stimuli, speech perception, and the semantic value of speech. These confounds leave lingering questions as to the exact mechanics of neural activity during speech production and are discussed here.

One complication to studying phonemes in cortex is the existence of co-articulation. Co-articulation, which describes when vocal tract position for one phoneme is altered in the presence of an adjacent phoneme, increases the difficulty in analysis of a single phoneme and its neural correlates. A common example is that the pronunciation of the \n\ in “tenth” places the tongue tip at the upper teeth in preparation for the adjacent \θ\, in comparison to the standard tongue tip position at the alveolar ridge (e.g. \n\ in “ten”). Whereas both phonemes are intelligible as an \n\ to listeners, this difference in position of the articulator organs may increase the difficulty of isolating single phonemes for scientific research. Cortical activity related to phoneme production may therefore incorporate a specific state space or alternatively be intrinsically inseparable within the phonemic context of speech.

Further, speech perception can influence neural activity in areas traditionally considered speech production areas. ECoG processing of signals recorded at the superior temporal gyrus has demonstrated that speech frequency information can be roughly reconstructed (Pasley et al. 2012). Speech production may also suppress activity in auditory cortex (Flinker et al. 2010). Some phonemic information may exist within the temporal lobe (Turkeltaub and Coslett 2010),

[4] This includes 9 separate muscles for respiration, 8 muscles to control the mandible, 8 muscles for tongue movement, 15 muscles for lip movement, 6 muscles of the soft palate, 3 muscles of the pharynx, 12 extrinsic laryngeal muscles, and 7 intrinsic laryngeal muscles.

but it may also be not physically separated with areas that process whole word information (Flinker et al. 2011). Auditory processing has been investigated during self-vocalization (i.e. when a person hears their own voice during speech) as it compares to the same speech being played back via headphones, highlighting differences in how a person listens to their own voice (Greenlee et al. 2011). This process is even more complex when a voice cannot completely be muted, as audio frequencies can travel to the ear via bone conduction. Controlling for a subject's own speech, then, is a potential confound if not fully considered.

Another confounding factor for studying speech sensorimotor cortex is that speech perception influences areas associated specifically with speech production. The so-called McGurk effect reveals an increase in speech perception with when a visual stimulus of that speech is present, but also confusion when speech and visual stimuli conflict in terms of content (McGurk and MacDonald 1976). This effect has also been studied with fMRI, revealing areas of traditional motor cortex contain generalized representation of phonemes in speech perception. (Wilson et al. 2004; Pulvermüller et al. 2006; D'Ausilio et al. 2009) One possible explanation for this activity is that "mirror neurons" aid in comprehension. Thus the organization of functional speech production in cortex appears more nuanced, activating regions traditionally considered exclusive areas for speech perception, motor activity, and articulation.

A final confound is the inherent correlation of speech and language. An analysis of ECoG patterns subjects using free recall to generate words revealed patterns that correlated with semantic clustering of words in the English language (Manning et al. 2012b). Activation of Broca's area may occur during naming, and activation of Wernicke's area may occur during semantic interpretation of word stimuli.

Finally, although speech production is primarily a motor process, speech cannot easily be separated into motor and sensory activation. These distinctions matter greatly in phonetics research, as the “place of articulation” and thus the capability for sensory feedback dictates constriction location and thus phoneme designation. A somatotopic map for speech articulator organs exists in both sensory and motor cortex, but Penfield’s cortical stimulation only ever elicited movements of articulator organs and never actually elicited speech (W Penfield and Rasmussen 1949). Despite the most recent ECoG studies of speech production (Bouchard et al. 2013), it remains unclear to what extent the changes in neural activity in motor cortex predict subsequent speech movements.

These issues lead to questions about how speech production is functionally organized in cortex. Several theoretical models of speech exist, each with differing views on the exact structure of speech directly prior to articulation, when such motor commands sent along cranial nerves. One computational model of speech from Hickok hypothesizes that a hierarchical state feedback control outputs phonemic information directly to articulators (Hickok 2012). The direction into velocities of articulators (DIVA) model hypothesizes that velocity and position maps comprise the structure directly preceding motor execution (Guenther 2006).   An older model, classifying speech perception as highly correlated with motor movement, is the aforementioned motor theory of speech (Liberman and Mattingly 1985; D’Ausilio et al. 2009). These models differ in how speech production is represented cortically.

Further, cortical representation could primarily be based upon the *movement* of articulators, similar to force-specific of motor movement (Fagg et al. 2009). Cortical representation may be primarily structured by *place* of articulation of a particular speech sound (perhaps related to sensory encoding (Hatsopoulos and Suminski 2011)), or structured by *location* or *position* of motor activity (Georgopoulos, Schwartz, and Kettner 1986). It could also be based primarily on the *force* of articulator movements (SH Scott and Kalaska 1997). Functional organization of speech in cortex may therefore be determined by degree of success in classification of content of speech. A study of the speech production using ECoG could leverage the use of BCI classification to answer questions as to the functional organization of cortex.

3.3. Description of the thesis

The work herein seeks to address the various unknowns in the current speech BCI literature. Having already addressed the current limitations and barriers to widespread BCI adoption, I here seek to add to the engineering merit of the communicative BCI field with the analysis and results of three separate studies. These studies, outlined in the following three chapters, have either been submitted or are pending submission to academic journals for publication. In the first study, I seek to determine what factors contribute to successful phonemic decoding of an ECoG signal (Journal of Neural Engineering). In the second study, I seek to determine cortical representation of phonemic categorization in speech production (Journal of Neuroscience). In the third study, I leverage classification results to address the structure of cortical correlates of speech production (Nature Neuroscience). The result of these studies outlines a set of guidelines for future speech-BCI research that will work towards useful speech-BCI neuroprothetics.

II. CLASSIFICATION OF ELECTROCORTICOGRAPHIC INFORMATION DURING SPEECH ARTICULATION

Emily M. Mugler, James L. Patton, Robert D. Flint, Zachary A. Wright, Stephan U. Schuele, Joshua Rosenow, Jerry J. Shih, Dean J. Krusienski and Marc W. Slutzky

1. Introduction

Although brain-computer interfaces (BCI) can be used in several different ways to restore communication (see (Birbaumer and Cohen 2007; Wolpaw et al. 2002a) for review), communicative BCI has not approached the rate or success of natural human communication (Leuthardt, Cunningham, and Barbour 2013). One particular BCI approach is to classify and decode neural signals related to speech production, introduced by Kennedy and Bakay with invasive cortical electrodes (Kennedy and Bakay 1998), but research in this field has failed to approach the efficiency of speech. Advances in electrocorticographic (ECoG) technology, in which electrical field potentials are recorded directly from the surface of the cortex, may be able to address these limitations by recording from cortical speech areas. ECoG has been used to decode movement kinematics and kinetics (Schalk et al. 2008; W. Wang et al. 2013) and classify rapid cognitive processes (Manning et al. 2012a). ECoG recordings have precise temporal and spatial resolution (Slutzky et al. 2010) and enable recording of rapid processes over a wide area of cortex (Schalk 2010). ECoG can therefore facilitate mapping of the rapid neural changes related to speech production (Bouchard et al. 2013), which involves concurrent activation in a wide area of cortex.

Most studies of speech production using ECoG to date have been intentionally limited in scope for simplification. Studies that employ a *whole-word* approach, classifying cortical activation patterns primarily based upon the differences between full words, initially identified the cortical areas that are active during speech articulation (Pei et al. 2010). Classification of articulated words with micro-ECoG electrodes over facial motor cortex successfully identified at best less than half of 10 words in one patient (Kellis et al. 2010b). Another study classified pairings of initial and final consonants by comparing the ECoG activation relative to word onset, and achieved up to 45% classification of a single consonant pairing in one out of 8 subjects (Pei et al. 2011). These whole-word studies outline preliminary success in speech decoding, but ultimately such success rates cannot be extrapolated to more complex speech and fall short of performance levels necessary for speech communication efficiency.

An approach that specifically decodes the smallest isolated segments of speech, called phonemes, may yield higher results. Intracortical speech BCIs employing a *phoneme* approach have achieved up to 21% classification success of all phonemes (Brumberg et al. 2011), and demonstrated up to 70% classification success of discrimination of 3 imagined vowels in an individual with locked-in syndrome (Guenther et al. 2009). Similar studies with ECoG succeeded in classifying limited subsets of phonemes, isolated from the context of words (Blakely et al. 2008; Leuthardt et al. 2011). ECoG studies have demonstrated the ability to rapidly classify phoneme production, yielding an average 83.5% discrimination of 2 vowels for 2 subjects in real-time (Leuthardt et al. 2011). One recent study detailed an approximate "phonemotopic" map of the areas to target within motor cortex using intermediate-density ECoG, updating traditional somatotopic maps for motor cortex (W. Penfield and Boldrey 1937). These approaches demonstrate the potential to decode phonemes from cortical signals.

To our knowledge, no ECoG study has specifically investigated phonemes with respect to their onset within word production. Approaches classifying the event-related potentials of phoneme onset, precise to the millisecond level, have not been applied using ECoG. Further, no ECoG study has investigated classification of a *comprehensive* set of phonemes for a language.

Finally, the position of phonemes within a word may influence neural signals, and these signals may differ when producing isolated phonemes. If phonemes are consistent in their neural representation, words with rhyming structure could potentially reveal properties of individual phoneme production and leverage phoneme identification.

In this study, we investigated production of words using the entire set of phonemes in the General American accent of English using ECoG. The rationale for this study was that once speech articulation is related to corresponding cortical signals, the first critical step toward motor-based speech prosthetics would be established. We attempted to identify specific sources of failure in the classification process, as a detailed explanation of sources of failure could guide for future approaches. Furthermore, we hypothesized that precisely synchronizing analysis to each individual phoneme event is crucial for accurately discerning event-related cortical activity. This reveals speech production dynamics in cortex, enabling decoding of articulation patterns. Finally, we explored leveraging the structure of similar sounding (i.e. rhyming) words to enhance phoneme identification.

2. Methodology

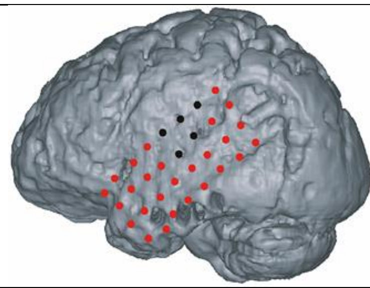
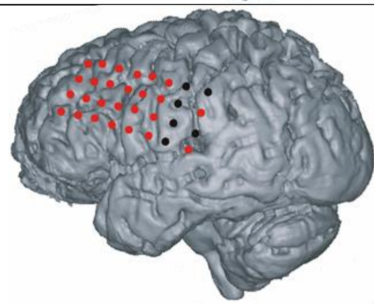
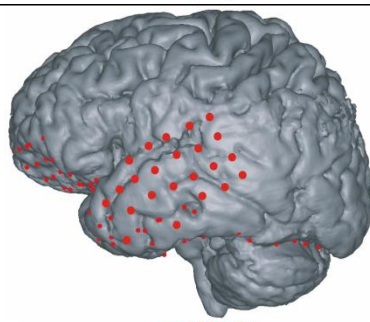
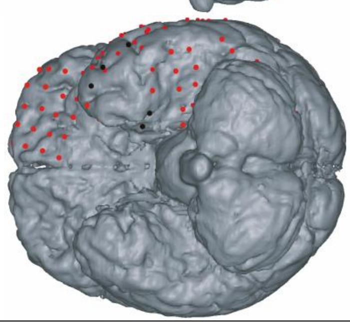
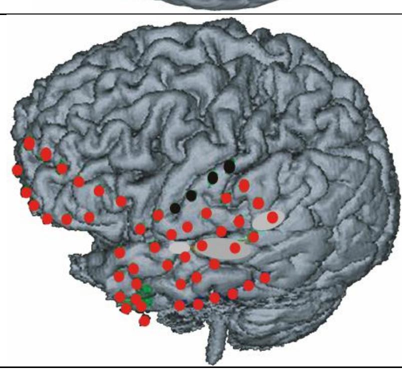
2.1. Subject Pool

Four subjects (mean age 42, 2 female) who required extraoperative ECoG monitoring for treatment of their intractable seizures gave informed consent to participate in this study (Table 2). Electrode coverage of cortex, determined by medical necessity, included some frontal and temporal areas in all subjects, although the degree of frontal coverage varied widely. This study was approved by the Institutional Review Boards at Northwestern and Old Dominion universities and at the Mayo Clinic.

Prior to explantation of the ECoG electrode grids, a functional map of cortical structures was determined using stimulation of electrode pairs by an epileptologist. Areas that, when stimulated, produced reading arrest were designated as being associated with language, and areas that produced motor movements of the tongue and articulators were designated as functional speech motor areas. This clinical mapping of eloquent cortex provided *a priori* knowledge of functionality of cortex and provided a "gold standard" for our analysis.

Three-dimensional reconstruction of ECoG electrode placement was determined using co-registration of pre-surgical structural magnetic resonance images and post-operational computed tomography scans (Miller et al. 2010; Hermes et al. 2010).

Table 2. Subject information and ECoG electrode locations. Electrode coverage varied due to each patient's clinical needs. Electrodes denoted in black contribute to improved classification performance.

Subject ID	**Grid location**	**Gender**	**Age**	**Trials Completed**
NU1		F	30	320
NU2		M	50	480
NU3	 	M	49	480
MC1		F	39	320

2.2. Data Acquisition

We simultaneously collected speech audio signal (sampled at 44.1 kHz) from a USB microphone (MXL) and 3 channels of EMG from the right masseter and both pharyngeal muscles (over the carotid triangle, Delsys Bagnoli-8). Speech and audio signals were collected using customized BCI2000 software (Schalk et al. 2004) and a Tucker-Davis Bioamp system, which was synchronized with ECoG signals recorded on a clinical system (Nihon Kohden for NU subjects and Natus XLTEK for the MC subject). ECoG signals were bandpass filtered from 0.5-300 Hz and sampled at 500 Hz for subject NU1, 1 kHz for subjects NU2 and NU3, and 9.6 kHz for MC1 (Figure 3). Differential cortical recordings compared to a reference ECoG electrode were exported for analysis with an applied bandpass filter (0.53 - 300 Hz) with a sensitivity of 75 μV.

2.3. Experimental Protocol

Prior to the start of the experiment, the subject was asked questions regarding accent and exposure to other accents. A survey comprised of four questions regarding accent, mother tongue, and foreign accent exposure was presented to the subject to ascertain subject accent and exposure. Accent was confirmed post-hoc.

The subject was presented with words on a The Modified Rhyme Test (MRT), consisting of 300 monosyllabic words, primarily with consonant-vowel-consonant structure (House et al. 1963). This stimulus set was chosen for its simplicity and the prevalence of rhyming structures and its frequency and variety of American English phonemes. The frequency of phonemes within the MRT set roughly approximates the phonemic frequency found in the English language (Mines, Hanson, and Shoup 1978), though there is unequal incidence of phonemes within the data set. Because the MRT did not include all phonemes present in the General American accent of English, 20 additional words, which included 4 phonemes excluded from the MRT (\ʒ\,\ j\, \ə\, and \aɪ\), were added to the stimulus set to create a comprehensive collection of General American phonemes in full words.

Using BCI2000, we presented words on a screen for 3 s, followed by a blank screen for 1 s. Subjects read each word aloud as soon as it appeared. Due to restrictions on time spent with the subjects, each recording block took 10.5 minutes to display 160 trials, in which a trial is one word stimulus. Total trials per subject varied from 320 words (Subjects NU1 and MC1) to 480 words (in which the first 160 words of the stimulus set were repeated) (Subjects NU2 and NU3).

2.4. Data Preprocessing

In order to enable classification by reducing the massive size of the data set for each trial, data was reduced to time-frequency features for each trial and further separated by phoneme (Figure 3). We used visual and auditory inspection of auditory spectral changes to manually label the onset of each phoneme in the speech signal (APPENDIX B. Graphical User Interface for Phoneme Labeling). Phoneme assignment was determined using the CMU Pronouncing Dictionary, which assumes General American pronunciation.

ECoG signals were common-average referenced. Data was aligned across recording platform using a TDT pulse signal (TTL), and the onset times of phoneme in speech signal were extrapolated and marked to the clinical ECoG recordings. Signals were split into 4-s trials centered around word onset. Each trial therefore included records of onset of each phoneme and its IPA distinction of each word. A 1-second of baseline rest activity was identified at the start of each trial and ending 1 second prior to word utterance. Thus trial duration ranged from 4.15 to 4.50 seconds, with word onset occurring immediately prior to the midpoint of each trial. We computed short-time Fast Fourier Transforms (FFTs) via the Goertzel algorithm (Rabiner and

Schafer 1978). in 50 ms windows of ECoG relative to baseline activity (Matlab). A 2 Hz frequency step size was used to create the frequency bins of the FFT.

To enable classification and reduce the massive quantity of data for each trial, spectrotemporal features were created for each phoneme using 7 frequency bands and 16 time bins per electrode. To create features in the frequency domain, we isolated and summed power changes in specific frequency bands from the FFT data to create frequency features. The high-gamma band, most commonly used in ECoG research due to its proven correlation with neuromotor activity, has definitions that vary widely in the literature. The frequency band methodology employed successfully by Flint and colleagues to determine motor movement from cortical activity served as a starting point, in which the high-gamma band was defined as 70 to 200 Hz (Flint, Ethier, et al. 2012), though 3 separate portions of the high-gamma band that avoided the 60 Hz noise harmonics with a 10 Hz span (65-115 Hz, 125-175 Hz, and 185-250 Hz) were analyzed for full investigation of high-gamma band changes. Additional frequency bands standard in EEG research for their correlation with cortical function, including the delta (0-4 Hz), mu (7-13 Hz), beta (15-30 Hz) spectra, as well as the local motor potential, were analyzed. The band from 250-300 Hz was also investigated when sampling frequency exceeded 500 Hz. To create features in the time domain, we summed 50 ms segments of the FFT from 300 ms prior to and 300 ms after phoneme onset. This created separate time bins that summarized the neural signal directly preceding and throughout pronunciation of each phoneme.

The time-frequency power features were then sorted by phoneme. For the full 320 word stimulus set, we analyzed 981 phonemes; for subjects who completed 480 words, we analyzed 1470 phonemes. To reduce input space dimensionality, features were ranked according to p-values from an ANOVA across phonemes, and the top 140 features were selected for classification. Therefore if $\boldsymbol{x}$ ECoG channels were recorded and converted to $\boldsymbol{y}$ frequency bands features and $\boldsymbol{z}$ time bins features, there would then be $\boldsymbol{x} * \boldsymbol{y} * \boldsymbol{z}$ features for each analyzed phoneme. These features could then be used for classification and analysis.

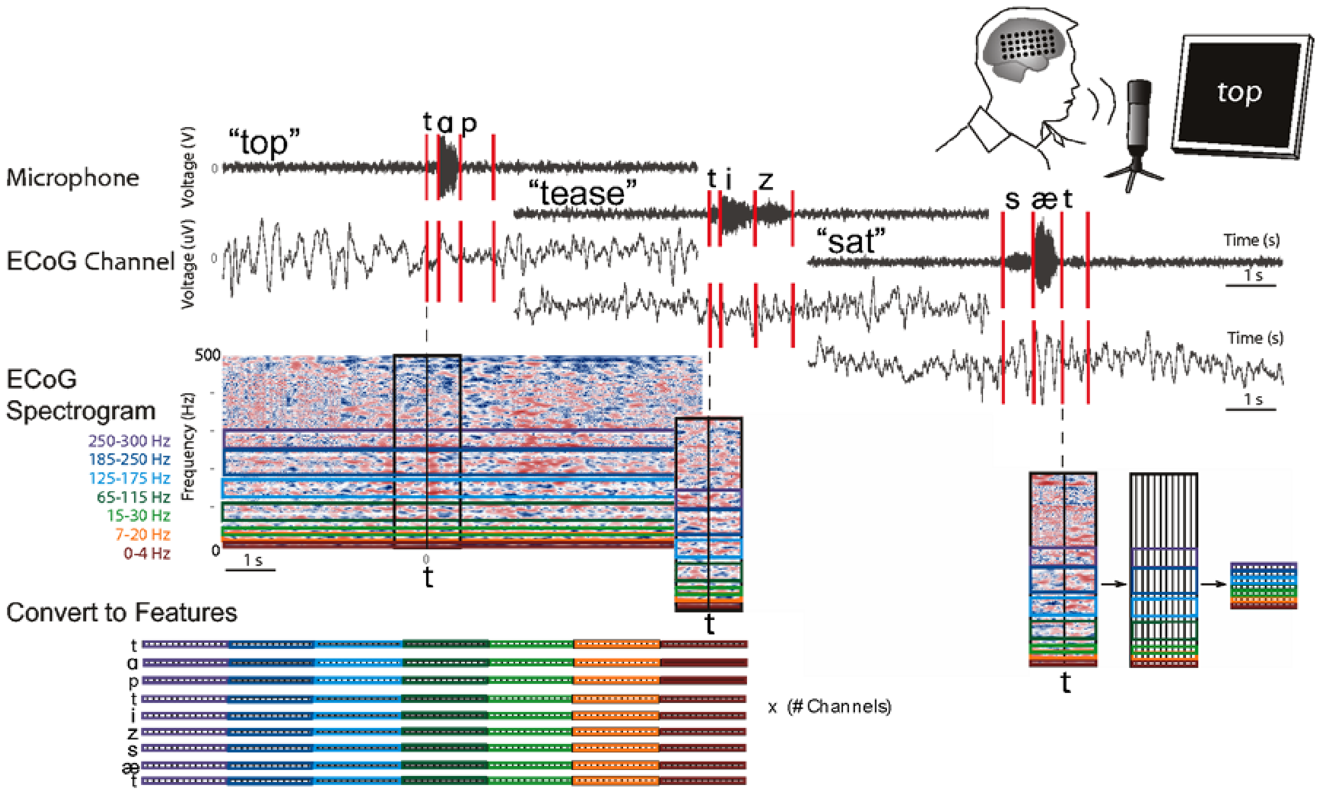


Figure 3. Overview of Preprocessing Methodology. Speech signal is recorded simultaneously with ECoG signal (apparatus inset). These signals are marked according to onset of phoneme time and aligned with to the ECoG signal. An short-time Fourier Transform is performed on the ECoG signal, and converted into features according to frequency band and time relative to phoneme onset time.

2.5. Classification

The 140 features with the lowest p-values between classification sets were selected to classify phonemes using linear discriminant analysis (LDA) (Slutzky et al. 2011; Flint, Lindberg, et al. 2012). LDA was selected primarily because it allowed us to identify the features that led to successful categorization. We used 10-fold cross-validation (with randomly-selected test sets) to compute success rates. The classifier was therefore trained on 90% of the phonemic data and tested on the remaining 10%; this process was repeated 10 times. A result was successfully classified if it had determined the correct phonemic label applied to the test set of data. Chance classification percentages were determined by randomizing phoneme labels and re-classifying; and computing the mean result of 100 repeats.

2.6. Rhyme Reduction

We hypothesized that the structure of words could be used to leverage a more fine-tuned analysis of each phoneme. By comparing similar words to isolate a single phonemic difference, we predicted that we could further isolate the changes in neural signal that elicited that particular phoneme. If ECoG signal were singularly robust and repeatable for each phoneme, words with similar structure (e.g. “cup” (\kəp\) and “cut” (\kət\)) could be analyzed together, and the elements related to the differing signal could be compared and isolated (\p\ and \t\, respectively). Assuming a similar time course for words of similar structure, this phonemic isolation could lead to better identification of neural correlates of phonemes.

Our stimulus set, which included combinations of similar-sounding consonant-vowel-consonant structure words, uniquely enabled investigation of this process. Words either belonged to one of 25 sets which shared a *rhyme*, defined as a final vowel-consonant pair, or a *body*, defined as the initial consonant-vowel pair. In this process, which we termed *rhyme reduction*, we compared differences in neural activation by aligning similar words to the onset of rhyme or body. This approach was attempted in two separate techniques: first with subtraction of an average signal, the second with a comparison of z-scores to an average signal. The subtractive rhyme reduction approach can be described by:

$$\boldsymbol{n} = \boldsymbol{n}\boldsymbol{m}_{\mathbf{1}:\boldsymbol{k}} - \frac{\mathbf{1}}{\boldsymbol{p}}\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{p}} \boldsymbol{n}_{\boldsymbol{i}}\,\boldsymbol{m}_{\mathbf{1}:\boldsymbol{k}} \qquad (1)$$

$$\boldsymbol{n} = \boldsymbol{m}_{\mathbf{1}:\boldsymbol{k}}\boldsymbol{n} - \frac{\mathbf{1}}{\boldsymbol{p}}\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{p}} \boldsymbol{m}_{\mathbf{1}:\boldsymbol{k}}\boldsymbol{n}_{\boldsymbol{i}} \qquad (2)$$

where *n* is the phoneme of interest at the beginning for a rhyming word or at the end of the body word; *m* is a phoneme that partially comprises the rhyme or body of the word, *k* is the number of phonemes that exist in the word excluding n, and *p* is the number of words that correspond to this rhyme or body.

An underlying assumption of this technique is that a robust response is elicited by the rhyme or body of the word, and that "white noise" of neural signal can be approximated by:

$$\text{noise} \approx \frac{1}{p}\sum_{i=1}^{p} /n_i / \tag{3}$$

The rhyme reduction approach was also evaluated using z-scores of features to ascertain if differences in phoneme utterance were proportional (instead of subtractive) relative to baseline utterance. We executed this approach, building off of the subtractive rhyme reduction process, by creating a template of the rhyme or the body of the word. The difference in frequency bandpower features of a word were then compared to the features of the template. Instead of assuming typical signal to be "white noise", the z-score methodology compares the features of a word to the features of average activity for a given rhyming structure.

$$z_{word} = \frac{x_{word} - \bar{x}_{template}}{CI_{template}} \tag{4}$$

where *n* is the isolated phoneme of interest, z_{word} is the z-score for a particular word as compared to its template components, x_{word} is a feature of the word of interest, $\bar{x}_{template}$ is a mean of feature values for that rhyming set and $CI_{template}$ is the confidence interval of the mean of the template. The isolated phoneme of each word is then a comparison of the z-scores of the word to the z-scores of the template. These z_{word} features were then fed into the LDA classifier.

We converted these signals to features, used ANOVA for feature selection, and classified using LDA. Features were aligned to the onset of the rhyme or the body of each word prior to the execution of the rhyme reduction. If performance revealed no increase with isolated phonemes, it could indicate either that the rhymes or bodies are not typical in their time course of articulation or that phoneme representation in cortex could be susceptible to effects of neighboring “coarticulated” phonemes.

2.7. Estimation of Information Transfer Rate

The goal for this technology is to decode phonemic information during speech production, but phonemes exist in combinations within words. We therefore analyzed how phonemic decoding of combinations of consonants could be applied to identify words of the data set. We further investigated this performance when constraining predictions of phoneme combinations to those existing in the stimulus set in the order of posterior probability. To calculate information transfer rate (ITR) from these results, we first calculated average word duration (520 ms) and phoneme duration (176 ms). ITR was then determined by multiplying the information capacity (in bits/phoneme) by classification success and rate of speech production (Wolpaw et al. 2002b). This procedure was extrapolated on other results reported in the literature using speech duration times from our results. Conversion to words per minute from bits per second was estimated using bit rates for syllable production of speech (Reed and Durlach 1998).

3. Results

3.1. Classification performance

Vowels and consonants were analyzed separately. Results varied widely over subjects, largely due to the wide variation in coverage of face motor areas (Figure 4). Subject NU2 had the highest overall performance, in which 36.10% of consonant phonemes were correctly classified. The maximum performance for classifying any one phoneme was 63% (\k\ for Subject NU2). Averaged across all subjects, 20.42 (±9.8)% of all phonemes were classified correctly, significantly greater than chance decoding (7.4%, $p<0.001$, t-test). Average classification performance for vowels across all subjects was 19.24 (±3.7)% also significantly greater than chance (12.9%, $p<0.01$, t-test), with the best performance in NU2 of 23.91%. Vowels that were most successfully classified were primarily those produced by the same articulatory structure that produced the consonants which were best classified for that particular subject. For example, with Subject NU2, the classifications of highest accuracy were the vowel \i\, which involves a high back tongue, and \k\, which has a constriction location close to where the \i\ is articulated.

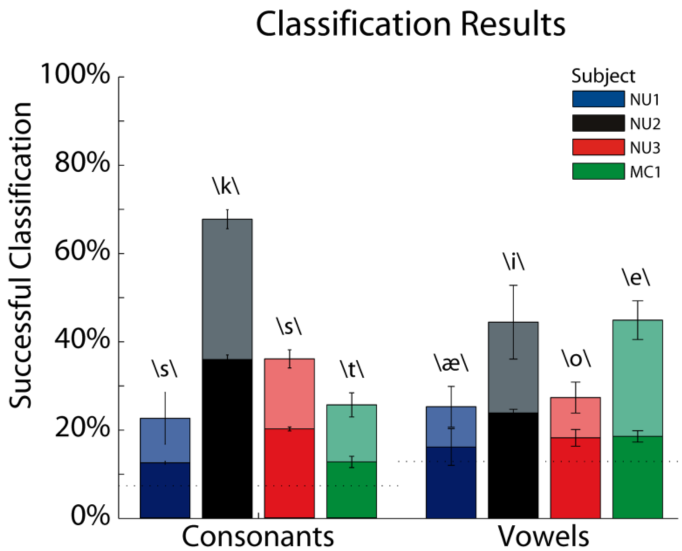


Figure 4. Classification results for phonemes for each subject. Chance percentage for each category is highlighted by a dotted line. Shaded bars indicate best performance of a single phoneme for that subject.

Interestingly, when a phoneme was misclassified, it was typically classified as its nearest neighbor within the International Phonetic Alphabet (IPA) chart of pulmonic consonants (Brown 2013), the authority for phonology research. Confusion matrices of results can demonstrate the exact misclassification of a classification algorithm, as variation from the diagonal reveals misclassification. In Figure 5, color denotes the percentage of successful classification of a given category of phoneme. In these confusion matrices, greater error in classification occurred within the thick black lines demarcating the place of articulation of phonemes. This classification failure therefore reveals properties of speech production that the International Phonetic Association has independently categorized. Further, notable exceptions to this rule reveal other properties of IPA organization. Phonemes that do not exist on the chart of pulmonic consonants, which include affricates (\ʧ\ in **ch**ip, \ʤ\ in **j**ump) as well as approximants (w in **w**in), had greater confusion with areas that would be considered their closest relatives within the IPA organization. Thus \w\, a labialized velar approximant, was confused with both labial phonemes and velar phonemes. Finally, results were most often distorted with phonemes that were not represented frequently within this data set, suggesting increasing the frequency of occurrence of these phonemes within the stimulus sets could improve classification results.

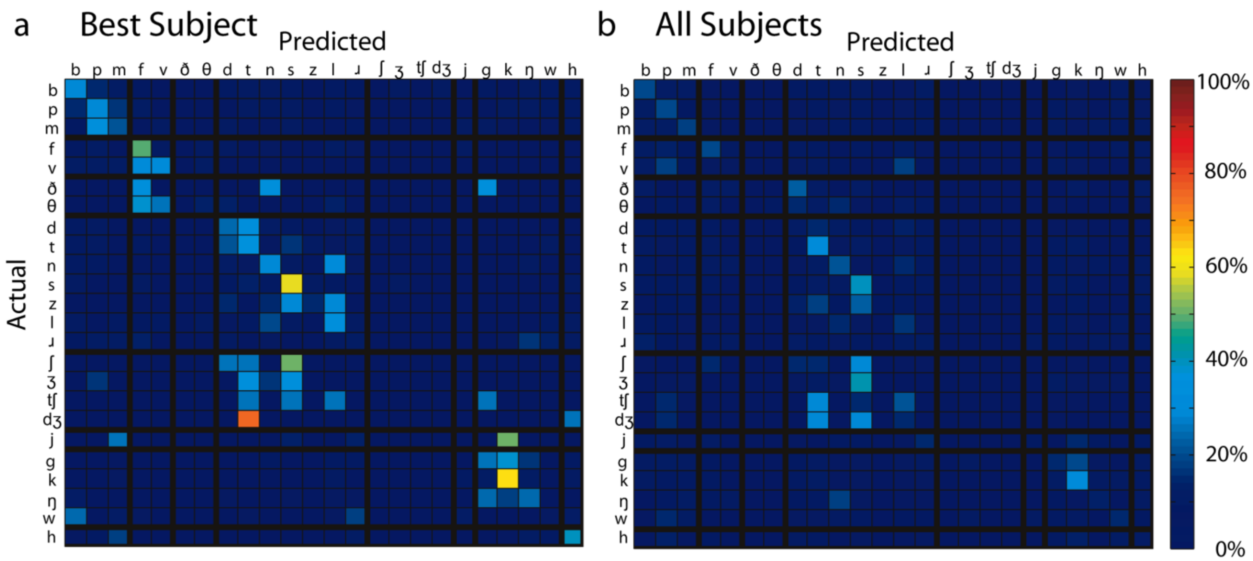


Figure 5. Confusion matrices of decoding results of consonant phonemes for the (a) best performing subject (NU2) and (b) an average across all subjects. Phonemes are sorted along the IPA pulmonic consonants place of articulation axis with thick black lines dividing place of articulation segments. Erroneous classification of phonemes often classified them as nearby phonemes according to IPA designation.

We further analyzed how performance varied due to the number of phonemes, the number of phoneme samples, and the accuracy of phoneme onset (Figure 6). To investigate the degree to which the quantity of data has affects results, classification was restricted to a subset of phonemes included in order of the frequency of their occurrence in the data set. This process yielded a maximum performance of 72.38% using 4 phonemes for Subject NU2. For all subjects, performance decreased until approximately 15 phonemes were included, when each additional phoneme had less than 10 samples. To investigate how classification varied due to quantity of phoneme samples, we restricted classification to subsets increasing in proportion of samples of the data set. Performance correlated with number of samples for all subjects, suggesting results could improve had there been more data. To evaluate classification dependence on the alignment to phonemic onset, we repeatedly reanalyzed the data after altering the onset time by random quantities drawn from a normal distribution. Performance sharply decreased with 100ms standard deviation of noise, which is less than the 176 ms average length of a phoneme in time (ranging from 75 ms for \b\ to 282 ms for \s\). These results demonstrate the critical need for precision in phonemic analysis, as performance sharply decreases as timing offset noise increases.

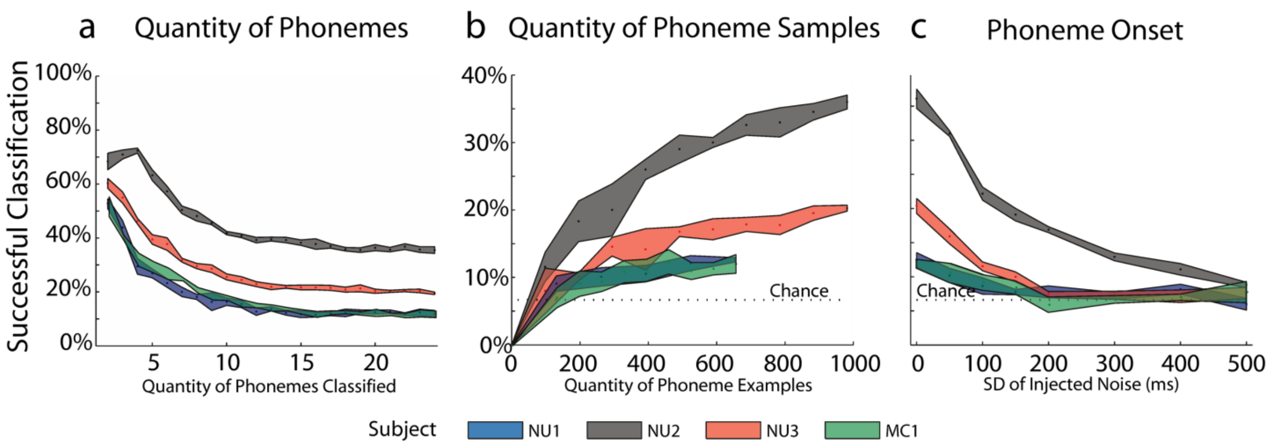


Figure 6. Dependence of classification performance on (a) quantity of phonemes, (b) quantity of phoneme samples, and (c) accuracy of altered phoneme onset. Traces correspond to classification results for each subject, plotting the mean and 1 standard deviation for 5 repetitions of randomized classification.

3.2. Analysis of Feature Contribution

To determine the factors that most influenced performance, we investigated the effects of the frequency bands, time bins, and quantity of electrodes used as features. Feature selection demonstrated high-gamma bandpower throughout the 65 – 250 frequency band strongly contributed to successful decoding. For Subject NU2, performance actually decreases with inclusion of frequency bandpower features outside of the high-gamma range. For Subject NU1, successful classification reached a maximum when limited to features in the 65 – 115 Hz frequency band; inclusion of the full range of high-gamma may therefore demonstrate overfitting. Further, other frequency bands – notably the mu and beta bands – are not critical to decoding performance.

A further investigation of what time bins most contributed to successful decoding indicated variability in importance of time among the subjects. Each time bin is 50 ms, and a minimum of 200ms (4 time bins) was necessary to classify phonemes for all subjects. Particularly, there range from 200 ms prior to phoneme onset to 200ms post-onset seemed most critical to decoding success, with causal features (prior to phoneme onset) being most critical. No evidence of a decrease in improvement of performance was noticeable with the inclusion of more time bins, suggesting that overfitting to time bins doesn't exist within our time boundary limits. Finally, because the time bin analysis is constrained to the relationship to phoneme onset, these results are normalized despite the range of phoneme length (75 ms average for \b\, 282 ms average for \s\), these results ignore differences in phoneme length. This process therefore evaluates phoneme onset classification fairly by not using information of phoneme length to improve classification results.

Primarily, performance was best when only incorporating data from the electrodes located over primary sensorimotor cortex. Although we discovered atypical functional organization for decoding phonemes for subject NU3, these subtemporal areas also responded to facial sensation and facial motor activity when directly stimulated. Although unique functional categorization may be related to epilepsy (G. A. Ojemann 1979; Springer et al. 1999), electrode contribution from all subjects supported facial motor activation. Finally, exploration of electrode dropping identified channels with casual activity related to speech production. All subjects had at least 5 or 6 electrodes that corresponded to speech articulation, and features were primarily selected from these channels for each subject (Table 1). Due to the widely varying placement of electrode grids in cortex, averaging results across subjects does not supplement data analysis but instead dilutes potential findings. Despite the standardized algorithmic process of classification, these subject-to-subject differences highlight best possible performance and factors that may have attributed that performance, particularly pertaining to electrode placement. Thus results are reported for each subject here separately.

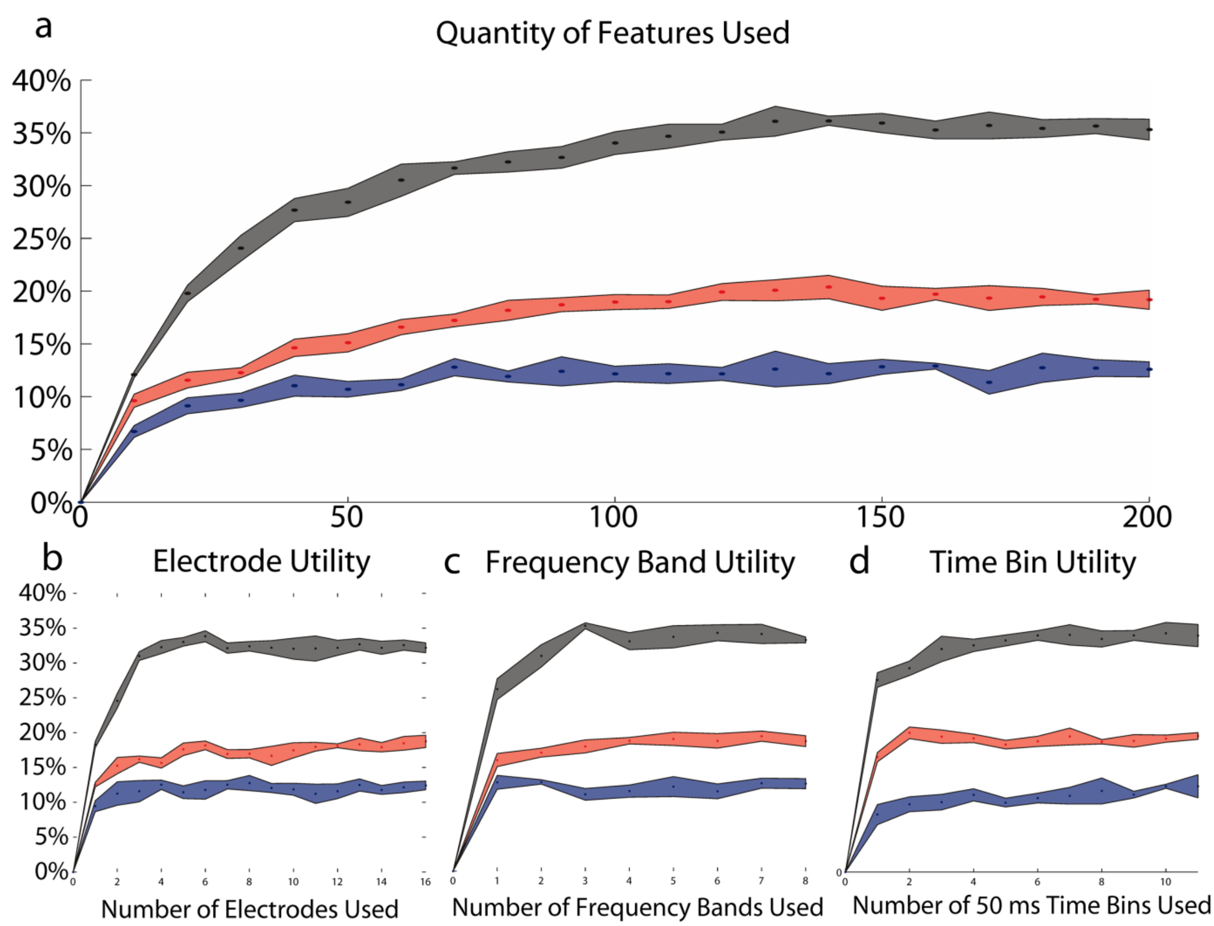


Figure 7. Feature Sensitivity for NU subjects. Successful classification results for (a) quantity of features used, (b) quantity of electrodes used, (c) quantity of frequency bands used, and (d) quantity of time bins used. Traces correspond to classification results for each subject, plotting the mean and 1 standard deviation for randomized classification.

3.3. Rhyme Reduction

The process of rhyme reduction did not improve decoding results, in both its subtractive (mean classification: 17.7%) and z-score (mean classification: 13.5%) formulations. The z-score approach was able to more accurately determine the features that contributed only when features were in the extremes and far from median activation values.

3.4. Information Transfer Rate

We calculated the classification success of all consonant combinations in our stimulus set with our best performing subject (NU2). We successfully identified 14.8% of these combinations *without having ever trained our algorithm to decode words* (chance = 0.83%, $p < 0.0001$, t-test). This investigation of phonemes within words notably outperforms simple joint probability of phoneme classification. When we constrained the predictions of phonemes for a whole word to only words used in the stimulus set, results improved to 18.8%. This rate equates to a gross information transfer rate of 3.0 bits/sec (equivalent to 33.6 words per minute) for a hypothetical BCI system.

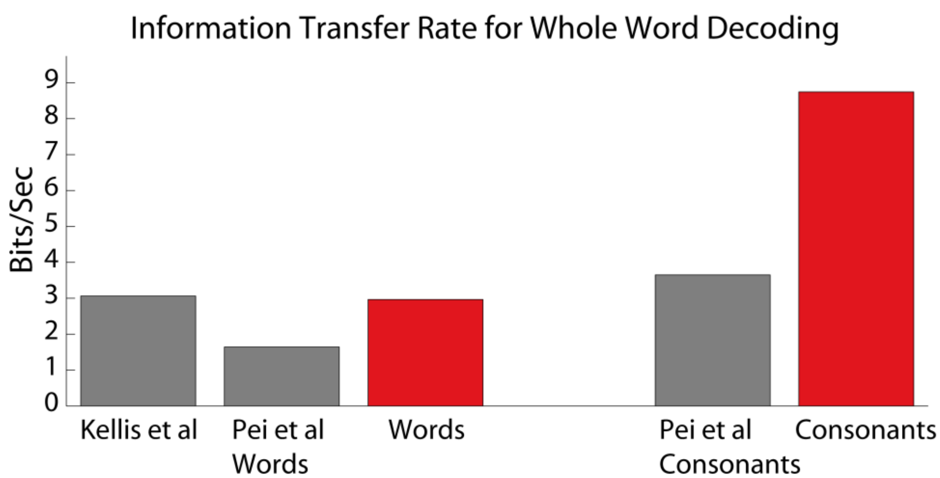


Figure 8. A comparison of gross information transfer rate for whole words and for phonemes compared with the relevant ECoG speech production literature.

4. Discussion

To our knowledge, this study is the first to decode the entire set of phonemes from American English using ECoG. It is also the first to successfully analyze and classify individual phonemes within word production. We found that that an event-related methodology enabled us to decode phonemes within words by aligning to the onset of each speech sound. Although other ECoG studies have classified phoneme production by comparing words with similar phonemes, by analyzing phonemes directly in context, we reveal properties of speech production that corroborate decades of phonetics research and can decode speech information efficiently.

To analyze neural activity during phoneme production, phonemes need to be precisely identified as events (Figure 6c). The high temporal precision required to accurately decode phonemes using this method suggests that there will be challenges for translating these algorithms to a real-time BCI for locked-in patients. Other methods may be necessary to detect onset of attempted speech production. This finding is further supported by the lack of improvement in classification with rhyme reduction. This negative result may indicate that the nature of phonemes changes in the context of position within words, and changes in word duration may affect decoding performance.

Our results suggest specific spatiotemporal guidelines for future endeavors into speech decoding using ECoG, advancing the science behind speech BCI development. Recording with higher electrode density over or neighboring sensorimotor cortices likely would improve decoding performance substantially. Although Kellis and colleagues showed that 5 electrodes over facial motor cortex with 1 mm spacing yielded best results for their 10-word stimulus set (Kellis et al. 2010b), our results demonstrate the likelihood that at least a 4 cm mediolateral span of speech sensorimotor cortex is necessary to decode phonemes articulated by separate articulator muscles. Thus, a high-density (1-2 mm) electrode array over an area of at least 4 cm of speech motor cortex may be optimal for decoding speech. Investigation of frequency content showed the high-gamma band provided the most information about speech motor activity. This is consistent with prior studies on hand and arm movements (Stark and Abeles 2007; T. Ball et al.

2008; Flint, Lindberg, et al. 2012; Flint, Ethier, et al. 2012). Finally, although speech production and speech reception are strongly related (e.g. McGurk effect (McGurk and MacDonald 1976)), the presented results are unlikely to be related to speech reception. Phonemes were primarily classified using causal features according to IPA designations for articulation, not auditory factors or frequency components of speech.

Comparisons with other speech ECoG studies are difficult due to limited data, differing analyses and electrode coverage, we can approximate comparisons by computing the efficiency (ITR) of our system (Figure 8). Our results compare favorably with those of Kellis et al. (Kellis et al. 2010b), which identified 10 words with 48% success using their 5 best micro-ECoG electrodes, and Pei et al (Pei et al. 2011), which identified 4 vowels and 9 consonant pairs at 40.7% (Figure 8). Although we did not directly train our decoders on whole words, we successfully identified phonemes of 14.8% of our 320 word set on a first attempt. Our best-performance volitional control of a single /k/ channel for Subject NU2, computed similarly, could yield a theoretical ITR of 32 words per minute, which is higher than many current BCI communication systems. Finally, speech recognition algorithms could be applied to phonemic results to leverage the frequency of phonemes within words in the English language to exclude impossible scenarios (e.g. words starting with \ŋ\).

Although word identification was not sufficient for communication purposes at a mere 18.8%, it is notable that words can be identified from phonemic analysis alone. This result outperforms joint probability of isolated phoneme prediction and indicates that classification across some words is better than others. It is important to note that we are not strictly decoding words, as we have prior information as to where phonemic onset occurs. However, the correct identification of phonemes within a word is a concrete step toward whole word decoding.

Despite these limitations, this study proves that decoding phonemic information from spoken words is not only possible, but follows guidelines of the phonetics research, confirming that ECoG is capable of decoding the fine motor components of speech production. While performance is good, we anticipate substantial improvement in classification with more phoneme repetitions, often limited due to clinical constraints. With further investigation and refinements in techniques, phonemic decoding using an ECoG-based BMI may provide efficient and intuitive communication. Similar to the ways in which a keyboard can provide a higher information transfer rate than a mouse, such interfaces may prove useful in areas of neuropathology and convert communication, and build toward implementable technology for individuals with locked-in syndrome and other communication disorders.

III. CORTICAL REPRESENTATION OF PHONEMIC CATEGORIZATION IN SPEECH PRODUCTION

Emily M. Mugler, James L. Patton, Felix Huang and Marc W. Slutzky

1. Introduction

Although investigators have proposed several different neural mechanisms of speech production, scientific inquiry has been unable to exactly identify these neural correlates. Attempts to determine such correlates using functional magnetic resonance imaging (fMRI) cannot record the rapid components and wide spatial area of activation of speech in cortex. Exploration of speech articulation with recent development in electrocorticographic (ECoG) recording have elucidated spatiotemporal neural mechanisms. One successful study demonstrated that speech production maintains some organization related to articulating musculature and can approximate categoric information of phonemes, the smallest separable parts of speech sounds (Bouchard et al. 2013). Although such phoneme distinctions can be roughly identified, the extent to which they can be incorporated and utilized in real-world applications remains unknown. Such a potential neural-speech interface application, if it could accurately decode the neural information related to speech production, could be used to facilitate speech communication in functionally "locked-in" individuals with advanced neuromuscular disorders. A deeper investigation therefore is required to analyze which properties of speech are robust in cortical organization and can further be leveraged to recognition ECoG speech of control.

Investigations decoding speech with ECoG have demonstrated some success at identifying sample groups of words (Kellis et al. 2010b), pairs of consonants of words (Pei et al. 2011), vowels (Leuthardt et al. 2011; Tankus, Fried, and Shoham 2012), and phonemes from within words (Chapter II). Phoneme decoding may surpass whole word decoding in terms of resulting efficiency of a speech application, represented by the information transfer rate of speech production. Moreover, we believe decoding phonemes would have the advantage of enabling transcription of any word instead of a limited set of words in a perfect decoding scenario. One potential confound to successful classification, however, is the existence of 39 separate phonemes in American English, which may overwhelm machine learning classification technqiues. Moreover, many phonemes are *homoganic*, sharing active articulator organs during speech production, potentially distracting the classifier and clouding results.

Linguistics and phonology have provided us with descriptions of exact mechanics of speech production, and have described separable, articulated phonemes in the Interational Phonetic Alphabet (IPA) (Ladefoged 1990; Brown 2013). We argue that a better approach would be to leverage the categorical designations of the IPA, potentially simplifying classification algorithms often confused by phonemes with similar designations. We investigated this mechanism in a broader approach that attempted to discern these categorical designations of phonemes to identify the phonemes themselves. In simpler terms, we sought to classify a phoneme by *first* classifying its descriptive IPA designations.

The IPA designations for consonants essentially define a state-space for each phoneme that is mutually exclusive according to three separate categories, which we can approximately consider as three dimensions of activation. The first dimension corresponds to the *place* of articulation. This refers to the point of primary constriction along the vocal tract, from the lips to the layrnx. The second degree is that of *manner* of articulation, related to the degree of constriction along the vocal tract from a plosives (complete constriction) to liquids (close to vowels). A final dimension for *vocalization* determines degree of phonation or voicing. This essentially acts as a binary indicator of one of two states, vocalized or not vocalized.

In this study, we evaluate how consideration of phonemic dimensions (place, manner, and voicing) might affect classification results and how these dimensions relate to the spatial and temporal aspects of corresponding cortical signals. We expand on previous efforts to identify specific phoneme by identifying phonemes using the distinctions of the IPA in order to leverage properties of phonation to identify exact articulation. Further, we identify which spectrotemporal features most contribute to successful decoding of phonemic classes. Finally, we compare these results to methodology used in automatic speech recognition technology.

2. Methodology

Four subjects (mean age = 42, 2 female), undergoing treatment for drug-resistant epilepsy, required extraoperative ECoG monitoring for epileptogenesis. These individuals volunteered for our study during their hospital stay and provided informed consent to participate in our study. Electrode coverage of cortex, determined by medical necessity, included frontal and temporal areas in all subjects.

This experiment involved protocol of visual presentation of words to be read aloud by the subject and data acquisition to align simultaneously recorded speech and ECoG signal. Experimental protocol consisted of presentation of words on a screen with a 4s inter-stimulus interval. The 300 words of the Modified Rhyme Test (House et al. 1963) were supplemented by 20 additional words to incorporate all General American English phonemes to the stimulus set. Subjects read each word immediately after visual presentation. Data collection was performed using BCI2000 software to present visual stimuli of words on a screen.

Speech was recorded with a MXL USB microphone at 44.1 kHz and synched to the RZ2; a randomized pulse was sent from the TDT system to the clinical system throughout the duration of each recording block in order to synch the data recording with the clinical Nihon-Kohden software. Sampling frequency was 500 Hz to 1 kHz to 9.6 kHz. Differential cortical recordings compared to a reference ECoG electrode were exported for analysis with an applied bandpass filter (0.53 - 300 Hz) with a sensitivity of 75 µV. Apparatus, and experimental set-up, and processing outline are depicted in Figure 9.

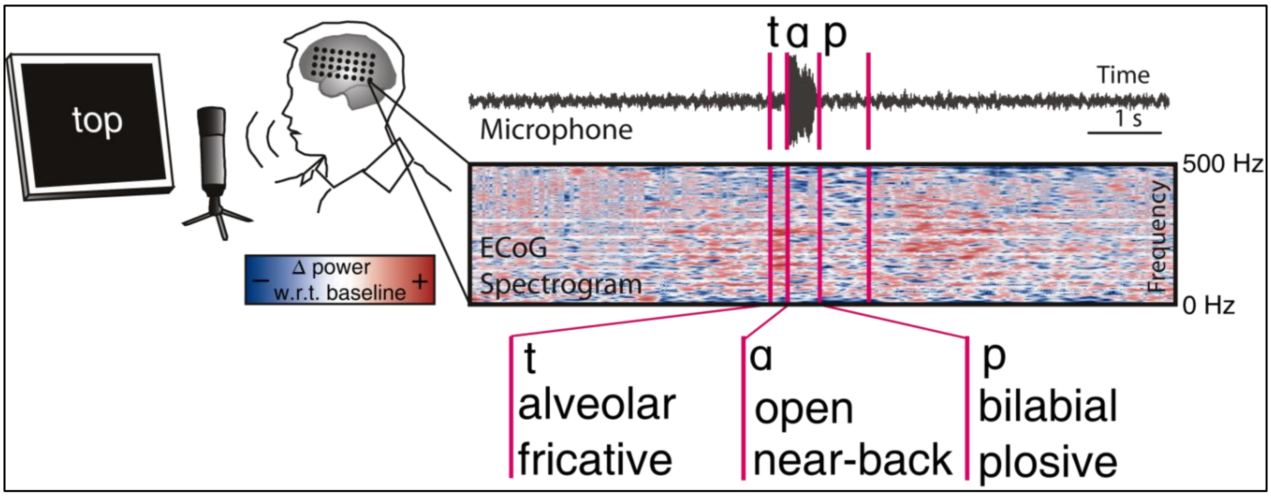


Figure 9. Experimental protocol and labeling for analysis. Subject reads words presented on a screen. Aligned microphone and ECoG Short-time Fourier Transform are labeled according to onset of phoneme articulation. Categorical descriptions, such as those depicted here, are assigned to the labeled data.

Data analysis primarily included alignment to phoneme onset, reduction to spectrotemporal features, and linear discriminant analysis on those features. Data was organized according to phoneme onset and converted to spectrotemporal features from 300 ms before phoneme onset to 300 ms post-onset. Conversion of data from Short Time Fourier Transforms (STFT) to simplified features consisted of summation over 50 ms time STFT segments relative to phoneme onset, as well as summation of bandpower changes in specific bandpower frequencies. Samples of features were organized according to their phonemic distinctions and classified using linear discriminant analysis (LDA). Features and data from the phonemic classification were relabeled, and multiple classifiers were simultaneously trained and tested on different dimensional labels within the phonemic test set. Thus for each 10-fold test set, the training labels on the data were categorized to fit either the *manner* of articulation or the *place* or articulation of each consonant. At the end of each 10-fold test set, for each phonemic consonant, one classifier predicted its manner of articulation and the other predicted its place of articulation. Classification of either categorical dimension was considered successful if it correctly identified the phonemic IPA designation of the given consonant.

In this study, two types of LDA classifier were used. *Multi-class* classification consisted of analyzing multiple phoneme classes in a single classification process to identify a single phoneme class as the likely category for a phoneme sample. In *one-versus-rest* classification (sometimes also referred to as *one-versus-the-rest* in the literature), one phoneme class is specified, and phonemes not in that class are combined into a “rest” group. In this scenario, specific features that lead to successful classification of any given phoneme distinction can then be specifically identified.

Bayesian statistics, in which conditional probabilities (dependence of one event upon the occurance or condition of another) are combined to predict events, have been used to describe many neural processes (Pouget et al. 2013). We applied Bayesian statistics to investigate whether categorizing IPA distinctions attributed to improved results for individual phoneme identification. The posterior probability for each phoneme dimension was therefore extracted during classification and used to evaluate the certainty with which the classifier could identify each given class. Posterior probability vectors for IPA dimensions were multiplied for each phoneme sample, and the combined probability of greatest certainty was considered to be the correct phonemic selection by the classifier and compared with the actual phoneme label. If the most probable pairing of predicted manner and predicted place of articulation could not possibly be combined to form an IPA consonant in American English consonant, an algorithmic constraint was applied to prevent such phonemically impossible results. In simpler terms, if the best guess resulting from the Bayesian simultaneous LDA classifier algorithm was not an American English phoneme, the classification with the next-best combined probability that satisfied the rule was the resulting predicted class.

For further elaboration on methodology, refer to Chapter II (see also Mugler et al, 2013).

3. Results

3.1. Classification of IPA Designations

Classification by phonemic distinction significantly outperforms chance classification across subject for manner and place of articulation ($p < 0.001$, t-test) (Figure 10). A maximum of 62.75% of phonemes were correctly classified by place of articulation for Subject NU2, with an ultimate high performance of 76.45% of categorized for any single constriction location (velar for Subject NU2). Identification of place of articulation had higher accuracy across all subjects. When classification results of place of articulation were combined across subjects, 41.18% (±13.0%) of phonemes were classified correctly, also a significant result ($p<0.05$, t-test).

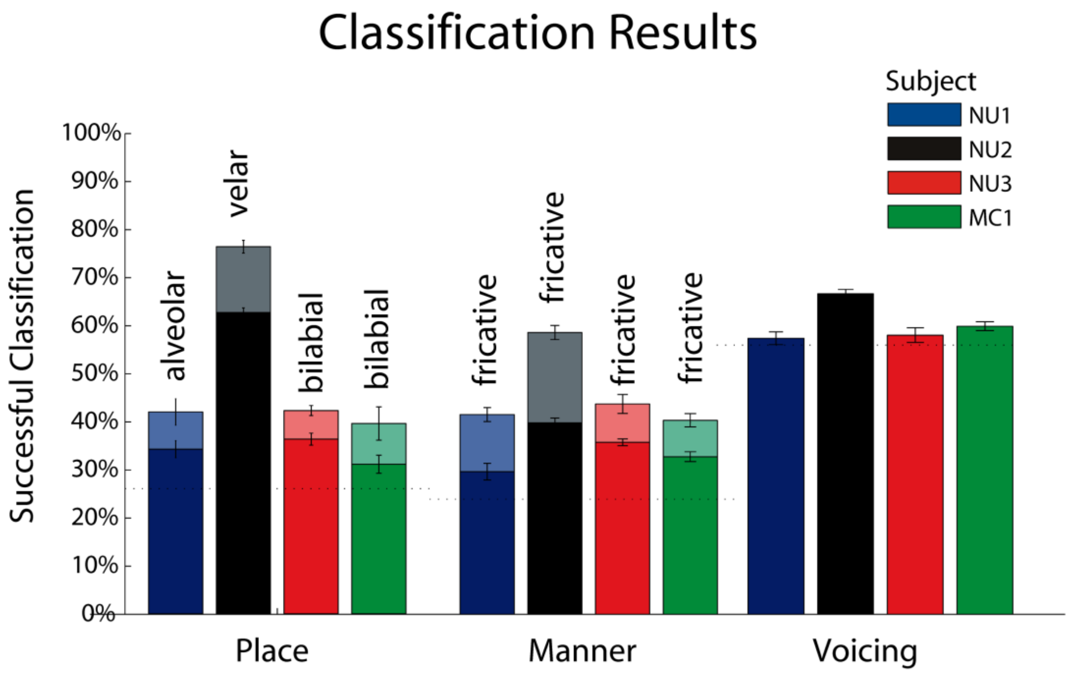


Figure 10. Classification results for phonemic distinction (place of articulation, manner of articulation, and voicing of articulation) for each subject. Chance percentage for each category is highlighted by a dotted line (place = 26.5%, manner = 24.3%, voicing = 56.0%). Shaded bars indicate best performance of a single class for each subject where applicable.

The best-performing place of articulation varied across subject – bilabial for Subjects NU3 and MC1, alveolar for Subject NU1, and velar for subject NU3. Interestingly, the fricative class of manner of aritculation performed best for all subjects, reaching a maximum performance of identifying at least half of all fricatives in Subject NU2. Potential contributing factors for this result include the fact that fricatives have the highest functional frequency in the data set, and therefore influence the LDA classifier more during classifier training. Further, these results also significantly outperform subjects' simple phoneme decoding performance ($p < 0.01$, t-test). These results reveal that wide inter-subject variability in phoneme category classification may exist.

For each of the two significantly performing classification categories, results represent a continuum of phoneme distinction represented in the data (Figure 11). If phoneme distinction was incorrectly identified for our best performing subject (Subject NU2), the nearest place of articulation or manner of articulation was typically the confused party. This confusion also suggests that the organization of the International Phonetic Alphabet can be independently substantiated using data from neural activity.

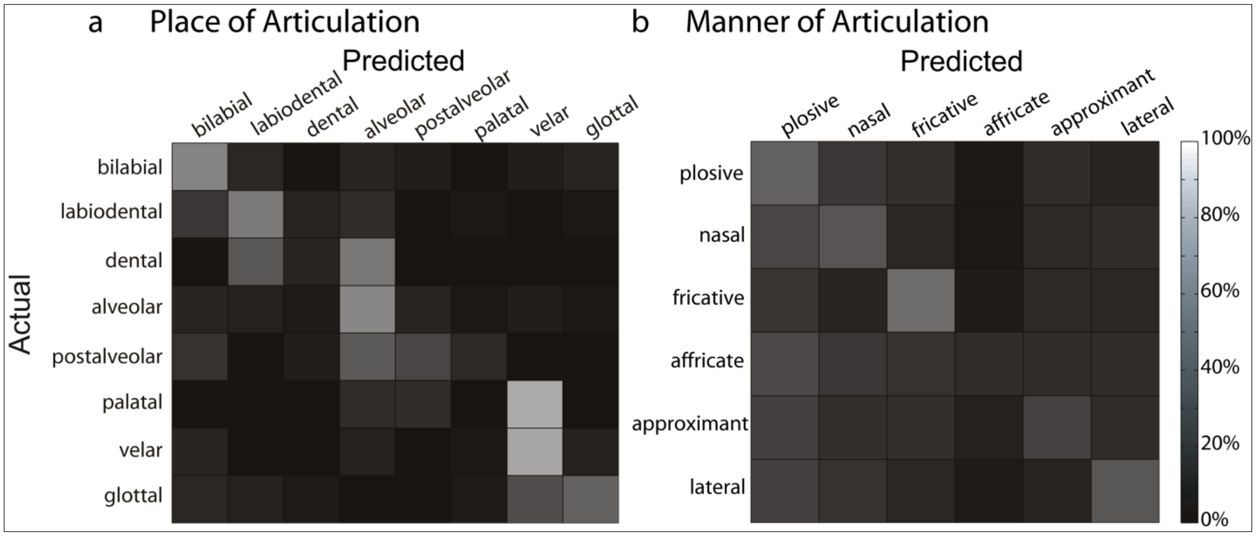


Figure 11. Phonemic classification based on IPA designations for Subject NU2. (A) Place of articulation and (B) manner of articulation for each phoneme are successfully classified along a continuum of phoneme distinction.

3.2. Bayesian classification

We investigated utilizing IPA categorical distinctions to improve results for individual phoneme identification. In 47.96% of our initial results, the LDA process originally misclassified one analyzed dimension (e.g. the manner of phoneme articulation) but classified the other dimension correctly (e.g. the place of phoneme articulation). In many instances, the most probable pairing of predicted manner and predicted place of articulation could not possibly be combined to form an IPA consonant in American English consonant. Our algorithmic constraint prevented phonemically impossible results, which initially occurred in 34.7% of phoneme samples.

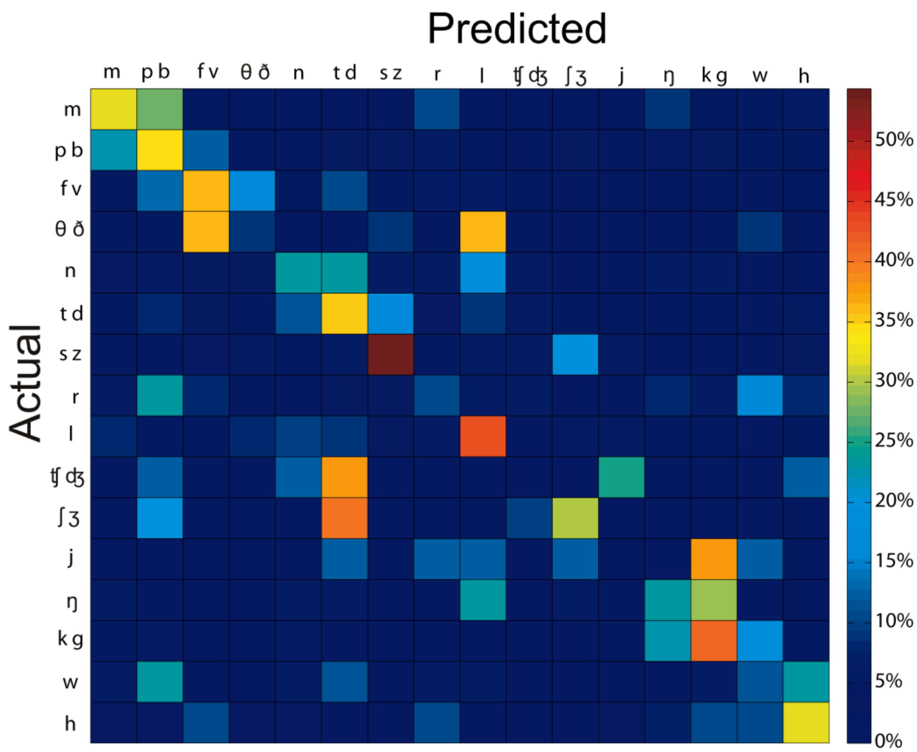


Figure 12. Bayesian phoneme classification with IPA constraints for manner and place of articulation - results for best-performing Subject NU2.

Multi-class LDA results using only manner and place demonstrate performance on par with previously reported LDA results identified by phoneme alone (overall success rate: 35.20%, peak: 54.39% at “s/z”) (Figure 12). The existence of most of the data categorization occuring along the diagonal indicates successful classification with confusion occuring in the direction of neighboring phonemes that share similarities of place of articulation. Further, the high values for confusion of “ʧ/ʤ” with “t/d” (> 40%) demonstrates that the first phoneme of an affricate may be more influential to the neural production of that phoneme.

These aforementioned Bayesian multi-class LDA results neglect the voicing dimension of consonants, the third dimension of pulmonic consonants as designated by the IPA. An example of the difference along this dimension is the activation of the larynx in articulation of the phoneme \b\ and the supression of laryngeal activity during utterance of \p\. Incorporating this dimension could further enable differentiation between such phonemes, which cannot be completed from the current organization. When all three dimensions (manner, place and vocalization) of phonemes are classified, performance of phoneme prediction decreases to a best-performance 25.51% and average 15.77% (±7.6%) across subjects (Figure 13). This result still performs significantly better than chance (7.7%, $p < 0.05$, t-test), despite the large decrease in phoneme prediction compared with manner-and-place classification. In comparision to previously work decoding by phoneme label (Chapter II), classification along IPA dimensions decreases in performance when vocalization was included in the Bayesian computation. However, the performance of mutli-class LDA classification on phonemes succeeds without the vocalization stipulation and can leverage the categorical dimensionality of LDA designations.

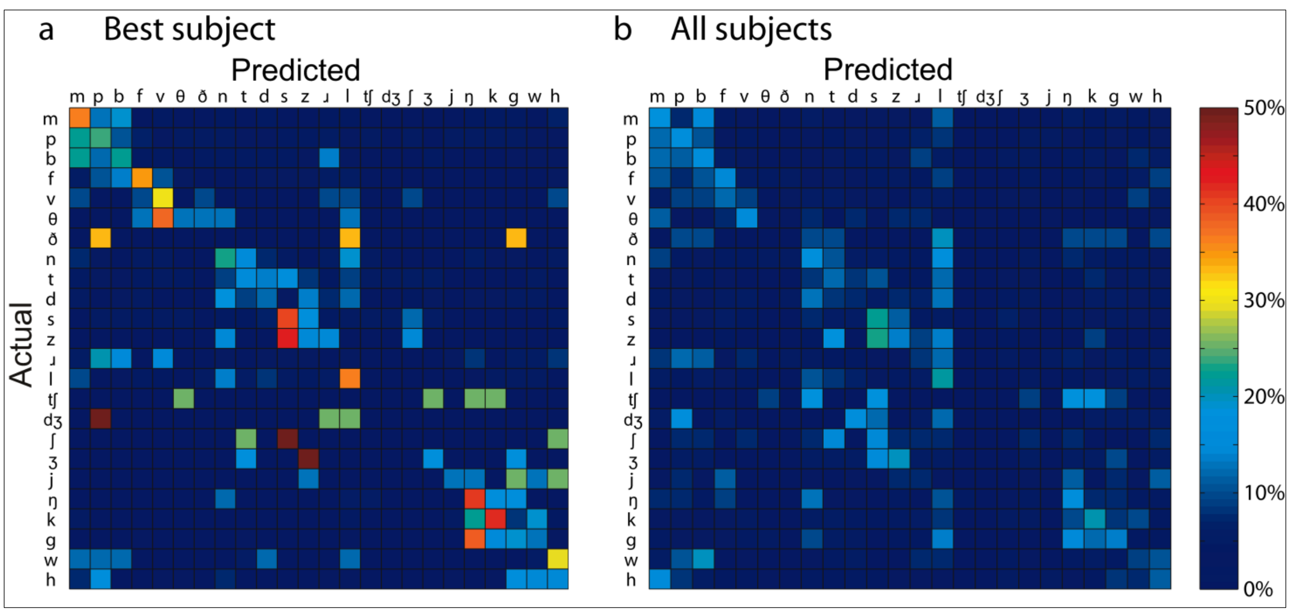


Figure 13. Confusion matrices for Bayesian-classified phonemes for (A) Subject NU2 and (B) all subjects. Classification confusion is predominant along place of articulation designations.

3.3. Phonemotopic Organization of Sensorimotor Cortex

By using linear discriminant analysis to compare one categorical group to the rest, we isolated and determined features that contribute to successful classification. By employing a one-versus-rest LDA classification process, we highlighted such features and find a general somatotopic pattern of cortical recordings. Here, we depict the features used on the expanded channels over sensorimotor cortex, and a time-frequency map of the features that contribute to successful one-versus-rest decoding (Figure 14**)**. The intensity of the color correlates with the significance of that feature in discriminating an isolated group compared to the rest of the data (one-way ANOVA). Specific electrodes often contributed to specific group decoding, and these electrodes are highlighted in this figure. These results depict a continuum of utterances represented along what is traditionally thought of as the Penfield speech sensorimotor cortex. The continuum of phonemic decoding previously reported for multi-class decoding (Figure 11) is therefore also represented along the physiology of the cortex. These results therefore support in part a validation of the Penfield motor map as it relates to active speech. It thus reveals a "phonemotopic" map in which position in sensorimotor cortex designates correspondence with phoneme category.

Although there are some channels which involve activation during articulation of any phoneme, activation recorded by channels can still reveal isolated differences between IPA classes. For example, for Subject NU2, one electrode's features exclusively decoded bilabial consonants, yet the channel was activated indiscriminantly during phonation (noted in red in Figure 14). Indeed, most phonemes involve lip movement in some way. However, the amplitude of the activation on this channel indicated the degree of involvement of that feature. Also important to note is the existence of multiple channels contributing to successful decoding of specific categories. Notably, the labiodental and alveolar distinctions feature activation on 2 neighboring channels in Subject NU2.

Because cortical activation for speech production correlates with the place of articulation, we further investigated electrode placement. In this study, subjects' coverage spanned an approximate total area of 46 (±8) cm^2, but electrode coverage contributing to best performance specifically included 4 cm frontotemporal span, likely covering sensorimotor cortex. Results indicate that to decode all parts of speech, electrode coverage must include cortical areas spanning lip motor to glottal motor areas. Because coverage varies so greatly from subject to subject and corresponding neural signals are so distinct in their location, we can conclude that electrode coverage uniquely contributes to successful classification of phonemic group.

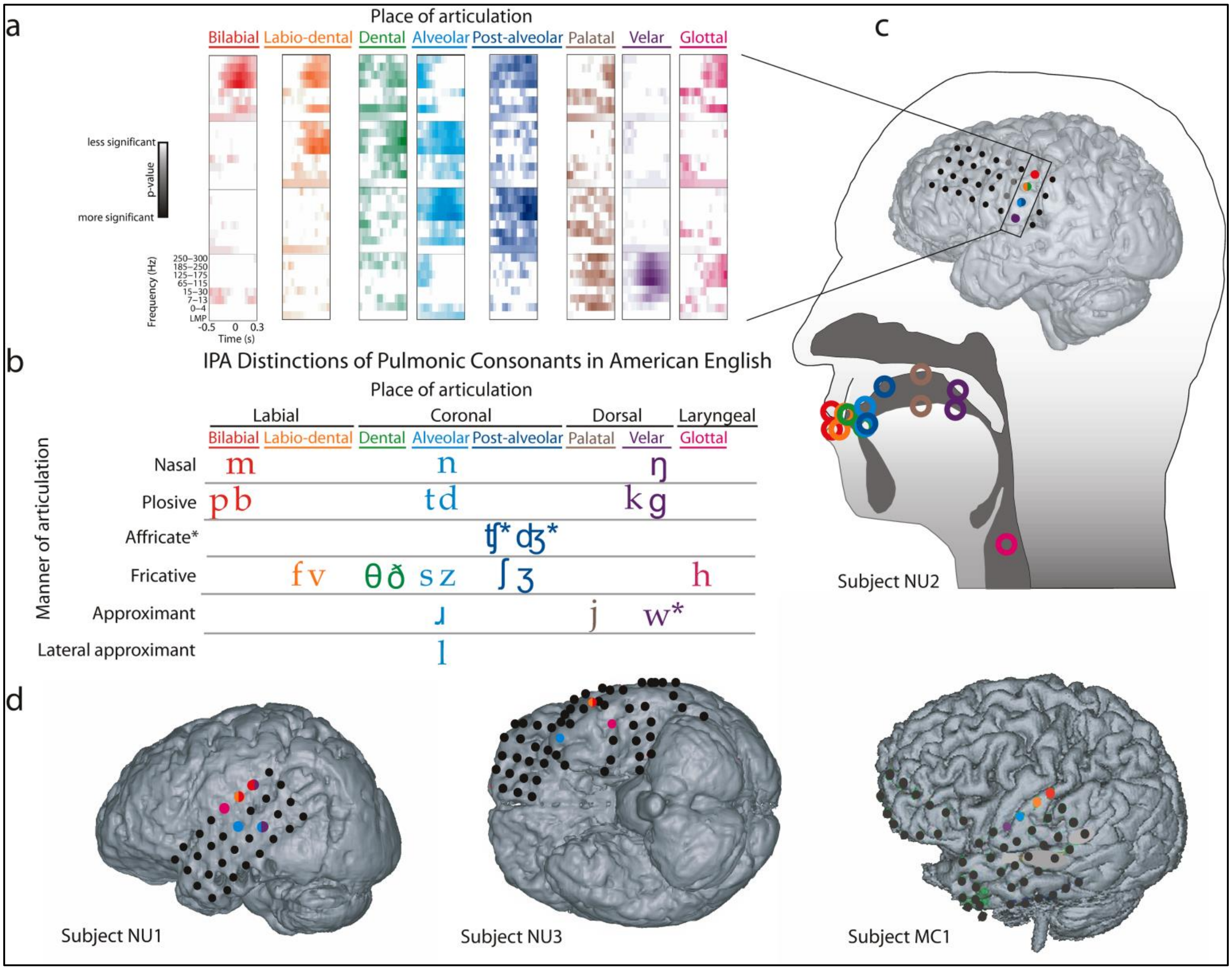


Figure 14. Phonemotopic organization of cortex. (a) Z-score increases in 50 ms frequency bands for each phonemic group in comparison with other phonemics groups of features for each highlighted channel. Each of the four small boxes is a mini-FFT of a given channel (b) Chart of pulmonic consonants adapted from International Phonetic Alphabet (Brown 2013; Ladefoged 1990). Only pulmonic consonants in the General American accent of English are represented and highlighted in colors corresponding with location of articulation in (c). Non-pulmonic consonants (approximants and affricates) are denoted with an asterisk in location of nearest pulmonic relation (c) Anatomically accurate tracing of Subject NU2 with superimposed 3D reconstruction of the brain (Hermes et al. 2010). Electrodes with features significant to decoding corresponding articulatory movements highlighted; some channels share dominance on a specific channel. High-gamma band activity increases during pronunciation for these channels, supporting a somatotopic hypothesis.

3.4. Computation of Information Transfer Rate

The ultimate goal of this technology is to not only reveal mechanisms of speech production, but also to enable classification of speech sounds during articulation. To leverage the classification of IPA categorical definitions during speech articulation, we predicted the actual phoneme according to Bayesian statistics of predicted phonemic categories. This two-step process of predicting categorical designations and then leveraging statistics to investigate probable phoneme was then used to calculate the efficiency of the speech articulation classification in terms of rate of information transfer.

In order to investigate the actual word decoding rates, phonemic predictions were re-sorted according to their original placement within a stimulus word. This enabled some identification of complete words based upon the combinations of resulting LDA predicted categories. 9.2% of words were able to be decoded by this methodology (chance = 0.83%, $p < 0.001$, t-test). This result, converted into gross information transfer rate of the system, is 0.37 bits per second, roughly equivalent to 2.5 words per minute.

4. Discussion

This study demonstrated the extent to which phonemes could be accurately classified by their linguistic properties from ECoG recordings during speech articulation. Our results indicate that the organization of the International Phonetic Alphabet is paralleled in patterns of neural activation. Classification by IPA category outperforms traditional phoneme decoding (Figure 10), likely because the existence of fewer phoneme categories simplifies performance of the LDA classifier. Place of articulation is the strongest predictor of classification success, followed by manner of articulation, and then followed by voicing of a phoneme. This indicates that our methods more accurately specify the place of articulation of phonemes during speech production, but it may also imply that the neural activation based upon where constriction occurs in the physiology of the vocal tract is more predominant in cortical representation.

Phonemic categories were combined using Bayesian conditional probability to predict phoneme (Figure 13), but results do not differ greatly from previously reported LDA classification results simply trained on phonemic consonants (36.1% best performance) (see Chapter II). This lack of improvement suggests that rarer categories (e.g. post-alveolar, lateral approximant) are more frequently neglected in broader categorical classification, possibly due to less frequent incidence within the data set, and may combine to further neglect the less-common phonemes. The similarity in classification confirms that Bayesian combination of IPA category can be leveraged to identify phoneme and builds toward decoding words from training on phonetic designations. These results confirm that using phonetic training for a classifier can lead to full word identification.

Phonemotopic organization was revealed by utilizing LDA in a one-versus-rest classification paragdigm (Figure 14). Identification of phonemes based on places of articulation showed activation on channels spanning speech sensorimotor cortex. This result confirms the Penfield motor map during speech articulation of full words, which to the best of our knowledge is the first study of its kind to achieve. These results expand on the ECoG-speech literature depicting an organ-based map during phoneme articulation (Bouchard et al. 2013). While their work indicated there may be atypical maps for vocalization components, we found no such atypical organization as it applied to laryngeal activation during vocalization. This also demonstrates the application of LDA to enable isolation of specific spatial areas of activation, which could potentially be used to design a better speech interface. Finally, these results confirm the utility of ECoG electrodes to access and identify neural signal related to phoneme production during speech articulation.

Placement of electrodes strongly impacted classification success, elucidating how critical electrode placement is to phoneme class prediction. Epileptologists selected grid locations specifically for medical observation of onset of epileptic seizures, and location of implantation therefore varied for each patient (see Figure 14). With our one-versus-rest classification paradigm, it is therefore possible to infer that classification differences in place and manner of phoneme articulation are indicative of areas of coverage. An apt analogy for decoding speech sounds without full speech sensorimotor coverage would therefore be typing with a keyboard with only a small selection of a subset of working keys. Accordingly, classification results from Subject NU2 showcase proper grid placement and set the benchmark for future work in speech-BCI decoding. We therefore emphasize results of our best subject to show what is possible, and not the average across all subjects, though it is important to note that both results are significantly better than chance. The methodology of pre-processing and classifying the data can therefore adequately discern properties of speech during production despite non-ideal electrode placement. Further, this methodology also confirms that electrode placement critical for future endeavors to decode speech information from ECoG.

The information transfer rate (ITR) of our Bayesian speech decoding paradigm is comparable to previously reported results (see Chapter II), and could be modified to further increase communication efficiency. Bayesian phoneme prediction using phonemic dimensions requires a multi-step process, as a BCI that uses only IPA designations of speech would not be capable of directly transcribing speech. A greater ITR could be achieved by including additional speech-specific channels for subjects where electrodes did not cover active areas (e.g. bilabial or velar channel). For example, a place-of-articulation brain-computer interface might not be able to successfully decode active speech, but it could still serve as a multi-channel interface, potentially more efficient than many current/conventional communicative BCI paradigms. Despite these issues, the ITR of our Bayesian multi-step process achieved 2.5 words per minute. Outside of previously-reported ITR of phoneme decoding 3.0 bits/sec (33.6 words per minute) (Chapter II), this ITR surpasses traditional BCI paradigms and approaches the fastest previously-reported communicative BCI that employ visual evoked potentials (Chapter I.2, Table 1). Full word prediction of words within the test set can be achieved with some success with both phoneme and phoneme category approaches.

Finally, the aforementioned results achieve comparable success to traditional automatic speech recognition (ASR) mechanisms. ASR algorithms typically analyze a single microphone channel to infer the phonemic information. Current models succeed when trained on a massive quantity of speech data. However, these formulae perform poorly when trained on small data sets. Accordingly, our results are roughly equate with those reported in ASR research produced in the 1990s, prior to creation of massive audio databases, or those reported in the ASR literature on specifically reduced data set sizes (34.7% success for 12 minutes of speech data) (Moore 2003). More study is therefore needed to evaluate whether database use would similarly improve ECoG classification results. Further, the methods here leverage classic ASR techniques of training a classifier phonemically to allow for whole word prediction.

The implications of this work, identifying neural correlates of speech production, affect neuroscience and brain-computer interface research in addition to linguistics. Application of one-versus-rest LDA classification enables spatial isolation of features that contribute to speech production and point to sources of failure in classification. Despite conditional properties of many neural processes, Bayesian application of linguistic principles to decoding phoneme production does not improve results of phoneme investigation on their own. Nevertheless, the results of this study rule out certain possible approaches, advancing communicative brain-computer interface research and raise new issues in decoding the nature of speech production in cortex.

IV. STRUCTURE OF CORTICAL CORRELATES OF SPEECH PRODUCTION

Emily M. Mugler, James L. Patton, Matt Goldrick and Marc W. Slutzky

1. Introduction

Speech is a complex human activity, involving synchronization of neural processes in a wide array of cortical regions. Current models of speech production have been created using data from several sources, including results of functional magnetic resonance imaging (fMRI) studies revealing areas of activation, diffusion tensor imaging studies depicting pathways of activation, and natural speech errors such as slips of the tongue. Despite such advances in classification of neural signal of speech, the functional architecture of speech processing during production remains poorly understood. Electrocorticography (ECoG) has recently enabled identification of neural activity during the speech process (E M Mugler et al. 2013; Bouchard et al. 2013; Kellis et al. 2010b; Pei et al. 2011; Blakely et al. 2008; Roland et al. 2010; Leuthardt et al. 2011). Bouchard and colleagues, for example, demonstrated that functional activity in speech motor cortex generally follows the classic somatotopic motor map of organs of articulation (W. Penfield and Boldrey 1937). ECoG technology is capable of recording from vast areas of cortex on a fine time scale compared to other technologies (Slutzky et al. 2010), which we argue will be crucial for understanding rapid fine motor processes such as articulation. While these recent studies have described cortical activity during speech production, the data from these investigations have not yet been leveraged to scientifically evaluate the validity of models of speech production.

We sought to compare and contrast three leading models to uncover the nature of organization of speech sound in cortex. New modalities of neural signal analysis, such as ECoG, enable comparison and evaluation of theoretical models of cortical activity. For our purposes, we consider models which describe functional neural mechanisms and demonstrate predictive power of cortical function. These models all differ in the hypothesized structures of speech directly preceding production. The first hypothesis, identified by Bouchard and colleagues, is that *organs of articulation* are most indicative of organization in cortex (Bouchard et al. 2013). Contrastingly, the second model hypothesizes that *phonemes* are the predominant representation of cortical organization (Hickok 2012). Finally, a third model hypothesizes that activity in speech motor cortex would be most identifiable by a *gestural* model (Browman and Goldstein 1992). Though such models have been described in theoretical and simulated models, they have not been used to evaluate or verify in fine electrophysiological recordings of cortical signal.

Models of neural function can be quantitatively compared with data to confirm or reject their related hypotheses. Such neurophyiological signals have long been used to identify cortical correlates of human activity. For example, in decoding directional tuning from neurons, it is unclear whether motor cortical representation is correlated primarily with force (SH Scott and Kalaska 1997) or with position of movement (Georgopoulos, Schwartz, and Kettner 1986). This uncertainty is due in part to the fact that force and motion share many common features, increasing the difficulty in distinguishing differences in functional organization of the data. Moreover, the scientific hypotheses about force versus position dominance in motor cortex have been evaluated by the extent to which each paradigm or model can predict and identify cortical signal. More simply, competing hypotheses can be substantiated or rejected by specifically looking for and comparing degree of representation within the neural recordings.

We propose to extrapolate the aforementioned idea to the domain of speech, another neuromotor process. In fact, because the models described above intend to be comprehensive in their descriptions of how speech production is represented in the cortex (e.g. all phonemes, all articulatory gestures, all motor organs of articulation), they can be directly compared by how well they can be decoded from the data. We can evaluate whether speech motor cortex has a predominantly *phoneme*-based (Hickok), *gesture*- or articulatory-action-based (Browman & Goldstein), or *organ*-based functional representation. This investigation of competing hypotheses parallels methodology used to identify the nature of directional tuning found in cortex. We seek to deconstruct neural correlates of the cortical signal controls the complex mechanisms of speech production.

In this study, we used classification of electrophysiological signal during speech production as a tool to investigate to what extent to which electrocorticographic signals parallel the findings of the literature from linguistics and phonology (Table 3). Using linear discriminant analysis (LDA) with a simple exclusive decoding paradigm (e.g. {b} or {not b}), we evaluated the degree to which categorical designations within each model could be discriminated. ECoG signal and speech were recorded simultaneously during speech articulation by 4 subjects of simplistic words from the Modified Rhyme Test. Data analysis was precisely aligned to onset of each phoneme, as our results from previous studies (see Chapters II and III) have demonstrated the need for exact phoneme onset times for successful classification of ECoG signal. The International Phonetic Alphabet distinctions was used identify phonemes used in speech production. The Task Dynamic model of inter-articulator coordination was used as a basis for the gestural model and to identify organ distinctions (Saltzman and Munhall 1989). Each model's categorical designations were classified and cross-validated on spectrotemporal features of phoneme-segmented samples relative to rest activity, and this process was repeated 5 times to bootstrap data. Z-scores were calculated for each category by comparing classification results to mean and standard deviation of chance results. To calculate chance values, labels were randomized for each exclusive category and classified on the data set 10 times. Results, when compared across comprehensive models, would determine the extent to which a model was successful in describing functional organization by performance across all model categories.

Table 3. Comparison of speech production models and how they represent production of the word “kit”. Abbreviations include tongue body (TB), tongue tip (TT) and glottis (G).

	Representation of “kit”
Phoneme Model	$\boldsymbol{k} + \text{ɪ} + \boldsymbol{t}$
Bayesian Phoneme Model	$(\boldsymbol{velar}\vert\boldsymbol{plosive}) + (\boldsymbol{near\ front}\vert\boldsymbol{near\ close}) + (\boldsymbol{alveolar}\vert\boldsymbol{plosive})$
Articulator organ model	$(\boldsymbol{TB}) + (\boldsymbol{Larynx}) + (\boldsymbol{TT})$
Gestural Model	$((\boldsymbol{TB_{close}})_{\boldsymbol{velar}} + \boldsymbol{TB_{release}} + \boldsymbol{G_{wide}} + \boldsymbol{V_{close}}) + ((\boldsymbol{TB_{v}})_{\boldsymbol{palatal}}) + ((\boldsymbol{TT_{close}})_{\boldsymbol{alveolar}} + \boldsymbol{TT_{release}} + \boldsymbol{G_{wide}} + \boldsymbol{V_{close}})$

2. Results

2.1. Gestural model best identifies cortical organization

Model evaluation and comparison was executed by calculating the extent to which each distinguishing speech category could be exclusively decoded from the data set. Quantitative comparison of each model was calculated by determining the percentage each category could be exclusively classified from the whole of the speech data. Thus, the more a category was included in the data set, the greater representation it had in the evaluation. In this way, an entire model, and how well it maps out each of its categorical designations from the data, can be represented as a circle. The percentage of the circle (i.e. θ for each segment) correlates with the relative frequency of that category within the data set. This novel twist on a traditional plotting method enables visualization and comparison of model success in classification.

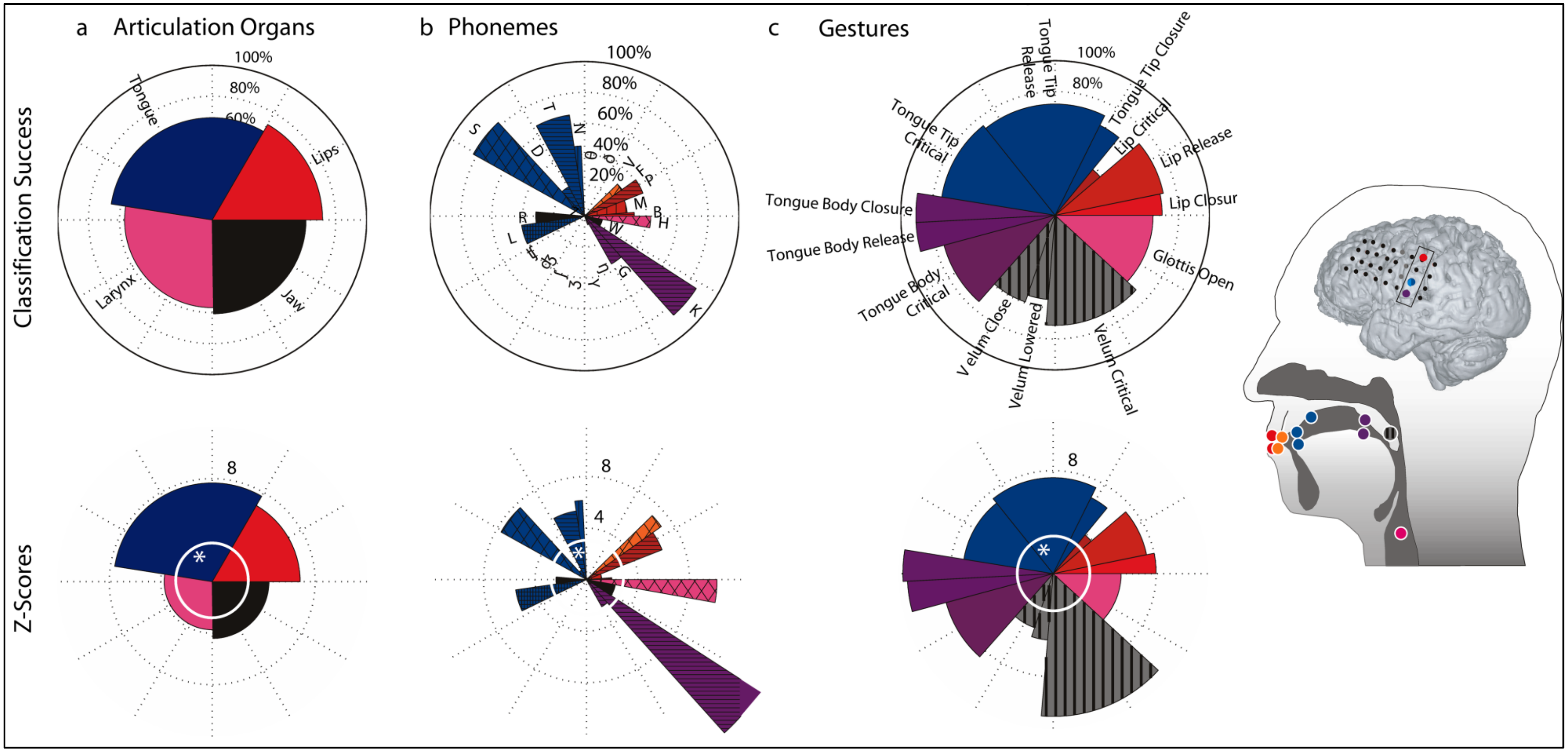


Figure 15. Results for three competing hypotheses for the subject with most complete electrode coverage of speech sensorimotor cortex. Top row indicates success of exclusive classification ({b} versus {not b}) of each category of each model; (a) 40.04% for articulation organ model, (b) 13.07% total coverage for phoneme model, and (c) for ultimate best 49.51% gestural model. Percentage of the whole (θ for each segment) represents relative frequency of a given category within the data set. Bottom row indicates statistical significance for classification of each category; the extent of each radius represents the z-score for exclusive classification of each category. Statistical significance is identified on these plots for each plots in the this row by a white circle ($p < 0.01$) At right, color-coordinated map of vocal tract constriction locations and electrodes determined to have correlation with those constriction locations.

Results are depicted for each model using data from one subject in Figure 15 for each model. The gestural model demonstrated the most total representation within the cortical signal analysis, explaining a total of 49.51% of the information in cortex. Contrastingly, total proportional area for organs of articulation was 40.04%, and a phonemic model could only classify 13.07% of the data. The second row of figures demonstrates the z-scores (and significance) of each model. Note that for all categories of the gestural model, classification results are statistically significant. Although data classification results may vary slightly across subjects, the gestural model surpasses classification success for other models (33.2%, compared to 29.8% and 7.0%), (Figure 16) though comparative model success to articulator organ model is not statistically significant.

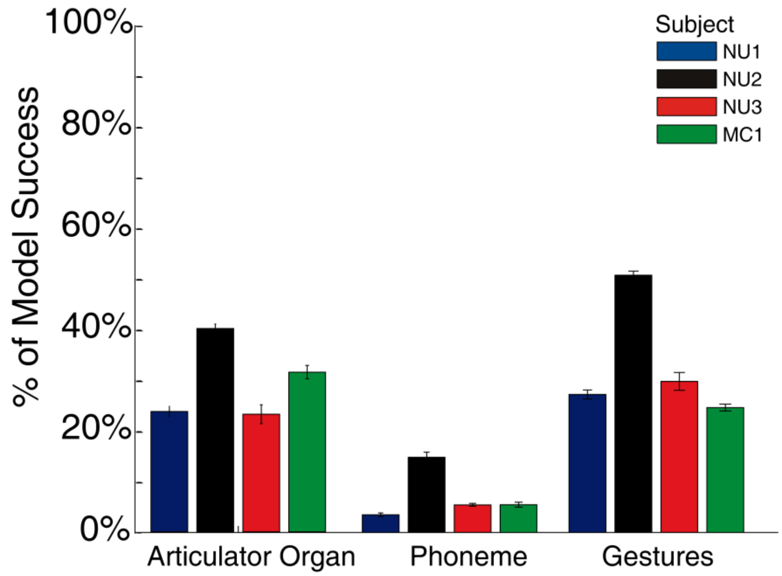


Figure 16. Percent successful description of each model for all subjects. Mean performance of gestural model outperforms both articulator organ model and the phoneme model.

2.2. Specificity of models affect performance

LDA classification can relatively penalize classification when large quantities of classes are compared. To investigate the degree to which the methodology may have unfairly evaluated less specific models, we investigated the phoneme model to a greater extent with broader categories. We specifically investigated the degree to which the *manner* of articulation of phonemes and the *place* of articulation of phonemes could be exclusively decoded from the data set. We investigated this broader categorization, because previously reported results have indicated that these designations, outlined by the International Phonetic Association, can be decoded from ECoG signal during speech (see Chapter 3). Results demonstrate that even with the broader categorization (Figure 17), the phonemic model still does not surpass the gestural model. Total proportional area for each model in this broad paradigm was 33.37% for place of articulation representation and 33.80% for manner of phoneme articulation.

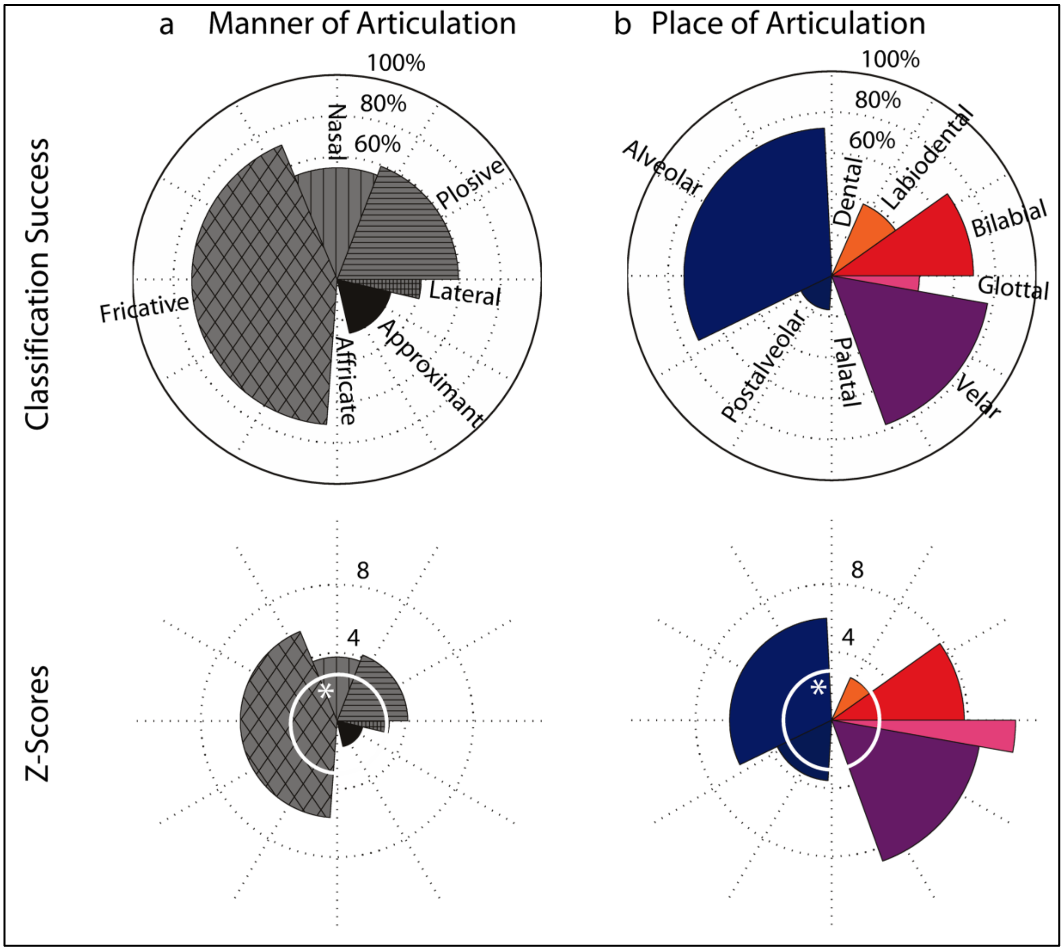


Figure 17. Broad representation of the phonemic model in cortex. The representation in cortex for Subject NU2 that encapsulates the exclusive decoding of the manner and the place of articulation as identified by the International Phonetic Alphabet. 33.37% of the data was explained by the place of articulation of phonemes and 33.80% of the data was explained by manner of phoneme articulation.

2.3. Prediction of speech works best with gestural model

A true test of the degree to which the data can be explained by a specific model is the degree to which we can predict patterns of neural activation. We investigated the degree to which we could predict speech based upon our best-performing model. Just as phoneme was extrapolated to infer gestures or articulator organ, we combined classification results of articulatory properties to predict actual phoneme during our recordings. Phoneme predictions were then used to predict what word was used in our stimulus word set. Using this strict method, only 1% of original articulated words were correctly identified by their analyzed categorical components (chance =

0.83%). This non-significant result implies that a more rigorous multi-step Bayesian process using categorical distinctions for the gestural model is needed to predict original phonemes (Chapter III). Word prediction methodology could not be similarly extrapolated to the articulatory organ model with any certainty.

3. Discussion

To our knowledge, this study is the first of its kind to utilize electrocorticographic signal to investigate the categorical organization of cortical representation during speech production. Our novel exclusive classification approach enabled evaluation of contrasting models of speech production developed over decades of interdisciplinary research. Our results indicate that gestural models, such as the task dynamic model of inter-articulator coordination (TADA) and the directions-into-velocities of articulators (DIVA model) (Guenther, Ghosh, and Tourville 2006), best describe the activation and functional organization in speech sensory motor cortex. Although phonemic (13.0% of data explained) and organ-of-articulation (40.0%) models are capable of identifying specific groups ($p<0.01$), the gestural model accounts for a greater percentage of classification success in comparison (49.5% of data explained) (Figure 15). This strongly indicates that a gesture-based model dominates cortical activity during phonation. This conclusion holds even when more specific models are analyzed according to broader categorical designations (mean: 33.5% of data explained) (Figure 17). Analyzing broader classification of a phoneme model does not improve model performance beyond that of competing hypotheses. In fact, the best-performing model has 3.25 times the quantity of categories of the articulator organ model. The speech sensorimotor cortex is considered to be the final cortical area in the speech production pathway, and these results point to gestures as the format of its final execution.

By analyzing the extent to which contrasting models can account for the percentage of rapid ECoG data recorded from cortex during speech production, we answer questions as to the nature and fundamental organization of speech as it is produced. The prominence of the gestural model in cortical representation may suggest why some speech errors, such as phonemic confusion, tongue twisters (e.g. “proper copper coffee pot”) and spoonerisms (e.g. “fight a liar”, “light a fire”), are confused in speech production. We suggest that such errors arise because they may require more processing in cortex to execute, or alternatively, that cortical processing of similar actions is already active and cannot easily initiate new activation patterns. In fact, this result may mirror other models of human performance, such as the motor selection interference concept in human motor control research, in which confusion is more likely to occur in areas of similar activation than in areas with great distance between activation. (Ivry et al. 2004). More concrete work in future ECoG investigations of speech therefore could therefore specifically test tongue twisters or alliterative samples of speech to determine activity on channels associated with specific gestures. Our work suggests that this result could further apply outside of the speech domain with other repetitive fine motor gestures in close proximity, such as with neurophysiological correlates of typographical errors.

This gestural hypothesis result may also supplement related research in cortical motor prostheses for fine motor control, as gestures may explain more cortical information than position and direction of movement. In speech, exact postures and gestures are honed for a specific language for many years of an individual’s life, augmented by auditory, somatosensory, and listener feedback (SK Scott, McGettigan, and Eisner 2009). If similar gestures are also performed with grasp, it may be possible that neural representation of hand motor areas would better align with gestural classification algorithms. Thus, study of hand force or velocity may perform better in an ECoG classification paradigm than hand position or posture (Flint, Lindberg, et al. 2012). Finally, relating back to the analogy of force versus position in motor cortex (Georgopoulos, Schwartz, and Kettner 1986; SH Scott and Kalaska 1997), we can consider a gestural model closer to a force or velocity model in speech cortex (Guenther 2006). Phonemes, contrastingly, could be better envisioned by position models in cortex, as their place and manner of articulation correlate more with locations along the vocal tract, and not with velocities or forces.

Neural interfaces, which provide hope for “locked-in” individuals unable to communicate by normal means, could leverage gestural representation to create efficient cortical speech prostheses. Our approach enabled exploration of limitations of the phoneme model, which has thus far been predominant in speech brain-computer interface literature. Results of this investigation would require modification for true practical implementation in communicative brain-computer interface. One long-term goal of such work is to decode the intended speech of individuals and reproduce words, not limited to parts of words or phonemes. The information transfer rate of our gestural system applied to our speech stimulus set could impact speech decoding in a potential BCI. Such an increase in communication efficiency could dramatically improve quality of life in locked-in individuals. Identification of word production from ECoG recordings is possible using phonemic (Chapter II) models, but the combination of gestural classifications could outperform such results. Although the phonemic models can produce an ITR of 33.6 words per minute, we predict a similar gestural BCI would surpass the fastest reported BCI communication speed by a factor 6, similar to the average speed of compositional typing. Word prediction becomes more complex when analyzing broadest categories of articulator organ. In a small, closed set of words, a system that identifies speech by jaw movement could potentially enable identification of words, but in a wider range of complex words, or words with similar activation patterns, the classifier would likely require extensive training on word production. Contrastingly, neurophysiological signals directly related to the motor gestural components of speech production enabled decoding of the physical components of speech sounds, which could enable decoding beyond small, closed sets of words. The methodology employed here could therefore further impact the field of automatic speech recognition. Accurate classifications of the multi-model gestures of speech, instead of audio recordings of speech alone, may improve automatic speech transcription in dictation systems. An automatic speech that could access and classify gestural components, perhaps using computer vision or other optical methods, might supplement classification algorithms, particularly when speech data contains significant ambient noise. These results therefore guide future investigations into the next critical step in development of communicative brain-computer interfaces.

4. Methodology

For further elaboration on general methodology, refer to Chapter II. For elaboration on methodological constraints to our stimulus set, refer to Chapter III.

Four subjects (mean age = 42, 2 female), undergoing treatment for drug-resistant epilepsy, required extraoperative ECoG monitoring for epileptogenesis. These individuals volunteered for our study during their hospital stay and provided informed consent to participate in our study. Electrode coverage of cortex, determined by medical necessity, included frontal and temporal areas in all subjects.

Experimental protocol consisted of presentation of words on a screen with a 4s inter-stimulus interval. The 300 words of the Modified Rhyme Test (House et al. 1963) were supplemented by

20 additional words to incorporate all General American English phonemes to the stimulus set. Subjects read each word immediately when presented. Data collection was performed using BCI2000 software to present visual stimuli of words on a screen.

Speech was recorded with a MXL USB microphone at 44.1 kHz and synched to the RZ2; a randomized pulse was sent from the TDT system to the clinical system throughout the duration of each recording block in order to synch the data recording with the clinical Nihon-Kohden software. Sampling frequency was 500 Hz to 1 kHz to 9.6 kHz. Differential cortical recordings compared to a reference ECoG electrode were exported for analysis with an applied bandpass filter (0.53 - 300 Hz) with a sensitivity of 75 µV. Apparatus, and experimental set-up, and processing outline are depicted in Figure 9.

Data analysis primarily included alignment to phoneme onset, reduction to spectrotemporal features, and linear discriminant analysis on those features (Chapter II, Chapter III). Data was organized according to precise, manually-denoted phoneme onset and converted to spectrotemporal features from 300 ms before phoneme onset to 300 ms post-onset. Conversion of data to simplified features consisted of summation over 50 ms time segments relative to phoneme onset, as well as summation of bandpower changes in specific bandpower frequencies. Samples of features were organized according to their phonemic distinctions, and features were selected based on significance to categorical designations (ANOVA). Samples were subsequently classified using linear discriminant analysis (LDA) and 10-fold cross-validation (shuffled data, 90% training set, 10% testing set).

To evaluate contrasting models, features and data from the phonemic classification were relabeled according to respectively differing categorical designations. For articulatory organ model labeling, methods established by Bouchard and colleagues were extended to the data used in our data set (Bouchard et al. 2013). For the gestural model, the designations specified by the task dynamic model (Browman and Goldstein 1992). Notably, although phonemic model categories are mutually exclusive, gestural and articulator organ models' categories often occur in combination. Therefore for the gestural model, data is labelled by multiple categories that exclusively describe each phoneme in conjunction (e.g. \b\ becomes the combination of three states {lip closure},{lip release}, {velum close}). For each model, multiple classifiers were simultaneously trained and tested on different dimensional labels within the phonemic test set. The posterior probability for each classification was extracted and used to evaluate the certainty with which the classifier could identify each given category within a model.

V. GENERAL DISCUSSION AND CONCLUSIONS

The body of work in this dissertation described the electrocortical signal of speech production and the ways in which that signal can be classified as a step towards a communicative brain-computer interface (BCI). This project not only solves practical problems of speech decoding for BCI, but develops neural engineering methodology to enable reproduction of this in the future. Moreover, this project increases the scientific understanding of a complex and uniquely human neural process of speech production.

1. Contributions to Neural Engineering

Prior to the work contributed here, speech sounds had not successfully been isolated in speech production of full words for use in brain-computer interface. Outlined in this section are the specific contributions framed in the context of the science of BCI.

1.1. Information transfer rate

This dissertation has mapped out rates of human communication and outlined strategies for improving neural interface methodology to parallel natural human communication levels. Moreover, the work presented here is the first of its kind to predict information transfer rate (ITR) using speech of whole words. We have demonstrated that use of speech to control BCI has the dramatic potential to improve speed of communication for locked-in users. This potential improvement might be akin to the use of a typewriter in comparison to a joystick to type out a message. Although decoding performance was clearly imperfect, our results surpass previously documented BCI methods (much like a keyboard with some percentage of working keys still would outperform a joystick). The advancement of phonemic decoding is depicted in Figure 18, literally raising the bar for communicative BCI performance.

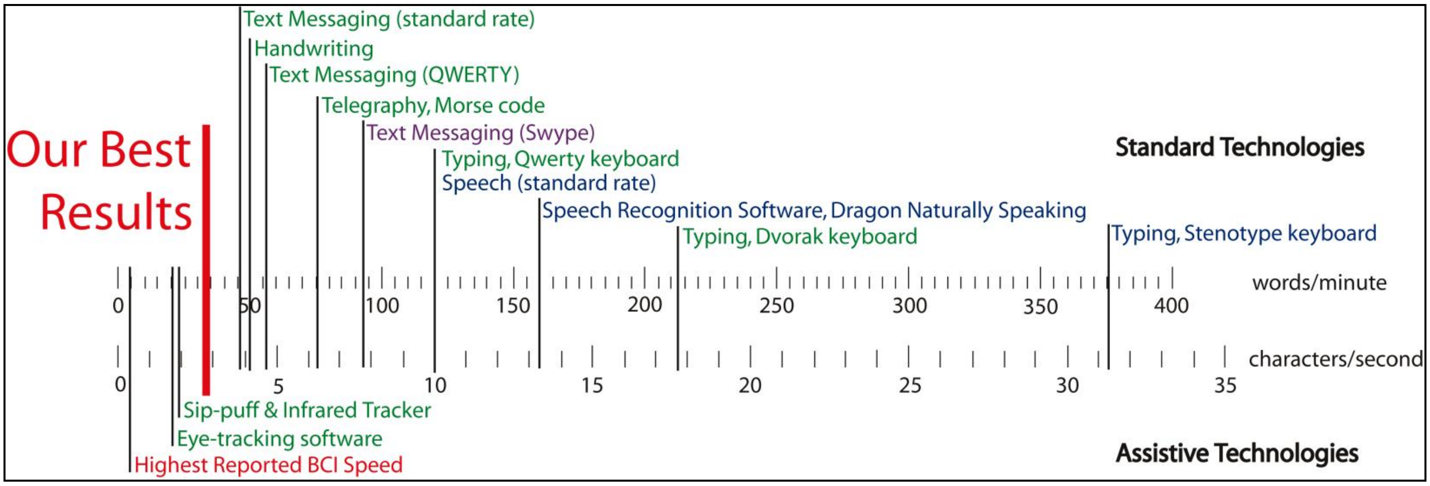


Figure 18. Speed of media in the English language, revisited. An updated version of Figure 2 depicting the contribution of this work in advancing brain-computer interface information transfer. Our highest calculated information transfer rates result from methods performed in Chapter II, in which we used consonant predictions of phoneme to identify word.

Although this work is not the first to investigate modalities of human communication to apply to BCI speeds (Schalk 2008), we use human communication speeds here as a model for ideal performance. Information transfer rate (ITR) is a widely reported metric in the BCI literature for comparison of non-equivalent systems. Here, we suggest that ITR of *human performance* should be the benchmark for BCI use. Ideally, with the advent of new technologies and BCI techniques, we could surpass this benchmark. We believe such advancements, like previously asserted (Section 1.3, Chapter I) would lead to wider adoption of BCI technology.

1.2. Failure-mode analysis

We assert that the first step in making prostheses for speech communication accessible is to determine what leads to success and failure in classification during healthy speech. This "failure-mode" approach elucidated some surprising dependencies in engineering factors, including phoneme onset, data recording time, number of phonemes analyzed, electrode utility, frequency band utility, and time bin utility.

We identified the degree to which precision to phoneme onset time affects classification results. This has important implications for future work with neural speech interface end-users, who would not likely be able to naturally produce phonemes. This finding therefore highlights a newly determined difficulty in future decoding speech sounds from such individuals with ECoG. Although this timing reliance may have been somewhat addressed by more invasive systems in locked-in individuals with ALS in the past (Brumberg et al. 2011), this issue of timing can now speak to how crucial it is. Inaccuracies in phonemic onset of 100 ms would perform dramatically poorly in comparison with a more precise phoneme onset analysis.

We identified the degree to which length of recording sessions may influence results. This suggests that more time with patients during recording sessions could improve classification results, and presents data to argue for longer recording sessions with the patient. Another approach would be to reduce the amount of rest time in between word trials, allowing more speech in our data collection sections and potentially also improving classification. Until this post-hoc power analysis on the data was performed, the recording session duration was not considered so critical for success.

Work presented here also identified that at least a span of 4 cm of speech sensorimotor cortex may be needed to identify all phonemes or types of phonemes. This highlights the advantage that may exist from a large EcoG electrode grid, in sharp contrast to depth electrodes inserted into deeper cortical structures that may not access all critical cortical areas (Guenther et al. 2009). Moreover, this contrasts with other BCI approaches demonstrated the usefulness of micro-ECoG electrodes or “mini-grids” with small inter-electrode differences to decode some differences in word production. Together with the results in the literature, we suggests specific guidelines of 4 cm length, 2 cm width, ≤4 mm spacing (“high density” spacing) for future electrode development to better address speech decoding. With such spacing, the estimated 3.1cm^2 of mouth sensorimotor cortex could be covered by as many as 20 separate ECoG electrodes (Fox et al. 2001). In relation to neuroanatomical structures, such an ECoG grid could additionally record from neighboring premotor and sensory areas. Our results indicate that recording with an electrode with these physical properties may dramatically improve speech-ECoG classification outcomes.

2. Contributions to Neuroscience

Only with the advent of technology that can specifically analyze signal with fine temporal and spatial resolution – like ECoG – does it become possible to classify and describe neural signal in such rapid fine motor processes like speech. We evaluated theoretical models of speech production based upon the degree they could explain representation of our ECoG speech data. Such methodology has been substantially used in the neuroscience literature with penetrating electrodes in cortex (Hubel and Wiesel 1962). Also, methodology refined here used for speech and neural engineering can be applied in the future with other technological development in accessing neural signal.

In fact, our data not only enables evaluation of current models, but our results may suggest new models and creates an experimental testbed for new models. With electrode grids with wider coverage and careful experimental design, ECoG could be used to investigate decoding of syllables, syntax, tone, and countless other properties of speech. This can significantly support research in cognitive neurolinguistics, which typically gets insight into neural processing mechanisms through speech errors, speech disorders, or fMRI results. Although these models have broader cortical scope beyond speech motor cortex and its organization, this area of cortex is one that we could directly evaluate using our experimental protocol.

3. Implications of this work

3.1. Clinical Implications

Although findings support high quality of life ratings in functionally locked-in patients when communication is retained, not all people may opt for implantable BCI technology. With such technology comes associated surgical risk, and for individuals that suddenly become locked-in (e.g. brainstem stroke), potential loss of remaining function may contribute to reluctance to undergo neurosurgery. In a recent study of long-term implantable ECoG grids in people with epilepsy, the majority of patients did not have serious device-related issues, and complication rates were on par with that of deep brain simulation implants (Cook et al. 2013). Only 4 of 15 patients experienced serious device-related adverse events (infection, seroma and device migration) in response to implanted ECoG electrodes. Nevertheless, a similar communicative BCI system must demonstrate utility and success at speech decoding past the point where benefits outweigh the risks before such a system would be adopted.

Notably, significant proportions (41%) of BCI users with ALS specifically report in surveys that they would prefer an implantable neural interface for communication (Huggins, Wren, and Gruis 2011; Blain-Moraes et al. 2012). One stated reason stemmed from belief that it would be less hassle for patient caregivers. Such surveys of the ultimate end-user are important to engineers designing the process, and the end-user must be kept in mind during development of neural interface technology in terms of long-term building toward a neural speech prosthetic device.

3.2. Research Implications

This work establishes successful methods for extracting speech from cortical signal, but moreover enables future methods which might be able to classify speech in real-time. The first steps of discrimination between simple phonemes in has already been used to control a 1-D cursor in real-time (Leuthardt et al. 2011). Whereas we are not yet decoding real-time speech, we have executed the critical steps to determine what methodology may be useful in accessing speech information from ECoG signal.

Moreover, this technology translates advances in intuitive control from neuromotor prosthetic development to speech communication. Such motor development has cleverly utilized motor imagery to reportedly provide a more direct and intuitive neural input for the user. Communication prostheses have been limited in their efficiency or intuitive use while motor prostheses have achieved decoding of multiple degrees of freedom for mechanical arm applications. By classifying signals associated with actual speech, use of a neural speech interface will likely feel more intuitive for end-users.

In the process of researching speech for communicative BCI, we have developed an algorithmic suite of Matlab code to enable synchronization of speech and cortical signal, rapid spectrotemporal signal processing and subsequent speech classification analysis.

3.3. Commercial Implications

The results here indicate that recording speech signal from sensorimotor cortex requires specific electrode placement. We assert that commercial BCI systems lacking fine spatial resolution, such as non-invasive EEG systems (Quasar, Neuralynx, Emotiv, Muse), may not succeed in their similar pursuits of speech signal. Functional Near-Infrared Spectroscopy (fNIRS) systems, which can access changes in the blood oxygenation levels through the skull, would also lack the ability to reproduce results presented in this work. Notably, this process is far removed from “mind-reading”, often claimed by BCI companies, and results reported here currently necessitate active speech. Similarly, although an early BCI paradigm was referred to as a “thought translation device” (Kübler et al. 1999), “thought” is a simplistic term for a broad group of higher-level cognitive process. Imagined speech may be adjacent to or included in the traditional definition of thought, but we here refrain from claiming that speech-ECoG BCI is decoding thoughts. Instead, we state that we are decoding motor-related neural activity of intended, overt speech.

4. Limitations of this work

4.1. Subject population

One of the biggest hindrances to completion of this research is the rarity of ECoG subjects with frontotemporal electrode coverage. Subjects are extremely rare even when placement of electrodes could cover speech sensorimotor cortex, though as stated earlier, placement of electrodes depends entirely on medical necessity. Moreover, even when patients are eager to participate, they may be feeling ill, tired, or not want to complete the full experimental protocol. In the 3 years of working with patients, only 6 patients participated in this study and only 4 subjects completed the minimum experimental protocol at Northwestern Memorial Hospital. Over the course of this period, we have expanded to collaborate with a scientist at Old Dominion University (Norfolk, VA) and a neurosurgeon at the Mayo Clinic (Jacksonville, Florida). The code we developed for BCI2000 software works with other hardware systems. Collaboration with other hospitals and research groups is therefore a strategy for dealing with such a limited subject pool, so that each subject is not functionally a case study.

Another more general limitation to this research is potential atypical functional organization in this patient population (G. A. Ojemann 1979; Springer et al. 1999). Atypical activation patterns were evident in results from Subject NU3, in which subtemporal areas (sometimes associated with object naming in healthy individuals) correlated with functional facial motor and sensory activity when stimulated. Excluding this subject, electrode contribution from subjects supported traditional definitions of speech motor areas. If atypical functional organization exists, it would be difficult to ascertain whether it was caused by the epilepsy or whether epilepsy created the strange functional patterns. It was therefore an assumption of our experiment that association with speech sensorimotor cortex, determined by cortical stimulation mapping, was indicative of functional speech areas for each subject.

Moreover, the subjects all had non-disordered speech that we could synchronize with neural signal. This enables conclusions about the nature and functional organization of speech, but may hinder development of techniques for the ultimate end-user of a speech BCI, would have some issue impacting natural speech production.

4.2. Speech

Pronunciation varied across subject and slightly within subject. In this research, a 2 of our 4 subjects were bilingual and had an accent that was not the General American English accent. Moreover, one subject originally spoke Dutch and English was not the first language. For all subjects, pronunciation was sometimes inconsistent within each trial, particularly during vocal stops (i.e. not annunciating the \t\ at the end of \pit\). Despite this variability in pronunciation, slight differences - predominantly manifested during phonation of vowels - were assumed to be consistent across the phonemes that comprised our stimulus set.

Moreover, properties of speech make it difficult to determine the extent to which one phoneme influences another in our data set. Vowel classification was influenced by the consonants that occurred coincidently. This could be improved by increase in data collection duration and by increasing the variety and complexity of words used in the data set to diversify the coincident combinations of phonemes.

4.3. Hardware

The ECoG electrode grids used in this research had a 1 cm inter-electrode spacing (Ad-Tech Corp). Micro-ECoG grids (≤4 mm inter-electrode spacing) may provide more utility in localizing neural signal indicative of speech-related function in cortex (see Chapter V, Section 1.2).

Electrocorticographic grid density is rapidly increasing with advancements in sensor development and implantable circuit technology, reaching inter-electrode distances of 500-μm (Viventi et al. 2011). As electrode grids get denser, it may be possible to investigate activity in cortex on a finer functional level, an exciting future direction for ECoG research. Other potential advancements in electrode conduit material, such as silk, polyimide, and other flexible materials, provide an increase in biocompatibility and therefore a potentially longer-lasting neural interface. Finally, fully implantable devices using wireless transmission technology can reduce risk of transcutaneous infection. Although the work presented here was limited in spacing and long-term implantation, potential improvements in hardware introduce a suite of new possibilities to increase practicality of BCI implementation.

Sampling frequencies for recording ECoG signal varied for different data collection sessions. Due to human error, sampling frequency was set as low as 500 Hz, which needed to be specified prior to experimental start by medical technicians in the Epilepsy Monitoring Unit. This prevented off-line analysis of frequency bandpower changes related to articulation in frequencies greater than 250 Hz for 2 patients. However, these frequency features did not improve successful classification results in patients recorded on the same system at a 1 kHz sampling rate, so this is not considered a detriment to our conclusions.

Finally, the ECoG signal, recorded on the clinical monitoring system of the hospital, was not recorded on the same computer with the speech signal, primarily because speech – which contains higher frequencies – needed to be sampled at a higher sampling frequency. An electronic pulse was emitted between computing systems to synchronize signals offline for 3 patients. We therefore assumed a minimal or negligible difference in time courses along these two signals, but recording both signals on a single system would prevent uncertainty in signal synchronization.

4.4. Biology

Electrodes used in this study were designed for and approved for temporary use (less than 29 days) by the Food and Drug Administration in the United States. ECoG research has made enormous strides in decoding cortical activity related to function in this short-term status, especially when it comes to robotic arm control (W. Wang et al. 2013). However, in order to extrapolate our results for a long-term implanted speech BCI, the longevity of such systems must be evaluated for patient safety. New studies in Australia explored the extent of long-term implantation of ECoG grids in humans for epileptic monitoring and seizure prediction, but sadly the clinical trial was halted due to lack of funding (Cook et al. 2013). In this study, 11 patients out of 15 had device-related adverse events during the clinical trial, and 2 serious adverse events occurred within 1 year of implantation. Adverse events were primarily related to immunological response to the implanted device. Two patients, however, refused explantation of the devices and still have a functioning ECoG implant predicting seizures (>2 years). If the utility of such a surgical process was validated for individuals with locked-in disorders, benefits may exceed risks for a potential ECoG speech interface. In this case, the work of this dissertation provides the groundwork for design of an ideal system.

There is no guarantee that individuals with complete locked-in syndrome will have typical motor cortex representation for speech motor areas. Metabolic activation in motor cortex appears equivalent in quadriplegics who cannot execute peripheral motor activity as compared to healthy controls (Shoham et al. 2001), and this may also hold for individuals with locked-in syndrome. It is possible that cortical deterioration develops in late-stage ALS, but this is poorly understood due to the rapid progression of the disease (Bach JR. 1994). A recently approved study in the European Union (the Netherlands) will investigate long-term implantation of ECoG grids in a locked-in subject with the intent of classifying and transcribing intended speech. Results of this clinical trial will indicate feasibility of a speech-BCI using ECoG in the target population. Even if speech motor cortex demonstrates normal “healthy” neural activity in locked-in individuals, extra scientific steps are likely necessary to enable neural speech communication. In the work presented in this dissertation, our formulas require training an algorithm on overt speech. Therefore overt speech is the most direct and most possible BCI application of this research.

Finally, as epilepsy drugs improve, drive for electrocorticographic monitoring for seizure onset decreases, diminishing the rare opportunities to use ECoG to study functional activation of cortex. It is therefore vital to properly design experimental protocol for these studies so that off-line analysis can continue should a ready subject pool of individuals with implanted electrodes disappear.

4.5. Signal processing

Many methods exist for converting signal to frequency bandpower features, and many choices exist in application of signal processing choices to ECoG signal. In this work, we used a Goertzel algorithm version of the Short-Time Fourier Transform. We acknowledge that other methods exist for reduction to features (e.g. Hilbert Huang Transforms, Synchrosqueezing, Discrete Hermite Transforms), and we did not exhaustively investigate each process. We used the STFT as a starting point for feature reduction. We used duration of audio signal of phonemes to guide our choices for window size for our STFT. For frequency analysis, we investigated all ranges or bands that were reported in the BCI literature. Improvements in signal processing methods that can precisely determine changes in frequency bandpower over 50 ms windows of time would likely improve classification results.

4.6. Software

The process of segmenting time data into phoneme onset times is still manual, and is the rate-limiting process on data analysis. This time-stamping process is performed manually (APPENDIX B. Graphical User Interface for Phoneme Labeling), as current methods of automation lack the specificity needed for accurate labeling of phonemes. The required specificity to phoneme onset time therefore highlights the degree to which manual labeling of the data is crucial prior to LDA classification.

The limitations preventing real-time processing are the biggest hindrance to our conclusions on information transfer rate. An assumption of our ITR calculations was that no pauses exist between words. We therefore state that the calculated ITR is the gross information transfer, and a more precise version could be calculated in future iterations of this work, in which words are decoded in succession.

Finally, computational processing is limited by the massive quantity of data collected during each experiment. With the current dramatic increases in rapid access memory and computational power, data could be converted to finer spatiotemporal features. If data did not have to be reduced to features in order to run classification algorithms, it would further enable more complex, real-time processing of signals. However, feature reduction will be required as long as the quantity of data features outnumbers the number of speech samples analyzed.

4.7. Speech Recognition and Identification

In this dissertation, we estimate the degree to which we can predict words from our word stimulus set by the categorical results computed during LDA classification. However, in this process, we intentionally limit our word identification algorithms used in this work to simplify the mechanisms. These mechanisms are meant to serve as a starting point for later speech classification algorithms. Moreover, we restricted this prediction mechanism to the consonants used during decoding, which limits full identification of words that share similar consonant pairs but have a differing vowel. However, there is more identifying information in consonants than in vowels (Shannon 1951), so this actually may serve the purposes of word identification in a fair yet beneficial manner.

5. Potential Improvements and Future Directions

Many areas for expansion and investigation – even by simply restructuring analysis using the current data set – could answer neuroscience questions and advance communicative brain-computer interface. Coarticulation – the blending of separate speech sounds in context – and the contextual dependence of speech sounds have huge implications for application in an ECoG-speech BCI. Moreover, we can investigate the effects of coarticulation during quicker speech in the context of sentences. For 2 subjects, full, phonetically-balanced sentences were recorded in addition to whole words. We have further established collaboration with another research group that has recorded long sections (paragraphs) of overt reading in over 10 subjects. Analysis of the ways in which speech sound categorical designations are altered by contextual information of surrounding text could answer questions of neural processing.

Algorithmic detection of a “go signal” for speech using this data set would enable an effective on-off switch for a neurally-controlled speech prosthesis. Classification between speech and rest periods could help identify phoneme onset. Until our methods uncovered the degree to which phoneme onset time was required for accurate speech analysis, it was unclear that this was an area that would augment speech communication through BCI.

Electromyographic signal was collected throughout all data collection, bilaterally at the carotid triangle and over the right masseter muscle, to determine motor movement associated with speech. This data was collected simultaneously with ECoG and microphone signal. Ideally, this data modality could be further synched with 3D video imaging, such as that created with a Microsoft Kinect device. Results from such 3D and EMG motor analysis could lead to developments of life-like active speech models that expand on the task dynamic application (TADA) model. This could particularly aid speech therapies and perhaps also in digital simulations of speech, an area where current simulations noticeably fall along the “uncanny valley” and typically triggers discomfort in users.

Classification of this EMG signal recorded during speech could also potentially provide new aides in communication to individuals with healthy speech communicating in adverse environments. Building on the success of previous EMG–speech decoding paradigms (B. Betts and Jorgensen 2005; Lee 2008; Schultz and Waibel 1998; Jou and Schultz 2008; Jou, Schultz, and Waibel 2007; Schultz and Wand 2010; Arjunan et al. 2006), we can apply methodology and tools developed in this dissertation to other speech-biosignal interfaces. In conjunction with high-density EMG grids, we can then mechanistically investigate factors that contribute to speech classification success. Moreover, results could indicate the potential benefit over ECoG BCI and identify the degree of muscular control needed to control an EMG interface. This could therefore identify the point at which a surgical procedure would be necessary to enable a speech interface for individuals with severe speech motor impairments.

The most dramatic advancement would result from converting our cortical signal processing software suite into a real-time process. Real-time speech decoding would reveal the extent to which we can predict words as they are phonated. Even pseudo-real-time, in which the word would be decoded shortly after utterance, could be useful for those who currently lack natural means of communication. Real-time decoding of speech would also build toward closed-loop BCI paradigms of this work, the results of which would indicate potential success in individuals with LIS. Speech recognition algorithms can and should be further developed to work in conjunction with cortical signal classification results to predict actual speech. In a real-time decoding scenario, the functional frequencies of phonetic information in language could be used to better identify speech. Important to note is that listeners may be capable of understanding speech despite some errors in transcription. A type of Turing test for interpretation of output could be used to determine successful classification of speech (Turing 1950); the benchmark for success of such a system would be human understanding of speech-BCI output.

We could optimize the language of the speech-ECoG to improve classification results. For example, we could apply speech classification algorithms other human languages, with relative frequencies of gestures, to optimize what a classification decoder can successfully select. Languages with more sparse phonemes, such as Hawaiian (with 8 pulmonic consonant phonemes), would likely improve the success of the classifier. An “Esperanto” for speech-ECoG could also be created, optimized for the neurophysiology of each user. Alternatively, the speech interface could potentially limit the set of the full scope of words to improve accuracy. With extended use, users could potentially adapt their neural signals to better facilitate speech interface output. Alterations to the language of the speech interface could therefore potentially improve decoding,

Finally, investigation of neural motivation behind dynamic learning of a speech motor task, such as pronunciation of foreign phonemes or beatboxing (Proctor et al. 2013), could reveal further mechanisms of cortical dynamics. Speech is a fine motor movement process, and how cortical signal changes and adapts to learn a foreign language or a new task such as beatboxing could have wider implications for motor control research.

6. Summary

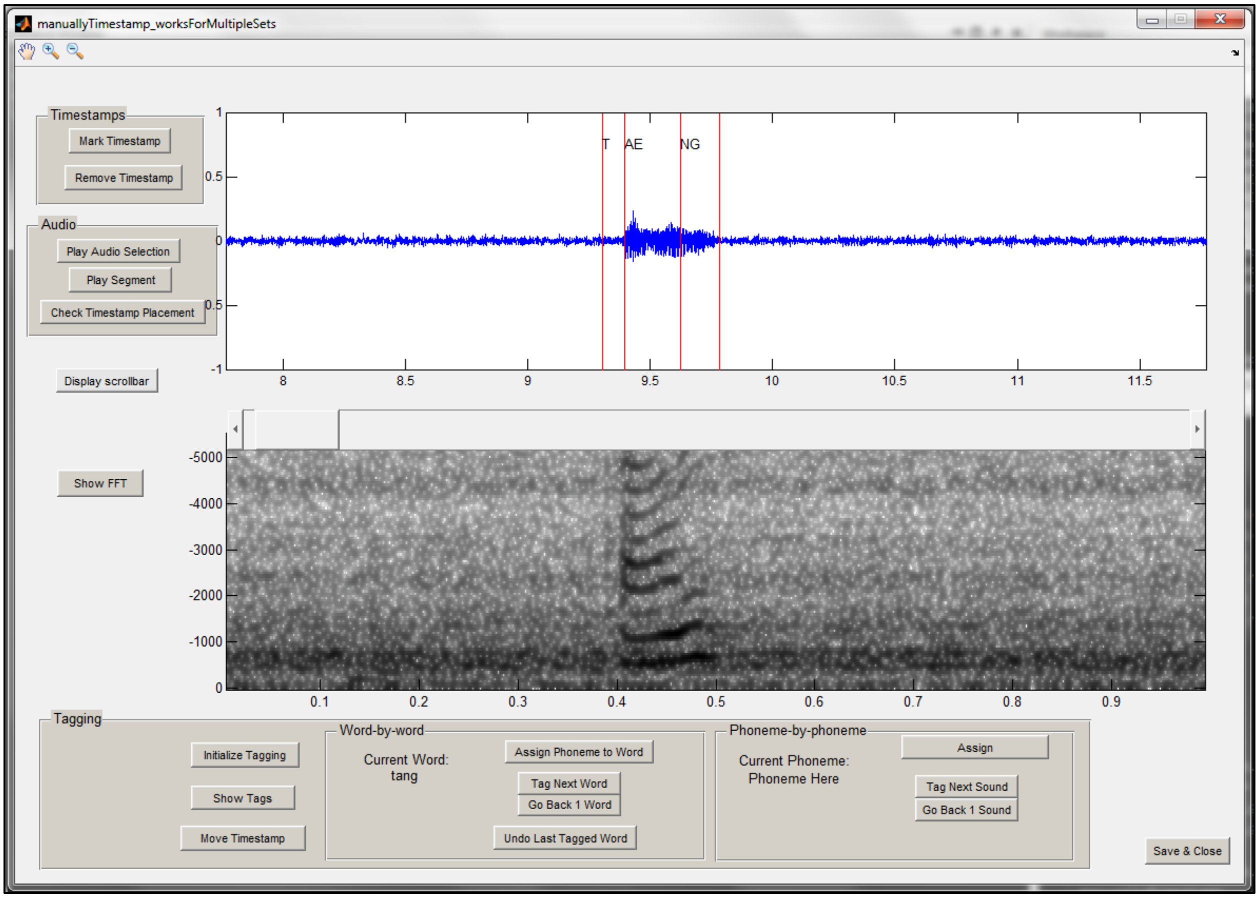
This investigation of speech for communicative brain-computer interface approaches supports the use of speech for future development in neural interface design for communication. The work of this dissertation advances the efficiency of communicative brain-computer interface approaches, establishes a methodological testbed for evaluation of neuroscience models and hypotheses, and demonstrates predominance of gestures in a functional model of speech production in cortex. Although classification of speech using ECoG signal demonstrated statistically significant performance, perfect speech identification from cortex remains a distant but worthy goal. In any case, methods developed here should guide future approaches to speech brain-computer interface design, development, and clinical application.

APPENDIX A. Word Stimulus Set

The Modified Rhyme Test (House et al. 1963)

went	sent	bent	dent	tent	rent
hold	cold	told	fold	sold	gold
pat	pad	pan	path	pack	pass
lane	lay	late	lake	lace	lame
kit	bit	fit	hit	wit	sit
must	bust	gust	rust	dust	just
teak	team	teal	teach	tear	tease
din	dill	dim	dig	dip	did
bed	led	fed	red	wed	shed
pin	sin	tin	fin	din	win
dug	dung	duck	dud	dub	dun
sum	sun	sung	sup	sub	sud
seep	seen	seethe	seek	seem	seed
not	tot	got	pot	hot	lot
vest	test	rest	best	west	nest
pig	pill	pin	pip	pit	pick
back	bath	bad	bass	bat	ban
way	may	say	pay	day	gay
pig	big	dig	wig	rig	fig
pale	pace	page	pane	pay	pave
cane	case	cape	cake	came	cave
shop	mop	cop	top	hop	pop
coil	oil	soil	toil	boil	foil
tan	tang	tap	tack	tam	tab
fit	fib	fizz	fill	fig	fin
same	name	game	tame	came	fame
peel	reel	feel	eel	keel	heel
hark	dark	mark	bark	park	lark
heave	hear	heat	heal	heap	heath
cup	cut	cud	cuff	cuss	cud
thaw	law	raw	paw	jaw	saw
pen	hen	men	then	den	ten
puff	puck	pub	pus	pup	pun
bean	beach	beat	beak	bead	beam
heat	neat	feat	seat	meat	beat
dip	sip	hip	tip	lip	rip
kill	kin	kit	kick	king	kid
hang	sang	bang	rang	fang	gang
took	cook	look	hook	shook	book
mass	math	map	mat	man	mad
ray	raze	rate	rave	rake	race
save	same	sale	sane	sake	safe
fill	kill	will	hill	till	bill
sill	sick	sip	sing	sit	sin
bale	gale	sale	tale	pale	male
wick	sick	kick	lick	pick	tick
peace	peas	peak	peach	peat	peal
bun	bus	but	bug	buck	buff
sag	sat	sass	sack	sad	sap
fun	sun	bun	gun	run	nun

APPENDIX B. Graphical User Interface for Phoneme Labeling



Axis above with speech microphone waveform outlined with specific phoneme labels applied. Timestamps are marked in red. Axis below indicative of corresponding spectrogram of above signal.

CITED LITERATURE

Arjunan, Sridhar P, Hans Weghorn, Dinesh Kant Kumar, and Wai C Yau. 2006. "Vowel Recognition of English and German Language Using Facial movement(SEMG) for Speech Control Based HCI." *ACM International Conference Proceeding Series*: 13–18. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Vowel+recognition+of+English+and+German+language+using+Facial+movement(SEMG)+for+Speech+control+based+HCI#0.

Bach JR. 1994. "Amyotrophic Lateral Sclerosis. Communication Status and Survival with Ventilatory Support." *Am J Phys Med Rehabil* 73 (3): 218.

Ball, Laura J, David R Beukelman, and Gary L Pattee. 2004. "Communication Effectiveness of Individuals with Amyotrophic Lateral Sclerosis." *Journal of Communication Disorders* 37 (3): 197–215. doi:10.1016/j.jcomdis.2003.09.002. http://www.ncbi.nlm.nih.gov/pubmed/15063143.

Ball, Tonio, Evariste Demandt, Isabella Mutschler, Eva Neitzel, Carsten Mehring, Klaus Vogt, Ad Aertsen, and Andreas Schulze-Bonhage. 2008. "Movement Related Activity in the High Gamma Range of the Human EEG." *NeuroImage* 41 (2) (June): 302–10. doi:10.1016/j.neuroimage.2008.02.032. http://www.ncbi.nlm.nih.gov/pubmed/18424182.

Bartels, Jess, Dinal S Andreasen, Princewill Ehirim, Hui Mao, Steven Seibert, E. Joe Wright, and Philip Kennedy. 2008. "Neurotrophic Electrode: Method of Assembly and Implantation into Human Motor Speech Cortex." *Journal of Neuroscience Methods* 174 (2): 168–176. doi:10.1016/j.jneumeth.2008.06.030. http://linkinghub.elsevier.com/retrieve/pii/S0165027008003865.

Bensch, Michael, Ahmed a Karim, Jürgen Mellinger, Thilo Hinterberger, Michael Tangermann, Martin Bogdan, Wolfgang Rosenstiel, and Niels Birbaumer. 2007. "Nessi: An EEG-Controlled Web Browser for Severely Paralyzed Patients." *Computational Intelligence and Neuroscience* 2007: 71863. doi:10.1155/2007/71863. http://www.ncbi.nlm.nih.gov/pubmed/18350132.

Betts, B J, Kim Binsted, and Charles C Jorgensen. 2006. "Small-Vocabulary Speech Recognition Using Surface Electromyography." *Interacting with Computers* 18 (6): 1242–1259. doi:10.1016/j.intcom.2006.08.012. http://linkinghub.elsevier.com/retrieve/pii/S0953543806001342.

Betts, BJ, and C Jorgensen. 2005. *Small Vocabulary Recognition Using Surface Electromyography in an Acoustically Harsh Environment*. *Moffett Field, CA, NASA, Ames Research Center*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.7060&rep=rep1&type=pdf.

Bin, Guangyu, Xiaorong Gao, Yijun Wang, Yun Li, Bo Hong, and Shangkai Gao. 2011. "A High-Speed BCI Based on Code Modulation VEP." *Journal of Neural Engineering* 8 (2) (April): 025015. doi:10.1088/1741-2560/8/2/025015. http://www.ncbi.nlm.nih.gov/pubmed/21436527.

Birbaumer, Niels, and Leonardo G Cohen. 2007. "Brain-Computer Interfaces: Communication and Restoration of Movement in Paralysis." *The Journal of Physiology* 579 (Pt 3): 621–36. doi:10.1113/jphysiol.2006.125633. http://www.ncbi.nlm.nih.gov/pubmed/17234696.

Blain-Moraes, Stefanie, Riley Schaff, Kirsten L Gruis, Jane E Huggins, and Patricia a Wren. 2012. "Barriers to and Mediators of Brain-Computer Interface User Acceptance: Focus Group Findings." *Ergonomics* 55 (5) (May): 516–25. doi:10.1080/00140139.2012.661082. http://www.ncbi.nlm.nih.gov/pubmed/22455595.

Blakely, Timothy M, Kai J Miller, Rajesh P N Rao, Mark D Holmes, and Jeffrey G Ojemann. 2008. "Localization and Classification of Phonemes Using High Spatial Resolution Electrocorticography (ECoG) Grids." *Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* 2008 (January): 4964–7. doi:10.1109/IEMBS.2008.4650328. http://www.ncbi.nlm.nih.gov/pubmed/19163831.

Blankertz, Benjamin, Guido Dornhege, Christin Schäfer, Roman Krepki, Jens Kohlmorgen, Klaus-Robert Müller, Volker Kunzmann, Florian Losch, and Gabriel Curio. 2003. "Boosting Bit Rates and Error Detection for the Classification of Fast-Paced Motor Commands Based on Single-Trial EEG Analysis." *IEEE Transactions on Neural Systems and Rehabilitation Engineering : a Publication of the IEEE Engineering in Medicine and Biology Society* 11 (2) (June): 127–31. doi:10.1109/TNSRE.2003.814456. http://www.ncbi.nlm.nih.gov/pubmed/12899253.

Bochkarev, V. V., A.V. Shevlyakova, and V. D. Solovyev. 2012. "Average Word Length Dynamics as Indicator of Cultural Changes in Society." *arXiv Preprint arXiv: ...*. http://arxiv.org/abs/1208.6109.

Bohland, Jason W, and Frank H Guenther. 2006. "An fMRI Investigation of Syllable Sequence Production." *NeuroImage* 32 (2) (August): 821–41. doi:10.1016/j.neuroimage.2006.04.173. http://www.ncbi.nlm.nih.gov/pubmed/16730195.

Bouchard, Kristofer E., Nima Mesgarani, Keith Johnson, and Edward F. Chang. 2013. "Functional Organization of Human Sensorimotor Cortex for Speech Articulation." *Nature* (February 20). doi:10.1038/nature11911. http://www.nature.com/doifinder/10.1038/nature11911.

Browman, C P, and L Goldstein. 1992. "Articulatory Phonology: An Overview." *Phonetica* 49 (3-4) (January): 155–80. http://www.ncbi.nlm.nih.gov/pubmed/1488456.

Brown, Adam. 2013. “International Phonetic Alphabet.” *The Encyclopedia of Applied Linguistics*. doi:10.1002/9781405198431.wbeal0565. http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal0565/full.

Brumberg, Jonathan S, Alfonso Nieto-Castanon, Philip R Kennedy, and Frank H Guenther. 2010. “Brain-Computer Interfaces for Speech Communication.” *Speech Communication* 52 (4) (April): 367–379. doi:10.1016/j.specom.2010.01.001. http://www.ncbi.nlm.nih.gov/pubmed/20204164.

Brumberg, Jonathan S, E Joe Wright, Dinal S Andreasen, Frank H Guenther, and Philip R Kennedy. 2011. “Classification of Intended Phoneme Production from Chronic Intracortical Microelectrode Recordings in Speech-Motor Cortex.” *Frontiers in Neuroscience* 5 (May) (May 12): 1–12. doi:10.3389/fnins.2011.00065. http://www.frontiersin.org/Neuroprosthetics/10.3389/fnins.2011.00065/abstract.

Chan, Adrian D C, Kevin Englehart, Bernard Hudgins, and D F Lovely. 2001. “Hidden Markov Model Classification of Myoelectric Signals in Speech.” *Coating*: 23–26.

Cheng, Ming, Xiaorong Gao, Shangkai Gao, Senior Member, and Dingfeng Xu. 2002. “Design and Implementation of a Brain-Computer Interface With High Transfer Rates.” *October* 49 (10): 1181–1186.

Collinger, Jennifer L, Brian Wodlinger, John E Downey, Wei Wang, Elizabeth C Tyler-Kabara, Douglas J Weber, Angus J C McMorland, Meel Velliste, Michael L Boninger, and Andrew B Schwartz. 2013. “High-Performance Neuroprosthetic Control by an Individual with Tetraplegia.” *Lancet* 381 (9866) (February 16): 557–64. doi:10.1016/S0140-6736(12)61816-9. http://www.ncbi.nlm.nih.gov/pubmed/23253623.

Cook, Mark J, Terence J O’Brien, Samuel F Berkovic, Michael Murphy, Andrew Morokoff, Gavin Fabinyi, Wendyl D’Souza, et al. 2013. “Prediction of Seizure Likelihood with a Long-Term, Implanted Seizure Advisory System in Patients with Drug-Resistant Epilepsy: a First-in-Man Study.” *Lancet Neurology* 12 (6) (June): 563–71. doi:10.1016/S1474-4422(13)70075-9. http://www.ncbi.nlm.nih.gov/pubmed/23642342.

Coyle, Shirley M, Tomás E Ward, and Charles M Markham. 2007. “Brain-Computer Interface Using a Simplified Functional Near-Infrared Spectroscopy System.” *Journal of Neural Engineering* 4 (3) (September): 219–26. doi:10.1088/1741-2560/4/3/007. http://www.ncbi.nlm.nih.gov/pubmed/17873424.

Crone, N E, L Hao, J Hart, D Boatman, R P Lesser, R Irizarry, and B Gordon. 2001. “Electrocorticographic Gamma Activity During Word Production in Spoken and Sign Language.” *Neurology* 57 (11) (December 11): 2045–53. http://www.ncbi.nlm.nih.gov/pubmed/11739824.

D’Ausilio, Alessandro, Friedemann Pulvermüller, Paola Salmas, Ilaria Bufalari, Chiara Begliomini, and Luciano Fadiga. 2009. “The Motor Somatotopy of Speech Perception.”

Current Biology : CB 19 (5) (March 10): 381–5. doi:10.1016/j.cub.2009.01.017. http://www.ncbi.nlm.nih.gov/pubmed/19217297.

DaSalla, Charles S, Hiroyuki Kambara, Makoto Sato, and Yasuharu Koike. 2009. “Single-Trial Classification of Vowel Speech Imagery Using Common Spatial Patterns.” *Neural Networks : the Official Journal of the International Neural Network Society* 22 (9) (November): 1334–9. doi:10.1016/j.neunet.2009.05.008. http://www.ncbi.nlm.nih.gov/pubmed/19497710.

Denby, B., Tanja Schultz, K. Honda, T. Hueber, J.M. Gilbert, and J.S. Brumberg. 2009. “Silent Speech Interfaces.” *Speech Communication* (August). doi:10.1016/j.specom.2009.08.002. http://linkinghub.elsevier.com/retrieve/pii/S0167639309001307.

Epstein, Melissa, Narineh Hacopian, and Peter Ladefoged. 2002. “Dissection of the Speech Production Mechanism.” In *UCLA Working Papers in Phonetics*, edited by Melissa Epstein, Narineh Hacopian, and Peter Ladefoged, 102:1–14.

Fagg, Andrew H, Gregory W Ojakangas, Lee E Miller, and Nicholas G Hatsopoulos. 2009. “Cortical Ensembles” 17 (5): 487–496.

Farwell, L. A., and E. Donchin. 1988. “Talking Off the Top of Your Head: Toward a Mental Prosthesis Utilizing Event-Related Brain Potentials.” *Electroencephalography and Clinical Neurophysiology*.

Flinker, Adeen, Edward F Chang, Heidi E Kirsch, Nicholas M Barbaro, Nathan E Crone, and Robert T Knight. 2010. “Single-Trial Speech Suppression of Auditory Cortex Activity in Humans.” *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience* 30 (49) (December 8): 16643–50. doi:10.1523/JNEUROSCI.1809-10.2010. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3010242&tool=pmcentrez&rendertype=abstract.

Flinker, Adeen, Edward F. Chang, N M Barbaro, M S Berger, and Robert T Knight. 2011. “Sub-Centimeter Language Organization in the Human Temporal Lobe.” *Brain and Language* 117 (3) (June): 103–9. doi:10.1016/j.bandl.2010.09.009. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3025271&tool=pmcentrez&rendertype=abstract.

Flint, Robert D, Christian Ethier, Emily R Oby, Lee E Miller, and Marc W Slutzky. 2012. “Local Field Potentials Allow Accurate Decoding of Muscle Activity.” *Journal of Neurophysiology* 108 (1) (July): 18–24. doi:10.1152/jn.00832.2011. http://www.ncbi.nlm.nih.gov/pubmed/22496527.

Flint, Robert D, Eric W Lindberg, Luke R Jordan, Lee E Miller, and Marc W Slutzky. 2012. “Accurate Decoding of Reaching Movements from Field Potentials in the Absence of Spikes.” *Journal of Neural Engineering* 9 (4) (June 25): 046006. doi:10.1088/1741-2560/9/4/046006. http://www.ncbi.nlm.nih.gov/pubmed/22733013.

Foerster, Otfrid. 1931. "The Cerebral Cortex in Man." *The Lancet* Volume 218 (Issue 5632): Pages 309–312. doi:10.1007/SpringerReference_4764.

Fox, P T, a Huang, L M Parsons, J H Xiong, F Zamarippa, L Rainey, and J L Lancaster. 2001. "Location-Probability Profiles for the Mouth Region of Human Primary Motor-Sensory Cortex: Model and Validation." *NeuroImage* 13 (1) (January): 196–209. doi:10.1006/nimg.2000.0659. http://www.ncbi.nlm.nih.gov/pubmed/11133322.

Furdea, A, S Halder, D J Krusienski, D Bross, F Nijboer, N Birbaumer, and A Kübler. 2009. "An Auditory Oddball (P300) Spelling System for Brain-Computer Interfaces." *Psychophysiology* 46 (3) (May): 617–25. doi:10.1111/j.1469-8986.2008.00783.x. http://www.ncbi.nlm.nih.gov/pubmed/19170946.

Georgopoulos, A., A. Schwartz, and R. Kettner. 1986. "Neuronal Population Coding of Movement Direction." *Science* 233 (4771) (September 26): 1416–1419. doi:10.1126/science.3749885. http://www.sciencemag.org/cgi/doi/10.1126/science.3749885.

Glenday, Craig, ed. 2013. *Guinness World Records 2013*. 2013th ed. Random House Digital, Inc., 2013.

Greenlee, Jeremy D W, Adam W Jackson, Fangxiang Chen, Charles R Larson, Hiroyuki Oya, Hiroto Kawasaki, Haiming Chen, and Matthew a Howard. 2011. "Human Auditory Cortical Activation During Self-Vocalization." *PloS One* 6 (3) (January): e14744. doi:10.1371/journal.pone.0014744. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3048394&tool=pmcentrez&rendertype=abstract.

Grozea, Cristian, Catalin D Voinescu, and Siamac Fazli. 2011. "Bristle-Sensors--Low-Cost Flexible Passive Dry EEG Electrodes for Neurofeedback and BCI Applications." *Journal of Neural Engineering* 8 (2) (April): 025008. doi:10.1088/1741-2560/8/2/025008. http://www.ncbi.nlm.nih.gov/pubmed/21436526.

Guenther, Frank H. 2006. "Cortical Interactions Underlying the Production of Speech Sounds." *Journal of Communication Disorders* 39 (5): 350–65. doi:10.1016/j.jcomdis.2006.06.013. http://www.ncbi.nlm.nih.gov/pubmed/16887139.

Guenther, Frank H, Jonathan S Brumberg, E Joseph Wright, Alfonso Nieto-Castanon, Jason a Tourville, Mikhail Panko, Robert Law, et al. 2009. "A Wireless Brain-Machine Interface for Real-Time Speech Synthesis." *PloS One* 4 (12) (January): e8218. doi:10.1371/journal.pone.0008218. http://www.ncbi.nlm.nih.gov/pubmed/20011034.

Guenther, Frank H, Satrajit S Ghosh, and Jason a Tourville. 2006. "Neural Modeling and Imaging of the Cortical Interactions Underlying Syllable Production." *Brain and Language* 96 (3): 280–301. doi:10.1016/j.bandl.2005.06.001. http://www.ncbi.nlm.nih.gov/pubmed/16040108.

Hardcastle, R.A., and C.J. Matthews. 1991. “Speed of Writing.” *Journal of the Forensic Science Society* 31 (1): 21–29. http://dx.doi.org/10.1016/S0015-7368(91)73114-9.

Hatsopoulos, Nicholas G, and Aaron J Suminski. 2011. “Sensing with the Motor Cortex.” *Neuron* 72 (3) (November 3): 477–87. doi:10.1016/j.neuron.2011.10.020. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3263767&tool=pmcentrez&rendertype=abstract.

Herff, C, D Heger, F Putze, C Guan, and T Schultz. 2013. “Self-Paced BCI With NIRS Based on Speech Activity”: 2–3. doi:10.3217/978-3-85125-260-6-111.

Hermes, Dora, Kai J Miller, Herke Jan Noordmans, Mariska J Vansteensel, and Nick F Ramsey. 2010. “Automated Electrocorticographic Electrode Localization on Individually Rendered Brain Surfaces.” *Journal of Neuroscience Methods* 185 (2) (January 15): 293–8. doi:10.1016/j.jneumeth.2009.10.005. http://www.ncbi.nlm.nih.gov/pubmed/19836416.

Hermes, Dora, Kai J Miller, Mariska J Vansteensel, Erik J Aarnoutse, Frans S S Leijten, and Nick F Ramsey. 2012. “Neurophysiologic Correlates of fMRI in Human Motor Cortex.” *Human Brain Mapping* 33 (7) (July): 1689–99. doi:10.1002/hbm.21314. http://www.ncbi.nlm.nih.gov/pubmed/21692146.

Hickok, Gregory. 2012. “Computational Neuroanatomy of Speech Production.” *Nature Reviews Neuroscience* 13: 135–145. http://www.nature.com/nrn/journal/v13/n2/full/nrn3158.html?WT.ec_id=NRN-201202.

Hochberg, Leigh R, Mijail D Serruya, Gerhard M Friehs, Jon a Mukand, Maryam Saleh, Abraham H Caplan, Almut Branner, David Chen, Richard D Penn, and John P Donoghue. 2006. “Neuronal Ensemble Control of Prosthetic Devices by a Human with Tetraplegia.” *Nature* 442 (7099) (July 13): 164–71. doi:10.1038/nature04970. http://www.ncbi.nlm.nih.gov/pubmed/16838014.

Hochberg, Leigh R., Daniel Bacher, Beata Jarosiewicz, Nicolas Y. Masse, John D. Simeral, Joern Vogel, Sami Haddadin, et al. 2012. “Reach and Grasp by People with Tetraplegia Using a Neurally Controlled Robotic Arm.” *Nature* 485 (7398) (May 16): 372–375. doi:10.1038/nature11076. http://www.nature.com/doifinder/10.1038/nature11076.

Hochberg, LR, and DM Taylor. 2007. “Intuitive Prosthetic Limb Control.” *The Lancet*: 345–346. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Intuitive+prosthetic+limb+control#0.

Hoffmann, Ulrich, Jean-Marc Vesin, Touradj Ebrahimi, and Karin Diserens. 2008. “An Efficient P300-Based Brain-Computer Interface for Disabled Subjects.” *Journal of Neuroscience Methods* 167 (1): 115–25. doi:10.1016/j.jneumeth.2007.03.005. http://www.ncbi.nlm.nih.gov/pubmed/17445904.

House, A.S., C. Williams, M.H.L. Hecker, and K.D. Kryter. 1963. "Psychoacoustic Speech Tests: A Modified Rhyme Test." *The Journal of the Acoustical Society of America* 35: 1899. http://link.aip.org/link/?JASMAN/35/1899/4.

Hubel, DH, and TN Wiesel. 1962. "Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex." *The Journal of Physiology*: 106–154. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1359523/.

Huggins, Jane E, Patricia A Wren, and Kirsten L Gruis. 2011. "What Would Brain-Computer Interface Users Want? Opinions and Priorities of Potential Users with Amyotrophic Lateral Sclerosis." *Amyotrophic Lateral …* 12 (5): 318–324. doi:10.3109/17482968.2011.572978.What. http://informahealthcare.com/doi/abs/10.3109/17482968.2011.572978.

Ivry, Richard, Jörn Diedrichsen, Rebecca Spencer, Eliot Hazeline, and Andras Semjen. 2004. *A Cognitive Neuroscience Perspective on Bimanual Coordination and Interference*. Edited by S. Swinnen & J. Duysens. *… of Interlimb Coordination*. Boston: Kluwer Academic Publishing. http://link.springer.com/chapter/10.1007/978-1-4419-9056-3_10.

Jackson, M M, K Ozawa, K Kido, I Mcclendon, and R Kerwin. 2013. "Field Study of an fNIR-Based Brain-Computer Interface for Communication": 6–7. doi:10.3217/978-3-85125-260-6-101.

Jou, Szu-Chen Stan, and Tanja Schultz. 2008. "Automatic Speech Recognition Based on Electromyographic Biosignals." *BIOSTEC*: 305–320.

Jou, Szu-Chen Stan, Tanja Schultz, and Alex Waibel. 2005. "Whispery Speech Recognition Using Adapted Articulatory Features." In *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference On*, 1:1009–1012. IEEE. http://ieeexplore.ieee.org/iel5/9711/30650/01415287.pdf?arnumber=1415287.

———. 2007. "Continuous Electromyographic Speech Recognition with a Multi-Stream Decoding Architecture." *Communication*: 401–404.

Kellis, Spencer, K. Miller, Kyle Thomson, Richard Brown, Paul House, and Bradley Greger. 2010a. "Classification of Spoken Words Using Surface Local Field Potentials." In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, 3827–3830. IEEE. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5627682.

Kellis, Spencer, Kai Miller, Kyle Thomson, Richard Brown, Paul House, and Bradley Greger. 2010b. "Decoding Spoken Words Using Local Field Potentials Recorded from the Cortical Surface." *Journal of Neural Engineering* 7 (5) (October): 056007. doi:10.1088/1741-2560/7/5/056007. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2970568&tool=pmcentrez&rendertype=abstract.

Kennedy, P R, and R a Bakay. 1998. “Restoration of Neural Output from a Paralyzed Patient by a Direct Brain Connection.” *Neuroreport* 9 (8) (June 1): 1707–11. http://www.ncbi.nlm.nih.gov/pubmed/9665587.

Kronegg, Julien, S Voloshynovskiy, and T Pun. 2005. “Analysis of Bit-Rate Definitions for Brain-Computer Interfaces.” *CSREA HCI*. http://cvml.unige.ch/publications/postscript/2005/Kronegg2005__Analysis_of_bit_rate_definitions_for_BCIs.pdf.

Kübler, Andrea, Boris Kotchoubey, Thilo Hinterberger, Nimr Ghanayim, Juri Perelmouter, Margarete Schauer, Christoph Fritsch, Edward Taub, and Niels Birbaumer. 1999. “The Thought Translation Device: a Neurophysiological Approach to Communication in Total Motor Paralysis.” *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale* 124 (2) (January): 223–32. http://www.ncbi.nlm.nih.gov/pubmed/9928845.

Kübler, Andrea, and Nicola Neumann. 2004. “Predictability of Brain-Computer Communication.” *Journal of ...* 18: 121–129. doi:10.1027/0269-8803.18.2. http://psycnet.apa.org/journals/jop/18/2-3/121/.

Kübler, Andrea, Nicola Neumann, Jochen Kaiser, Boris Kotchoubey, Thilo Hinterberger, and Niels P. Birbaumer. 2001. “Brain-Computer Communication: Self-Regulation of Slow Cortical Potentials for Verbal Communication.” *Archives of Physical Medicine and Rehabilitation* 82 (11) (November): 1533–1539. doi:10.1053/apmr.2001.26621. http://linkinghub.elsevier.com/retrieve/pii/S0003999301531901.

Kübler, Andrea, Femke Nijboer, and J Mellinger. 2005. “Patients with ALS Can Use Sensorimotor Rhythms to Operate a Brain-Computer Interface.” *Neurology* 64 (10). doi:10.1212/01.WNL.0000158616.43002.6D. http://www.ncbi.nlm.nih.gov/pubmed/15911809.

Kushler, Clifford A., and Randal J. Marsden. 2006. “System and Method for Continuous Stroke Word-Based Text Input”. USA.

Ladefoged, P. 1990. “The Revised International Phonetic Alphabet.” *Language* 66 (3): 550–552. http://www.jstor.org/stable/414611.

Lee, Ki-Seung. 2008. “EMG-Based Speech Recognition Using Hidden Markov Models with Global Control Variables.” *IEEE Transactions on Bio-Medical Engineering* 55 (3) (March): 930–40. doi:10.1109/TBME.2008.915658. http://www.ncbi.nlm.nih.gov/pubmed/18334384.

Leuthardt, Eric C, John Cunningham, and Dennis Barbour. 2013. “Towards a Speech BCI Using ECoG.” In , edited by Christoph Guger, Brendan Z. Allison, and Günter Edlinger, 93–110. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-36083-1. http://link.springer.com/10.1007/978-3-642-36083-1.

Leuthardt, Eric C, Charles Gaona, Mohit Sharma, Nicholas Szrama, Jarod Roland, Zac Freudenberg, Jamie Solis, Jonathan Breshears, and Gerwin Schalk. 2011. “Using the Electrocorticographic Speech Network to Control a Brain-Computer Interface in Humans.” *Journal of Neural Engineering* 8 (3) (June): 036004. doi:10.1088/1741-2560/8/3/036004. http://www.ncbi.nlm.nih.gov/pubmed/21471638.

Leuthardt, Eric C, Gerwin Schalk, Jonathan R Wolpaw, Jeffrey G Ojemann, and Daniel W Moran. 2004. “A Brain–computer Interface Using Electrocorticographic Signals in Humans.” *Journal of Neural Engineering* 1 (2) (June): 63–71. doi:10.1088/1741-2560/1/2/001. http://www.ncbi.nlm.nih.gov/pubmed/15876624.

Liberman, a M, and I G Mattingly. 1985. “The Motor Theory of Speech Perception Revised.” *Cognition* 21 (1) (October): 1–36. http://www.ncbi.nlm.nih.gov/pubmed/4075760.

Long, GS, and JL Berke. 2010. “Functions of the Larynx and Production of Sounds.” In *Handbook of Mammalian Vocalization*, edited by Stefan M. Brudzynski, 419–426. Elsevier. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:HANDBOOK+OF+MAMMALIAN+VOCALIZATION#4.

MacKenzie, I. Scott, and R. William Soukoreff. 2003. “Phrase Sets for Evaluating Text Entry Techniques.” *CHI ’03 Extended Abstracts on Human Factors in Computer Systems* - *CHI '03*: 754. doi:10.1145/765968.765971. http://portal.acm.org/citation.cfm?doid=765891.765971.

Maier-Hein, Lena. 2005. “Speech Recognition Using Surface Electromyography.”

Manning, Jeremy R, Michael R Sperling, Ashwini Sharan, Emily a Rosenberg, and Michael J Kahana. 2012a. “Spontaneously Reactivated Patterns in Frontal and Temporal Lobe Predict Semantic Clustering During Memory Search.” *The Journal of ...* 32 (26): 8871–8878. doi:10.1523/JNEUROSCI.5321-11.2012. http://www.jneurosci.org/content/32/26/8871.short.

———. 2012b. “Spontaneously Reactivated Patterns in Frontal and Temporal Lobe Predict Semantic Clustering During Memory Search.” *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience* 32 (26) (June 27): 8871–8. doi:10.1523/JNEUROSCI.5321-11.2012. http://www.ncbi.nlm.nih.gov/pubmed/22745488.

Mathematica, Wolfram. “Average English Word Length.” *Wolfram|Alpha*.

McGurk, Harry, and John MacDonald. 1976. “Hearing Lips and Seeing Voices.” *Nature* 264. http://www.nature.com/nature/journal/v264/n5588/abs/264746a0.html.

McWhirter, Norris, ed. 1985. *The Guinness Book of World Records*. 23rd US ed. New York, New York, USA: Sterling Publishing Co., Inc.

Mellinger, J, T Hinterberger, M Bensch, M Schröder, and N Birbaumer. 2003. “Surfing the Web with Electrical Brain Signals : The Brain Web Surfer (BWS) for the Completely Paralysed”: 1–7.

Metze, Florian, Lena Maier-Hein, Tanja Schultz, and Alex Waibel. 2005. “Session Independent Non-Audible Speech Recognition Using Surface Electromyography.” *In Proc. ASRU, San Juan, Puerto Rico.*

Miller, Kai J, Adam O Hebb, Dora Hermes, Marcel den Nijs, Jeffrey G Ojemann, and Rajesh N P Rao. 2010. “Brain Surface Electrode Co-Registration Using MRI and X-Ray.” *Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* 2010 (January): 6015–8. doi:10.1109/IEMBS.2010.5627597. http://www.ncbi.nlm.nih.gov/pubmed/21097113.

Mines, M a, B F Hanson, and J E Shoup. 1978. “Frequency of Occurrence of Phonemes in Conversational English.” *Language and Speech* 21 (3): 221–41. http://www.ncbi.nlm.nih.gov/pubmed/15641442.

Moore, RK. 2003. “A Comparison of the Data Requirements of Automatic Speech Recognition Systems and Human Listeners.” *INTERSPEECH*: 2581–2584. ftp://ftp.dcs.shef.ac.uk/home/spc/ES030106.PDF.

Morse, Samuel. 1849. “Improvement in Electric Telegraphs”. USA.

Mugler, E M, R D Flint, Z A Wright, S U Schuele, J Rosneow, and J L Patton. 2013. “Decoding Articulatory Properties of Overt Speech from Electrocorticography.” *Proceedings of the Fifth International Brain-Computer Interface Meeting 2013, Pacific Grove, CA, June 3-7, 2013*. 4–5.

Mugler, Emily, Michael Bensch, Sebastian Halder, Wolfgang Rosenstiel, Martin Bogdan, Niels Birbaumer, and Andrea Kübler. 2008. “Control of an Internet Browser Using the P300 Event-Related Potential” 10 (1): 56–63. http://ijbem.ws.hosei.ac.jp/volume10/number1/100108.pdf.

Mugler, Emily M, Carolin A Ruf, Sebastian Halder, Michael Bensch, and Andrea Kübler. 2010. “Design and Implementation of a P300-Based Brain-Computer Interface for Controlling an Internet Browser.” *IEEE Transactions on Neural Systems and Rehabilitation Engineering : a Publication of the IEEE Engineering in Medicine and Biology Society* (c) (August). doi:10.1109/TNSRE.2010.2068059. http://www.ncbi.nlm.nih.gov/pubmed/20805058.

Nijboer, Femke, Eric W Sellers, J Mellinger, M a Jordan, Tamara Matuz, A Furdea, Sebastian Halder, et al. 2008. “A P300-Based Brain-Computer Interface for People with Amyotrophic Lateral Sclerosis.” *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology* 119 (8): 1909–16. doi:10.1016/j.clinph.2008.03.034. http://www.ncbi.nlm.nih.gov/pubmed/18571984.

Nykopp, Tommi. 2001. "Statistical Modelling Issues for the Adaptive Brain Interface." *Helsinki University Of*. http://uais.lzu.edu.cn/uploads/soft/101110/1-101110205255.pdf.

Ojemann, George A. 1979. "Individual Variability in Cortical Localization of Language." *Journal of Neurosurgery* 50 (2) (February): 164–9. doi:10.3171/jns.1979.50.2.0164. http://www.ncbi.nlm.nih.gov/pubmed/430127.

———. 1991. "Cortical Organization of Language." *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience* 11 (8) (August): 2281–7. http://www.ncbi.nlm.nih.gov/pubmed/1869914.

Ojemann, George, Jeff Ojemann, E Lettich, and M Berger. 1989. "Cortical Language Localization in Left, Dominant Hemisphere." *Journal of Neurosurgery* 71: 316–326. http://thejns.org/doi/abs/10.3171/JNS/2008/108/2/0411.

Ostrach, Teresia R. 2005. "Typing Speed : How Fast Is Average."

Pasley, Brian N, Stephen V David, Nima Mesgarani, Adeen Flinker, Shihab a Shamma, Nathan E Crone, Robert T Knight, and Edward F Chang. 2012. "Reconstructing Speech from Human Auditory Cortex." *PLoS Biology* 10 (1) (January): e1001251. doi:10.1371/journal.pbio.1001251. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3269422&tool=pmcentrez&rendertype=abstract.

Pei, Xiaomei, Dennis L Barbour, Eric C Leuthardt, and Gerwin Schalk. 2011. "Decoding Vowels and Consonants in Spoken and Imagined Words Using Electrocorticographic Signals in Humans." *Journal of Neural Engineering* 8 (4) (July 13): 046028. doi:10.1088/1741-2560/8/4/046028. http://www.ncbi.nlm.nih.gov/pubmed/21750369.

Pei, Xiaomei, Eric C Leuthardt, Charles M Gaona, Peter Brunner, Jonathan R Wolpaw, and Gerwin Schalk. 2010. "Spatiotemporal Dynamics of Electrocorticographic High Gamma Activity During Overt and Covert Word Repetition." *NeuroImage* (October). doi:10.1016/j.neuroimage.2010.10.029. http://www.ncbi.nlm.nih.gov/pubmed/21029784.

Penfield, W, and T Rasmussen. 1949. "Vocalization and Arrest of Speech." *Archives of Neurology and …*. http://archneurpsyc.ama-assn.org/cgi/reprint/61/1/21.pdf.

Penfield, W., and E. Boldrey. 1937. "Somatic Motor and Sensory Representation in the Cerebral Cortex of Man as Studied by Electrical Stimulation." *Brain: A Journal of Neurology* 60: 389–443. doi:10.1093/brain/60.4.389.

Penfield, Wilder, and Lamar Roberts. 1959. *Speech and Brain Mechanisms*. Princeton, NJ, USA: Princeton University Press.

Perrone-Bertolotti, M., J. Kujala, J. R. Vidal, C. M. Hamame, T. Ossandon, O. Bertrand, L. Minotti, P. Kahane, K. Jerbi, and J.-P. Lachaux. 2012. "How Silent Is Silent Reading?

Intracerebral Evidence for Top-Down Activation of Temporal Voice Areas During Reading.” *Journal of Neuroscience* 32 (49) (December 5): 17554–17562. doi:10.1523/JNEUROSCI.2982-12.2012. http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.2982-12.2012.

Polich, J, and a Kok. 1995. “Cognitive and Biological Determinants of P300: An Integrative Review.” *Biological Psychology* 41 (2) (October): 103–46. http://www.ncbi.nlm.nih.gov/pubmed/8534788.

Popescu, Florin, Siamac Fazli, Yakob Badower, Benjamin Blankertz, and Klaus-R Müller. 2007. “Single Trial Classification of Motor Imagination Using 6 Dry EEG Electrodes.” *PloS One* 2 (7) (January): e637. doi:10.1371/journal.pone.0000637. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1914378&tool=pmcentrez&rendertype=abstract.

Pouget, Alexandre, Jeffrey M Beck, Wei Ji Ma, and Peter E Latham. 2013. “Probabilistic Brains: Knowns and Unknowns.” *Nature Neuroscience* 16 (9) (September): 1170–8. doi:10.1038/nn.3495. http://www.ncbi.nlm.nih.gov/pubmed/23955561.

Pritchard, W S. 1981. “Psychophysiology of P300.” *Psychological Bulletin* 89 (3) (May): 506–40. http://www.ncbi.nlm.nih.gov/pubmed/7255627.

Proctor, Michael, Erik Bresch, Dani Byrd, Krishna Nayak, and Shrikanth Narayanan. 2013. “Paralinguistic Mechanisms of Production in Human ‘Beatboxing’: a Real-Time Magnetic Resonance Imaging Study.” *The Journal of the Acoustical Society of America* 133 (2) (February): 1043–54. doi:10.1121/1.4773865. http://www.ncbi.nlm.nih.gov/pubmed/23363120.

Pulvermüller, Friedemann, Martina Huss, Ferath Kherif, Fermin Moscoso del Prado Martin, Olaf Hauk, and Yury Shtyrov. 2006. “Motor Cortex Maps Articulatory Features of Speech Sounds.” *Proceedings of the National Academy of Sciences of the United States of America* 103 (20) (May 16): 7865–70. doi:10.1073/pnas.0509989103. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1472536&tool=pmcentrez&rendertype=abstract.

Rabiner, L.R., and R.W. Schafer. 1978. *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall.

Reed, Charlotte M., and Nathaniel I. Durlach. 1998. “Note on Information Transfer Rates in Human Communication.” *Presence: Teleoperators <html_ent Glyph=“@amp;” Ascii=“&”/> Virtual Environments* 7 (5) (October): 509–518. doi:10.1162/105474698565893. http://www.mitpressjournals.org/doi/abs/10.1162/105474698565893.

Roland, Jarod, Peter Brunner, James Johnston, Gerwin Schalk, and Eric C Leuthardt. 2010. “Passive Real-Time Identification of Speech and Motor Cortex During an Awake

Craniotomy." *Epilepsy & Behavior : E&B* 18 (1-2) (May): 123–8. doi:10.1016/j.yebeh.2010.02.017. http://www.ncbi.nlm.nih.gov/pubmed/20478745.

Rowland, L P, and N a Shneider. 2001. "Amyotrophic Lateral Sclerosis." *The New England Journal of Medicine* 344 (22) (May): 1688–700. doi:10.1056/NEJM200105313442207. http://www.ncbi.nlm.nih.gov/pubmed/11386269.

Saltzman, EL, and KG Munhall. 1989. "A Dynamical Approach to Gestural Patterning in Speech Production." *Ecological Psychology*: 38–68. http://www.tandfonline.com/doi/abs/10.1207/s15326969eco0104_2.

Schalk, Gerwin. 2008. "Brain-Computer Symbiosis." *Journal of Neural Engineering* 5 (1) (March): P1–P15. doi:10.1088/1741-2560/5/1/P01. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2722922&tool=pmcentrez&rendertype=abstract.

———. 2010. "Can Electrocorticography (ECoG) Support Robust and Powerful Brain-Computer Interfaces?" *Frontiers in Neuroengineering* 3 (June) (January): 9. doi:10.3389/fneng.2010.00009. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2903308&tool=pmcentrez&rendertype=abstract.

Schalk, Gerwin, Eric C Leuthardt, Peter Brunner, Jeffrey G Ojemann, Lester a Gerhardt, and Jonathan R Wolpaw. 2008. "Real-Time Detection of Event-Related Brain Activity." *NeuroImage* 43 (2) (November): 245–9. doi:10.1016/j.neuroimage.2008.07.037. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2603617&tool=pmcentrez&rendertype=abstract.

Schalk, Gerwin, Dennis J McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R Wolpaw. 2004. "BCI2000: a General-Purpose Brain-Computer Interface (BCI) System." *IEEE Transactions on Bio-Medical Engineering* 51 (6) (June): 1034–43. doi:10.1109/TBME.2004.827072. http://www.ncbi.nlm.nih.gov/pubmed/15188875.

Scheme, Erik J, Bernard Hudgins, and Phillip A Parker. 2007. "Myoelectric Signal Classification for Phoneme-Based Speech Recognition." *IEEE Transactions on Biomedical Engineering* 54 (4): 694–699.

Schultz, Tanja, and Alex Waibel. 1998. "Language Independent and Language Adaptive Large Vocabulary Speech Recognition." In *Fifth International Conference on Spoken Language Processing*. Citeseer. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.4.2176&rep=rep1&type=pdf.

Schultz, Tanja, and Michael Wand. 2010. "Modeling Coarticulation in EMG-Based Continuous Speech Recognition." *Speech Communication* 52 (4) (April): 341–353.

doi:10.1016/j.specom.2009.12.002. http://linkinghub.elsevier.com/retrieve/pii/S0167639309001770.

Scott, SH, and JF Kalaska. 1997. "Reaching Movements with Similar Hand Paths but Different Arm Orientations. I. Activity of Individual Cells in Motor Cortex." *Journal of Neurophysiology* (Mi). http://jn.physiology.org/content/77/2/826.short.

Scott, SK, C McGettigan, and Frank Eisner. 2009. "A Little More Conversation, a Little Less Action—candidate Roles for the Motor Cortex in Speech Perception." *Nature Reviews Neuroscience* 10 (April). http://www.nature.com/nrn/journal/v10/n4/abs/nrn2603.html.

Sellers, Eric W, and Emanuel Donchin. 2006. "A P300-Based Brain-Computer Interface: Initial Tests by ALS Patients." *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology* 117 (3) (March): 538–48. doi:10.1016/j.clinph.2005.06.027. http://www.ncbi.nlm.nih.gov/pubmed/16461003.

Sellers, Eric W, Dean J Krusienski, Dennis J McFarland, Theresa M Vaughan, and Jonathan R Wolpaw. 2006. "A P300 Event-Related Potential Brain-Computer Interface (BCI): The Effects of Matrix Size and Inter Stimulus Interval on Performance." *Biological Psychology* 73 (3): 242–52. doi:10.1016/j.biopsycho.2006.04.007. http://www.ncbi.nlm.nih.gov/pubmed/16860920.

Shannon, Claude E. 1951. "Prediction and Entropy of Printed English." *Bell System Technical Journal*. http://languagelog.ldc.upenn.edu/myl/Shannon1950.pdf.

———. 1961. "Two-Way Communication Channels." *Proc. 4th Berkeley Symp. Math. Stat. Prob*. http://home.ustc.edu.cn/~zuoyuan/download/DSP/Two-way communication channels.pdf.

———. 2001. "A Mathematical Theory of Communication." *ACM SIGMOBILE Mobile Computing and …* 5 (I): 3–55. http://dl.acm.org/citation.cfm?id=584093.

Shoham, S, E Halgren, E M Maynard, and R a Normann. 2001. "Motor-Cortical Activity in Tetraplegics." *Nature* 413 (6858) (October 25): 793. doi:10.1038/35101651. http://www.ncbi.nlm.nih.gov/pubmed/11677592.

Simonyan, Kristina, and Barry Horwitz. 2011. "Laryngeal Motor Cortex and Control of Speech in Humans." *The Neuroscientist* 17 (2): 197–208. doi:10.1177/1073858410386727.Laryngeal. http://nro.sagepub.com/content/17/2/197.short.

Slutzky, Marc W, Luke R Jordan, Todd Krieg, Ming Chen, David J Mogul, and Lee E Miller. 2010. "Optimal Spacing of Surface Electrode Arrays for Brain-Machine Interface Applications." *Journal of Neural Engineering* 7 (2) (April): 26004. doi:10.1088/1741-2560/7/2/026004. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2844916&tool=pmcentrez&rendertype=abstract.

Slutzky, Marc W, Luke R Jordan, Eric W Lindberg, Kevin E Lindsay, and Lee E Miller. 2011. "Decoding the Rat Forelimb Movement Direction from Epidural and Intracortical Field Potentials." *Journal of Neural Engineering* 8 (3) (June): 036013. doi:10.1088/1741-2560/8/3/036013. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3124348&tool=pmcentrez&rendertype=abstract.

Sorger, Bettina, Joel Reithler, Brigitte Dahmen, and Rainer Goebel. 2012. "A Real-Time fMRI-Based Spelling Device Immediately Enabling Robust Motor-Independent Communication." *Current Biology : CB* (June 27): 1–6. doi:10.1016/j.cub.2012.05.022. http://www.ncbi.nlm.nih.gov/pubmed/22748322.

Springer, J a, J R Binder, T a Hammeke, S J Swanson, J a Frost, P S Bellgowan, C C Brewer, H M Perry, G L Morris, and W M Mueller. 1999. "Language Dominance in Neurologically Normal and Epilepsy Subjects: a Functional MRI Study." *Brain : a Journal of Neurology* 122 (Pt 1 (November): 2033–46. http://www.ncbi.nlm.nih.gov/pubmed/10545389.

Stark, Eran, and Moshe Abeles. 2007. "Predicting Movement from Multiunit Activity." *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience* 27 (31) (August 1): 8387–94. doi:10.1523/JNEUROSCI.1321-07.2007. http://www.ncbi.nlm.nih.gov/pubmed/17670985.

"Talk About a Fast Show: Londoner Rattles Off Raglan in 25 Seconds." 2011. *Irish Independent*, June 17. http://www.independent.ie/irish-news/talk-about-a-fast-show-londoner-rattles-off-raglan-in-25-seconds-26743158.html.

Tankus, Ariel, Itzhak Fried, and Shy Shoham. 2012. "Structured Neuronal Encoding and Decoding of Human Speech Features." *Nature Communications* 3 (January): 1015. doi:10.1038/ncomms1995. http://www.ncbi.nlm.nih.gov/pubmed/22910361.

Terumitsu, Makoto, Yukihiko Fujii, Kiyotaka Suzuki, Ingrid L Kwee, and Tsutomu Nakada. 2006. "Human Primary Motor Cortex Shows Hemispheric Specialization for Speech." *Neuroreport* 17 (11) (July): 1091–5. doi:10.1097/01.wnr.0000224778.97399.c4. http://www.ncbi.nlm.nih.gov/pubmed/16837833.

Towle, Vernon L, Hyun-Ah Yoon, Michael Castelle, J Christopher Edgar, Nadia M Biassou, David M Frim, Jean-Paul Spire, and Michael H Kohrman. 2008. "ECoG Gamma Activity During a Language Task: Differentiating Expressive and Receptive Speech Areas." *Brain : a Journal of Neurology* 131 (Pt 8) (August): 2013–27. doi:10.1093/brain/awn147. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2724904&tool=pmcentrez&rendertype=abstract.

Turing, A. M. 1950. "I.—Computing Machinery and Intelligence." *Mind* LIX (236): 433–460. doi:10.1093/mind/LIX.236.433. http://mind.oxfordjournals.org/cgi/doi/10.1093/mind/LIX.236.433.

Turkeltaub, Peter E, and H Branch Coslett. 2010. "Localization of Sublexical Speech Perception Components." *Brain and Language* 114 (1) (July): 1–15. doi:10.1016/j.bandl.2010.03.008. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2914564&tool=pmcentrez&rendertype=abstract.

Vidaurre, Carmen, and Benjamin Blankertz. 2010. "Towards a Cure for BCI Illiteracy." *Brain Topography* 23 (2) (June): 194–8. doi:10.1007/s10548-009-0121-6. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2874052&tool=pmcentrez&rendertype=abstract.

Viventi, Jonathan, Dae-Hyeong Kim, Leif Vigeland, Eric S Frechette, Justin a Blanco, Yun-Soung Kim, Andrew E Avrin, et al. 2011. "Flexible, Foldable, Actively Multiplexed, High-Density Electrode Array for Mapping Brain Activity in Vivo." *Nature Neuroscience* 14 (12) (December): 1599–605. doi:10.1038/nn.2973. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3235709&tool=pmcentrez&rendertype=abstract.

Walliczek, Matthias, Florian Kraft, Szu-Chen Stan Jou, Tanja Schultz, and Alex Waibel. 2006. "Sub-Word Unit Based Non-Audible Speech Recognition Using Surface Electromyography." *Test*: 1487–1490.

Wand, Michael, Christopher Schulte, Matthias Janke, and Tanja Schultz. "Array-Based Electromyographic Silent Speech Interface."

Wang, Wei, Jennifer L Collinger, Alan D Degenhart, Elizabeth C Tyler-Kabara, Andrew B Schwartz, Daniel W Moran, Douglas J Weber, et al. 2013. "An Electrocorticographic Brain Interface in an Individual with Tetraplegia." *PloS One* 8 (2) (January): e55344. doi:10.1371/journal.pone.0055344. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3566209&tool=pmcentrez&rendertype=abstract.

Wang, Yijun, Zhiguang Zhang, Xiaorong Gao, and Shangkai Gao. 2004. "Lead Selection for SSVEP-Based Brain-Computer Interface." *Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* 6 (January): 4507–10. doi:10.1109/IEMBS.2004.1404252. http://www.ncbi.nlm.nih.gov/pubmed/17271308.

Ward, David J., Alan F. Blackwell, and David J. C. MacKay. 2000. "Dasher---a Data Entry Interface Using Continuous Gestures and Language Models." *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology - UIST '00* 2: 129–137. doi:10.1145/354401.354427. http://portal.acm.org/citation.cfm?doid=354401.354427.

Wernicke, Carl. 1874. *Der Aphasische Symptomencomplex: Eine Psychologische Studie Auf Anatomischer Basis*. Max Cohn & Weigert.

Wilson, Stephen M, Ayşe Pinar Saygin, Martin I Sereno, and Marco Iacoboni. 2004. "Listening to Speech Activates Motor Areas Involved in Speech Production." *Nature Neuroscience* 7 (7) (July): 701–2. doi:10.1038/nn1263. http://www.ncbi.nlm.nih.gov/pubmed/15184903.

Wolpaw, Jonathan R, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan. 2002a. "Brain-Computer Interfaces for Communication and Control." *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology* 113 (6) (June): 767–91. http://www.ncbi.nlm.nih.gov/pubmed/12048038.

———. 2002b. "Brain-Computer Interfaces for Communication and Control." *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology* 113 (6) (June): 767–91.

Wolpaw, Jonathan R, Dennis J McFarland, Theresa M Vaughan, and Gerwin Schalk. 2003. "The Wadsworth Center Brain-Computer Interface (BCI) Research and Development Program." *IEEE Transactions on Neural Systems and Rehabilitation Engineering : a Publication of the IEEE Engineering in Medicine and Biology Society* 11 (2) (June): 204–7. doi:10.1109/TNSRE.2003.814442. http://www.ncbi.nlm.nih.gov/pubmed/12899275.

Zander, Thorsten Oliver, Moritz Lehne, Klas Ihme, Sabine Jatzev, Joao Correia, Christian Kothe, Bernd Picht, and Femke Nijboer. 2011. "A Dry EEG-System for Scientific Research and Brain-Computer Interfaces." *Frontiers in Neuroscience* 5 (May) (January): 53. doi:10.3389/fnins.2011.00053. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3103872&tool=pmcentrez&rendertype=abstract.

VITA

EDUCATION

2007–present **University of Illinois at Chicago**, Chicago, IL, USA
Ph.D. candidate in Neural Engineering, Bioengineering Department; 3.65 GPA

2006–2007 **Eberhard-Karls Universität Tübingen**, Tübingen, Germany
Visiting Fulbright Scholar in Brain-Computer Interface at the Institute of Medical Psychology and Behavioral Neurobiology

2002–2006 **Duke University**, Durham, NC, USA
Bachelor of Science in Biomedical Engineering and Certificate in Neuroscience; 3.180 GPA

1998–2002 **Hudson High School**, Hudson, OH, USA
Honors Diploma; 4.223 GPA

RESEARCH EXPERIENCE

2007–present **Doctoral Student** Bioengineering Department, University of Illinois at Chicago
Designed experimental protocols to investigate human motor movement and neural signal during overt speech; designed Matlab analytical framework and user interfaces for biosignal classification
(Advisors: Patrick J. Rousche, PhD (2007-2009), James L. Patton, PhD (2009-present))

2006–2007 **Fulbright Scholar** Eberhard-Karls University of Tübingen, Institute for Medical Psychology
Created an open-source internet browser for use with control without voluntary movement; adapted software for patients with amyotrophic lateral sclerosis; acted as a cultural diplomatic representative and created international research network. (Advisor: Andrea Kübler, PhD)

2005–2006 **Pratt Undergraduate Research Fellow** Wolf Electrophysiology Lab, Pratt School of Engineering, Duke University
Designed behavioral experiments for animal subjects; constructed electrical circuitry for positive reward feedback systems; prepared written reports for publication and oral presentations for the Pratt Board of Trustees; trained animal subjects following U.S. Government regulations. (Advisor: Patrick Wolf, PhD)

2004 **Research Intern** McGovern Institute for Brain Research, Moore Lab, Massachusetts Institute of Technology
Designed somatosensory experiments and apparatus for animal subjects; trained animal subjects using a manual positive reward feedback system; analyzed and documented somatosensory data for publication. (Advisor: Christopher I. Moore, PhD)

2002 **Research Intern** Walter A. Hoyt, Jr. Musculoskeletal Research Laboratory Summa Medical Center
Collected load-cell data on bones of rats; presented data and documented results in written reports; self-taught statistical analysis. (Advisor: Michael J. Askew, MD)

PUBLICATIONS

Peer-Reviewed Journal Articles

2013 **E. M. Mugler**, J. L. Patton, M. W. Slutzky, “Cortical organization and decoding of phonetic structure in speech motor cortex,” in prep.

2010 **E. M. Mugler**, C.A. Ruf, S. Halder, M. Bensch, A. Kübler, "Design and Implementation of a P300-Based Brain-Computer Interface for Controlling an Internet Browser," *IEEE Trans. on Neural Sys. & Rehab. Eng.*, Vol.18, No.6, pp.599-609, Dec. 2010

2008 **E. M. Mugler**, M. Bensch, S. Halder, W. Rosenstiel, M. Bogdan, N. Birbaumer, A. Kübler, “Control of an Internet Browser Using the P300 Event-Related Potential,” International Journal of Bioelectromagnetism. Vol. 10, No. 1, pp. 56 - 63, 2008.

Abstracts

2013 **E. M. Mugler**, R. D. Flint, Z. A. Wright, S. U. Schuele, J. Rosenow, J. L. Patton, M. W. Slutzky, "Decoding Articulatory Properties of Overt Speech from Electrocorticography," Proceedings of the Fifth International Brain-Computer Interface Meeting 2013, Pacific Grove, CA, June 3-7, 2013.

2013 **E. M. Mugler**, J. L. Patton, M. Goldrick, M. W. Slutzky. Functional categorization and contextual independence of cortical representation of speech. Society for Neuroscience 43rd Annual Meeting San Diego, CA.

2012 **E. M. Mugler**, M. W. Slutzky, J. L. Patton. Phonemic differences in cortical representation of overt speech. Society for Neuroscience 42nd Annual Meeting, New Orleans, LA.

2011 **E. M. Mugler**, M. W. Slutzky, J. L. Patton. Electromyographic-speech decoding with electrocorticographic correlations: Toward utilization of phoneme production for brain-computer interface. Society for Neuroscience 41st Annual Meeting, Washington, DC.

2010 **E. M. Mugler**, P. J. Rousche. Development of a laryngeal surface electromyographic biofeedback system for an efficient neurally-controlled communication interface. Fourth International Brain-Computer Interface Meeting 2010, Pacific Grove, CA.

2009 **E. M. Mugler**, C. Ruf, S. Halder, M. Bensch, A. Kübler. The P300-Brain-Computer Interface Browser: Development and Criteria for Evaluation. Berlin Brain-Computer-Interface 2009, Berlin, Germany.

2009 **E. M. Mugler**, P. J. Rousche. Speech rehabilitation using laryngeal electromyography feedback. Biomedical Engineering Society Annual Meeting, Pittsburgh, PA, October 7-10.

2003 M. J. Askew, GB. Schneider, K. J. Grecco, J. Hsu, **E. M. Mugler**, D. A. Noe, “Effect of Pharmaceutical Bone Growth Stimulation with Novel Anabolic Peptides: Biomechanical and Bone Density Measurements in a Rat Model,” Proceedings of IMECE ’03, 2003 ASME International Mechanical Engineering Congress & Exposition, Washington, D.C., November 16-21, 2003, IMECE2003-43044.

2003 J. A. Edwards, K. A. Greene, R. S. Davis, M. W. Kovacik, **E. M. Mugler**, D. A. Noe, “Measurement of Maximum Knee Flexion Following Total Knee Arthroplasty,” Mid-American Orthopaedic Association conference presentation, April 23-27, 2003.

HONORS AND AWARDS

2009–present Grant from the NSF's Integrative Graduate Education and Research Traineeship for thesis research

2011 Graduate Student Representative selected to represent the University of Illinois at Chicago - Computational Transportation Science IGERT in the National IGERT Trainee Poster Competition. (Chicago, IL, USA).

2010 Winner, Award for Innovation in the Student Poster Competition at the Fourth International BCI Meeting in Pacific Grove, CA.

2009 Graduate Student Representative selected to accompany UIC delegates to NSF-funded international intellectual exchange on "Nanoneuronics" at Trinity College Dublin (Dublin, Ireland) and University of Ulster (Ulster, Northern Ireland).

2008 - 2011 Graduate Student Representative of Biomedical Engineering Society (BMES)
Appointed to position by undergraduates; assisted in reviving UIC chapter

2008 Conference Travel Scholarship, 4th International Summer School on Emerging Technologies in Biomedicine, "Advanced Methods for the Estimation of Human Brain Activity and Connectivity, Applications to Rehabilitation Engineering," University of Patras (Patras, Greece).

2007 - 2009 Teaching Assistantship, 50% Appointment and Tuition Waiver

2006 - 2007 Fulbright Grant to Germany, Institute of International Education

2006 Dean's List, Pratt School of Engineering, Duke University (Durham, NC, USA).

2005 - 2006 Pratt Undergraduate Research Fellowship, Duke University (Durham, NC, USA).

2002 Dean's List, Pratt School of Engineering, Duke University (Durham, NC, USA).

2002 Emily and Donald Barlow Scholarship (Hudson, OH, USA).

2002 Hudson Bicentennial Scholarship (Hudson, OH, USA).

INVITED TALKS

February 19 2013 **Invited Speaker** (Host: Cara Stepp, PhD) "Advances for Efficient Communication in Brain-Computer Interface," Department of Speech, Language & Hearing Sciences, Sargent College of Health and Rehabilitation Sciences, Boston University (Boston, MA, USA).

October 3 2012 **Keynote Speaker** (Host: Bonnie Williams, PhD) "From Student to Researcher: My Journey as an Integrated BioScientist," STEM Speaker Series at The University of Akron (Akron, OH, USA).

October 4 2012 **Invited Speaker** (Host: Greg Smith, PhD) "Design and Implementation of a P-300-Based Brain-Computer Interface for Controlling an Internet Browser," Integrated Biosciences Colloquia, Biology Department, The University of Akron (Akron, OH, USA).

August 12 2011 **Graduate Student Speaker** (Host: W. Zev Rymer, MD, PhD) "EMG and ECoG Interfaces for Speech and Spoken Communication" Sensory Motor Performance Program Seminar, Rehabilitation Institute of Chicago (Chicago, IL, USA).

TEACHING EXPERIENCE

April 4 2013 **Guest lecturer** (Host: James L. Patton, PhD) "Poor signals and BMI prospects" Biocontrol, Bioengineering Department, University of Illinois at Chicago (Chicago, IL, USA).

August 31 2012 **Guest lecturer** (Host: James L. Patton, PhD) "How to get the most out of the BioE Seminar course", Bioengineering Department, University of Illinois at Chicago (Chicago, IL, USA).

TEACHING EXPERIENCE *(cont'd)*

February 16 2010 **Guest lecturer** (Host: John Hetling, PhD) "Fourier Transforms and Frequency Analysis", Bioengineering Department, University of Illinois at Chicago (Chicago, IL, USA).

Spring 2010 **Teaching Assistant** Bioinstrumentation and Measurements I, Department of Bioengineering, University of Illinois at Chicago (Chicago, IL, USA).

Spring 2009, Spring 2008 **Teaching Assistant** Introduction to Cell and Tissue Engineering, Department of Bioengineering, University of Illinois at Chicago (Chicago, IL, USA).

Fall 2009 **Teaching Assistant** Modeling Physiological Data and Systems, Department of Bioengineering, University of Illinois at Chicago (Chicago, IL, USA).

Fall 2008, Fall 2007 **Teaching Assistant** Introduction to Bioengineering, Department of Bioengineering, University of Illinois at Chicago (Chicago, IL, USA).

ACADEMIC SERVICE

2012-present **Reviewer** IEEE Engineering in Medicine and Biology Conference

April 2013 **Volunteer judge** Next Generation Innovators Challenge, Midwest Research Competition: Positive Impact, Wheeling High School, Wheeling, Illinois, USA.

August 2012 **Grant consultant** SBIR, Ensis Scientific Consulting

RELEVANT GRADUATE COURSEWORK

Neural Engineering I and II, Biological Signal Analysis, Brain Machine Interfaces: Theory and Practice, Models of the Nervous System, Biorobotics, Neural Networks, Sensory Prostheses, Materials in Bioengineering, Bioinstrumentation and Measurement, Machine Learning

OTHER TRAINING

Fluent in MatLab (Signal Processing Toolbox, Statistics Toolbox, Graphical User Interfaces), Macromedia, Adobe Illustrator, BCI2000 and Microsoft Office
Proficient in LabView, Simulink, JavaScript, Unix, Latex, Flash, Neuron, Google SketchUp and HTML
Fluent in German

PROFESSIONAL AFFILIATIONS

Society of Women Engineers - member
IEEE - Graduate Student Member
Chicago Acoustic Underground - Singer/Songwriter
Delta Gamma Fraternity - President of Beta Theta chapter, Duke University, 2005; alumna member
U.S. Lacrosse - Coach at high school level, member

SCHOLARLY INTERESTS

Brain-computer interface design and application; increasing information transfer rate of brain-computer interfaces; rehabilitation engineering; error augmentation and biofeedback; neuroscience of speech production

LONG-TERM GOAL

Establishing innovative brain-computer interface paradigms for communication to improve the quality of life for individuals with communication disorders as well as the general population