The Thesis Committee for Brian Yinbok Tsang
Certifies that this is the approved version of the following Thesis:

Neural Speech Decoding from Segregated Brain Regions' Neuromagnetic Signature

APPROVED BY
SUPERVISING COMMITTEE:

Jun Wang, Supervisor

Paul Ferrari

Abbas Babajani-Feremi

Neural Speech Decoding from Segregated Brain Regions' Neuromagnetic Signature

by

Brian Yinbok Tsang

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Engineering

The University of Texas at Austin

May 2022

Acknowledgements

I would like to first acknowledge my supervisor Dr. Jun Wang for his incredible support over the course of my research career starting back in my undergraduate years. The research opportunities Dr. Wang has given me over the years has helped tremendously in growing and sustaining my interest in research, culminating in this dissertation. I would like to also express my sincere appreciation for Debadatta (Dave) Dash, who provided me amazing mentorship, guidance, and feedback on the topics of this dissertation. I want to also thank my committee members, Dr. Paul Ferrari and Dr. Abbas Babajani-Feremi, for their helpful suggestions and feedback on my work throughout the process. Lastly, I would like to acknowledge the participants of this study for their time and effort spent providing the data used in this dissertation.

Abstract

Neural Speech Decoding from Segregated Brain Regions' Neuromagnetic Signature

Brian Yinbok Tsang, M.S.

The University of Texas at Austin, 2022

Supervisor: Jun Wang

Abstract: Locked-in syndrome (LIS) is a condition in which patients are in full-body paralysis but retain cognitive function. Improving the ability of these patients to communicate with others may significantly increase their quality of life. Since motor control is almost completely gone for LIS patients, a brain-computer interface (BCI) using neural speech decoding may be the only method to restore communication ability. As a non-invasive neuroimaging modality, magnetoencephalography (MEG) offers several strong attributes for potential speech-BCI usage. Its high spatial and temporal resolutions make it suitable for potential real-time speech decoding, and its non-invasive quality is crucial for accessibility. Neural speech decoding using MEG is a relatively new field, and there is little research on how different regions of the brain contribute to speech decoding. Additionally, the contributions of different brainwaves from individual brain regions towards MEG speech decoding hasn't been studied. Such information may be helpful for future MEG speech-BCI studies regarding feature selection and sensor placement. This dissertation investigates the degree in which MEG signals from the eight

individual brain regions corresponding to the lobes of the brain (left- and right- frontal, parietal, temporal, and occipital lobes) contribute towards neural speech decoding, as well as how individual brainwaves from these areas influence decoding performance. Both imagined and overt speech phrases were decoded using MEG signals collected from seven healthy participants. Results indicated that almost all regions provided neuromagnetic signals helpful for speech decoding. Additionally, both temporal regions offered relatively high individual decoding performances, and when combined, speech decoding performance was almost on-par with whole-brain performance. Lastly, frequency-specific analysis showed signs that helpful MEG signals for speech decoding may be encoded similarly across brain regions, with the delta band (0.3-4Hz) consistently providing the best information for speech decoding. Possessing flexibility in sensor count and placements, optically pumped magnetometer (OPM) MEG systems may find the results of this dissertation useful in determining sensor placements and reducing overall sensor count, which would lower the cost of the system.

Table of Contents

List of Tables

List of Figures

Introduction and Background

Amyotrophic Lateral Sclerosis (ALS) and Locked-In Syndrome (LIS)

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disease that causes motor neurons to gradually deteriorate over several years, eventually leading to death due to the failure of the respiratory system for most patients. There is currently no cure for ALS. Patients with ALS tend to live until around 2-5 years after diagnosis and suffer from rapid loss in motor functions during those years (Wijesekera & Nigel Leigh, 2009). One of the commonly lost motor functions is speech, which is impacted through the deterioration of bulbar motor neurons linked to speech function. As the disease progresses, speech decline can be characterized by decreases in both speaking rate and speech intelligibility (Ball et al., 2002; Green et al., 2013; Makkonen et al., 2018). Research testing performance of ALS patients on various pragmatic speech and comprehension tasks has shown that ALS may also negatively impact other important aspects of communication such as maintaining discourse topic and interpreting non-literal meanings (Bambini et al., 2016). With a decline in communication capabilities, ALS patients naturally experience a decline in quality of life (QoL) in tandem. In fact, verbal communication ability has been shown in research to be directly linked to QoL in ALS patients (Felgoise et al., 2015). Hence, it would be helpful for those suffering from ALS to receive some form of communication aid to improve their QoL. Current aids include speech generating devices (SGDs), which typically rely on movements still in control of the patient such as eye gaze to type out words (Beukelman et al., 2011). However, SGDs are particularly slow compared to verbal communication and hence can lead to the forgetting of information and decline in attention during conversation (Higginbotham et al., 2016). Therefore, a speech aid that can facilitate a more direct method of

communication would be an improvement over SGDs and can help ALS patients improve their overall QoL.

ALS patients can eventually progress into a condition called Locked-In Syndrome (LIS), which involves the total loss of control over voluntary muscles while maintaining consciousness. Besides ALS, LIS is can be caused by trauma or hemorrhage resulting in damage to the ventral pons. LIS patients are quadriplegic and antarthritic, meaning that they cannot move or communicate verbally at all (Plum & Posner, 1982). Some patients maintain control over eye movements and facial movements, which can be utilized in nonverbal communication (Smith & Delargy, 2005). However, this form of communication is limited and can be fatiguing. Despite the severe limitations that come with the condition, research has actually shown that a significant portion of chronic LIS patients are satisfied with life and even attempt to go to social activities regularly (Laureys et al., 2005; León-Carrión et al., 2002). Additionally, LIS patients who medically stabilize within a year of brainstem damage tend to live for many years after, with one study showing that 83% lived more than 10 years after injury (Doble et al., 2003). Keeping these aspects of LIS in mind, it would be incredibly beneficial for the QoL of LIS patients if they were provided a method of communication that is fast and accurate. Since LIS patients are severely limited in motor function, most existing speech generation aids aren't effective, so the easiest and most direct option would be to utilize brain activity directly.

BRAIN-COMPUTER INTERFACES

One category of promising solutions for improving communication ability of LIS patients is the Brain-Computer Interface (BCI). BCIs are systems that measure neural signals from the user's brain and translate those signals into commands for an external

device. These systems consist of three main components: a data acquisition device to measure the user's brain activity, a translation algorithm that converts the input brain signals into output commands, and lastly the external device to be controlled (Wolpaw et al., 2000). Both the user and the translation algorithm are adaptive controllers in the system, so BCIs generally require training of the user as well as tuning of the algorithm to achieve optimal performance (Vaughan et al., 2003). Through directly interpreting the brain signals of users, BCIs bypass motor control and allow users with motor disabilities to operate devices (Millán et al., 2010). Research in the BCI field has explored numerous applications for users with motor disabilities such as operating cursors, prosthetic arms, humanoid robots, and even drones (Abiri et al., 2019).

BCI-Spellers

With the goal of improving the communication abilities of users who have speech hinderances, BCI-spellers are a type of BCI that allows users to pick out characters for spelling out sentences without the use of motor control. The major BCI-spellers use electroencephalograms (EEGs) to measure electrophysiological signals from users' brain. Through electrodes placed on the scalp, EEGs record electrical activity in the brain caused by neurons firing. EEGs are often used in BCI research due to its non-invasive nature and relative affordability, making it a practical candidate for commercial BCIs. Some BCI-spellers look specifically at the P300 wave, an event-related potential that can be spotted using EEG around 300ms after a specific event (McCarthy & Donchin, 1981). These spellers typically use visual or audio stimuli to evoke the P300 signal, with the most well-known BCI-speller flashing rows and columns of a letter matrix to determine the user's desired letter (Farwell & Donchin, 1988). Since these spellers have to show multiple instances of stimuli per desired letter, communication rate is fairly slow, with

most spellers hovering at a rate of around 2-5 characters per minute (Jin et al., 2012; Rezeika et al., 2018). A faster type of BCI-speller uses Steady-State Visual Evoked Potentials (SSVEP), which is a type of evoked potential that matches the frequency of a flashing visual stimulus. In this speller, different targets, such as letters, are assigned unique flickering frequencies, and the BCI determines the target by finding the frequency of the SSVEP. These spellers are faster than P300 spellers on average, with advanced SSVEP spellers hitting communication rates ranging from 25 to 60 characters per minute (Chen et al., 2015; Nakanishi et al., 2018; Spüler et al., 2012). However, these rates are still very slow, especially when compared to rates during normal speaking (around 200 words per minute). Additionally, users of both types of spellers must continually focus on targets during generation, which can fatigue the users in longer sessions.

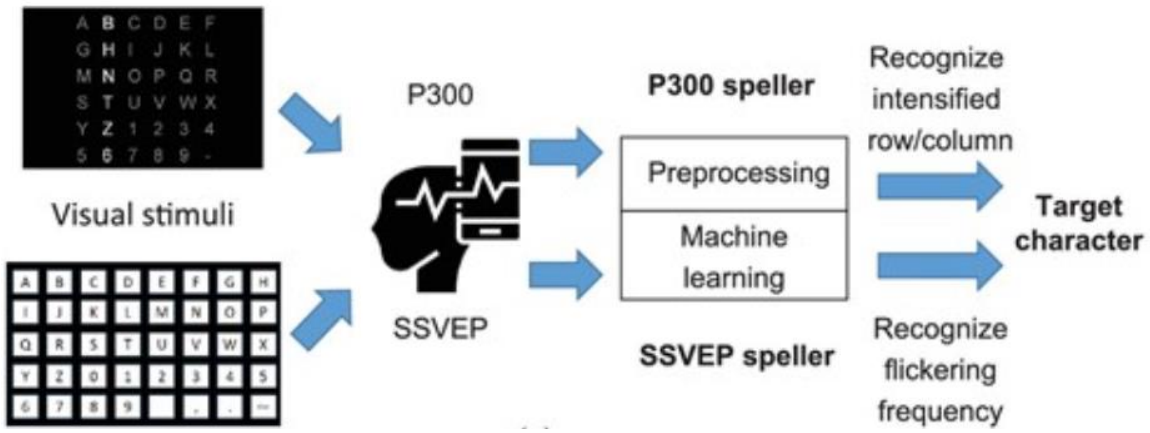


Figure 1: Processing pathway of P300 and SSVEP BCI-spellers, from (Zhang et al., 2021) Copyright @ 2020, Oxford University Press.

Speech-BCIs

Speech-BCIs aim to alleviate issues of BCI-spellers by decoding speech directly from the collected brain signals. This type of translation from neural signals to speech,

also known as neural speech decoding, is currently often achieved through a combination of signal processing and machine learning. If performance in the decoding can consistently stay at high levels, speech-BCIs would easily be faster than BCI-spellers by directly outputting speech instead of using BCI-spellers. Communication rate may even approach that of normal speech if systems become advanced enough. Generally, speech-BCIs would output text transcripts of the speech, but text-to-speech systems can be utilized after text generation to produce speech sounds. Overall, speech-BCIs offers a very promising solution for LIS patients as a communication aid, and research in the field has been advancing quickly, partially due to the large improvements in machine learning in recent years.

Speech-BCIs can be split into several types, two of such are those that decode for closed set vocabulary and those that decode for open set vocabulary. Closed set vocabulary decoding involves the use of a limited set of stimuli during data collection, such as a set of four syllables or a set of six words (Jahangiri & Sepulveda, 2018; Koizumi et al., 2018; Panachakel & Ramakrishnan, 2021). The decoding task was to simply distinguish the stimuli from each other using the brain activity corresponding to each stimulus. As such, closed set decoding is much easier than open set decoding, in which the model tries to decode the phonemes or words used in a variety of sentences. The difficulty in open set decoding is variable and can be roughly gauged by the number of unique words in the sentence data collected. Achieving high performance in open set decoding is the end goal for speech-BCIs, but due to its difficultly, many studies are focused on improving closed set vocabulary as a first step.

Another distinction that can be made for speech-BCIs are those that decode overt speech and those that decode imagined speech. Overt speech is speech generated through complete articulation of the speech sounds at a normal audible level. Imagined speech,

also sometimes known as inner speech, is speech that is imagined but not articulated, thereby producing no audible output. The final goal for most speech-BCIs is to decode imagined speech accurately, but overt speech decoding can be done to validate models as a means of progressing towards imagined speech decoding. Two other speech types, silent speech and whispered speech, are less common in speech-BCI studies. Silent speech involves the movement of the articulators without actual audible output. Whispered speech is articulated speech producing low-amplitude audible output. These two types of speech are more extensively studied in another field called silent speech interface, which focuses on the kinematic signals of articulators (e.g., tongue and lips) instead of brain signals (Cao et al., 2021; Denby et al., 2010; Gonzalez-Lopez et al., 2020). Silent and whispered speech are out of scope of this thesis. Table 1 displays the key characteristics of the four types of speech.

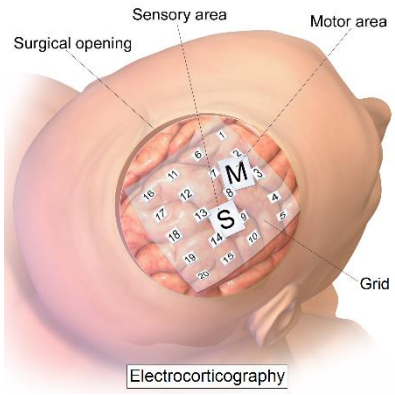
Speech Modality	**Articulation**	**Audible Output**
Overt	Yes	Yes
Imagined	No	No
Silent	Yes	No
Whispered	Yes	Limited

Table 1: Key characteristics of speech modalities

Several neuroimaging modalities have been researched for Speech-BCI applications. The most well-studied ones are EEG and electrocorticography (ECoG). As mentioned earlier, EEG is a non-invasive modality that collects electrophysiological data using electrodes on the scalp. On the other hand, ECoG requires a grid of electrodes to be implanted on the cortical surface. This invasive procedure is done for epileptic patients to localize seizures and scope out a potential resection surgery later. If the grid happens to

be near language areas of interest, patients may be asked to participate in neural speech studies with consent. Since this data is directly from the brain surface, signal-to-noise ratio (SNR) for ECoG data is high, making it attractive for neural speech decoding studies. An advantage both modalities possess is that they have millisecond-level temporal resolution, which is high amongst neuroimaging modalities. This level of temporal resolution is crucial for decoding speech, as it allows the imaging device to track changes in neural activity within the duration of even phonemes. Additionally, EEGs are particularly suited for speech-BCIs due to its availability and non-invasive characteristic. One downside is that non-invasive modalities have a lower signal-to-noise ratio (SNR) than invasive modalities, making speech decoding with EEG more difficult.

EEG



ECoG

Figure 2: Depictions of an EEG helmet, from (Sheerman-Chase, 2012), and a subdural electrode grid used in ECoG, from (Blausen.com staff, 2014)

Current Speech-BCI Progress

As neural speech decoding is generally harder using non-invasive modalities, current EEG speech decoding research is mainly focused on improving closed set vocabulary decoding, and most studies are still focused on relatively small sets of words,

phonemes, or syllables (Panachakel & Ramakrishnan, 2021). However, performances have been improving recently, and some studies have shown some promising results. (Bakhshali et al., 2020) managed to achieve an impressive average accuracy of 90.25% while classifying a set of four words of imagined speech. For a relatively larger set of eight imagined words, (Li et al., 2021) attained a 54.31% average accuracy, which is considerably higher than the chance level. (Koizumi et al., 2018) achieved an average accuracy of 82.9% for a set of six imagined words. (Chengaiyan et al., 2020) managed to identify the vowels from imagined consonant-vowel-consonant words at an average accuracy of 80%, which is high considering that five vowels were possible. Given the current lack of studies successfully decoding larger sets of vocabulary, there is still more progress to be made before open set decoding and speech-BCI is possible with EEG.

Since ECoG data offers higher SNR than EEG data, studies in neural speech decoding with ECoG have attempted decoding for larger vocabulary sets. (Sun et al., 2020) tried to decode open sets (1200 to 1900 unique words) of overt speech and managed to achieve word error rates (WERs) 10% or lower for most participants. However, when the models were tested on unseen data, WERs averaged much higher at 80%. For reference, Google API's commercial automatic speech recognition system has shown to have around a 20% WER in practice (Filippidou & Moussiades, 2020). The results on unseen data showed that the ECoG speech-BCI system in the study was still limited and wouldn't function in a real-world setting. (Moses et al., 2019) attempted to decode speech in a question-and-answer scenario, presenting 9 questions with 24 possible answers. Decoding accuracies ranged from 30%-61%, which is impressive considering chance level (7%). The study showed that the potential of using ECoG speech-BCI in a simple conversational setting. A recent work using ECoG actually tried to decode speech from a subject with anarthria (Moses et al., 2021). Impressively, real-time decoding on an

open set (50 unique words) resulted in a median WER of 25.6%. These results display the potential for speech-BCIs in helping those with anarthria.

MAGNETOENCEPHALOGRAPHY (MEG)

Magnetoencephalography (MEG) is a non-invasive neuroimaging modality which uses special sensors to detect magnetic fields produced by the electrical activity of neurons (Baillet, 2017). These magnetic signals are formed mainly from primary currents caused by post-synaptic potentials during neuronal activity. Since the magnetic field generated by an electrical current is perpendicular to current flow, only fields produced by tangential currents can be detected outside the head. Tangential currents exist in the sulci of the cortex, so MEGs mainly measure cortical activity from the folds (Vrba & Robinson, 2001). Due to the magnetic signals generated being very weak, MEGs employ very sensitive magnetometer sensors and superconducting interference device (SQUID) technology. These sensors are housed in thermally insulated containers filled with liquid helium, and the system is also commonly magnetically shielded to avoid picking up magnetic signals from external sources.

MEG offers strong qualities for potential speech-BCI applications. Firstly, it is completely non-invasive and hence is much more accessible than invasive modalities such as ECoG. Secondly, MEG imaging has millisecond-level temporal resolution, and data can be processed in real time. This allows neural speech decoding to be eventually done in real time, which is the end goal for most speech-BCIs. Lastly, MEG offers better spatial resolution than EEG while maintaining similar temporal resolution, and magnetic signals passing through the scalp are less distorted than electrical potentials from the same source, making MEG data generally easier to clean and interpret (Baillet, 2017). This potentially gives MEG an edge over EEG for neural speech studies and decoding.

Overall, MEG possesses qualities that make it a strong potential candidate for eventual speech-BCI application.

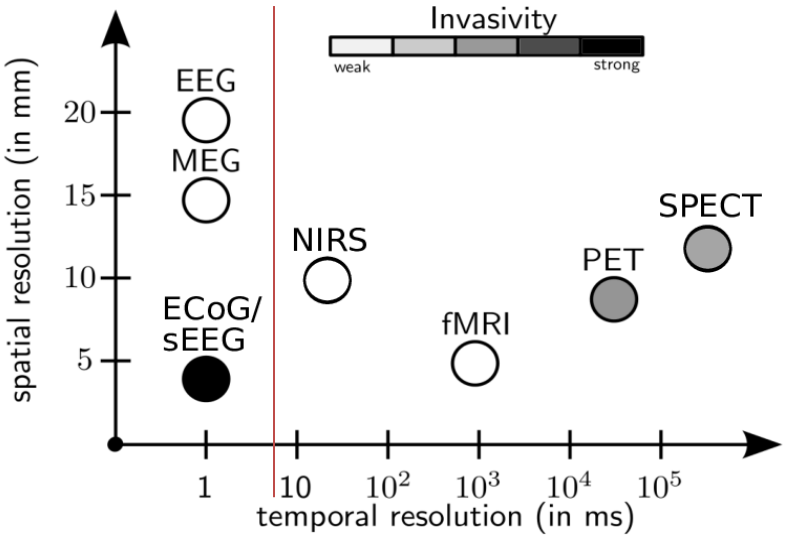


Figure 3: Comparison of several neuroimaging modalities, from (Olivi, 2011)

MEG Speech-BCI

Compared to EEG or ECoG speech-BCI literature, MEG speech-BCI research is a relatively new field. Despite this, recent progress in the area has been promising. A pilot study done with three healthy participants speaking five overt phrases resulted in an average decoding performance of 89.6%, indicating the speech decoding potential using MEG (Dash et al., 2018). A later study involving eight healthy participants impressively achieved an average decoding accuracy of 96.65% on five overt phrases and an average decoding performance of 93.24% on five imagined phrases (Dash, Ferrari, & Wang, 2020). Considering the high accuracy on imagined phrases, these results are promising for the future of MEG speech-BCIs. A study on the same dataset focused on sensor selection found that only nine sensors were needed to achieve an average accuracy of

81.67% on overt speech, indicating that MEG speech-BCIs in the future may be able to utilize smaller sets of sensors to lower costs (Dash, Wisler, et al., 2020). Using MEG data from three ALS participants, a pilot ALS speech decoding study managed to decode five overt phrases spoken by those participants at an average accuracy of over 80%, showing signs that MEG speech-BCIs may eventually be able to help ALS patients as a communication aid (Dash, Ferrari, Hernandez, et al., 2020).

OPM-MEG

While MEG technology holds promise for potential use in speech-BCI applications, there are some drawbacks that could limit its practicality. Multiple downsides stem from the requirement of cooling with liquid helium; firstly, a gap between the sensor and scalp must be maintained for thermal insulation. This distance forces sensors to pick up magnetic signals further away from the scalp, where the signals are weaker. Additionally, since the same helmet apparatus is used for every subject, variance in head shape and size between subjects makes the gap inhomogeneous and can impact signal quality across subjects. Head movements within the helmet may also be problematic for signal quality. Secondly, liquid helium is expensive and will potentially face shortages in the future as natural resources are limited. Another expensive aspect of a MEG system is the magnetic shielding, which often is applied to the room where the system is housed. Besides the requirement for the shielding, another downside for this setup is that it is not portable, which would greatly hinder the usability of a speech-BCI.

An emerging technology called optically pumped magnetometers (OPMs) aims to address some of these issues. OPMs measure the same magnetic fields as SQUID gradiometers but through the monitoring of spin polarization in alkali-atom vapor (Allred et al., 2002). Unlike SQUID gradiometers, OPMs do not require thermal cooling and

hence can be placed on the scalp of the individual. This allows the sensors to capture magnetic signals of higher quality and may lead to higher spatial resolution compared to traditional MEG (Borna et al., 2020). The new advancement also led to the development of wearable helmets holding the sensor array (Boto et al., 2017; Hill et al., 2020; Iivanainen et al., 2019), which allows for head movement of the participant. This improves the practicality of the system and is a step towards a potentially portable apparatus. The smaller device also allows for a smaller magnetic shielding room to be utilized, which significantly reduces initial cost. The elimination for the need for cooling also lowers the cost of the device, thereby improving accessibility. However, OPM-MEG systems still face the necessity for a shielded room and therefore will not be portable until a solution is found. Despite this, OPM-MEG technology is headed in a positive direction towards a more accessible, adaptable, and wearable MEG system.

Figure 4: Examples of a magnetic shielded room for OPM-MEG (left) and an OPM-MEG helmet (right), courtesy of Dr. Davenport, UT Southwestern Medical Center

MOTIVATION FOR DISSERTATION

OPM-MEG technology allows for customizable sensor arrays where the number of sensors and the locations of these sensors may be altered. Looking ahead at a potential speech-BCI using OPM-MEG technology, the utilization of a smaller number of sensors may be advantageous if decoding performance remains comparable. Fewer sensors would decrease the cost of the system and make the wearable device lighter, thereby improving the accessibility and practicality of the speech-BCI. If the optimal sensor locations are known for neural speech decoding, a cheaper OPM-MEG speech-BCI could be possible. However, it is also known in neural decoding research that sensor contribution to decoding tends to vary between participants, partially due to natural physiological variance. Hence, for a more generalized speech-BCI device, it may be beneficial to decrease specificity and find optimal brain lobes for decoding instead.

Speech Production Mechanisms in the Brain

The cerebral cortex of the brain consists of the right and left hemisphere, and each hemisphere can be split into four lobes: frontal, parietal, temporal, and occipital. These lobes are visualized in Figure 5. Brain functions generally span multiple lobes, although some lobes or smaller regions may be more crucial than others for certain functions. Two examples of this are the two major language areas of interest: Broca's area, located roughly at the left inferior frontal lobe, and Wernicke's area, typically defined near the left posterior temporal lobe. Both areas were discovered in the 19th century through the study of patients with lesions in those areas; the patients studied by Broca exhibited a loss in speech articulation ability, and those studied by Wernicke showed difficulty in language comprehension. (Geschwind, 1965) proposed a neurobiology model for language centered around these two language areas and the acuate fasciculus, a white-

matter pathway that connects the two regions. However, modern neuroscience views this model, now commonly known as the Classic Model, as too limited in its characterization of language function in the brain (Tremblay & Dick, 2016).

Fundamental understanding of speech production highlights several areas of importance in the cortex. Through a meta-analysis of 58 word production neuroimaging studies, (Indefrey & Levelt, 2000) summarized the core areas of the brain active in speech production. Speech production is often believed to start from lexical phonological information retrieval, which is generally observed to correlate with activity in the left superior and middle temporal gyri. Most definitions of Wernicke's area cover parts of both these gyri, indicating the importance of Wernicke's area in speech production. Broca's area is generally believed to be responsible for processing the phonological information into phonetical encodings. There is also some evidence that the left mid superior temporal gyrus is sometimes involved in this encoding process. Motor areas can use this phonetical information as input for articulation. Some known motor areas involved in speech articulation are the supplementary motor area, premotor cortex, and the primary motor cortex, all of which are located in the posterior area of the frontal lobe. Other regions of the brain outside of the cortex such as the basal ganglia and cerebellum have also been found to be very important for the motor control of articulation (Eickhoff et al., 2009; Wildgruber et al., 2001). Lastly, the somatosensory cortex in the parietal lobe processes motor feedback, and the primary auditory cortex in the temporal lobe processes audio feedback (Golfinopoulos et al., 2010). Despite knowledge of some of these core processes in speech production, current neuroscience is still uncertain about the true neurological workings of speech, and recent findings continue to highlight the complicated nature of language function in the brain (Tremblay & Dick, 2016).

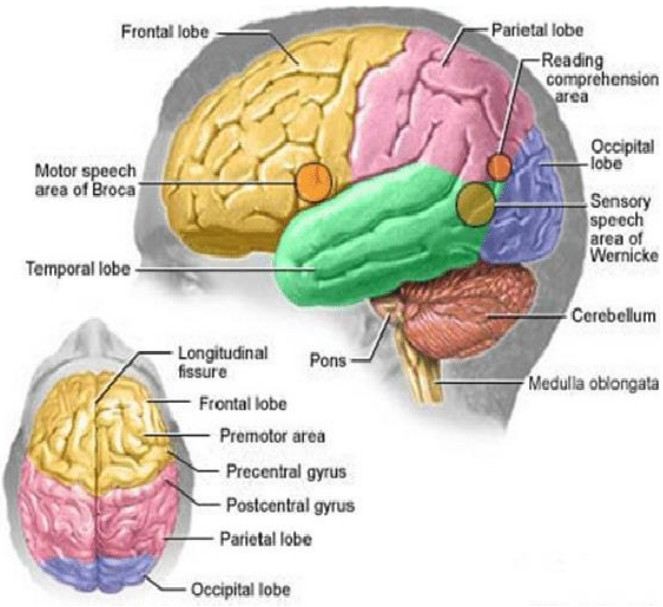


Figure 5: Graphic showing the four different lobes as well as some areas of interest for language, from (Guerrero-Mosquera et al., 2012) Copyright @ 2012, Licensee IntechOpen.

Goals and Hypothesis

This dissertation aims to investigate the contribution of MEG signals from different brain lobes towards neural speech decoding. Studies using fMRI neuroimaging during language comprehension and production have shown neural activity in multiple lobes from both hemispheres (Alderson-Day et al., 2016; Behroozmand et al., 2015; Huth et al., 2016; Newman et al., 2010; Price, 2010). EEG speech decoding studies have also found signals from multiple regions in both hemispheres to be helpful in speech decoding (Kumar et al., 2018; Li et al., 2021; Mirkovic et al., 2015). Looking at this previous research, we expect multiple lobes of the brain from both hemispheres to contribute helpful MEG signals towards speech decoding. Due to the importance of the Broca's and Wernicke's area in language function, we also expect signals from the left frontal lobe and left temporal lobe contribute to decoding more than the other regions. Some EEG

speech decoding studies have found these two lobes to be the best data sources for decoding (Nguyen et al., 2017; Zhao & Rudzicz, 2015), but others have found signals from other lobes such as the parietal lobes and right temporal lobe to be very important (Kumar et al., 2018; Li et al., 2021; Mirkovic et al., 2015). Thus, we also believe that we might find other lobes to contain information very important to MEG speech decoding besides the left frontal and left temporal lobes. Overall, this study aims to identify lobes of importance for MEG neural speech decoding with OPM-MEG in mind. The results of our study may be useful in guiding future OPM-MEG speech decoding studies with regards to which brain lobes should be prioritized for sensor placements.

Decoding Speech Using Neuromagnetic Signals from Individual Brain Regions

METHODS

MEG Data Collection

The MEG data analyzed in this study were collected in two MEG labs, one in Dell Children's Medical Center of Austin, TX, USA and the other in Cook Children's Medical Center in Fort Worth, TX, USA. Both MEG labs utilized the Elekta Neuromag TRIUX MEG device (MEGIN US LLC), which holds 204 gradiometer sensors and 102 magnetometer sensors. To avoid external magnetic interference, both machines were housed in magnetically shielded rooms. Short recordings were done before participants enter to check noise levels in the environment; experimentation only began if the noise level was deemed low enough. Additionally, proper alignment and calibration of the sensors were ensured before data collection. To remove potential noise caused by cardiac artifacts, cardiac artifact data was collected using two electrocardiograph sensors placed near the collarbone. Stimuli were shown visually using STIM2 software (Compumedics Ltd.) and a DLP projector that displayed the stimuli roughly 90cm in front of the participant. Participants removed any magnetic materials before entering the room to avoid potential magnetic distortions. To minimize the chances of head movement within the device, sufficient time was given for participants to adjust in the collection device until comfortable, and instructions to avoid any head movements were given. Participants were required to give written informed consent before data collection and were monetarily compensated. The data of seven healthy subjects (three males, four females; age = 41±14 years) was used for this study. All participants were right-handed, fluent in English, and had no history of language or cognitive disorders. The experiments were

approved by the University of Texas at Dallas (IRB# 15-92) and the University of Texas at Austin (IRB# 2015-09-0042).

Figure 6: Elekta Neuromag MEG device with participant inside

Experimental Protocol

Both imagined speech and overt speech data were collected within the same time-locked protocol, pictured in Figure 7. First, the participants were given a pre-stimuli period of half a second, which was then followed by the display of a visual stimulus for one second. In each trial, one of the five selected English phrases were shown: “Do you understand me”, “That’s perfect”, “How are you doing”, “Good-bye”, and “I need help”. In total, each phrase was shown 100 times per experiment and were displayed in pseudo-random order. Following the display of the phrase, the participants were given one second to imagine and prepare for speaking of the phrase. A fixation cross is displayed on the screen to indicate this period. After the disappearance of the cross, participants overtly articulated the phrase, which generally took 2-2.5 seconds. In between trials,

participants were given around one second as a non-movement baseline before the next trial begins. The experiment generally took around 45 minutes to complete.

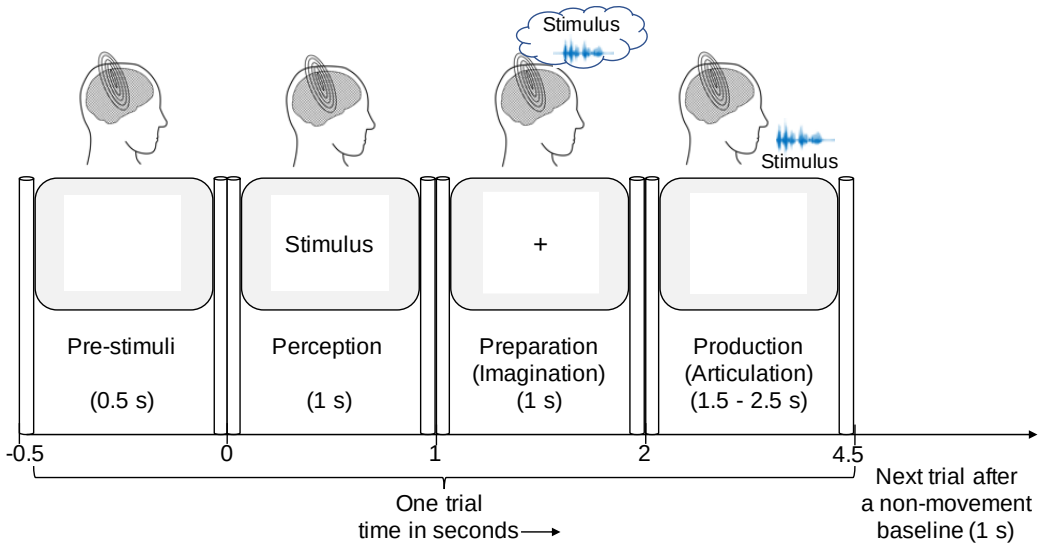


Figure 7: Time-locked experimental protocol for data collection

Data Cleaning and Preprocessing

For data cleaning, the data in each trial was visually inspected for artifacts or mistimed actions in the protocol, such as an early start in articulation. Trials with these features were removed. The MEG device sampled at a 4 kHz and had an online filter of 0.03 – 1300 Hz. After collection, the MEG signals were resampled to 1kHz, and a 4^{th} order Butterworth filter with a low-pass threshold of 250Hz was used to filter the signals. Power line noise at 60 Hz and its harmonics were filtered out using a 2^{nd} order IIR notch filter. Sensors with abnormally noisy or completely flat responses were removed from the dataset. On average, 25% of the trials were rejected per participant during data cleaning. For decoding purposes, 60 of the clean trials were selected for each phrase and participant, totaling to 300 phrases per participant.

Feature Selection and Decoding Approach

This study aimed to decode the five phrases, either spoken or imagined, using MEG signals collected from different lobes of the brain. The MEG data for imagined phrases were taken from the one second to two second portion of each trial, visualized in Figure 7. Imagined speech was selected for decoding since imagined speech decoding is important for eventual speech-BCI applications. For overt phrases, MEG data was taken starting from the two second mark and ending at the end of the trial. Since overt speech decoding generally performs better than imagined speech decoding, overt speech decoding may elucidate patterns in performance that may otherwise be unclear in the imagined speech decoding results. It would also be interesting to investigate differences between the two types of decoding. Perception was not the focus of this dissertation, and thus it was not decoded in this work.

The overall sensor set of 204 gradiometer sensors was separated out into eight sets that covered the eight individual lobes. It is important to note that source reconstruction was not used, so this study was not estimating the activity at the source from the lobes. Instead, these eight sensor sets were sensor sets approximately corresponding to the eight lobes. Later usages of the term ‘lobe’ will refer to these sensor sets. These sets were used to provide signals for feature generation and eventual decoding. In addition to decoding using the individual lobe sensor sets, the study also decoded using sensor data from each hemisphere as well as the entire brain. Results from this decoding may serve as references for the individual lobe results, and the left and right hemispheres can also be directly compared with each other in terms of decoding performance.

Previous MEG speech decoding studies indicated that using the root mean square (RMS) values of the preprocessed signals from each sensor served as informative

features, and given those inputs, linear discriminant analysis (LDA) provided solid baseline performance for MEG neural speech decoding (Dash, Ferrari, Hernandez, et al., 2020). Hence, most of the results in the study utilized RMS values as features and LDA as the decoder. Besides competitive performance, LDA also offered the advantage of being fast to train. This advantage was particularly important for this study since many models had to be trained; a model had to be trained for each of the eight lobes, for each participant, and for the two types of phrases. Later analysis also included training models using specific frequency bands within each lobe. Neural network models generally had to be used to beat LDA performance in MEG neural speech decoding (Dash, Ferrari, Hernandez, et al., 2020; Dash, Ferrari, & Wang, 2020), but those models take much longer to train than LDA and hence were not utilized.

During training, the LDA coefficient threshold and regularization factor were selected using Bayesian optimization search. Previous studies have indicated that the cognitive variance between the participants was high (Dash et al., 2019), so separate models were trained for each participant. Decoding performance was represented by average five-fold cross-validation accuracy. Since this study's goal was to compare the discriminatory power of MEG signals coming the individual brain lobes, using cross-validation accuracy as an indication of performance was sufficient. In other studies focused on improving the overall speech-BCI accuracy, testing performance is typically preferred as the performance metric. Lastly, support vector machine (SVM) decoders were also used to replicate the study and verify if the LDA results were model dependent. SVMs have seen success in neural speech decoding in EEG studies (Lopez-Bernal et al., 2022; Matsumoto & Hori, 2014; Paul et al., 2018), and previous MEG speech decoding studies have indicated similar decoding performances between SVM and LDA decoders (Dash, Ferrari, Hernandez, et al., 2020). The SVM models for this study utilized a 2nd

order polynomial kernel, and the ‘C’ and kernel size hyperparameters were selected using Bayesian optimization search.

To investigate how the role of brainwaves in decoding may differ across individual lobes, the study also decoded speech using six individual brainwaves from each individual lobe. These brainwaves consist of the delta (0.3 – 4 Hz), theta (4 – 8 Hz), alpha (8 – 15 Hz), beta (16 – 30 Hz), gamma (30 – 59 Hz), and high gamma (61 – 119 Hz) frequency bands. Band-power was calculated for each band to serve as input features, also utilized in previous MEG speech decoding studies (Dash et al., 2021; Dash, Ferrari, Hernandez, et al., 2020). Decoding was first done using LDA, and SVM was used later to verify model dependency.

After observing results from the above experiments, decoding was also performed using the bilateral lobe sensor sets created by combining the left and right sensor sets of each lobe together. RMS features and LDA classification were used again. This decoding served to investigate if the bilateral lobes had significant differences in performance with each other and if the individual left and right lobes contained complementary information for speech decoding.

Statistical Analysis

Statistical tests were used to investigate the differences in decoding performances and efficacy of decoding. These comparison tests were run on the subject level and hence consider the variance in decoding performances between participants. Each individual brain lobe and hemisphere performance was compared against all-sensor performance for significance using one-tailed, two-sample T-Tests. Additionally, similar T-Tests were run between the individual performances and chance level, found to be 20% through bootstrapping, to test if the brain lobes performed significantly above chance level. One-

way ANOVA on the eight individual brain lobes was performed to determine if the decoding performances were significantly different. For individual brainwave decoding results, two-way ANOVA was performed to determine if there were significant differences between the results with respect to the two variables: brain region and brainwave. If significance was found, Post-hoc Tukey Tests were used to determine which individual regions or brainwaves differed from each other.

RESULTS

Decoding Performance of the Eight Separated Brain Lobes using LDA

Decoding results using separated sensor sets based on brain lobes and LDA as the decoding model is visualized in Figure 8. The decoding results were averaged across participants, and the error bars indicate standard deviation in performance. Average accuracy when using all sensors was 73.4% for overt phrases and 45.5% for imagined, both being the highest in their category. The right temporal lobe sensors performed the best out of the single lobes for both overt (55.7% accuracy) and imagined (39.76% accuracy) phrases. The results for the T-Tests of regional performances against all-sensor performance and chance level performance are tabularized in Table 2. All lobes performed significantly above chance level ($p<0.05$) except the left frontal lobe for imagined speech. For overt speech, only the right temporal lobe did not perform significantly worse than when using all sensors ($p=0.07$), and for imagined speech, only the left frontal lobe performed significantly worse than using all sensors ($p=0.02$). Right hemisphere performance was not significantly different from left hemisphere performance for either mode of speech ($p>0.05$, two-tailed two-sample T-Test). One-way ANOVA on the performances of the eight individual brain lobes failed to reject null for

both types of speech ($p>0.05$), indicating that the performances of the eight lobes may not be significantly different.

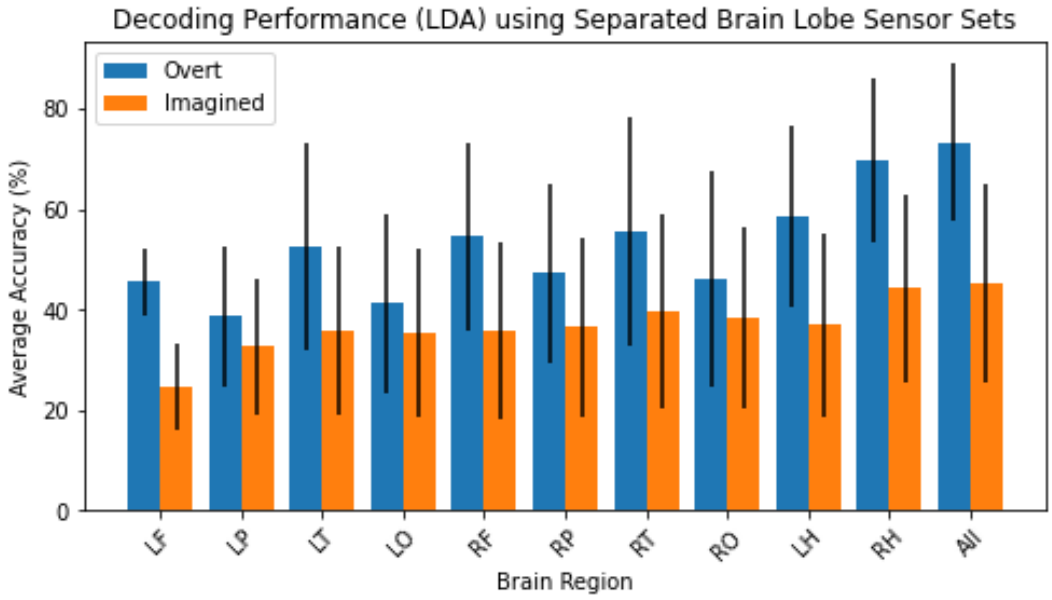


Figure 8: Average accuracies of overt and imagined speech decoding using sensor sets from different lobes of the brain and a LDA decoder. “LH” contains all left hemisphere sensors, and “RH” contains all right hemisphere sensors. “All” includes all sensors. Error bars indicate standard deviation across participants.

Speech Type	T-Test (one-tailed, two-sample)	LF	LP	LT	LO	RF	RP	RT	RO	LH	RH
Overt	P-value (vs. all)	<.01	<.01	0.04	0.00	0.04	0.01	0.07	0.01	0.08	0.35
	P-value (vs. chance)	<.01	<.01	<.01	0.01	<.01	<.01	<.01	0.01	<.01	<.01
Imagined	P-value (vs. all)	0.02	0.11	0.19	0.18	0.19	0.21	0.31	0.26	0.23	0.46
	P-value (vs. chance)	0.10	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.02	<.01

Table 2: Table showing p-values of one-tailed, two-sample T-Tests of the individual brain lobe performances versus all-sensor performance and chance-level performance.

Decoding Performance of Individual Brainwaves within Lobes

To investigate the role of brainwaves in neural speech decoding for the different brain lobes, we utilized the spectral band-power of the individual brainwaves of each brain lobe and hemisphere as inputs for decoding. Figures 9 and 10 show the decoding results when using these individual brainwave features. The brainwaves performed in a similar pattern between the brain lobes, with delta usually performing similarly well to using all brainwaves. When decoding overt speech, high gamma and theta brainwaves generally performed worse than delta but better than the rest of the brainwaves. For imagined speech, all the brainwaves besides delta seemed to perform similar well.

Two-way ANOVA on these results for overt speech decoding was significant ($p<0.05$) for both the brain lobe and brainwave variables. Post-hoc Tukey Tests indicated that the right temporal lobe performed significantly better ($p<0.05$) than the left frontal, left parietal, and left occipital lobes. Delta decoding results were also found to be significantly higher ($p<0.05$) than all other brainwaves' results. Lastly, interaction between the two variables was insignificant ($p=1.00$). For imagined speech results, two-way ANOVA was significant ($p=0.02$) for the lobes but insignificant ($p=0.77$) for the frequency bands. Post-hoc Tukey Tests indicated that the left frontal lobe performed significantly worse ($p<0.05$) than the right temporal, right parietal, and right occipital lobes. Interaction between the two variables was again deemed insignificant ($p=1.00$).

Figure 9: Overt speech decoding performances, grouped by brainwave (a) and grouped by brain region (b), using individual brainwaves within the brain lobes.

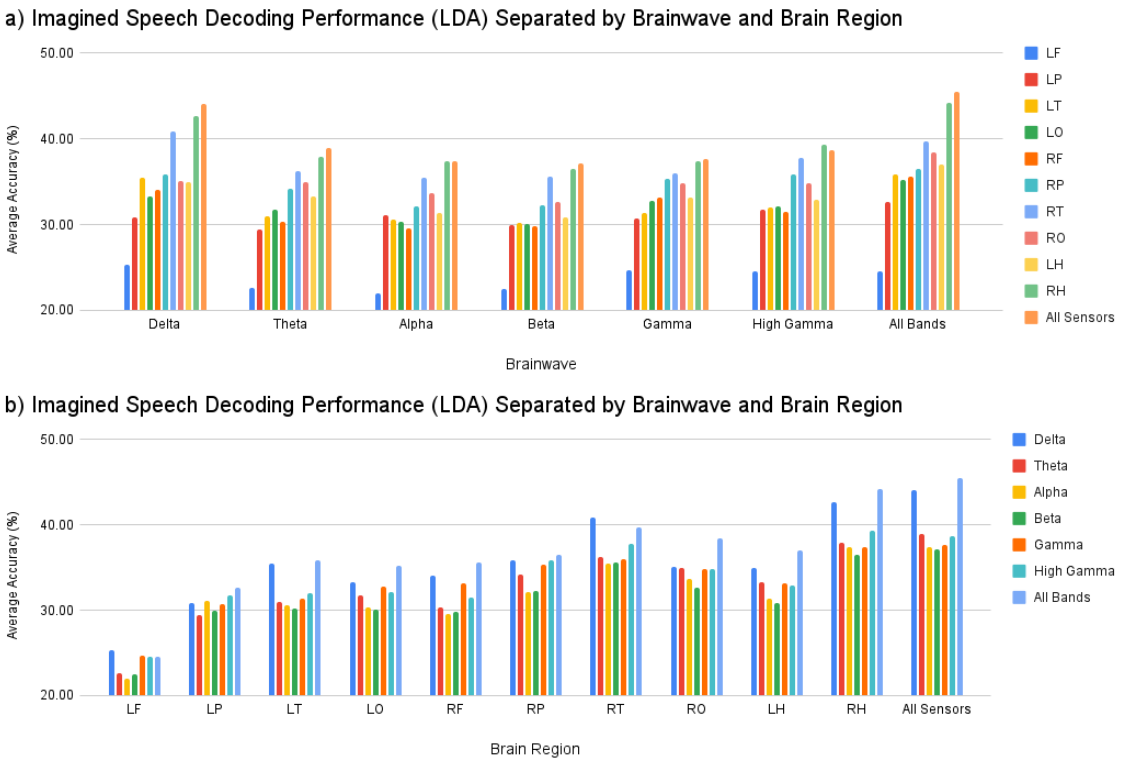


Figure 10: Imagined speech decoding performances, grouped by brainwave (a) and grouped by brain region (b), using individual brainwaves within the brain lobes.

Decoding Performance of Bilateral Lobes

Seeing that right and left hemisphere decoding performance and patterns seemed to be similar, it would be interesting to combine the sensor sets bilaterally to see if there are overall differences between the bilateral lobes. Figure 11 displays these decoding results using a LDA decoder. The bilateral temporal lobe performed best for both overt and imagined speech, reaching accuracies close to the all-sensor performances (67.30% vs. 73.41% for overt, 44.47% vs. 45.47% for imagined). Looking at Table 3, the bilateral temporal lobe had relatively high p-values when tested against all-sensor performance

(.28 for overt, .47 for imagined), further supporting the closeness of temporal lobe performance to whole-brain performance. The bilateral frontal lobe was the only other sensor set that did not perform significantly worse than whole-brain accuracy in overt speech decoding (p=0.07). All bilateral lobes performed significantly above chance level for both modes of speech. Table 4 compares the bilateral results with individual left and right results for overt speech decoding. Notably, combining the temporal lobes laterally led to a 11.6% increase in performance, implying that the two lobes contain complementary information. This is not as apparent in the three other lobes as we only see small increases in accuracy for the bilateral results.

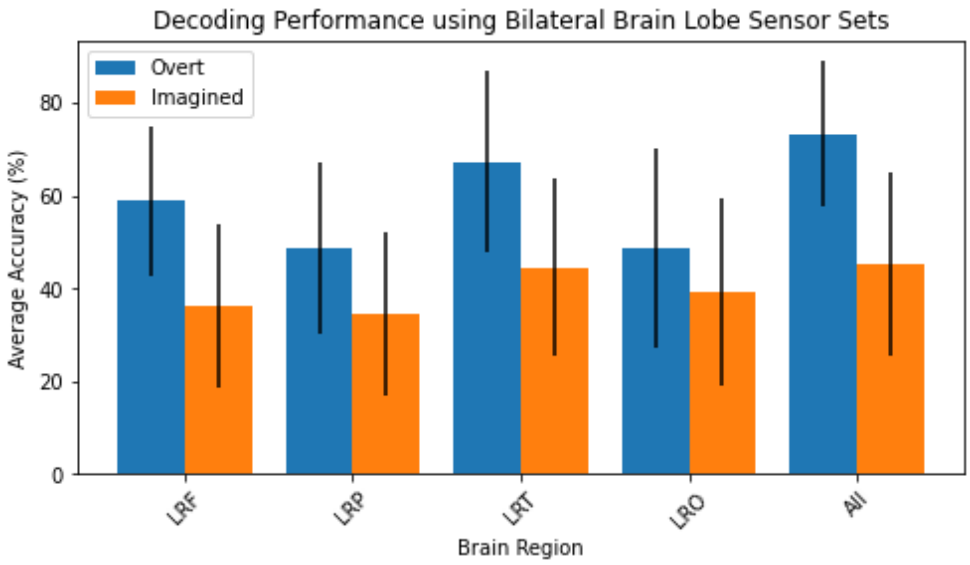


Figure 11: Average accuracies of overt and imagined speech decoding using sensor sets from bilateral lobes of the brain and a LDA decoder.

Speech Type	T-Test (one-tailed, two-sample)	LRF	LRP	LRT	LRO
Overt	P-value (vs. all)	0.07	0.01	0.28	0.02
	P-value (vs. chance)	<.01	<.01	<.01	<.01
Imagined	P-value (vs. all)	0.21	0.16	0.47	0.30
	P-value (vs. chance)	0.02	0.03	0.00	0.02

Table 3: P-values of one-tailed, two-sample T-Tests of the bilateral brain lobe performances versus all-sensor performance and chance-level performance.

Lobe	Left	Right	Combined
Frontal	45.51%	54.66%	58.98%
Parietal	38.80%	47.28%	48.78%
Temporal	52.56%	55.70%	67.30%
Occipital	41.23%	46.06%	48.73%
All	58.64%	69.73%	73.41%

Table 4: Comparison of overt speech decoding performance between the two hemispheres.

Decoding Performance Using SVM

To verify if the earlier results using LDA were model-dependent, the above experiments were replicated using a SVM as decoder. Figure 12 compares very similarly to Figure 8 in terms of accuracies and relative performances between the lobes. Table 5 and Table 2 also contain similar p-values, with right temporal lobe again being the only individual lobe that did not perform significantly worse (p=0.14) than whole-brain

performance in overt speech decoding. Like when using LDA, all lobes performed significantly above chance level for both modes of speech except in the case of the left frontal lobe for imagined speech. Figures 13 and 14 also parallel Figures 9 and 10 respectively in terms of brainwave performance patterns across lobes. As with the LDA results, the delta brainwave performed very similarly to all-brainwaves in overt speech decoding, and alpha and high gamma have higher accuracies than the remaining frequency bands. For imagined speech, delta seemed to perform best again and the rest had similar accuracies. Overall, the SVM results matched up closely with the earlier LDA results, implying that the LDA results were not model-dependent.

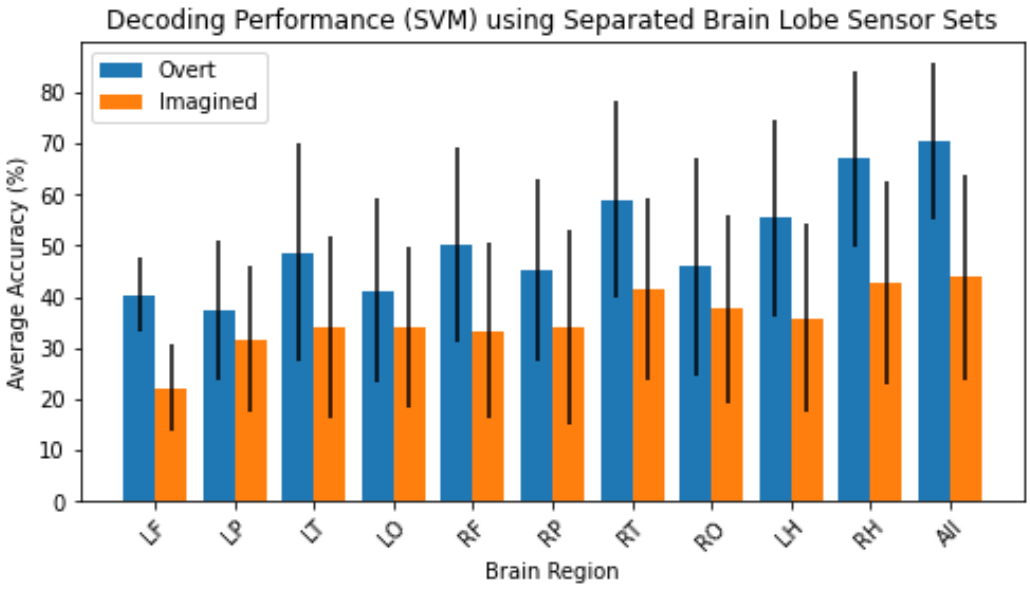


Figure 12: Average accuracies of overt and imagined speech decoding using sensor sets from individual lobes of the brain and a SVM decoder.

Speech Type	**T-Test (one-tailed, two-sample)**	**LF**	**LP**	**LT**	**LO**	**RF**	**RP**	**RT**	**RO**	**LH**	**RH**
Overt	P-value (vs. all)	<.01	<.01	0.03	0.01	0.03	0.01	0.14	0.02	0.08	0.36
	P-value (vs. chance)	<.01	<.01	<.01	0.01	<.01	<.01	<.01	0.01	<.01	<.01
Imagined	P-value (vs. all)	0.02	0.12	0.19	0.18	0.18	0.20	0.42	0.29	0.24	0.46
	P-value (vs. chance)	0.27	0.03	0.04	0.02	0.04	0.05	0.01	0.02	0.03	0.01

Table 5: P-values of one-tailed, two-sample T-Tests of the individual brain lobe performances using SVM versus all-sensor performance and chance-level performance.

Figure 13: Overt speech decoding performances, grouped by brainwave (a) and grouped by brain region (b), using individual brainwaves within the brain lobes and SVM as decoder.

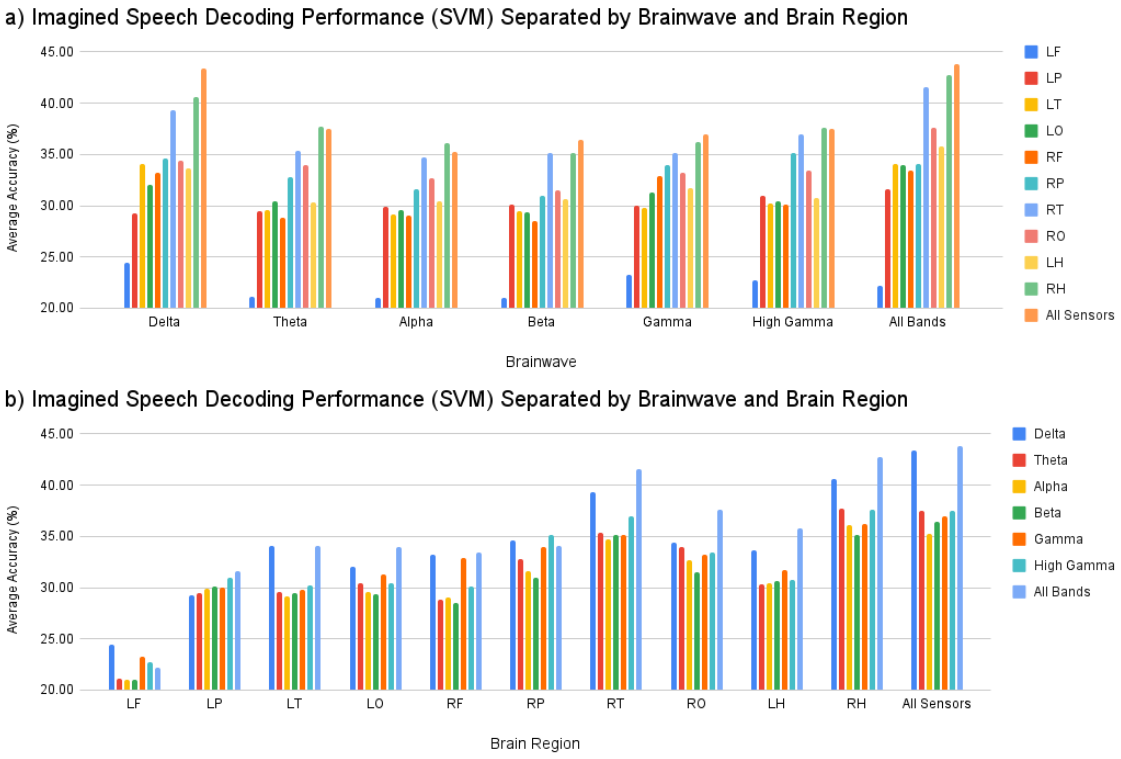


Figure 14: Imagined speech decoding performances, grouped by brainwave (a) and grouped by brain region (b), using individual brainwaves within the brain lobes and SVM as decoder.

DISCUSSION

Discussion on the results will be focused around the LDA results as the SVM results were very similar and helped verify that the LDA results were not model-dependent. Additional focus will be on overt speech decoding results since imagined speech decoding performance was generally low, making it hard to draw strong conclusions.

The first interesting aspect of the results was the fact that almost all lobes achieved results significantly higher than chance level, indicating that helpful neural activity for speech decoding may be spread across most of the brain. This supports

previous neuroimaging studies which found neural activation during speech production in many areas of the brain outside of just the major language and motor regions (Alderson-Day et al., 2016; Behroozmand et al., 2015; Newman et al., 2010; Price, 2010). In the case of overt speech decoding, all lobes except the right temporal lobe performed significantly worse than whole-brain performance, which implies that the neuromagnetic information collected from the different lobes was not completely redundant for decoding. This is even more convincing considering that the p-value for the T-Test comparison between whole-brain and the right temporal lobe was only 0.07, just barely above significance threshold of 0.05. Additionally, the number of subjects was small and performance variance between subjects was large, which makes it difficult to draw strong conclusions for p-values just above the threshold. For imagined speech decoding, most lobes performed similarly to whole-brain performance. However, overall performance of the decoding was not high, so there's a chance this pattern may not hold with stronger decoding results. Overall, the overt speech decoding results seem to indicate that it would be best to utilize sensors from more than one lobe of the brain.

One-way ANOVA on results indicated that the lobes performed similarly well with each other. However, the small subject set and large variance in performance between subjects, visualized in the error bars of Figure 8, made it difficult for comparison tests to result in significance. A larger set of subjects would be preferred before one can conclude confidently that all lobes contribute similarly to speech decoding.

Another interesting result was how the right hemisphere performed similarly to the left hemisphere for decoding both speech types ($p>0.05$, two-tailed two-sample T-Test). In fact, the right hemisphere had an average accuracy 11.51% higher than the left hemisphere, and each lobe in the right hemisphere outperformed their left hemisphere counterpart by an average of 8.23%. This observation supports previous fMRI and MEG

research showing significant right-side neural activation during speech production (Alderson-Day et al., 2016; Alexandrou et al., 2017; Behroozmand et al., 2015; Price, 2010) as well as EEG speech decoding research that found right-side sensors to contribute greatly to decoding (Kumar et al., 2018; Mirkovic et al., 2015). In particular, the bilateral temporal lobes were mentioned frequently in the above studies, and those lobes turned out to be two of the best performing lobes in terms of average accuracy in this study. The bilateral decoding results for the temporal lobes resulted in a high accuracy of 67.30%, just a notch under the accuracy when using all sensors (73.40%). For both types of speech decoding, T-Tests confirmed that bilateral temporal lobe performance was not significantly worse than whole-brain performance ($p=0.28$ for overt, $p=0.47$ for imagined). This suggests that sensor sets covering the bilateral temporal lobes may be able to substitute for whole-brain sensor sets for MEG speech decoding, especially if there is incentive to decrease sensor number. Lastly, Table 4 showed that combining the temporal sensor sets laterally resulted in a notable increase in performance (+11.30%) over using the two lobes individually. This suggests that the two temporal lobes provided complementary information for speech decoding, and sensors should cover both regions for speech decoding purposes. Interestingly, the other bilateral lobes did not see large increases in performance over their individual counterparts, suggesting there may have been mostly redundant information bilaterally for those lobes.

Results from the individual brainwave analysis showed that the distribution of brainwave performance was similar across the brain lobes. This observation could imply that helpful neuromagnetic signals for speech decoding is encoded similarly in the different lobes. A previous study with the same dataset had already shown that the delta band performed the best out of the brainwaves for both types of speech decoding (Dash et al., 2021), but it is interesting to see that this is consistent across the brain lobes. Jaw

motion during articulation has been suspected as a possible reason for delta contribution, but a previous study with the same dataset had combined the delta brainwaves with corresponding jaw motion data for increased overall decoding performance, implying that the delta brainwaves contained important neural information outside of possible jaw motion information (Dash et al., 2021). Delta brainwaves have been shown to be active during the speech perception of sentences and is possibly associated the tracking of linguistic structures in sentences such as syllables and phrases (Ding et al., 2016; Ding & Simon, 2012; Doelling et al., 2014). Since this study utilized sentences as stimuli, subjects' delta activity may reflect the processing of sentences from auditory feedback and hence be useful in decoding. However, the observation that delta performance is the best across almost all lobes, even in the case of imagined phrases, seems to suggest that there may be additional underlying reasons for this phenomenon that could warrant further investigation.

Conclusion

FINDINGS

For the purposes of finding the brain lobes that provide the best information for MEG neural speech decoding, this study decoded overt and imagined speech using sensor sets corresponding to each individual lobe. Decoding results found the right temporal lobe to be the best performing individual lobe, and decoding performance using the bilateral temporal lobes was comparable to whole brain decoding performance for both speech types. For MEG speech decoding applications, the temporal lobe regions could be prioritized for sensor coverage. This information may benefit OPM-MEG speech-BCIs that have limited sensor sets or want to minimize sensor usage. Additionally, improvement in bilateral temporal lobe decoding results over the individual temporal lobes seems to imply that the two lobes contain complementary information for speech decoding. Combined with the fact that right hemisphere results were better than that of the left hemisphere, this also suggests that the right hemisphere should not be ignored for speech decoding purposes. Interestingly, the left frontal lobe, which contains Broca's area, did not stand out in speech decoding performance. Overall, speech-BCI applications should be wary of placing sensors on just Broca's and Wernicke's areas as using signals only from those areas may not always be sufficient in achieving decoding performance comparable to whole-brain performance.

Individual brainwave decoding showed similar patterns of brainwave performance across all lobes, implying some level of similarity in how the lobes encode helpful information for speech decoding. It would be interesting to see if future MEG speech decoding studies find a similar pattern or if this phenomenon is specific only to this dataset. The usefulness of the delta brainwave would also be important to look out for in

future MEG speech decoding studies, particularly in imagined speech decoding where possible reasons such as jaw motion rhythm and audio feedback can be ruled out. Given that the delta brainwave was by far the best performing brainwave in this study and a previous one (Dash et al., 2021), MEG speech decoding studies should consider setting their lower thresholds for bandpass filtering relatively low to retain delta brainwave information.

CONTRIBUTION AND NOVELTY

As far as we know, this study is the first to investigate the role of different brain lobes in MEG neural speech decoding. With the advent of OPM-MEG technology, the findings of this study may be helpful in guiding future OPM-MEG speech decoding studies with regards to sensor placement prioritization. This study also investigated individual brainwave performance in the different lobes, finding that the lobes seemed consistent in how the different frequency bands performed. The delta brainwave stood out in performance for almost all lobes for both overt and imagined speech decoding, which could warrant further investigation if the phenomenon is consistent with future MEG speech decoding studies. Lastly, the study supports previous neuroimaging and speech decoding work that highlight the importance of the right hemisphere activity in speech production. Despite the left lateralization of language in most people, it is becoming apparent that speech decoding studies should not just focus on the left hemisphere.

FUTURE WORK

Several avenues of future work may further enhance or validate the work presented in this dissertation. Firstly, a source reconstruction approach may better

elucidate the specific regions of the brain that contribute to decoding and make spatial analysis of the brain across subjects more accurate and comparable. It would also help reduce the influence of brain activity from multiple regions on a single sensor. Secondly, a larger dataset with more participants will help validate the results from this experiment. The current participant count is low and variance in performance between the participants is high, making it difficult to draw conclusions using significance testing. Additionally, a larger vocabulary dataset will help improve the results' relevancy to eventual open-set vocabulary decoding. Thirdly, this experiment should be replicated using OPM-MEG speech data for validation. This study decoded speech using SQUID gradiometer data, which may produce results different than those when using OPM-MEG magnetometer data. Validation from a study using OPM-MEG would make the results more directly applicable for OPM-MEG speech-BCIs. Lastly, including ALS participants in the study may help uncover possible differences between ALS and healthy participants, specifically with regards to the spatial distribution of helpful neuromagnetic information for speech decoding. Analysis on brainwave performance might also be informative for ALS speech decoding science.

References

Abiri, R., Borhani, S., Sellers, E. W., Jiang, Y., & Zhao, X. (2019). A comprehensive review of EEG-based brain–computer interface paradigms. *Journal of Neural Engineering*, *16*(1), 011001. https://doi.org/10.1088/1741-2552/aaf12e

Alderson-Day, B., Weis, S., McCarthy-Jones, S., Moseley, P., Smailes, D., & Fernyhough, C. (2016). The brain's conversation with itself: Neural substrates of dialogic inner speech. *Social Cognitive and Affective Neuroscience*, *11*(1), 110–120. https://doi.org/10.1093/scan/nsv094

Alexandrou, A. M., Saarinen, T., Mäkelä, S., Kujala, J., & Salmelin, R. (2017). The right hemisphere is highlighted in connected natural speech production and perception. *NeuroImage*, *152*, 628–638. https://doi.org/10.1016/j.neuroimage.2017.03.006

Allred, J. C., Lyman, R. N., Kornack, T. W., & Romalis, M. V. (2002). High-Sensitivity Atomic Magnetometer Unaffected by Spin-Exchange Relaxation. *Physical Review Letters*, *89*(13), 130801. https://doi.org/10.1103/PhysRevLett.89.130801

Baillet, S. (2017). Magnetoencephalography for brain electrophysiology and imaging. *Nature Neuroscience*, *20*(3), 327–339. https://doi.org/10.1038/nn.4504

Bakhshali, M. A., Khademi, M., Ebrahimi-Moghadam, A., & Moghimi, S. (2020). EEG signal classification of imagined speech based on Riemannian distance of correntropy spectral density. *Biomedical Signal Processing and Control*, *59*, 101899. https://doi.org/10.1016/j.bspc.2020.101899

Ball, L. J., Beukelman, D. R., & Pattee, G. L. (2002). Timing of speech deterioration in people with amyotrophic lateral sclerosis. *Journal of Medical Speech-Language Pathology*, *10*(4), 231–235.

Bambini, V., Arcara, G., Martinelli, I., Bernini, S., Alvisi, E., Moro, A., Cappa, S. F., & Ceroni, M. (2016). Communication and pragmatic breakdowns in amyotrophic lateral sclerosis patients. *Brain and Language*, *153–154*, 1–12. https://doi.org/10.1016/j.bandl.2015.12.002

Behroozmand, R., Shebek, R., Hansen, D. R., Oya, H., Robin, D. A., Howard, M. A., & Greenlee, J. D. W. (2015). Sensory–motor networks involved in speech production and motor control: An fMRI study. *NeuroImage*, *109*, 418–428. https://doi.org/10.1016/j.neuroimage.2015.01.040

Beukelman, D., Fager, S., & Nordness, A. (2011). Communication Support for People with ALS. *Neurology Research International*, *2011*, e714693. https://doi.org/10.1155/2011/714693

Blausen.com staff. (2014). *Medical gallery of Blausen Medical 2014*. WikiJournal of Medicine. 10.15347/wjm/2014.010

Borna, A., Carter, T. R., Colombo, A. P., Jau, Y.-Y., McKay, J., Weisend, M., Taulu, S., Stephen, J. M., & Schwindt, P. D. D. (2020). Non-Invasive Functional-Brain-Imaging with an OPM-based Magnetoencephalography System. *PLOS ONE*, *15*(1), e0227684. https://doi.org/10.1371/journal.pone.0227684

Boto, E., Meyer, S. S., Shah, V., Alem, O., Knappe, S., Kruger, P., Fromhold, T. M., Lim, M., Glover, P. M., Morris, P. G., Bowtell, R., Barnes, G. R., & Brookes, M.

J. (2017). A new generation of magnetoencephalography: Room temperature measurements using optically-pumped magnetometers. *NeuroImage*, *149*, 404–414. https://doi.org/10.1016/j.neuroimage.2017.01.034

Cao, B., Sebkhi, N., Bhavsar, A., Inan, O. T., Samlan, R., Mau, T., & Wang, J. (2021). Investigating Speech Reconstruction for Laryngectomees for Silent Speech Interfaces. *Interspeech 2021*, 651–655. https://doi.org/10.21437/Interspeech.2021-1842

Chen, X., Wang, Y., Nakanishi, M., Gao, X., Jung, T.-P., & Gao, S. (2015). High-speed spelling with a noninvasive brain–computer interface. *Proceedings of the National Academy of Sciences*, *112*(44), E6058–E6067. https://doi.org/10.1073/pnas.1508080112

Chengaiyan, S., Retnapandian, A. S., & Anandan, K. (2020). Identification of vowels in consonant–vowel–consonant words from speech imagery based EEG signals. *Cognitive Neurodynamics*, *14*(1), 1–19. https://doi.org/10.1007/s11571-019-09558-5

Dash, D., Ferrari, P., Hernandez, A., Heitzman, D., Austin, S. G., & Wang, J. (2020). Neural Speech Decoding for Amyotrophic Lateral Sclerosis. *Interspeech 2020*, 2782–2786. https://doi.org/10.21437/Interspeech.2020-3071

Dash, D., Ferrari, P., Malik, S., & Wang, J. (2018). Overt Speech Retrieval From Neuromagnetic Signals Using Wavelets And Artificial Neural Networks. *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 489–493. https://doi.org/10.1109/GlobalSIP.2018.8646401

Dash, D., Ferrari, P., & Wang, J. (2020). Decoding Imagined and Spoken Phrases From Non-invasive Neural (MEG) Signals. *Frontiers in Neuroscience, 14.* https://www.frontiersin.org/article/10.3389/fnins.2020.00290

Dash, D., Ferrari, P., & Wang, J. (2021). Role of Brainwaves in Neural Speech Decoding. *2020 28th European Signal Processing Conference (EUSIPCO)*, 1357–1361. https://doi.org/10.23919/Eusipco47968.2020.9287714

Dash, D., Wisler, A., Ferrari, P., Davenport, E. M., Maldjian, J., & Wang, J. (2020). MEG Sensor Selection for Neural Speech Decoding. *IEEE Access*, *8*, 182320–182337. https://doi.org/10.1109/ACCESS.2020.3028831

Dash, D., Wisler, A., Ferrari, P., & Wang, J. (2019). Towards a Speaker Independent Speech-BCI Using Speaker Adaptation. *Interspeech 2019*, 864–868. https://doi.org/10.21437/Interspeech.2019-3109

Denby, B., Schultz, T., Honda, K., Hueber, T., Gilbert, J. M., & Brumberg, J. S. (2010). Silent speech interfaces. *Speech Communication*, *52*(4), 270–287. https://doi.org/10.1016/j.specom.2009.08.002

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158–164. https://doi.org/10.1038/nn.4186

Ding, N., & Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, *107*(1), 78–89. https://doi.org/10.1152/jn.00297.2011

Doble, J. E., Haig, A. J., Anderson, C., & Katz, R. (2003). Impairment, Activity, Participation, Life Satisfaction, and Survival in Persons With Locked-In Syndrome for Over a Decade: Follow-Up on a Previously Reported Cohort. *The Journal of Head Trauma Rehabilitation*, *18*(5), 435–444.

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, *85*, 761–768. https://doi.org/10.1016/j.neuroimage.2013.06.035

Eickhoff, S. B., Heim, S., Zilles, K., & Amunts, K. (2009). A systems perspective on the effective connectivity of overt speech production. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *367*(1896), 2399–2421. https://doi.org/10.1098/rsta.2008.0287

Farwell, L. A., & Donchin, E. (1988). Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, *70*(6), 510–523. https://doi.org/10.1016/0013-4694(88)90149-6

Felgoise, S., Zaccheo, V., Duff, J., & Simmons, Z. (2015). Verbal communication impacts quality of life in patients with amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *17*, 1–5. https://doi.org/10.3109/21678421.2015.1125499

Filippidou, F., & Moussiades, L. (2020). A Benchmarking of IBM, Google and Wit Automatic Speech Recognition Systems. *Artificial Intelligence Applications and Innovations*, *583*, 73–82. https://doi.org/10.1007/978-3-030-49161-1_7

Geschwind, N. (1965). Disconnexion Syndromes In Animals And Man. *Brain*, *88*(2), 237. https://doi.org/10.1093/brain/88.2.237

Golfinopoulos, E., Tourville, J. A., & Guenther, F. H. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *NeuroImage*, *52*(3), 862–874. https://doi.org/10.1016/j.neuroimage.2009.10.023

Gonzalez-Lopez, J. A., Gomez-Alanis, A., Martín Doñas, J. M., Pérez-Córdoba, J. L., & Gomez, A. M. (2020). Silent Speech Interfaces for Speech Restoration: A Review. *IEEE Access*, *8*, 177995–178021. https://doi.org/10.1109/ACCESS.2020.3026579

Green, J. R., Yunusova, Y., Kuruvilla, M. S., Wang, J., Pattee, G. L., Synhorst, L., Zinman, L., & Berry, J. D. (2013). Bulbar and speech motor assessment in ALS: Challenges and future directions. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *14*(7–8), 494–500. https://doi.org/10.3109/21678421.2013.817585

Guerrero-Mosquera, C., Trigueros, A. M., & Navia-Vazquez, A. (2012). EEG Signal Processing for Epilepsy. In *Epilepsy—Histological, Electroencephalographic and Psychological Aspects*. IntechOpen. https://doi.org/10.5772/31609

Higginbotham, J., Fulcher, K., & Seale, J. (2016). *Time and timing in interactions involving individuals with ALS, their unimpaired partners and their speech generating devices*.

Hill, R. M., Boto, E., Rea, M., Holmes, N., Leggett, J., Coles, L. A., Papastavrou, M., Everton, S. K., Hunt, B. A. E., Sims, D., Osborne, J., Shah, V., Bowtell, R., & Brookes, M. J. (2020). Multi-channel whole-head OPM-MEG: Helmet design and a comparison with a conventional system. *NeuroImage*, *219*, 116995. https://doi.org/10.1016/j.neuroimage.2020.116995

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458. https://doi.org/10.1038/nature17637

Iivanainen, J., Zetter, R., Grön, M., Hakkarainen, K., & Parkkonen, L. (2019). On-scalp MEG system utilizing an actively shielded array of optically-pumped magnetometers. *NeuroImage*, *194*, 244–258. https://doi.org/10.1016/j.neuroimage.2019.03.022

Indefrey, P., & Levelt, W. J. M. (2000). The Neural Correlates of Language Production. In *The new cognitive neurosciences* (2nd ed., pp. 845–865). MIT Press.

Jahangiri, A., & Sepulveda, F. (2018). The Relative Contribution of High-Gamma Linguistic Processing Stages of Word Production, and Motor Imagery of Articulation in Class Separability of Covert Speech Tasks in EEG Data. *Journal of Medical Systems*, *43*(2), 20. https://doi.org/10.1007/s10916-018-1137-9

Jin, J., Sellers, E. W., & Wang, X. (2012). Targeting an efficient target-to-target interval for P300 speller brain–computer interfaces. *Medical & Biological Engineering & Computing*, *50*(3), 289–296. https://doi.org/10.1007/s11517-012-0868-x

Koizumi, K., Ueda, K., & Nakao, M. (2018). Development of a Cognitive Brain-Machine Interface Based on a Visual Imagery Method. *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1062–1065. https://doi.org/10.1109/EMBC.2018.8512520

Kumar, P., Saini, R., Roy, P. P., Sahu, P. K., & Dogra, D. P. (2018). Envisioned speech recognition using EEG sensors. *Personal and Ubiquitous Computing*, *22*(1), 185–199. https://doi.org/10.1007/s00779-017-1083-4

Laureys, S., Pellas, F., Van Eeckhout, P., Ghorbel, S., Schnakers, C., Perrin, F., Berré, J., Faymonville, M.-E., Pantke, K.-H., Damas, F., Lamy, M., Moonen, G., & Goldman, S. (2005). The locked-in syndrome: What is it like to be conscious but paralyzed and voiceless? In S. Laureys (Ed.), *Progress in Brain Research* (Vol. 150, pp. 495–611). Elsevier. https://doi.org/10.1016/S0079-6123(05)50034-7

León-Carrión, J., Eeckhout, P. van, & Domínguez-Morales, M. del R. (2002). Review of subject: The locked-in syndrome: a syndrome looking for a therapy. *Brain Injury*, *16*(7), 555–569. https://doi.org/10.1080/02699050110119466

Li, F., Chao, W., Li, Y., Fu, B., Ji, Y., Wu, H., & Shi, G. (2021). Decoding imagined speech from EEG signals using hybrid-scale spatial-temporal dilated convolution network. *Journal of Neural Engineering*, *18*(4), 0460c4. https://doi.org/10.1088/1741-2552/ac13c0

Lopez-Bernal, D., Balderas, D., Ponce, P., & Molina, A. (2022). A State-of-the-Art Review of EEG-Based Imagined Speech Decoding. *Frontiers in Human Neuroscience*, *16*. https://www.frontiersin.org/article/10.3389/fnhum.2022.867281

Makkonen, T., Ruottinen, H., Puhto, R., Helminen, M., & Palmio, J. (2018). Speech deterioration in amyotrophic lateral sclerosis (ALS) after manifestation of bulbar symptoms. *International Journal of Language & Communication Disorders*, *53*(2), 385–392. https://doi.org/10.1111/1460-6984.12357

Matsumoto, M., & Hori, J. (2014). Classification of silent speech using support vector machine and relevance vector machine. *Applied Soft Computing*, *20*, 95–102. https://doi.org/10.1016/j.asoc.2013.10.023

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science (New York, N.Y.)*, *211*(4477), 77–80. https://doi.org/10.1126/science.7444452

Millán, J. del R., Rupp, R., Mueller-Putz, G., Murray-Smith, R., Giugliemma, C., Tangermann, M., Vidaurre, C., Cincotti, F., Kubler, A., Leeb, R., Neuper, C., Mueller, K., & Mattia, D. (2010). Combining Brain–Computer Interfaces and Assistive Technologies: State-of-the-Art and Challenges. *Frontiers in Neuroscience*, *4*. https://www.frontiersin.org/article/10.3389/fnins.2010.00161

Mirkovic, B., Debener, S., Jaeger, M., & Vos, M. D. (2015). Decoding the attended speech stream with multi-channel EEG: Implications for online, daily-life

applications. *Journal of Neural Engineering*, *12*(4), 046007. https://doi.org/10.1088/1741-2560/12/4/046007

Moses, D. A., Leonard, M. K., Makin, J. G., & Chang, E. F. (2019). Real-time decoding of question-and-answer speech dialogue using human cortical activity. *Nature Communications*, *10*(1), 3096. https://doi.org/10.1038/s41467-019-10994-4

Moses, D. A., Metzger, S. L., Liu, J. R., Anumanchipalli, G. K., Makin, J. G., Sun, P. F., Chartier, J., Dougherty, M. E., Liu, P. M., Abrams, G. M., Tu-Chan, A., Ganguly, K., & Chang, E. F. (2021). Neuroprosthesis for Decoding Speech in a Paralyzed Person with Anarthria. *New England Journal of Medicine*, *385*(3), 217–227. https://doi.org/10.1056/NEJMoa2027540

Nakanishi, M., Wang, Y., Chen, X., Wang, Y.-T., Gao, X., & Jung, T.-P. (2018). Enhancing Detection of SSVEPs for a High-Speed Brain Speller Using Task-Related Component Analysis. *IEEE Transactions on Biomedical Engineering*, *65*(1), 104–112. https://doi.org/10.1109/TBME.2017.2694818

Newman, A. J., Supalla, T., Hauser, P., Newport, E. L., & Bavelier, D. (2010). Dissociating neural subsystems for grammar by contrasting word order and inflection. *Proceedings of the National Academy of Sciences*, *107*(16), 7539–7544. https://doi.org/10.1073/pnas.1003174107

Nguyen, C. H., Karavas, G. K., & Artemiadis, P. (2017). Inferring imagined speech using EEG signals: A new approach using Riemannian manifold features. *Journal of Neural Engineering*, *15*(1), 016002. https://doi.org/10.1088/1741-2552/aa8235

Olivi, E. (2011). *Coupling of numerical methods for the forward problem in Magneto- and Electro-Encephalography* [Phdthesis, Université Nice Sophia Antipolis]. https://tel.archives-ouvertes.fr/tel-00838707

Panachakel, J. T., & Ramakrishnan, A. G. (2021). Decoding Covert Speech From EEG-A Comprehensive Review. *Frontiers in Neuroscience*, *15*. https://www.frontiersin.org/article/10.3389/fnins.2021.642251

Paul, Y., Jaswal, R. A., & Kajal, S. (2018). Classification of EEG Based Imagine Speech Using Time Domain Features. *2018 International Conference on Recent Innovations in Electrical, Electronics Communication Engineering (ICRIEECE)*, 2921–2924. https://doi.org/10.1109/ICRIEECE44171.2018.9008572

Plum, F., & Posner, J. B. (1982). *The Diagnosis of Stupor and Coma*. Oxford University Press.

Price, C. J. (2010). The anatomy of language: A review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, *1191*(1), 62–88. https://doi.org/10.1111/j.1749-6632.2010.05444.x

Rezeika, A., Benda, M., Stawicki, P., Gembler, F., Saboor, A., & Volosyak, I. (2018). Brain–Computer Interface Spellers: A Review. *Brain Sciences*, *8*(4), 57. https://doi.org/10.3390/brainsci8040057

Sheerman-Chase, T. (2012). *EEG Brain Scan* [Photo]. https://www.flickr.com/photos/tim_uk/8135749317/

Smith, E., & Delargy, M. (2005). Locked-in syndrome. *BMJ*, *330*(7488), 406–409. https://doi.org/10.1136/bmj.330.7488.406

Spüler, M., Rosenstiel, W., & Bogdan, M. (2012). Online Adaptation of a c-VEP Brain-Computer Interface(BCI) Based on Error-Related Potentials and Unsupervised Learning. *PLOS ONE*, *7*(12), e51077. https://doi.org/10.1371/journal.pone.0051077

Sun, P., Anumanchipalli, G. K., & Chang, E. F. (2020). Brain2Char: A deep architecture for decoding text from brain recordings. *Journal of Neural Engineering*, *17*(6), 066015. https://doi.org/10.1088/1741-2552/abc742

Tremblay, P., & Dick, A. S. (2016). Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain and Language*, *162*, 60–71. https://doi.org/10.1016/j.bandl.2016.08.004

Vaughan, T. M., Heetderks, W. J., Trejo, L. J., Rymer, W. Z., Weinrich, M., Moore, M. M., Kübler, A., Dobkin, B. H., Birbaumer, N., Donchin, E., Wolpaw, E. W., & Wolpaw, J. R. (2003). Brain-computer interface technology: A review of the Second International Meeting. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *11*(2), 94–109. https://doi.org/10.1109/tnsre.2003.814799

Vrba, J., & Robinson, S. E. (2001). Signal Processing in Magnetoencephalography. *Methods*, *25*(2), 249–271. https://doi.org/10.1006/meth.2001.1238

Wijesekera, L. C., & Nigel Leigh, P. (2009). Amyotrophic lateral sclerosis. *Orphanet Journal of Rare Diseases*, *4*(1), 3. https://doi.org/10.1186/1750-1172-4-3

Wildgruber, D., Ackermann, H., & Grodd, W. (2001). Differential Contributions of Motor Cortex, Basal Ganglia, and Cerebellum to Speech Motor Control: Effects

of Syllable Repetition Rate Evaluated by fMRI. *NeuroImage*, *13*(1), 101–109. https://doi.org/10.1006/nimg.2000.0672

Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., Donchin, E., Quatrano, L. A., Robinson, C. J., & Vaughan, T. M. (2000). Brain-computer interface technology: A review of the first international meeting. *IEEE Transactions on Rehabilitation Engineering*, *8*(2), 164–173. https://doi.org/10.1109/TRE.2000.847807

Zhang, X., Wu, D., Ding, L., Luo, H., Lin, C.-T., Jung, T.-P., & Chavarriaga, R. (2021). Tiny noise, big mistakes: Adversarial perturbations induce errors in brain–computer interface spellers. *National Science Review*, *8*(4), nwaa233. https://doi.org/10.1093/nsr/nwaa233

Zhao, S., & Rudzicz, F. (2015). Classifying phonological categories in imagined and articulated speech. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 992–996. https://doi.org/10.1109/ICASSP.2015.7178118