Causal Processes Underlying Unimodal and Multimodal Language

by

EunSeon Ahn

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in the University of Michigan
2023

Doctoral Committee:

Associate Professor David Brang, Chair
Associate Professor Jonathan Brennan
Assistant Professor Taraz Lee
Professor Thad Polk

EunSeon Ahn

eunahn@umich.edu

ORCID ID: 0009-0006-5434-2008

Dedication

To my mom, dad, Dan, family, and friends,

Who made this journey worthwhile

Acknowledgements

Many wonderful and talented individuals have directly or indirectly contributed to the completion of this dissertation. I am immensely grateful for all my mentors, colleagues, family, and friends who have helped me along the way.

First, I would like to thank my advisor, David Brang, for his 7 years of mentorship, expertise, and guidance. David, I have learned so much during my time in your lab and I am thankful for all the exciting projects that I have been able to be a part of. You will always be the face of all things research in my mind.

To my dissertation and advisory committee, thank you for taking the time to provide valuable feedback that helped improve the quality of my work and engage in discussions that challenged me to think more broadly. I would like to especially thank Taraz Lee, who have painstakingly trained me on TMS. Thank you to Shawn Hervey-Jumper for spending many hours answering my questions about tumors and providing thoughtful feedback on my tumor lesion work.

To my current and former lab members of the Multisensory Perception Lab, thank you so much for being so incredibly generous with your time, support, and knowledge. Without your help, I would still be working on my dissertation. Anu, Olivia, Karthik, Cody, and Areti—you guys have brought me so much joy (and food) during my graduate school years. I truly treasure all the times we've been able to spend together both inside and outside of lab and I am very lucky to call you guys my friends. And of course, my CCN friends outside of the lab, Dalia,

Hyesue, Erin, and Madison, I will always cherish the many snack runs, late night work-sessions, cooking/baking dates, and decorating parties we shared.

To Kari, Cece, Anu, and Madison, who have bombarded me with nearly too much faith in my abilities, thank you for always being there to share in my joys and provide comfort when I need it, you guys are wonderfully amazing.

Most importantly, I'd like to express a huge thanks to my family, especially my parents, who always believed that I could accomplish anything in the world and did their best to help me succeed. You are the true source of all my accomplishments. Lastly, Dan, thank you for all the love, support, and coke slushies. You have filled my life with so much laughter and happiness. Thank you.

Table of Contents

List of Tables

List of Figures

Abstract

Language, including speech production and perception, is a major cognitive function necessary for a healthy social and vocational outlook. It is reported that approximately 5-10% of the American population experience communication disorders which can manifest as hearing impairments, difficulty speaking, speech impairments such as stuttering, and more complex language disorders (Ruben, 2009). Given the high prevalence of communication disorders in the United States and the crucial role that language plays in everyday life, it is important to investigate the underlying neural processes and mechanisms that support this social function as well as the brain regions and networks that are involved. A deeper understanding of the mechanisms and structural correlates can help identify the numerous ways in which these functions may be impaired in individuals through disorder, disease, or injury. By understanding which specific components of the process are impacted by neural damage, researchers may gain greater insight into new ways to treat and rehabilitate language impairments as well as to promote the development of devices that can assist in living with these deficits.

In this dissertation, I focus on two important aspects of language that are relevant to clinical deficits: semantic naming and audiovisual speech integration. Focusing on these two critical components of language, I discuss three lines of research that examine the causality of the brain regions involved in these unimodal and multimodal language functions. In Study 1, I employ a causal method, voxel lesion symptom mapping, in intrinsic brain tumor patients to show that the left middle temporal gyrus (MTG) is the primary locus of semantic naming. This finding is consistent with established findings in the stroke lesion literature and demonstrates the

validity of the brain tumor model in lesion mapping. In Study 2, I extend the scope of language causality to audiovisual speech integration using the same brain tumor model. Audiovisual speech integration is a highly relevant form of multisensory integration. It allows the merging of information from various unisensory modalities into a single coherent percept and is an important part of how the brain processes sensory information. Using lesion mapping, I examine which brain regions are critically responsible for audiovisual speech integration behaviors to dissociate whether the merging of conflicting audiovisual speech and the processing of congruent audiovisual speech rely on the same audiovisual integration mechanism. This study challenges the widely held underlying assumption that these two forms of audiovisual processing reflect the same integration mechanism. Lastly, in Study 3, I extend the test of the causal brain regions involved in audiovisual speech to healthy individuals. In this study, single-pulse transcranial magnetic stimulation was applied to disrupt the cortical activity in the left posterior superior temporal sulcus (pSTS), a region largely believed to be the hub of multisensory speech processing. I show that inhibitory stimulation to this multisensory zone can disrupt the fusing of conflicting audiovisual speech while having no effect on the processing of congruent audiovisual speech. These findings point to a dissociation in neural mechanisms between the two audiovisual integration processes and demonstrate that the pSTS reflects only one of the multiple critical areas necessary for audiovisual speech interactions.

Chapter 1 Introduction

A large number of neuroimaging studies have examined the brain areas and networks subserving speech production and perception. While this research has greatly advanced our understanding of the brain areas implicated in language functions, the majority of studies employ correlational neuroimaging methods such as functional Magnetic Resonance Imaging (fMRI), electroencephalography (EEG), and magnetoencephalography (MEG). Although these methods allow us to easily study language non-invasively in healthy populations, they provide correlations between brain structure and function rather than evidence of causal structure-to-function relationships. As a result, correlative measures may reveal non-causally relevant activations in certain brain regions due to domain-general or separate task-relevant dimensions. This is especially relevant when investigating complex processes that involve a large network of brain areas like language. Therefore, such methods are unable to provide a clear distinction between the brain areas that are critical versus the areas that provide support functions.

To better understand how the brain enables language perception and production, we must quantify the unique and causal contributions of each component. Specifically, we focused on the following two important aspects that are relevant to clinical deficits in language: semantic naming and audiovisual speech integration. Semantic naming involves the ability to recognize and retrieve the lexicon for a pictured object and is necessary for identifying and describing physical objects in our environment. On the other hand, audiovisual (AV) speech perception involves the ability to process concurrent auditory and visual speech information, thereby improving speech recognition. Audiovisual integration serves the important function of

recovering language information from visual cues and is especially pertinent for individuals with language deficits who may need compensatory mechanisms to accurately process speech. Focusing on these two important components of language, we sought to address the following three aims.

The first aim of the dissertation was to validate causality in a classic language process like semantic naming, using lesions caused by intrinsic brain tumors. Lesion studies allow us to draw causal, rather than correlational inferences, between a brain area and a cognitive function. Instead of merely establishing the involvement of a brain area in the execution of a function by studying the activation of brain patterns or waves, lesion studies identify which areas are necessary to maintain particular cognitive functions. In Study 1, we investigated the causally necessary brain structures that support the semantic naming process in patients with a tumor lesion. Semantic naming performance was measured via the picture naming task, a well-established language paradigm used to assess word retrieval. Using fMRI, researchers have implicated the left prefrontal and middle temporal areas during semantic naming functions (Binder et al., 2009; Hoffman & Morcom, 2018; Xu et al., 2020). Moving beyond the correlative methods, many language researchers used lesion studies to identify crucial regions for various language functions. However, the majority of lesion studies focus on the post-stroke aphasic patient population, with very few studies (Fekonja et al., 2021; Habets et al., 2019) looking at lesions in tumor patients. Because no single patient population group can serve as the optimal model for examining healthy brain functions, studying distinct clinical populations allows us to ensure that the findings generalize beyond the specifically targeted clinical group. Additionally, using a different clinical population can introduce new dimensions that may explain some of the variance in established models. For example, in our study, we investigated whether different sub-

regions of the tumor lesion such as the core (made up of necrotic tissue and other tumor-infiltrated masses) and peritumoral edema (the swelling caused by fluid around the tumor core) can differentially contribute to observed impairment in semantic naming function. Therefore, Study 1 discusses the work using tumor lesions as an alternative clinical model to examine the robustness of causal language findings in a separate disease model.

The second aim of this dissertation was to extend the scope of language causality to audiovisual speech integration using tumor lesions. Having validated the use of the tumor lesion model in Study 1 and having identified which sub-regions of the tumor best model the relationship between lesions and symptoms, Study 2 expanded on this unimodal approach by examining the brain regions causally linked to multimodal speech perception. Concurrent visual cues help enhance our auditory speech perception by providing complementary temporal and phonemic information (Luo et al., 2010; Plass et al., 2020; Schroeder et al., 2008). When the auditory system only extracts limited speech information, the concurrent processing of visual information becomes especially valuable as it can recover some of this lost speech information. Thus, given that language perception naturally occurs in multisensory environments, it is important to examine how the brain enables vision to support language processes. Toward this goal, in Study 2, we investigated the causal brain structures that are necessary for audiovisual speech integration observed at the behavioral level through tumor lesion mapping.

The third aim of our dissertation was to examine the effect of electrical disruption applied to the posterior superior temporal sulcus (pSTS) on audiovisual speech integration in healthy subjects. In Study 3, we extended our test of the causal processes involved in audiovisual speech established through our lesion mapping work to healthy individuals. While lesion mapping provides compelling evidence of the critical regions of the brain necessary to maintain a

cognitive function, it is limited to patient populations like stroke and tumor patients whose brain functions may have been otherwise compromised by the chronic disease beyond the impact of the lesion. Due to the inevitable gap in time between the onset of lesion formation and the time of behavioral testing, it also leaves the possibility that the brain develops compensatory mechanisms to support the original cognitive function. Additionally, within a single patient, lesions can develop in many different areas, making the localization of cognitive functions less precise.

Transcranial magnetic stimulation (TMS) is a noninvasive brain stimulation method that involves the application of magnetic pulses to targeted brain areas that results in either temporary excitation or inhibition of neuronal functions. Using TMS, we can control exactly which area is targeted and inhibited and is thus necessary to support a specific behavioral function. Therefore, we used TMS with neurotypical individuals to expand upon the findings from our audiovisual integration lesion work. By applying single pulse stimulation during individual trials, we disrupted the cortical excitability of the brain areas believed to be involved in audiovisual integration and examined whether this temporary disruption creates behavioral impairment in audiovisual integration performance similar to those observed in tumor patients that present with lesions in the same area. This TMS study allowed us to investigate audiovisual processing in healthy subjects in a more controlled setting outside the hospital and enabled us to further establish the generalizability of our causal relationships to the general population.

Chapter 2 Validation of Causality in Semantic Naming Function Using Intrinsic Brain Tumor Model

2.1 Abstract

The ability to accurately name an object and retrieve related semantic information is critical for everyday communications. Research into the neural origins of semantic naming has implicated temporal and frontal regions. However, a majority of prior research has used correlative and non-invasive methods to study semantic naming tasks, limiting our understanding of speech processing and the influence of brain lesions on clinical outcomes.

Voxel-lesion symptom mapping (VLSM) has been used extensively to examine the causal relationship between behavioral deficits and specific brain region function by analyzing the continuous behavioral performance at each lesioned voxel. When applied to patients with a lesion caused by a stroke, numerous VLSM studies have suggested a causal link between anterior and mid-temporal regions and semantic language function. However, due to the etiology of stroke, such stroke lesion studies may be biased towards identifying critical areas around major neuro-vasculature, preventing the generalization of the findings to the normal population. In contrast, tumor lesions are more uniformly distributed throughout the brain and occur in a wider age demographic. While tumor models may validate and expand upon stroke lesion findings, few studies have used VLSM in the brain tumor population, leading to knowledge gaps concerning which tumor features account for variance in VLSM models.

The objectives of the present study are twofold: (1) to replicate findings from stroke VLSM research by identifying regions that are associated with semantic naming deficits in non-

vascular lesions, thereby validating the use of tumor lesions in VLSM and (2) to investigate how unique aspects of tumor lesions can further account for variance in VLSM models. Results from our multivariate support vector regression lesion-symptom mapping (SVR-LSM) analyses of n = 109 tumor patients validate stroke VLSM research by demonstrating a causal relationship between semantic naming and damage to the left temporal lobe, most notably at middle and inferior temporal regions. Subsequently, we examined how tumor sub-regions (i.e., peritumoral edema or core) distinctly contributed to semantic naming impairment. While the presence of both peritumoral edema and core in temporal regions was associated with impairments in semantic naming, core lesions in this critical area translated to more significant deficits. Finally, consistent with stroke VLSM research, a larger core volume was associated with greater semantic naming impairment irrespective of location, emphasizing the importance of covarying this dimension. Our findings indicate that different tumor sub-regions have distinct impacts on semantic naming performance and that the uneven distribution of tumor cells found within the lesion is causally linked to varying levels of language deficit.

2.2 Introduction

Naming an object in a photograph, classically tested using a semantic naming or picture naming task, requires the ability to map a concept to a lexicon and is a fundamental component of human language production. Accurate picture naming performance is supported by cognitive processes consisting of, at the minimum, visual perception of the object, semantic processing and lexical retrieval of the percept, and motor articulation to produce the name of the object (for review see Levelt et al., 1998). Consequently, semantic naming abilities can be disrupted by impairments encountered along any of these stages and are believed to be maintained by a broad network of cortical and subcortical regions. While various brain regions are involved in

supporting this function, prior functional imaging, electrophysiological, and stimulation studies report that left prefrontal and middle to posterior temporal areas are essential to facilitate semantic naming functions (Levelt et al., 1998; Krieger-Redwood & Jefferies, 2014; Visser et al., 2010). Specifically, the involvement of left temporal regions is believed to be recruited during the phonological encoding stage of the process (Levelt et al., 1998). As one of the most widely studied functions in language literature, semantic naming task is believed to involve all stages of language production including conceptualization, formulation, and articulation (Levelt, 1998). As such, semantic naming is an optimal language process that can be used to replicate stroke lesion findings in tumor lesion to validate the causality of brain regions using brain tumor model.

Complementing the insights from functional neuroimaging and electrophysiological research, lesion studies draw causal inferences between the brain area and semantic naming ability. Specifically, combining traditional lesion-symptom mapping with advanced imaging methods, voxel-lesion symptom mapping (VLSM; Bates et al., 2003) enables the analysis of the brain regions implicated in behavioral performance by looking at the relationship between lesion status and the measure of a cognitive function at the individual voxel level. Traditionally, within the language domain, these lesion-symptom mapping studies examined patients with post-stroke aphasias to determine which brain voxels must remain intact, and are therefore functionally critical, for the maintenance of normal language function. In supporting the semantic naming function, stroke lesion studies typically report the involvement of left anterior and mid-temporal regions, although the specific regions vary according to types of errors and variables included in the analysis (Baldo et al. 2013; Schwartz et al., 2009).

Compared to a large number of post-stroke VLSM studies, few studies have investigated the relationship between tumor locations and their associated behavioral deficits using VLSM (van Grinsven et al., 2023). While patient populations often offer unique opportunities to study what processes occur in their healthy counterparts, no single patient population with neurological disorders can perfectly model healthy brain functions. However, investigating the relationship between the tumor location and behavior can bring additional insights into the established stroke lesion work, expand on the generalizability of prior findings, and introduce new dimensions that may explain a part of the variance of the established models. As a result, studying tumor patients can address some of the limitations universally faced by stroke lesion studies. For example, given the etiology of stroke, the distribution of stroke lesions is necessarily concentrated around major neural vasculature (Adam et al. 2018); on the other hand, tumor lesions are more widely distributed (Anderson et al., 1990; Herbet and Duffau, 2022). With a broader distribution of lesions, tumor lesions thereby offer greater power to detect brain regions that are causally associated with specific functions compared to stroke lesions. Additionally, stroke occurs more frequently in older adults with a mean age of around 70 (Kissela et al., 2012; Wang, Rudd, and Wolfe, 2013). On the other hand, brain tumors occur in a broader age distribution with a median age of primary glioma diagnosis at 38 years of age (Lin et al., 2020). As cognitive functions decline and change as a function of age, the findings from the stroke lesions may therefore be limited to the older, aging population. Replication and extension of the stroke lesion findings in the tumor population would further allow the extension of the significant findings to broader age groups.

Beyond substantiating the findings from the stroke lesion work, tumor lesion mapping can offer new features and parameters different from those of stroke lesions. These features can

then be used as additional variables to help explain the variance observed in the brain-behavior model. In particular, our study distinguished between the different sub-regions of the tumor to investigate whether the introduction of this distinction can more accurately model the relationship between lesion and behavioral symptoms. In 2015, Ji et al. reported that tumor lesions exhibit heterogeneous tissue morphology with cellular properties of the lesion varying as a function of distance away from the center of the lesion. This suggests that the ratio of healthy to damaged cells can vary within a single lesion. This is in contrast with the lesions resulting from ischemic stroke, which are caused by the disruption of blood supply to the brain, creating a largely uniform lesion with fairly homogenous tissue cellularity (Fussen et al., 2011). Therefore, we're likely to gain more information by segmenting tumor lesions into different sub-regions such as peritumoral edema, necrotic core, and enhancing and non-enhancing tumor core based on their phenotypic differences observed through different MR image contrast intensity (Banerjee and Mitra, 2020; Menze et al., 2015; Bakas et al., 2018). By including sub-regions as a variable, we allow for unexplained variance in the uniform lesion model to be absorbed into the distinction in the tumor sub-regions, should these regions have a variable impact on behavioral impairments.

For the purposes of this study, we were primarily interested in comparing the impact of the tumor core versus the impact of the peritumoral edema. The peritumoral edema is the region of the tumor that consists of vasogenic edema, the swelling caused by fluid around the tumor core, as well as infiltrating tumor cells (Rathore et al., 2018). Given the recent understanding that cancer cells remodel neurons and neuronal circuits, it is essential to understand the extent to which neuronal processing impacts cognition and behavior within tumor regions with dense tumor and few neurons (tumor core) and sparse tumor with many neurons (peritumor edema).

These two regions are particularly interesting because while the tumor cores, as the most aggressive region of the tumor, are maximally resected as part of the standard surgical treatment while trying to preserve critical cognitive functions (Parrish-Novak et al., 2015), there are no international surgical guidelines as to whether peritumoral edemas should be resected (Qin et al., 2021). However, many studies in the past decade have found that the volume and the presence of peritumoral edema negatively predict the overall survival rate of the patients and increase the likelihood of recurrence of gliomas (Schoenegger et al., 2009; Rathore et al., 2018; Qin et al., 2021; Wu et al., 2015). One hypothesis that may explain the discrepancy in the standard surgical protocol is the gradual transition from tumor tissues to healthy tissues between the sub-regions. The tumor cells begin to resemble normal healthy tissues the further it moves away from the core of the tumor to the peritumoral edema until the visually healthy cells are reached. In a 2015 study, the authors Ji et al. built a machine-learning classifier capable of detecting tumor infiltration from tissues using quantitative measures from stimulated Raman scattering images. Their data show that particular cellular properties, including cellularity, axonal density, and protein-to-lipid ratios, vary as a function of distance from the tumor core. The further the cells are located from the gross margins of the core, the more normal these measures become, more closely resembling healthy cells. In addition, several studies (Barajas et al., 2010; Barajas et al., 2012; Gill et al., 2014) provide further evidence that these sub-regions, specifically, the nonenhancing and enhancing part of the tumor (corresponding respectively to peritumoral edema and core in our study) have different histopathological features. Such differences in the cellular and molecular composition of tissues across tumor sub-regions may reflect the varying levels of damage observed in neuronal structure from the different levels of tumor infiltration. Expanding upon this idea that different levels of tissue damage may correlate to varying levels of preserved

cellular functionality, here we examine whether the location and the size of tumor core and edema disproportionately explain the observed semantic naming deficits in tumor patients.

Using data from a large sample of patients who completed a picture naming, or semantic naming task, here we test the hypotheses that: 1) Voxel lesion mapping with brain tumor lesions will identify cortical and subcortical regions similar to and overlapping with vascular stroke lesions; 2) given the broad non-vascular coverage of brain tumors, lesions will uncover novel temporal regions subserving semantic naming function; 3) different sub-regions of the tumor lesion will be associated with varying levels of behavioral impairment and allow us to account for additional variance in VLSM models. Given the inherent differences between stroke and tumor etiology, performing brain-behavior mapping in both tumor and stroke populations permits the possibility of exploring additional parameters. These new parameters may explain the observed behavioral deficits that another model cannot account for, while allowing us to strengthen and complement the conclusions made from stroke literature and consequently establish greater generalizability of the results to the normal population.

To test these hypotheses, we specifically used a multivariate VLSM approach called support vector regression lesion-symptom mapping (SVR-LSM). Compared to the SVR-LSM approach, the traditional VLSM approach relies on univariate analyses, meaning that lesion status and behavior are examined at the individual voxel level as though each voxel was independent of its neighboring voxels. Because SVR-LSM considers the role of all the voxels simultaneously by fitting a single regression model, this multivariate approach has been shown to be sensitive and specific in detecting the relationship between brain regions and behavior (Zhang et al., 2014). Through this approach, the investigation of the contributions of the different tumor sub-regions can reveal interesting relationships between lesion type and observed deficits, in

parallel with the current clinical understanding of the roles of these sub-regions, leading to valuable clinical implications for glioma patients.

2.3 Methods

2.3.1 Subjects

One hundred and nine patients with a diagnosis of World Health Organization (WHO) grade 2-4 diffuse glioma were included in the study. Patients were recruited from a prospective registry at the University of California San Francisco (UCSF). Patient demographics are shown in Table 1. There were no restrictions on the lesion location for the subjects to be included in the analysis; this was to localize the semantic naming function within the whole brain with maximal lesion coverage.

Table 1. *Summary of demographics and clinical information.*

Characteristic	Participants
Sex	65M, 44F
Mean Age	54.2
Handedness	99R, 10L
Education	
High school	14
Some college	16
Bachelors	24
Graduate	17
Professional	2
Unknown	36
Diagnosis	
Glioblastoma	78
Oligodendroglioma	16
Diffuse glioma	15
Tumor Grade	
WHO Grade II	10
WHO Grade III	18
WHO Grade IV	81
Tumor Laterality	34R, 75L

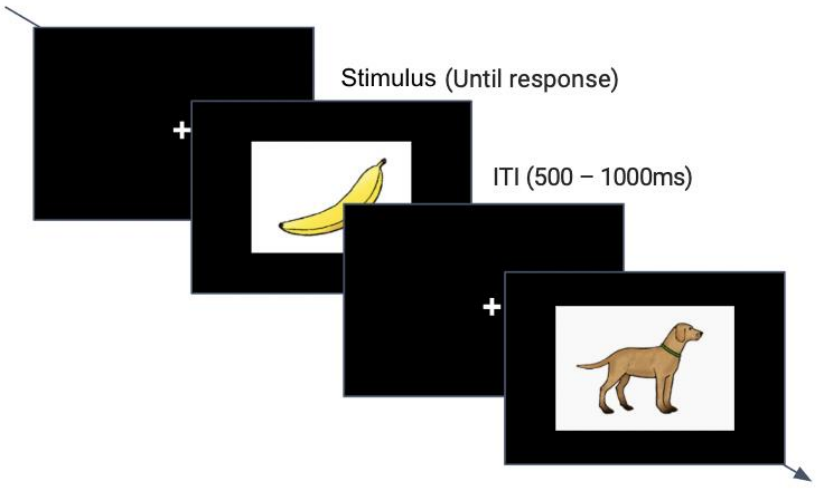
Note. Education was self-reported in patients who completed the Neuro-QOL assessment (Cella et al., 2012). Pathologic diagnoses were made by board-certified pathologists using the World Health Organization (WHO) Revised Classification of Tumors in the CNS.

2.3.2 Experimental Procedure

All subjects underwent a battery of cognitive and language tasks (see for a full list of tasks: Aabedi et al., 2021) including a semantic naming task one or two days prior to their tumor resection surgery. In this task, the patients were presented with 48 colored images on a 15.5-inch MacBook Pro placed approximately 2 ft. away while in a seated position in a clinical setting. The

stimuli were presented using PsychToolbox (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007) on *MATLAB 2018b*. The patients were asked to verbally name the presented object on the screen as quickly and accurately as they can without making any mistakes (Fig. 1). The experimenter advanced through the trials following the subject's verbal response to the stimuli. Patients' verbal responses were recorded and time-locked to the visual onset of each stimulus and manually scored. Two research assistants independently scored each subject's performance according to the Quick Aphasia Battery (QAB; Wilson et al., 2018). The research assistants were blind to the diagnosis and the MR images. Specifically, the following scoring guide was used to rate the subject's performance: 4- correct, accept reasonable alternative labels, 3-correct but delayed for more than 3 seconds or self-corrected, 2- at least half of the phonemes are correct or apraxia error on target, 1-some relation to the target, 0-unrelated response or no response given within 6 seconds. The overall score for the semantic naming task for each subject was obtained by taking an average score across all trials.

Figure 1. *A representative trial sequence of the semantic naming task.*



Note. The stimuli were presented using PsychtoolBox on MATLAB. Following the verbal response from the patients, the experimenter used a response box to progress to the next stimuli.

2.3.3 Image Acquisition

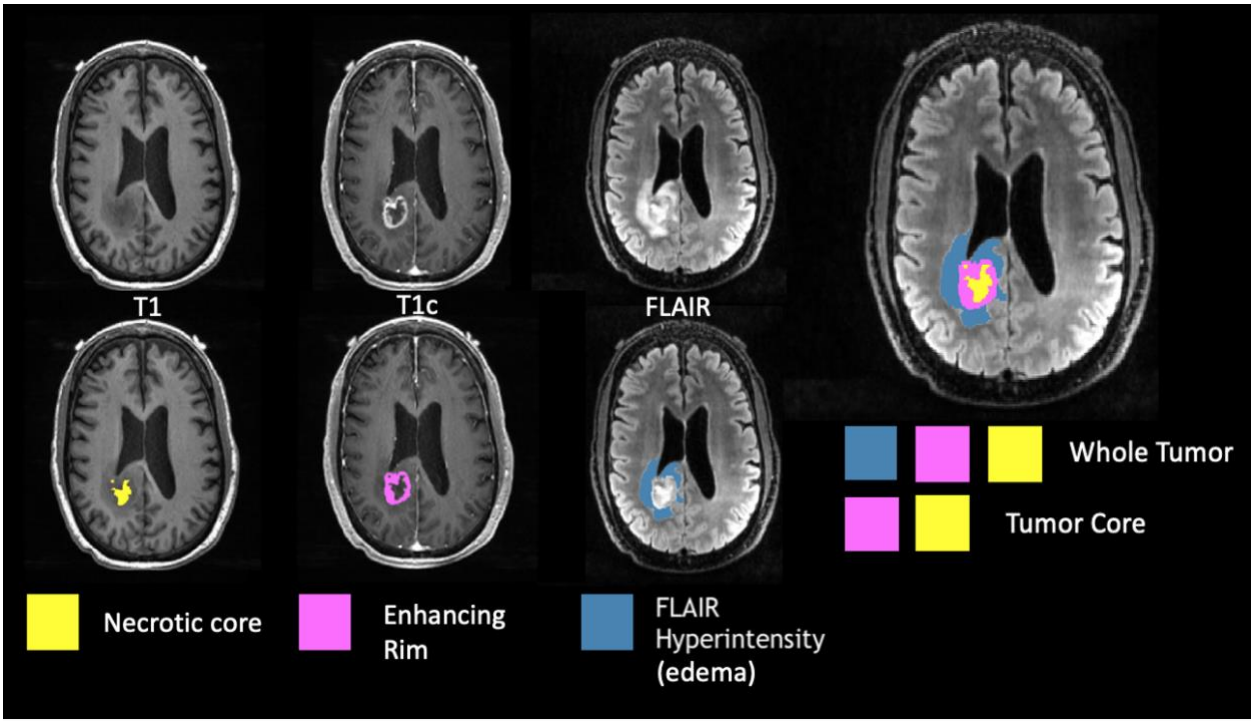
Pre-operative T1-weighted images (T1), post contrast-enhanced T1-weighted images (T1c), and T2-weighted fluid attenuation inversion recovery (FLAIR) images were collected using a 3 Tesla magnetic resonance (MR) machine as part of the patient's standard clinical care. In post contrast-enhanced T1 images, brain vasculature shows up bright due to the intravenous injection of the chemical gadolinium. This contrasting agent helps highlight the disruption in the blood brain barrier caused by the tumor mass (Larsson et al., 1990; Tofts & Kermode, 1991). FLAIR images, on the other hand, limit signals coming from the cerebral spinal fluid and is useful for identifying edema (Hawkins-Daarud et al., 2013; Husstedt et al., 2000; Watanabe et al., 1992), which is swelling of the brain caused by a buildup of fluid. These images were collected within 3 days of cognitive and language testing and used to outline the tumor boundaries.

2.3.4 Tumor Sub-regions

The boundaries of diffuse glioma were defined using the aforementioned MRI sequences (see Fig. 2). In T1 post contrast-enhanced images, the tumor core shows up as dark regions (which we also refer to as necrotic core), with its boundary highlighted by the bright enhancing tumor rim (which we also refer to as enhancing tumor or enhancing core). With respect to MRI-based features, we can describe the tumor core as being made up of this enhancing rim and necrotic core. Histologically, tumor core consists of active cancer cells, regions of hyper vascularity, and necrosis. On the other hand, FLAIR hyperintense regions are adjacent to and just outside of this T1 post contrast-enhanced tumor core and is composed of vasogenic edema and infiltrating tumor cells (Rathore et al., 2018). The FLAIR hyperintense regions, or edemas, are sometimes referred to as the nonenhancing tumor. Whereas T1 post contrast-enhanced core regions are histologically marked by many dense tumor cells and few neurons, FLAIR regions

have many neurons and few cancer cells. Despite these marked differences, the extent to which these radiographic and histologically diverse regions influence cognition and behavior remains incompletely understood. Therefore, we aimed to investigate the extent to which these two regions differentially correlate with semantic naming impairments. To aid this goal, our study partitioned the whole tumor into these two regions: 1.) T1 post contrast-enhanced core (which we may refer to as "core" or "tumor core" in this paper) and 2.) FLAIR hyperintensity (which we may synonymously refer to as "edema" in this paper). It should be noted that across different studies, the specific definitions and exact terminologies of the tumor sub-regions may vary.

Figure 2. *Tumor sub-region partition across different MRI sequences.*



Note. T1 post-contrast (T1c) images enhance the rim of the core (pink) and T2 FLAIRs enhance the FLAIR hyperintensity consisting of peritumoral edema (blue).

2.3.5 Lesion Delineation and Spatial Normalization

To delineate the tumor boundaries, we completed both manual and automatic tumor segmentation. Manual segmentation was used to outline the tumor into the tumor core and the

FLAIR hyperintensity. Automatic tumor segmentation was used to further delineate the boundaries within the tumor core separating the enhanced rim of the core and the necrotic core.

For manual segmentation, the T1 and T2-weighted images were co-registered using *SPM12* (http://www.fil.ion.ucl.ac.uk/spm/). The program *3D Slicer* (slicer.org) was then used to select the tumor areas visible under differently weighted MR images. This program was used to outline and "color" the images based on grouped pixel intensity values. For example, T1-contrast images highlight the boundary of the core of the tumor while T2-FLAIR images highlight the boundary of the peritumoral edema due to the high fluid levels found in the edema. From the T2-weighted FLAIR images, the boundaries of the FLAIR hyperintensity, which encompasses edema, can be unclear from the tumor core. Therefore, all FLAIR hyperintensity masks had their respective core masks (if present) subtracted to ensure that the voxels for FLAIR hyperintensity and core from the same subject were mutually exclusive. The resulting binarized masks for core and FLAIR hyperintensities denote whether each voxel is lesioned in a patient. The tumor core mask and FLAIR hyperintensity mask were combined to produce a total lesion mask for each subject. If specific components of the tumor were missing for any subject, only the available subcomponents of the tumor were added up to create the total lesion mask (i.e., if the subject had no FLAIR hyperintensity outside of the T1 post contrast-enhanced region, then total lesion consisted only of tumor core). The manual segmentation was carried out by research members who were blind to the behavioral performance of the subjects. All manually generated masks underwent a secondary confirmation by a board-certified neurosurgeon.

For auto segmentation, we utilized a software called DeepBraTumIA (https://www.nitrc.org/projects/deepbratumia/) which uses deep learning to automatically conduct voxel-wise segmentation of brain tumors. DeepBraTumIA segments the tumor into the

following sub-compartments: necrosis, contrast-enhancing tumor (which our study refers to as enhancing rim or contrast-enhancing rim), and edema. The resulting masks from the auto segmentation were only used for analysis comparing the contribution of necrotic core versus the enhancing rim in predicting the behavioral deficit. Since the manually segmented masks underwent a more rigorous quality-checking process, the analysis comparing the impact of the tumor core versus FLAIR hyperintensity on the behavioral measure utilized masks that were generated through manual tumor segmentation.

Prior to running the voxel-lesion symptom mapping (VLSM) analysis, all subjects' T1 images were spatially normalized to the standardized Montreal Neurological Institute template (MNI152) using SPM12. The resulting transformation matrix from the normalization process was then applied to the subject's lesion mask to match the standard space. The normalized masks ensured that all subjects' lesion masks can be overlaid on a standard brain.

2.3.6 Multivariate Voxel-Lesion Symptom Mapping (VLSM)

VLSM consists of looking at the lesion status of the voxels in the brain, i.e., whether the voxel is lesioned, to study its relationship to a continuous behavioral score (Bates, 2003). Early VLSM methods utilized a univariate approach where a t-test was performed at each voxel to study the significance of the correlation between lesion status and behavior of interest. While univariate VLSM offers a simple, straightforward analysis approach, it has been criticized for its inability to capture the pattern of voxels functionally involved in a behavior. This is because univariate VLSM only considers a single voxel at each analysis point and it assumes that each voxel is spatially independent of its neighboring voxels. This mass univariate approach requires that thousands of individual t-tests be run for each subject's lesion and results in a high rate of false positive tests. As alternatives, multivariate methods have been proposed by various research

groups (DeMarco & Turkeltaub, 2018; Mah, Husain, Rees, & Nachev, 2014; Pustina et al., 2018; Zhang, Kimberg, Coslett, Schwartz, & Wang, 2014) to detect the distributed network of brain regions and dependencies involved in the behavioral function of interest with greater sensitivity.

Accordingly, in the present study, a multivariate support vector regression lesion-symptom mapping was used to model the relationship between the spatial location of the glioma lesions and the degree of impairment in semantic naming. SVR utilizes the Support Vector Machine, which normally nonlinearly transforms the data into a high-dimensional feature space and searches for a hyperplane that best separates the different class labels, to predict a continuous variable. SVR-LSM is a multivariate machine-learning approach to lesion mapping proposed by Zhang et al. (2014) that utilizes the lesion status of all the included voxels to best predict the behavioral score. Therefore, it intrinsically models the dependencies between the voxels and identifies the pattern of voxel lesions that are involved in facilitating the cognitive function of interest. In the present study, DeMarco & Turkeltaub's (2018) MATLAB GUI implementation of Zhang et al. (2014)'s SVR-LSM approach was used to study which damaged regions in the brain contribute to deficits in the semantic naming task. To address how the two main different glioma sub-regions of interest differentially contributed to semantic naming impairment, 3 separate SVR-VLSM analyses were run, one including only the core masks, one with only the FLAIR hyperintensity masks (consisting of vasogenic edema), and one including the combined total lesion. By running these parallel analyses, we examined whether SVR-LSM can identify significant brain regions correlated with semantic naming function using either only the tumor core or only the FLAIR hyperintensities and if so, whether these identified regions would be the same across both sub-regions.

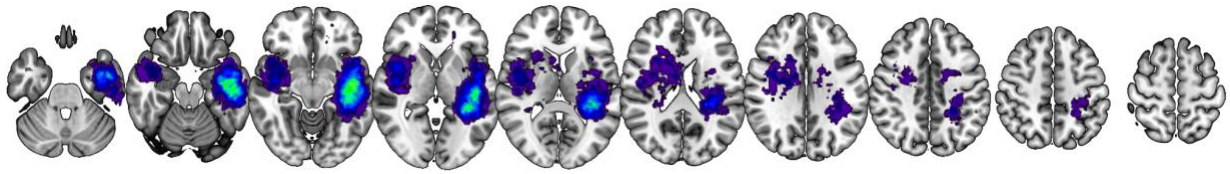
Lesion volumes were regressed out of both behavioral and lesion data to correct for lesion volume as larger lesions tend to be inherently associated with a greater degree of cognitive impairment regardless of the lesion location. To account for multiple comparisons, a permutation-based cluster-level threshold of $p < 0.05$ was used. The default hyperparameters of SVR, epsilon = 0.1, sigma = 0.447, gamma = 5, and cost = 30, were used to build our model.

2.3.7 Region of Interest (ROI) Analysis: T1-post contrast-enhanced core vs. FLAIR region

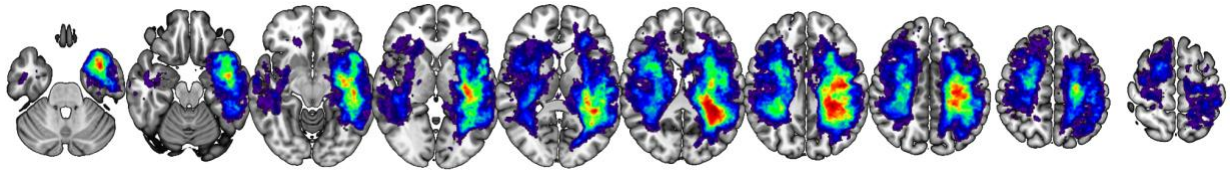
By design, the causally involved regions that can be identified through LSM analysis are necessarily limited by the observed distribution of included lesions. Since it is impossible for the model to predict how a voxel that is not lesioned in any of the included subjects contributes to the behavioral impairment, only voxels that are lesioned in at least 5 subjects (Damasio et al., 2004) were included in the analysis. The final overlay maps of which voxels were included in the analyses after applying this minimum subject threshold are shown in Figure 3 for each of the 3 analyses (T1 post-enhanced core, FLAIR hyperintensity, and total lesion).

Figure 3. *Overlay maps showing the distribution of lesions across all subjects.*

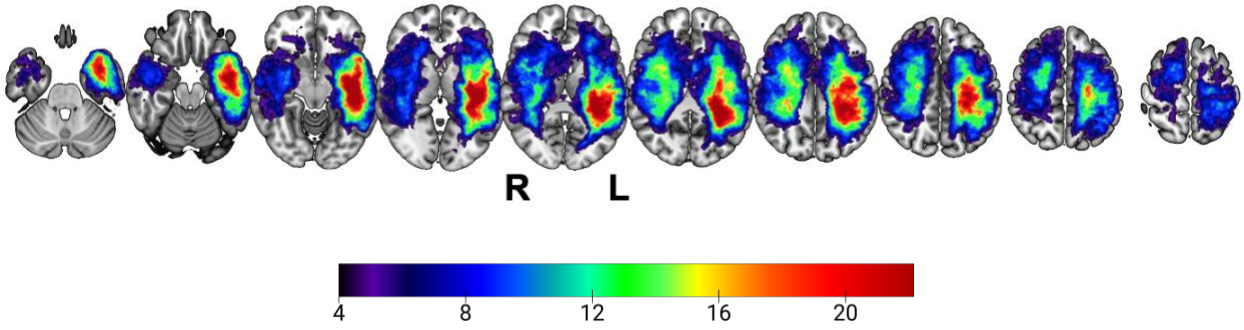
A. T1 post-enhanced core



B. FLAIR hyperintensity



C. Total Lesion



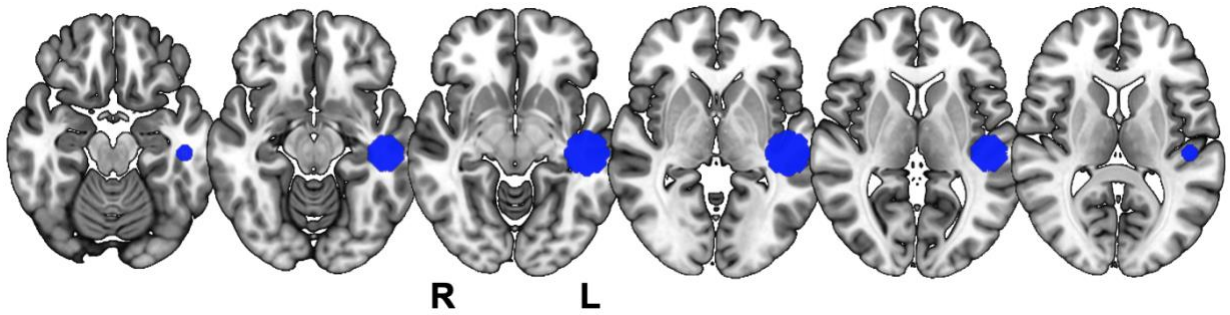
Note. The color hue denotes the number of subjects who have lesions in that voxel and the green outline denotes the boundary of voxels that had at least 5 subjects' lesions overlap. **(A)** Overlay map combining the total lesion (core + FLAIR hyperintensity) across all subjects. **(B)** Overlay map combining the core mask across all subjects. **(C)** Overlay map combining the FLAIR hyperintensity mask across all subjects. Note that left and right are flipped.

It is important to note that because most FLAIR hyperintensity regions are larger in volume than that of cores (if both sub-regions are present), the resulting number and the distribution of voxels that can be analyzed for core lesions are much more restricted than that of FLAIR hyperintensity or total lesions. Consequently, SVR-LSM analysis of the core is less statistically powered. Therefore, to better address the question of whether different tumor sub-regions of the tumor differentially impair the semantic naming performance, a region of interest

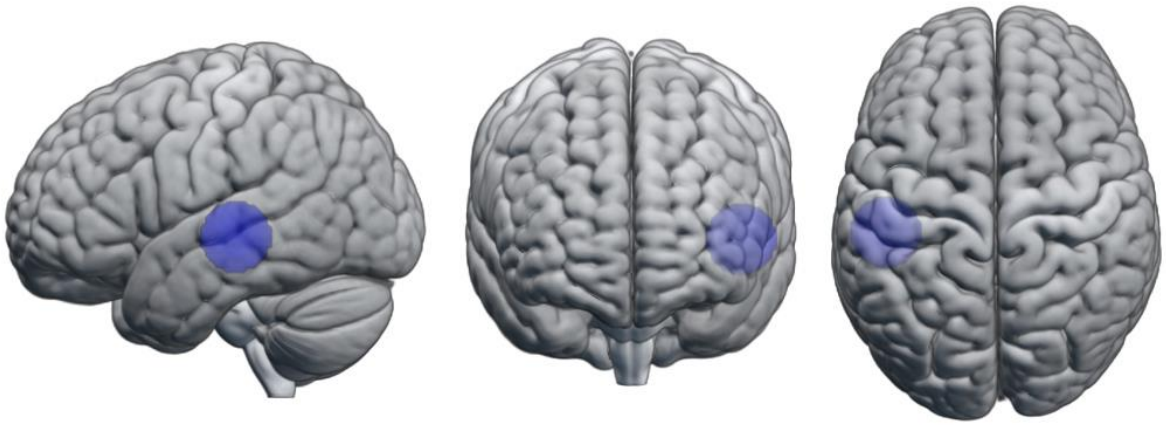
(ROI) analysis was conducted to focus on the region of the brain that has previously been linked to semantic naming function through stroke lesion studies: anterior and middle temporal regions (Baldo et al., 2013; Krieger-Redwood & Jefferies, 2014; Visser et al., 2010). This ROI analysis allowed for a more fine-grained analysis to better understand the effect of each sub-region of the tumor lesion on the behavioral deficit relative to an established control (i.e., stroke model). The set of coordinates of our ROI was derived from the results of a prior stroke lesion mapping study by Baldo et al. (2013). For their voxel lesion analysis, Baldo et al. (2013) measured 96 patients' performance on the Boston Naming Task (Kaplan et al., 2001), a comparable to our semantic naming task, making it a good baseline for our tumor lesion study while allowing us to select an ROI independent of our data. The authors report the MNI coordinates of (-48, 24, -4) as the most significant area associated with behavioral deficits in Boston naming. This coordinate is located around the middle temporal gyrus (MTG). Finally, this ROI was generated by creating a sphere centered around this coordinate with a radius of 15mm (Fig. 4). This 15 mm radius value was derived from Baldo et al.'s (2013) reported significant cluster volume of 9865.

Figure 4. *Location of the MTG ROI.*

A



B



Note. (A) A Cross-section showing the location of the 15mm spherical ROI centered at MNI coordinates (-48, 24, -4) on standard-space Montreal Neurological Institute template (MNI152). Left and right are flipped. (B) 3D rendering showing the location of the ROI

For the analysis, a Wilcoxon signed-rank test was used to examine whether the mean semantic naming performance differed between the subjects whose lesions overlapped with this spherical ROI versus the subjects whose lesions did not overlap with this ROI. Following this analysis, we then further explored if the type of overlap with the ROI, i.e., whether the overlap of the ROI was with the lesion's core, FLAIR hyperintensity, or both, was associated with different performance on the semantic naming task. Therefore, a one-way ANOVA test was conducted to compare the difference in semantic naming score for those with different types of overlap with

the ROI (no overlap, core overlap, FLAIR hyperintensity overlap, both core and FLAIR hyperintensity overlap).

Additionally, based on the prior work from stroke lesions studies that show lesion volume serves as a significant covariate in predicting the magnitude of the impairment, analysis of covariance (ANCOVA) was performed with the total lesion volume as a covariate to determine whether the total lesion volume alone could explain the relationship between the presence of a lesion in the ROI and symptom.

2.3.8 Region of Interest (ROI) Analysis: Necrotic core vs. Contrast-enhancing rim

In addition to investigating the differential contribution of the tumor core versus FLAIR hyperintensity on semantic naming, we wanted to further understand the relative contribution of glioma regions within the tumor core by separately analyzing the necrotic core versus the T1 post contrast-enhancing core rim. We hypothesized that because the necrotic core constitutes a greater extent of dead tissues relative to the T1 post-enhanced rim (Ji et al., 2015; Ye et al., 2020), its presence in the ROI may be associated with greater semantic naming impairments. As specified previously, this segmentation within the core was done automatically, using deep-learning based technique.

Initially, we intended to run the same type of analysis as the core versus FLAIR hyperintensity, focusing on whether the MTG ROI's binary overlap status with the respective tumor mask would result in significantly different behavioral measures. However, given the relatively small size of the core and consequently, the separate subregions of the core with respect to the ROI, there were very few subjects (n = 2) that could be separately grouped into the necrotic core overlap group versus the enhancing rim group to give us any meaningful insights about the respective contribution of each subregion to the behavioral deficit. Therefore, taking

these 2 segmented subregions of the tumor core (the enhancing core rim and the necrotic core), we calculated the number of voxels that overlaps between each of the two core subregions with the same spherical MTG ROI used in the previous ROI analysis. This volume of overlap was then used to calculate the log ratio of overlap between the two subregions with the ROI:

$$Ratio\ of\ Overlap\ =\ log(\frac{Volume\ of\ Enhancing\ Rim\ Overlap}{Volume\ of\ Necrotic\ Core\ Overlap})$$

The greater this ratio, the greater the volume of the enhancing rim that overlaps with MTG ROI compared to the volume of the necrotic core; the smaller the ratio, the greater the volume of the necrotic core compared to the volume of the contrast-enhancing rim. After obtaining this ratio of overlap, we then calculated Spearman's correlation between this log ratio and the semantic naming score to see if the presence of more necrotic tissues in the core was associated with greater impairments in behavior.

2.4 Results

2.4.1 SVR-VLSM localizes semantic naming functions in the left temporal lobe.

Three separate SVR-LSM analyses for core, FLAIR hyperintensity, and combined total lesion masks were conducted to identify which lesioned brain regions were associated with semantic naming performance. After the multiple comparisons correction of the family-wise cluster level threshold of $p < 0.05$ and regressing out the total lesion volume from the behavioral and lesion data, both the core and the total lesion analyses respectively identified a single cluster of voxels with a significant correlation to the semantic naming function (Fig 5). While the significant clusters for these two analyses varied in size ($t(108) = 6.259$, $p = 8.28E\text{-}9$), the two significant clusters were both centered in the left temporal region: core cluster centered at MNI coordinates (-41, -26, -5) and total lesion at MNI coordinates (-41, -27, -5). The significant

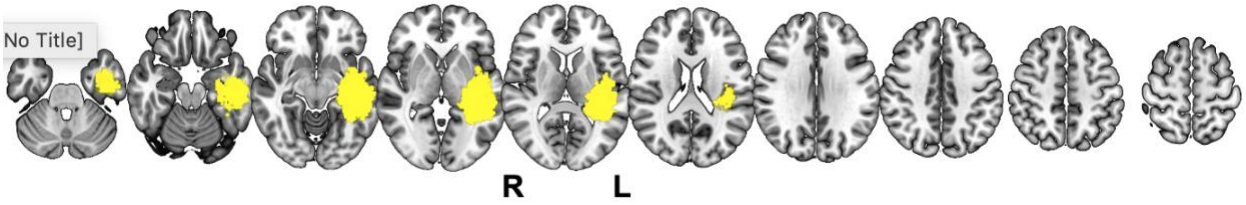
cluster from the total lesion analysis was approximately 74% larger than those identified from the core mask and largely encompassed the significant cluster from the core analysis. These results are consistent with what we would expect as the total lesion masks were made by combining the subject's core and FLAIR hyperintensity mask. Furthermore, the SVR-LSM analysis of the core revealed more restricted brain regions to be associated with semantic naming performance compared to the analysis of the total lesion; while the core analysis identifies regions more limited to the left middle temporal area, the analysis of the total lesion reveals the involvement of the entirety of the left temporal region at large.

Applying the same correction and normalization process, the SVR-LSM analysis of the FLAIR hyperintensity masks similarly identified significant clusters centered around the middle temporal area with the center MNI coordinates (-42, -32, -1). However, while the center of significant clusters in the FLAIR hyperintensity analysis points approximately to the left middle temporal region, the actual distribution of the clusters was distinct from those of core and total lesion analyses (Fig 5). For example, while the results of core and total lesion LSM identify significant regions to be within the middle temporal region, the results of FLAIR hyperintensity LSM reveal significant clusters bordering the middle temporal area, without having significant clusters inside this region. The results of the whole-brain VLSM analysis suggest that the FLAIR hyperintensity and core both affect semantic naming but in largely non-overlapping regions of the temporal lobe. While possible, this finding is likely due to the spatial correlation between core and edemas. Specifically, the tumor core typically resides close to the center of a larger volume of FLAIR hyperintensity (see Fig. 5 for reference). Consequently, they're not spatially independent of one another and the two analyses for FLAIR hyperintensities and core are not orthogonal. For example, even though the voxels for the core were not included in the FLAIR

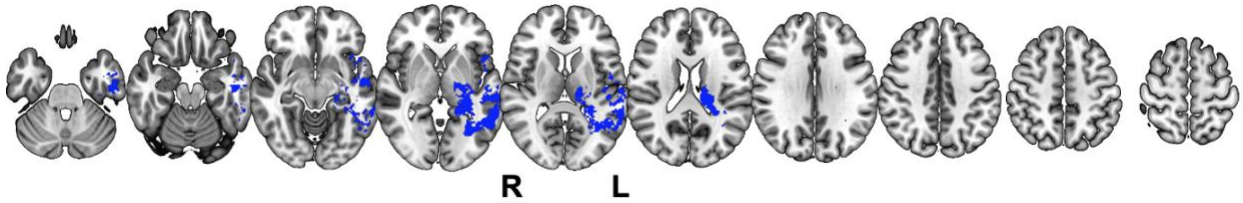
analysis, the subjects still had lesions in their cores near their FLAIR hyperintensities. Therefore, it's likely that the impact of the presence of the core is reflected in the results of the FLAIR VLSM.

Figure 5. *Brain regions identified by SVR-VLSM to be associated with semantic naming performance after a family-wise cluster level thresholding of $p < 0.05$*

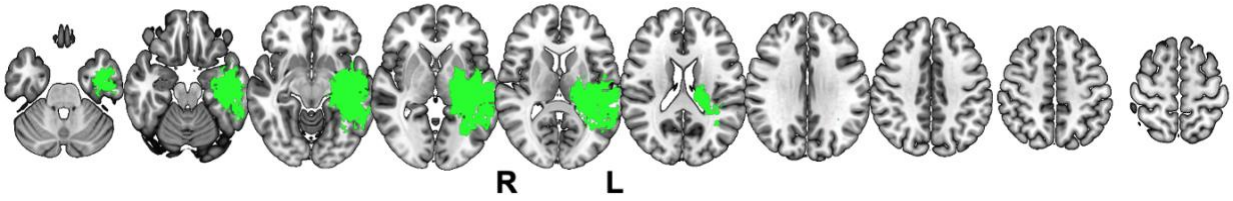
A. T1 post-enhanced core



B. FLAIR hyperintensity



C. Total Lesion



Note. **(A)** Results of VLSM with only the core masks passed in as overlay maps. **(B)** Results of VLSM with only the FLAIR hyperintensity masks passed in as overlay maps. **(C)** Result of VLSM with the total lesion (combination of core and FLAIR hyperintensity) passed in as overlay maps. Note that left and right are flipped.

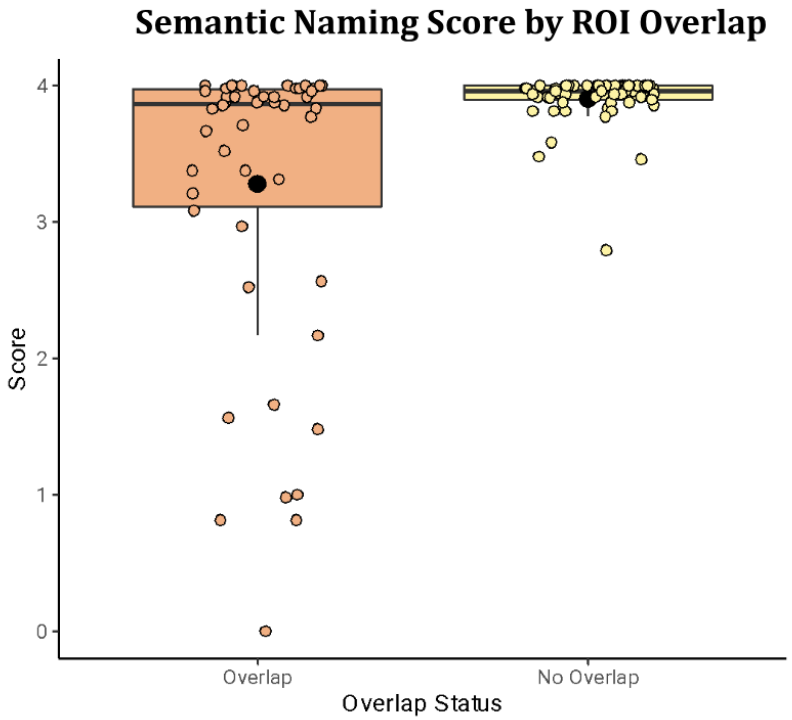
For a voxel to be included in the VLSM analysis to study its relationship to behavior, the voxel needs to be lesioned in multiple subjects (threshold varies across studies). This is because we cannot examine the contribution of a particular voxel to a behavioral deficit if none of the subjects presented with a lesion in that voxel. And because the tumor cores tend to be much

smaller in volume than FLAIR hyperintensities, core analyses offer a much more restricted number of voxels that can be included in the analysis. Therefore, we followed up our whole brain analysis with an ROI analysis to take a more targeted approach to increase our sensitivity to detect an effect in the left middle temporal region.

2.4.2 The presence of core in the MTG region predicts greater impairment in semantic naming

Using the pre-selected region of interest identified as critical in a prior stroke lesion study using Boston Picture Naming (Mah, Husain, Rees, and Nachev, 2014), we first confirmed that this ROI causally correlates with semantic naming in the brain tumor model. This was done by comparing semantic naming scores of subjects who had tumor lesions in the ROI versus subjects who did not. Unsurprisingly, subjects who have lesions overlapping the ROI show significantly lower semantic naming scores than subjects who have lesions elsewhere in the brain ($W = 852$, $p = 3e\text{-}4$; fig 6). This first analysis confirms that the MT region of the brain is causally involved in facilitating the semantic naming function, given that presence of lesions in this area is associated with impaired behavioral performance.

Figure 6. *Participant scores on the semantic naming task based on whether their lesion overlaps the ROI.*



Note. Tumor patients with lesions in the ROI demonstrate a significantly lower semantic naming score ($\underline{X}_{\text{Overlap}} = 3.28 \pm 1.09$) than patients with lesions elsewhere in the brain ($\underline{X}_{\text{No Overlap}} = 3.90 \pm 0.19$)

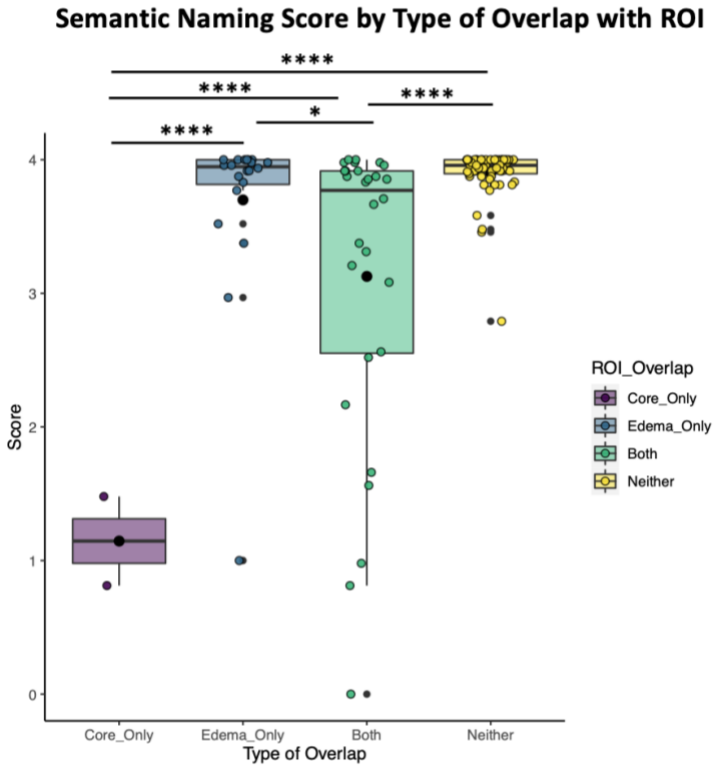
Following this high-level analysis, we next sought to identify whether the presence of different glioma imaging sub-regions (FLAIR hyperintensity and core) within this ROI contributed differentially to the observed behavioral impairments. Given previous literature showing that there are a greater number of necrotic cells and fewer viable tumor cells in the core than in the edema (Eidel et al., 2017), we hypothesized that the subjects whose core is present in the ROI would show more impaired semantic naming performance compared to subjects who do not have a core overlapping the ROI. Therefore, a one-way ANOVA test comparing the group means of the semantic naming score across different types of ROI overlap was conducted to test this hypothesis.

While there were only 2 subjects that had the core overlapping the ROI without the FLAIR hyperintensity also overlapping the FLAIR hyperintensity (see Fig. 7), the importance of the tumor core within the ROI as compared to FLAIR hyperintensity in predicting the cognitive deficit could be inferred from the comparing the behavioral score from the group of subjects who had both core and FLAIR hyperintensity overlap the ROI with those from the group of subjects who only had their FLAIR hyperintensity overlap the ROI. The results of the ANOVA revealed that there were indeed significant differences in semantic naming scores across subjects who had different types of overlap with the ROI ($F(3,106)=16.41$, $p<0.0001$; fig. 7). Specifically, the post-hoc Tukey's test revealed the following significant pairwise differences: 1.) Subjects with core but not FLAIR overlapping the ROI ($n=2$) demonstrate significantly impaired semantic naming performance compared to all other groups (see table 2 for each statistic). However, the importance of the core in predicting a behavioral impairment was further highlighted by the finding that 2.) subjects with both core and FLAIR hyperintensity in the ROI show greater impaired semantic naming function than subjects with only the FLAIR hyperintensity in the ROI ($M_{difference} = -0.57, 95\% CI_{adjusted} = [-0.06, -1.09], p_{adjusted} = 0.023$) as well as subjects with lesions outside of the ROI ($M_{difference} = -0.77, 95\% CI_{adjusted} = [-1.17, -0.37], p_{adjusted} < 0.0001$). Particularly, there was no difference in semantic naming performance between those with FLAIR hyperintensity in the ROI compared to those with no lesions at all in the ROI. These results suggest that the presence of a core in a critical brain region seems to be the primary driving factor in disrupting cognitive function of a brain region.

Table 2. *Result of post-hoc Tukey's pairwise test comparing the mean semantic naming score between different overlap types.*

Group	Control Group	Lower Limit	Difference	Upper Limit	P Value
Core	FLAIR	-3.854	-2.553	-1.253	8.124E-6
Core	Both	-3.265	-1.981	-0.697	0.0006
Core	None	-4.012	-2.751	-1.489	6.931E-7
FLAIR	Both	0.059	0.572	1.086	0.023
FLAIR	None	-0.652	-0.197	0.258	0.671
Both	None	-1.173	-0.770	-0.366	1.506E-5

Figure 7. *Participant scores on semantic naming task based on the different types of overlap the participant has with the ROI (core only, FLAIR hyperintensity only, both, and no overlap).*



Note. None of the subjects shows only core overlap with the ROI. Participants with both core and FLAIR hyperintensity present in the ROI show significantly lower semantic naming scores than those with only FLAIR hyperintensity present in the ROI or those with lesions outside of the ROI.

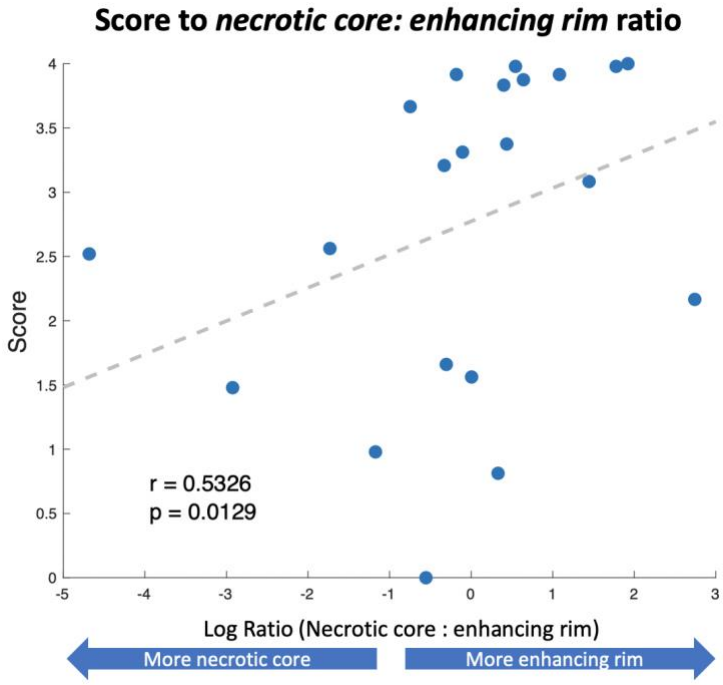
Prior work suggests that the total lesion volume serves as the covariate for observed behavioral deficit. And our ANOVA showed that those with the core present within the critical ROI showed significantly greater impaired behavioral performance compared with subjects without the core. Therefore, we next wanted to understand whether lesion volume is a confounding variable, i.e., larger regions are more likely to be present in any given ROI. We performed ANCOVA to test for the difference in semantic naming performance across different lesion overlap types while controlling for the total lesion volume of each subject. The analysis showed that the covariate, total lesion volume, was not significant ($F(1,115) = 0.814$, $p = 0.369$), suggesting that the total lesion volume does not influence the association between the different lesion ROI overlap types and the semantic naming scores.

2.4.3 Greater proportion of necrotic core within the tumor core predicts worse performance in semantic naming performance.

The extent to which diffuse gliomas differentially remodel neural networks thereby resulting in behavioral impairments remains poorly understood. In fact, when VLSM studies apply brain tumor models, prior investigations (Herbet and Duffau, 2022) commonly apply resection cavity volumes to address this gap in knowledge. Our analysis demonstrates that the presence of diffuse glioma infiltration in the left temporal lobe correlates with semantic naming impairments and the extent of impairments correlates most closely with T1 contrast-enhanced core. We then wanted to address whether further MR tissue classification within the core could better explain the behavioral impairment or whether the tumor core as a whole contributed equally to the observed behavioral damage. Therefore, we partitioned regions within the T1 post-necrotic tumor core and T1 post-enhancing tumor rim (see Fig. 2 for reference). These partitioned regions were then used to calculate a ratio of how much the necrotic core overlaps

with the same MTG ROI to how much the enhancing rim overlap with the MTG ROI (eq 1.). Using this ratio, we found that the overlap ratio positively correlates with the semantic naming performance ($r = 0.62$, $p < 0.01$, see figure 8): the larger, positive ratio denotes more enhancing rim overlap, and the negative, smaller ratio denotes more necrotic core overlap. This correlation suggests that when there is a greater proportion of necrotic core overlapping with the MTG ROI as compared to its enhancing rim, the more impaired subject's semantic naming performance. These data suggest that sub-regions within the T1 post-enhancing glioma core have a greater impact on semantic naming.

Figure 8. *The correlation between the log ratio of the overlap (volume of enhancing rim overlapping with MTG ROI : the volume of necrotic core overlapping with MTG ROI) and the semantic naming score.*



2.5 Discussion

Efforts to uncover the neural substrate causally responsible for cognitive and behavioral impairments have offered both mechanistic and clinical insights. In this study, we sought to

demonstrate the validity of lesion mapping in adult glioma patients and used VLSM and ROI analyses to understand how lesion heterogeneity is causally responsible for semantic naming abilities. Here, we challenged the underlying assumption that all lesioned areas contribute equally to behavioral impairments. To do this, we specified different sub-regions within the tumor and investigated their respective correlation with the magnitude of the behavioral impairments.

The first goal of this study was to determine the cortical and subcortical regions responsible for semantic naming using a diffuse glioma disease model. Through the whole brain analysis using SVR-VLSM and the more targeted region of interest analysis, we replicated findings from studies that used strokes as a lesion model, in which left temporal lobe lesions elicited significantly impaired performance in semantic naming tasks. Our findings further demonstrate the role of the left temporal lobe in chronic disease models including stroke, epilepsy, traumatic brain injury, and brain tumors (Baldo et al., 2013; Hillis et al., 2001; Schwartz et al., 2009; Walker et al., 2011, Binder et al., 2020; Snyder et al., 2023). On the other hand, while a number of prior fMRI studies also report the causal involvement of the fronto-temporal regions (see for review Vigneau et al., 2006; Hoffman & Morcom, 2018), we did not find any involvement of this region in semantic naming performance. This may be due to differences in lesion coverage and inclusion criteria across studies (Aabedi et al., 2021).

Recent evidence in the biology of diffuse gliomas has demonstrated gliomas synaptically integrate into neural circuits. Bidirectional interactions exist between neurons and glioma cells, with neuronal activity driving glioma growth and gliomas increasing neuronal excitability. (Aabedi et al., 2021; Krishna et al., 2023). Gliomas are by definition heterogenous with some regions of dense tumor infiltration and other regions with spare tumor. Therefore, within the

confines of glioma, the variable degrees of tumor infiltration are clinically delineated by MRI T1 post contrast enhancing and FLAIR regions. Using brain regions identified in prior stroke lesion work to be involved in the picture naming function (Mah et al., 2014), the result of our region-of-interest analysis showed that subjects with both core and FLAIR in the ROI show significantly lower semantic naming score than the subjects with only the FLAIR in the ROI or the subjects with lesions elsewhere that didn't overlap with the ROI. Our findings suggest that the presence of the tumor core in a brain region confers a greater impact on the associated behavioral impairment than the presence of FLAIR in the same region. While we had hypothesized that the location and the presence of the core will have greater contributions to semantic naming impairment, we were surprised to find that the presence of FLAIR alone in the ROI had no significant impact on the behavioral performance compared to those who didn't have any lesion present in the ROI. In other words, these results indicate that the location of the tumor core was the primary driving factor in establishing the lesion-symptom relationship in a glioma model. It is likely that within FLAIR regions, the extent of glioma infiltration is below the threshold required to identify a relationship between lesioned voxels and semantic naming errors. Alternatively, the deficits that are experienced by patients may be more subtle than can be captured with the task used in this study.

Our data support a model of glioma impairment in which the tumor core more strongly impacts behavior. This model has important methodological implications for future research using the tumor as a lesion model. Traditionally, tumor lesion studies have made no distinction between these two sub-regions of the lesion: by combining the core and FLAIR regions into a single mask, researchers assumed that the different sub-regions contribute equally to behavioral impairment. Given our findings that these regions differentially impair cognitive function, the

single lesion mask approach would result in lower sensitivity in identifying significant regions contributing to behavioral impairment in the analyses. For example, let's say voxel A is necessary for maintaining a behavior such that significant growth of tumor cells in this area would lead to a behavioral impairment. Now let's say that all the subjects in the analysis only had FLAIR regions inside this voxel. Based on our findings, it's likely that the lesion mapping analysis would fail to reveal a significant role of this voxel in the behavior in this case. Given the relative importance of core regions over FLAIR regions in predicting the magnitude of the deficit, it's advantageous to distinguish these regions so that the unexplained variance can be absorbed by the tumor sub-region variable.

These data provide evidence that glioma sub-regions vary not only by histopathological properties but also by behavioral outcome and manifestation. By employing a similar analytical approach to stroke lesions, researchers may gain valuable insights into the relationship between different types of neuronal damage and the observed behavioral or cognitive deficits. Consequently, stroke studies can potentially capitalize on the classification of stroke subtypes or lesion homogeneity as additional parameters to evaluate whether they better explain the observed behavioral deficits in stroke patients. In a 2019 study, Boss et al. found that the lesion homogeneity in acute ischemic stroke measured by diffusion-weighted imaging (DWI) is a strong predictor of a patient's degree of disability 3 months out. The authors suggest that the observed signal intensity through DWI used to determine the lesion homogeneity (which they then used as a binary measure) may be a reflection of the level of viable tissues following the lesion. Such findings offer a similar perspective to our study by demonstrating that lesion studies can benefit from the consideration of these additional parameters to capture how much functionality is remaining in the lesioned brain areas.

In our whole brain SVR lesion mapping analyses, we found that the core and the total lesion masks demonstrated significant results in a largely overlapping area of the left middle temporal lobe. While the analysis of the FLAIR data identified similar regions, the identified regions were much more limited in coverage. We believe this difference may be attributed to the non-orthogonal relationship between the core and FLAIR SVR-LSM analyses. Although our FLAIR hyperintensity analyses excluded core masks as inputs, most FLAIR regions present concurrently with a core; FLAIR typically encircles the tumor core, making the two masks within a single subject highly spatially correlated. Therefore, if FLAIR occurs in close proximity to a critical region (responsible for cognitive function), it's also highly probable that a core is also present in that critical region. This means that the significant regions found in the FLAIR analyses can be largely, if not entirely, driven by the effect of the core which is located nearby. While the reverse could hold true in which FLAIR hyperintensity is driving the effect found in the core analysis, we don't believe this is likely given our results. First, our ROI analysis shows a strong effect of the core location on the behavioral score compared to the minimal effect of the edema location. Second, despite the larger coverage of FLAIR hyperintensity in the brain compared to the core (see Fig. 3) with respect to both the voxels that meet the inclusion criteria and the number of lesioned subjects in each voxel, SVR-LSM results show a more limited causal region in the semantic naming task. Taken together, our results point to the conclusion that core locations are a better predictor of behavioral impairment than consideration of FLAIR hyperintensities or entire lesions. This distinction can be used to create a more precise and sensitive mapping of brain structure to function in tumor patients.

The findings from the study have important clinical implications for the management of patients with diffuse gliomas. Since 2021, the World Health Organization has sub-classified

diffuse gliomas according to both histological and molecular features (Hervey-Jumper et al., 2023; Molinaro et al., 2020). This emerging data demonstrate marked differences in survival outcomes across diffuse glioma subtypes. However, despite these differences, removal of T1 post contrast enhancing and FLAIR regions offers a survival benefit to patients. Therefore, an understanding of the capacity of glioma sub-regions to participate in neural computations is essential. These data suggest that the glioma core remains the most critical factor in determining the degree of impairment. Therefore, treatment approaches such as functional brain mapping should be employed, particularly within nonenhancing FLAIR regions.

Like all lesion and clinical studies, our study faces a number of limitations. First, lesion mapping analyses require a broad distribution of lesions to understand where damage either does or does not impair behavioral performance. The group of included subjects in our study only had a few lesions in the most anterior and posterior regions of the brain and therefore could not be used to examine the involvement of these regions in maintaining semantic naming function. Given that cores are typically much smaller in size compared to FLAIR regions, the whole brain coverage of core masks was even more limited. However, by following up the whole-brain SVR lesion mapping with the ROI analysis, we were able to conduct a more fine-grained analysis to better understand the effect of each sub-region on behavior.

Tumor lesion studies also face the criticism that compared to stroke lesions, its localization of brain functions may be less accurate. This criticism stems from the possible reorganization of neural functions following the growth of the tumor and the relatively mild nature of deficits observed in tumor patients compared to stroke patients (Anderson et al., 1990; Karnath & Steinbach, 2011). Critics argue that possible reorganization of brain function and the mild deficits observed in tumor patients make it hard to successfully localize the causal region

supporting the cognitive function of interest. However, Duffau (2011) provides a compelling counter argument noting that 1) reorganization of cognitive functions following a tumor is largely limited to low-grade gliomas that exhibit slow growth and 2) even if there were enough reorganization of functions in the brain such that it offsets the cognitive deficits caused by the tumor, there would be no observable behavioral deficits to establish any faulty relationship between brain structure and function. Furthermore, using several cognitive tasks thought to specifically involve the right posterior regions in the brain, Shallice et al. (2010) showed that their selected tasks previously known to involve parieto-occipital areas did indeed show patterns specific to the expected brain regions, differing based on targeted tasks. This finding demonstrates that the behavioral symptoms following the growth of tumors were not broad and generalized. Based on these findings, we argue that tumor lesions also allow valid conclusions to be drawn about the neural bases of cognitive functions and that tumor lesion data can be important to validating and extending the lesion mapping literature.

The second limitation of our study is that QAB scores in semantic naming tasks are negatively skewed with most participants scoring near or at the ceiling (a score of 4). As this measure is designed to identify clinically observable deficits, it may be less sensitive to more subtle, semantic naming impairments, including reaction times. Specifically, QAB scoring penalizes subjects for delayed responses of 3 seconds or longer, whereas minor neurological damage may not necessarily cause such a robust delay. However, given that we were able to replicate the results of prior stroke lesions, we do not believe this to be a major concern in our study.

Finally, our last major limitation is that while semantic naming is a clinically relevant behavior important for facilitating normal language function, it also recruits large regions and

networks in the brain. Future research may benefit from testing the behavioral contribution of the core vs. edema by examining damage to focal sensory regions. Similarly, we also believe that we could gain valuable insights from looking at the tumor as a function of the distance from the center of the tumor mass. Prior work of Ji et al. (2015) shows that microscopy measures using stimulated Raman scattering vary as a function of distance from the necrotic tumor core. Building off this finding, future research could address the questions of whether the impact of the lesion is binary or if it's graded such that the further the lesioned tissue is from the center of the mass, the less impact it has on the behavioral deficit.

To summarize, our data demonstrate that the diffuse glioma lesion work similarly replicates the findings of stroke lesion work and can be used to map the relationship between lesion and behavior. We also show that while lesion mapping traditionally considers all voxels within the lesion to have equal contribution to the behavior, different voxels can contribute disproportionately to the deficit based on which sub-region the voxel comes from. Specifically, our results show that tumor cores play a greater role in explaining the impairment observed in a tumor patient as compared to edema which are captured by FLAIR hyperintensities. Consideration of this distinction within a lesion can provide greater sensitivity in lesion mapping studies to better understand the relationship between brain and behavior.

Chapter 3 Identification of the Locus of Audiovisual Integration Using Intrinsic Brain Tumor Model

3.1 Abstract

The ability to understand spoken language is essential for social, vocational, and emotional health, but can be disrupted by environmental noise, injury, or hearing loss. These auditory deficits are often ameliorated by visual speech signals that convey redundant or supplemental speech information, but the neural mechanisms critically responsible for these audiovisual interactions remain poorly understood. Previous transcranial magnetic stimulation and lesion-mapping studies suggest that the left posterior superior temporal sulcus (pSTS) is causally implicated in the generation of the McGurk effect, an audiovisual illusion in which auditory and visual speech is perceptually 'fused'. However, research also suggests that the McGurk effect is neurally and behaviorally dissociable from other visual influences on speech perception and, therefore, may not provide a generalizable index of audiovisual interactions in speech perception more broadly. To examine which brain regions are critically responsible for audiovisual speech integration behaviors more generally, we measured the strength of the McGurk effect, congruent audiovisual performance, and incongruent audiovisual performance in patients with intrinsic brain tumors (n = 182). Region of interest analyses were conducted to examine the effect of a tumor within the left pSTS on audiovisual speech behaviors. Results identified a weak pattern of evidence demonstrating that a tumor within the pSTS selectively reduces the McGurk effect. We discuss the weak nature of this effect in the context of the dataset

used and provide suggestions for future research using brain tumors as a model to understand audiovisual speech perception.

3.2 Introduction

Disruptions in speech perception can occur through various means, from temporary disruptions such as being in a noisy environment to more serious, permanent disruptions such as damaging one's ears or the brain or suffering from age-related hearing decline. Regardless of the precise cause of the disruption, one important mechanism through which listeners can recover missing auditory speech information is through relying on visual speech information (Grant & Seitz, 2000; Sumby & Pollack, 1954). Given the importance of speech perception in everyday function and our natural tendency to depend on visual information when the auditory information is compromised, it is vital to understand how this exchange of information is facilitated in the brain. Yet, the neural mechanisms and brain regions that are causally involved in visual facilitation of speech perception are still poorly understood.

To study visual facilitation of speech perception, many researchers relied on the use of McGurk illusion, where an auditory phoneme is paired with the visual lip movements of a different phoneme to create the illusion of a fused phoneme (e.g., presenting auditory /ba/ with visual /ga/ tends to produce the percept /da/). Given the convincing nature of the illusion demonstrating that visual information has a strong effect on auditory perception to the point where perception could be entirely altered, McGurk processing became a widespread measure of audiovisual integration. Using McGurk stimuli, numerous studies found that left posterior superior temporal sulcus is the critical region allowing the integration of audiovisual information (Sekiyama et al., 2003; Bernstein et al., 2008; Benoit et al., 2010; Irwin et al., 2011; Nath et al., 2011; Szycik et al., 2012). A particularly strong evidence comes from a TMS study by

Beauchamp et al. (2010) in which the authors showed that electrical stimulation of the left pSTS, but not a control site, reduces the subject's likelihood of experiencing the McGurk illusion. Similarly, studies of stroke patients report that patients with lesions in this brain region show reduced susceptibility to McGurk percepts (Beauchamp et al., 2010; Hickok et al., 2018). Many studies, therefore, concluded that pSTS maintained healthy audiovisual speech integration.

However, even though McGurk effect does, in fact, demonstrate a strong influence of visual speech on auditory speech processing, researchers have started to raise concerns that focusing solely on McGurk effect may prevent us from fully understanding numerous facets of audiovisual integration process (Alsius et al., 2018; Ganesan et al., 2020; Hickok et al., 2018; Van Engen et al., 2017). For example, if McGurk perception requires listeners to process incongruent auditory and visual information, whereas normal everyday speech requires processing of congruent auditory and visual information. Therefore, McGurk perception may not truly reflect a general, everyday audiovisual speech processing. Consequently, the findings from research using McGurk stimuli may not generalize to natural audiovisual integration process. With increasing criticism of the widespread use of McGurk stimuli, it becomes necessary to make a clear distinction between the modulatory role that visual information plays in McGurk effect and the facilitatory role that visual information plays in congruent audiovisual speech. That is, McGurk processing relies on the ability of visual information to *change* the percept of auditory information while the congruent audiovisual processing focuses on the ability of visual information to *enhance* the information in the auditory domain.

If McGurk processing relies on the same integration process utilized by congruent audiovisual processing that's used to improve a listener's speech comprehension, we expect an individual's ability to understand audiovisual speech in noisy environments to be highly

correlated with the individual's susceptibility to McGurk effect. Yet, in a 2017 study by Van Engen et al., the authors found no significant relationship between the two measures, suggesting that there are unshared mechanisms unique to each of these processes. Similarly, Hickok et al. (2018) replicated this finding that there are no significant correlations between the amount of advantage an individual experiences from congruent visual information and their susceptibility to McGurk effect. Furthermore, studies show that listeners experience McGurk effect only when they're aware that they are listening to speech (Eskelund et al., 2011; Palmer & Ramsey, 2012; Tuomainen et al., 2005). Audiovisual speech facilitation, however, occurs even when the listener is not aware that they are listening explicitly to speech information (Faivre et al., 2014; Palmer & Ramsey, 2012; Plass et al., 2014).

In conjunction with these behavior findings pointing towards a dissociative mechanism between the two process, electrophysiological studies suggest that the two processes rely on distinct neural mechanisms. Using EEG, researchers report that McGurk processing involves different neural responses compared to that of congruent audiovisual processing. For example, Roa Romero et al. (2015) report observation of reduced N1 component as well as greater suppression of early and late beta-band power in McGurk processing compared to congruent audiovisual speech processing. This is likely attributed to the stronger integration process required for McGurk perception and the additional step needed in the cognitive process to resolve the audiovisual conflict. Similar reports by other studies denote different patterns of neurophysiological responses that distinguish the neural processes required to process congruent versus incongruent McGurk stimuli (Arnal et al., 2009; Fingelkurts et al., 2003; Lange et al. 2013). Importantly, evidence suggests that congruent audiovisual process rely largely on direct feedforward connection between visual to auditory areas, while incongruent audiovisual

processing, as in the case of McGurk, rely more heavily on a secondary feedback from the higher order STS region (Arnal et al., 2009; Arnal et al., 2011).

Given the growing converging behavioral and electrophysiological evidence that audiovisual facilitation and McGurk perception are two dissociated processes, it naturally follows that the two processes may also be causally supported by different brain regions. While there exists strong causal evidence that the left pSTS is critical for maintaining the McGurk illusion (Beauchamp et al., 2011; Hickok et al., 2018), it remains unclear whether this same area is necessary for experiencing the perceptual benefits of congruent audiovisual speech. A precise cortical mapping of the two processes can better elucidate the neural mechanisms that underlie each of these processes and is therefore, particularly important for expanding our understanding of different facets of audiovisual speech perception. In line with this goal, we sought to examine whether the left pSTS is causally required for the broad, general audiovisual speech processing, or instead, is more specific to the generation of the McGurk percept.

To address this gap in knowledge, we collected data from intrinsic brain tumor patients to examine brain areas causally linked to different measures of audiovisual speech perception. Specifically, we aimed to use a support vector regression voxel lesion symptom mapping (SVR-VLSM; Zhang et al., 2014), the approach used in the previous study. As a brief review of the method, SVR-VLSM is a multivariate version of the traditional univariate VLSM approach (Bates et al., 2003), which models the relationship between behavioral symptoms and subjects' lesions at the voxel level. The traditional univariate approach investigates the involvement of each voxel one by one, disregarding a given voxel's spatial dependence with neighboring voxels. However, the multivariate approach considers the relationship between all voxels

simultaneously, providing a more precise estimate of which voxels are critical for maintaining a behavioral function (Zhang et al., 2014; DeMarco & Turkeltaub, 2018).

Using this multivariate approach, we hoped to expand upon an earlier large-scale stroke lesion study by Hickok et al. (2018). This prior work demonstrated that 1.) patients with lesions in the posterior superior temporal lobe (gyrus and sulcus) exhibited a disrupted McGurk effect. And 2.) there was no significant correlation between audiovisual facilitation and McGurk processing measures, adding to the growing body of evidence that the two processes are distinct and dissociable.

Building upon these findings, we recruited a large sample of intrinsic brain tumor patients with glioma (n = 182) and collected multiple measures of audiovisual processing, including the congruity effect of auditory and visual stimuli (referred to as audiovisual advantage in Hickok et al., 2018), the incongruity effect of auditory and visual stimuli (reported in the methods section of Hickok et al., 2018 but excluded from the lesion mapping results), and McGurk fusion frequency. Using this method, we aimed to identify not only the brain regions associated with McGurk perception but also the regions associated with processing congruent audiovisual information and general incongruent audiovisual information (non-McGurk specific conflict effect) to better understand whether these processes rely on the same neural correlates.

3.3 Methods

3.3.1 Subjects

The study included 182 subjects (108 male, $\overline{age}$ = 52.8) with a diagnosis of diffuse glioma. Subjects consisted of patients undergoing tumor treatment who were scheduled to undergo a tumor resection surgery at the University of California San Francisco between 2017 - 2022. Patients participated as volunteers during the administration of other clinical procedures

and received no financial compensation. Similar to Study 1 focusing on the semantic naming function, we did not restrict subjects based on the location of their tumor. This broad inclusion criterion was to maximize the coverage of the lesions in the brain to be able to draw whole-brain inferences.

3.3.2 Audiovisual Speech Paradigm

The subjects were given the audiovisual speech task in a clinical setting at the University of California San Francisco one or two days prior to their tumor resection surgery. The patients were seated upright in a chair and the task was performed on a laptop using PsychToolbox (Brainard, 1997; Pelli, 1997). The patients used a Logitech F310 gamepad to provide their responses.

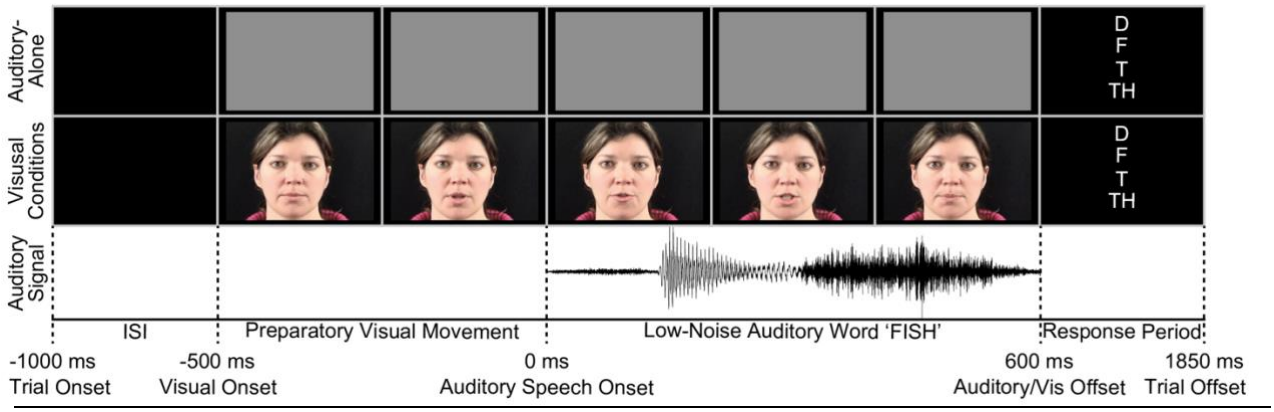
Our audiovisual speech task used the audio and movie stimuli adapted from a prior study by Ross et al. (2007). From the set, we selected 40 single syllable words starting with ‘b’, ‘f’, ‘g’, and ‘d’ and the second sound of the stimuli balanced across the 4 consonants. The movies were cut and adjusted such that they were 1100 ms in length and the audio onset was at 500 ms.

The task consisted of 3 visual conditions: no visual, congruent audiovisual, and incongruent audiovisual. In incongruent audiovisual condition, there was a mismatch in the initial consonant between the auditory word and the visual word that was played. For example, a trial might consist of speaker saying the auditory “bear” played with the video of the speaker mouthing “gear”. The task also consisted of two noise levels: low and high. Low noise level did not have any additional noise applied beyond the way the audio was recorded. This yielded SNR of 32.2 dB SPL. For high noise level, we added pink noise to the audio bringing the SNR of the signal to -6.4 dB SPL. In total, the task included a total of 120 trials with 40 trials per each of the 3 visual conditions, half of which were low noise and half were high noise trials. The 4

consonants were similarly divided across the visual conditions and noise levels. The audiovisual pairings occurred in equal frequencies across all conditions. This meant that half of the incongruent audiovisual trials consisted of combination that tend to elicit McGurk fusion responses and half that do not.

For each trial, the subjects were presented with a black screen, followed by either a gray block (in no visual condition) or a video of the preparatory lip movements (Figure 9). The audio started 500 ms after the onset of this gray block or video. Following the audio offset, the subjects were prompted to identify the starting letter of the word that they heard using their gamepad after being provided with 4 response options on the screen. Subjects were asked to choose the option that sounded closest to what they heard if they were unsure. They were informed that the word that they heard may not be real words. The 4 response options always included the initial consonant of the spoken word, the initial consonant of the viseme, as well as the two common McGurk fusion percepts. These four response options remained consistent across all stimuli and conditions, even if some options were not relevant to certain conditions (e.g., the 'viseme' option was still present in the auditory-alone condition). The subjects had 1.25 seconds to respond, after which the task would automatically advance to the next trial. Adapting to participants' performance, the maximum wait time after a missed trial increased by 0.25 seconds, up to a maximum of 2 seconds. On the other hand, if participants successfully responded before the maximum wait time for ten consecutive trials, this wait time would decrease in 0.25-second intervals until it reached a minimum value of 1.25 seconds.

Figure 9 *Trial schematic with the auditory and visual stimuli for the word 'fish'.*



Note. Each trial began with a black screen lasting between 375 to 625 m (uniformly randomized). 500 ms before the auditory onset, either a video of preparatory visual movement or a gray block was presented. No sound was played during this time until the auditory speech onset. After the presentation of the auditory stimuli, 4 response choices were displayed on the screen.

3.3.3 Magnetic Resonance Imaging and Lesion Delineation

The MRI protocol, lesion delineation, and multivariate VLSM approach used in the current study follow the same procedure as our previous lesion mapping study focusing on semantic naming. Short summaries of these methods are reiterated here.

All participants underwent magnetic resonance imaging in a 3 Tesla scanner within 3 days of completing the audiovisual task. The following MR sequences were obtained as part of the subjects' standard clinical care: pre-operative T1-weighted images (T1), post gadolinium enhanced T1-weighted images (T1c), and T2-weighted fluid attenuation inversion recovery (FLAIR).

The lesion delineation was performed either manually using 3D Slicer (http://slicer.org) or automatically using DeepBraTumIA (https://www.nitrc.org/projects/deepbratumia/). Contrast-enhanced T1 images were used to define the borders of the contrast-enhancing tumors and FLAIR images were used to define the borders of the non-enhancing tumor. These two masks

were combined to create a total lesion mask. For both manual and automatic lesion delineation approaches, all masks underwent a quality check process after the original mask had been generated. Any mistakes in the lesion boundaries were fixed at this stage.

To allow for comparisons across all subjects, the subjects' lesion masks generated through manual means were normalized to the standardized Montreal Neurological Institute (MNI152) template using the transformation matrix based on their anatomical T1-weighted images via SPM12 (http://www.fil.ion.ucl.ac.uk/spm/). Automatically segmented masks were spatially normalized to MNI space as a part of the auto-segmentation process in DeepBraTumIA.

3.3.4 Voxel Lesion Symptom Mapping

We aimed to use the same SVR-VLSM applied to the semantic naming VLSM study (Study 1) to locate the significant regions necessary for supporting audiovisual processing. However, as will be further discussed in the results section, we did not conduct any VLSM due to the limited coverage afforded by our subject pool.

3.3.5 Behavioral Measures

The following behavioral measures were calculated based on subjects' task performance: 1.) auditory accuracy, which captures the baseline auditory perception. 2.) congruent audiovisual accuracy, which captures the behavioral enhancement experienced by listeners from having congruent visual stimulus 3.) incongruent audiovisual accuracy, which captures the cognitive cost experienced by having to process incongruent visual stimuli 4.) frequency of reported McGurk fusion responses, which captures individual's susceptibility to McGurk illusion.

Using these behavioral measures, we hoped to understand which voxels are causally implicated in the facilitation of healthy audiovisual processing and integration. Auditory

accuracy was calculated as the proportion of trials in which participants reported correct responses in the auditory-only trials. Congruent audiovisual accuracy was calculated as the proportion of trials in which participants reported correct responses in the congruent audiovisual trials. Incongruent audiovisual accuracy was calculated as the proportion of trials in which participants reported correct responses in the incongruent audiovisual trials to quantify how much speech perception is impaired by the presence of incongruent visual stimuli. Finally, the frequency of McGurk fusion will measure how frequently people report a fusion response (the combined perception of the different auditory and visual information) when presented with McGurk audiovisual stimuli.

3.3.6 Region-of-Interest (ROI) Analysis

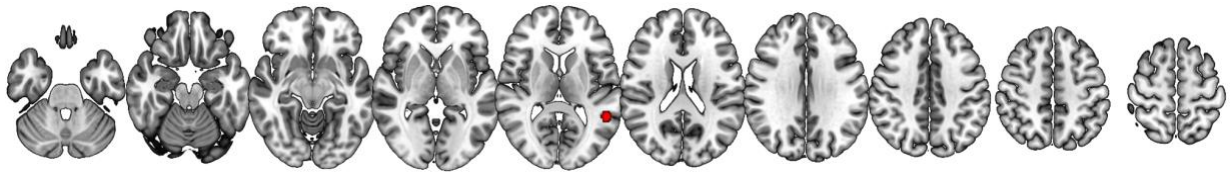
Based on prior literature demonstrating that the left posterior superior temporal sulcus plays a critical role in the experience of the McGurk effect (Beauchamp et al., 2010; Erickson et al., 2014; Hickok et al., 2018), we additionally conducted an ROI analysis of lesions to the pSTS. Our ROI was generated by taking the center of mass of the bank of the left hemisphere superior temporal sulcus, denoted by the Freesurfer's automated cortical parcellation (aparc+aseg) as 'lh-bankssts'. The mask for the ROI was generated by creating a sphere centered around this coordinate. In selecting the ROI size, we found that ROI with a radius of 15 mm tightly encompassed the lh_bankssts (left hemisphere banks of superior temporal sulcus). Therefore, we selected 20 mm radius as our target ROI size. By selecting the larger ROI that expands a little beyond the specified boundaries of the pSTS, we imposed more liberal inclusion criteria to account for spatial smoothing that occurs during normalization of subjects' masks. Following the main ROI analyses using 20 mm radius ROI, we then sought to better understand the dependence of our findings on the size of ROI. To do so, we created 8 differently sized ROIs by varying the

radii of the sphere from 5 mm to 40 mm in increments of 5 mm. The image of these ROIs overlaid on top of the MNI brain is illustrated in Figure 10.

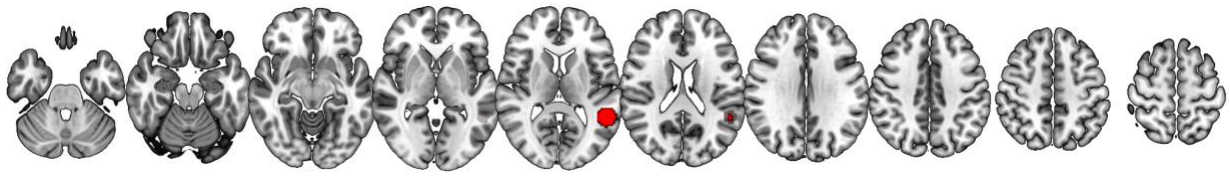
The goal of this ROI analysis was to investigate the role of left pSTS in finer detail to understand whether the various behavioral measures of audiovisual processing are affected by the lesion damage in the left pSTS. To this end, the measures of audiovisual facilitation effect, audiovisual conflict effect, and fusion response rates of participants whose tumor lesions did overlap with the left pSTS ROI were compared to the measures of participants whose tumor lesions did not fall within the pSTS.

Figure 10. *Cross-sections showing the location of the spherical ROIs centered at MNI coordinates (-53, -48, 10) on standard-space Montreal Neurological Institute template (MNI152) ranging from 5mm in radius to 40mm in radius.*

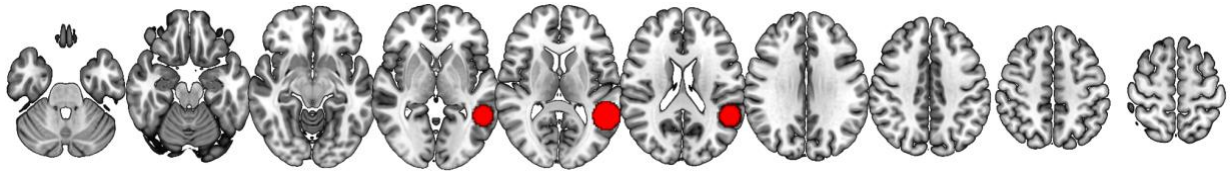
ROI with 5 mm radius



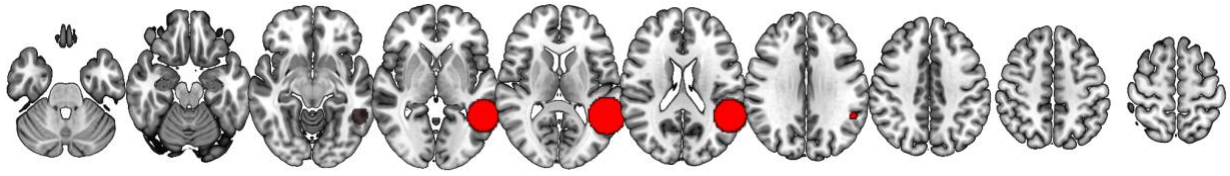
ROI with 10 mm radius



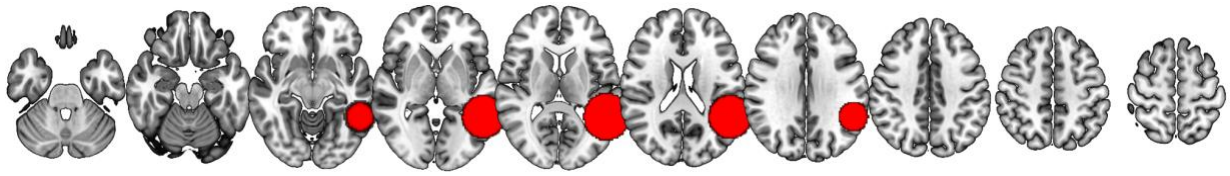
ROI with 15 mm radius



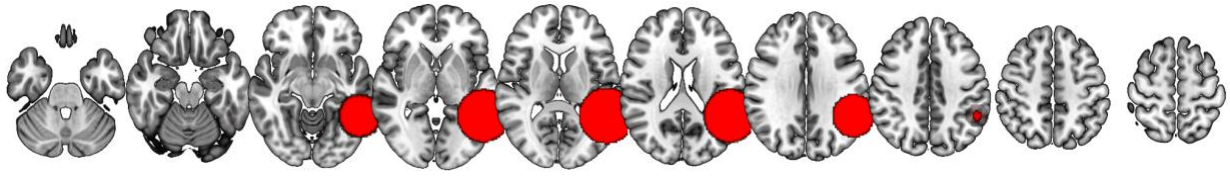
ROI with 20 mm radius



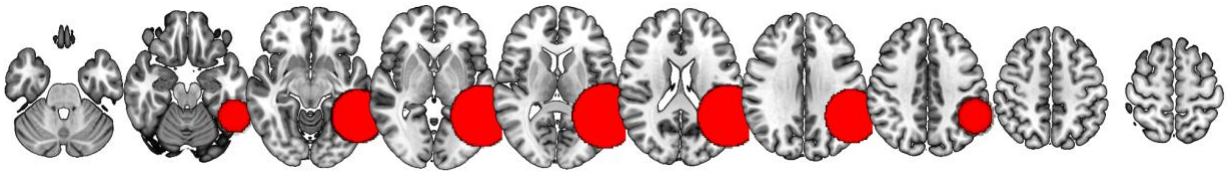
ROI with 25 mm radius



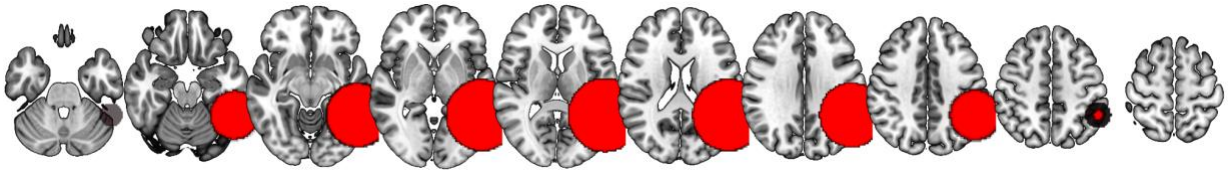
ROI with 30 mm radius



ROI with 35 mm radius



ROI with 40 mm radius



Note. Left and right are flipped

3.4 Results

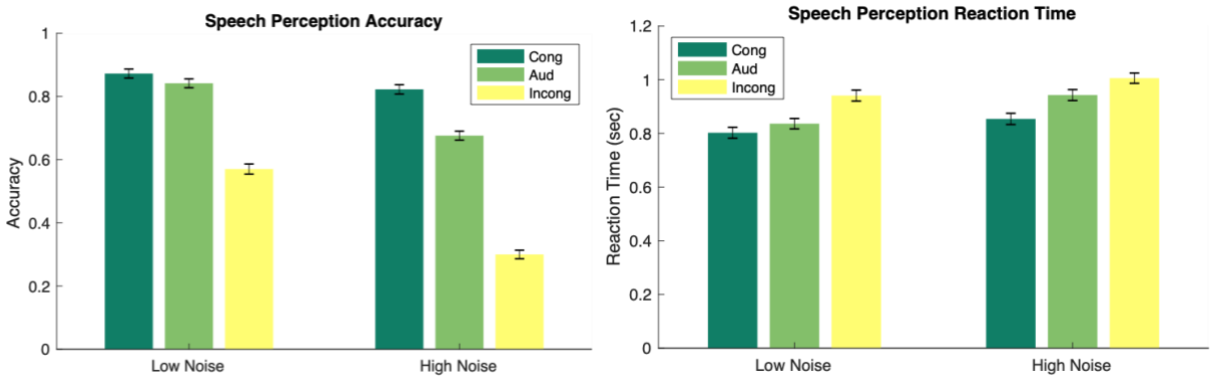
3.4.1 Behavioral Data

Without consideration for specific lesion locations, we first examined the subjects' average behavioral performance across the various audiovisual measures to ensure the validity of our task. Behavioral performances from all included subjects are shown in Figure 11. Accuracy and RT data show that high levels of noise reduced auditory accuracy but also resulted in greater usage of visual information (leading to improved performance when the visual signal was congruent and reduced performance when incongruent). Accuracy data (Figure 11a) revealed main effects of visual-type [$F(1.3,234.69) = 750.11$, $p = 4.94E\text{-}85$, $\eta_p^2 = .453$], noise-level [$F(1,181) = 492.79$, $p = 1.53E\text{-}53$, $\eta_p^2 = .147$], and a significant interaction between the two [$F(1.84,332.65) = 163.69$, $p = 1.96E\text{-}47$, $\eta_p^2 = .051$]. RT data (Figure 11b) mirror those of the accuracy data, with main effects of visual-type [$F(1.90,343.87) = 164.87$, $p = 3.43E\text{-}49$, $\eta_p^2 = 0.046$], noise-level [$F(1,181) = 150.762$, $p = 1.34E\text{-}25$, $\eta_p^2 = 0.018$], and a significant interaction between the two [$F(1.88,339.66) = 11.054$, $p = 3.53E\text{-}5$, $\eta_p^2 = .002$]. Results were inconsistent with a speed-accuracy tradeoff. For example, under high noise, congruent visual information

both improved accuracy [$t(361.59) = 7.16$, $p = 4.42E\text{-}12$, $d = 0.751$] and sped responses [$t(361.78) = 3.04$, $p = .0025$, $d = 0.3189$] relative to the auditory-alone condition. Consistent with past research (Van Engen et al., 2017), we did not observe a negative correlation between high noise congruent and incongruent audiovisual trials across participants which would have indicated the two multisensory effects (higher accuracy in congruent trials and lower accuracy in incongruent trials) reflect a shared mechanism. Moreover, the correlation we observed was significantly positive [$r_s = 0.264$, $p = 3.12E\text{-}4$] indicating that when subjects performed better on the congruent trials, they also performed better on incongruent trials, potentially reflecting general performance differences across subjects.

Figure 11. *Behavioral data across all subjects.*

A) **B)**



Note. **(A)** Accuracy in the main experimental conditions from the audiovisual speech paradigm (congruent audiovisual, auditory-alone, incongruent audiovisual in low and high levels of auditory noise). **(B)** RT data for the same conditions presented in panel a. Error bars reflect SEM. Lower accuracy in the incongruent audiovisual condition is consistent with greater McGurk illusion incidence

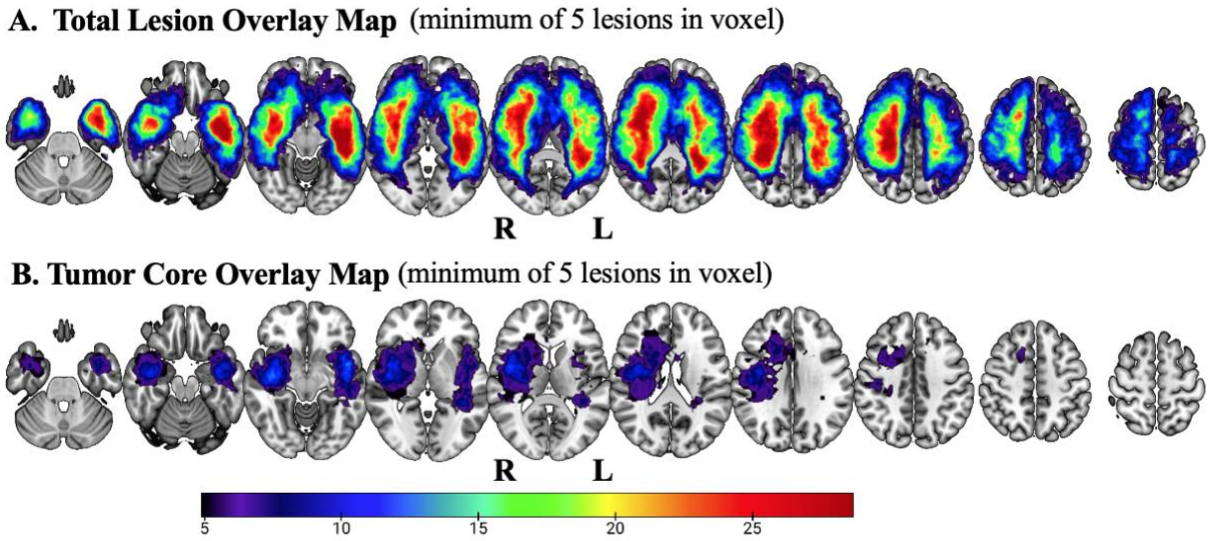
Finally, we examined our results as a function of word frequency. More frequently used words are perceived faster and more efficiently than less frequently used words (or even pseudo words). Therefore, we wanted to investigate whether certain audiovisual stimuli combinations resulted in greater McGurk fusion responses due to higher word frequency of the intended fused percept. To test this, we used a previously acquired behavioral dataset from healthy, college-aged

adults (n = 93) using the identical task used in the present study. $SUBTL_{WF}$ measure, from the SUBTLEX-US dataset (Brysbaert and New, 2009), was used as the word frequency metric. This metric specifically measures the word frequency per million words based on American English subtitles including 51 million words. Using this value, we ran a correlation test between the probability of fusion options reported by subjects against the $SUBTL_W$ value. The test revealed no significant relationship between the frequency of the fused word stimuli used in the English language and the likelihood of subjects reporting that specific fused response (r_s = -0.146, p = 0.369).

3.4.2 Voxel Lesion Symptom Mapping

Our behavioral analyses demonstrated the expected patterns of trial accuracies and RTs, showing the behavioral benefits of congruent visual information and the behavioral costs of incongruent information. We then assessed these behavioral scores in conjunction with their respective lesion masks to conduct SVR-LSM. Voxels with fewer than 5 subjects' lesions were excluded from the analyses (Damasio et al., 2004). This minimum lesion threshold varies across studies and is largely dependent on the trade-off between power and analysis coverage: too low of a threshold can result in poorly powered analysis while too high of a threshold can greatly limit the brain regions where you can study the relationship between the voxel and cognitive behavior. The combined lesion overlay of total lesions and T1-post enhanced cores are shown in Figure 12.

Figure 12. *Overlay maps showing the distribution of lesions.*



Note. The color hue denotes the number of subjects who have lesions in that voxel **(A)** Overlay map of all total lesions across all subjects (n=182). Voxels with less than 5 subjects' lesions overlapping were excluded from the analyses and are not shown in the overlay map. **(B)** Overlay map of all T1 post-enhanced core masks (142 of 182 subjects had core masks). Voxels with less than 5 subjects' lesions overlapping were excluded and are not shown in the overlay map. Note that left and right are flipped.

Our prior study using VLSM to identify the locus of semantic naming function indicated that tumor cores primarily predict the degree of behavioral impairments in the brain tumor models. Therefore, we aimed to focus specifically on the relationship between the location of tumor core to map our behavior rather than the location of the entire lesion in our VLSM analyses. As foreshadowed in the methods sections, we observed sparse coverage of the tumor cores in the left hemisphere (Fig. 12). All lesion mapping studies are subject to the same limitation that it is not possible to draw an inference about the role of voxels in which few lesions occurred. Consequently, our VLSM analyses of the core with various audiovisual measures would have been too poorly powered to detect any significant relationship between voxels in the core and our audiovisual measures. Given the limited spatial distribution of the core masks, we decided to forgo any further analyses using the SVR-LSM approach.

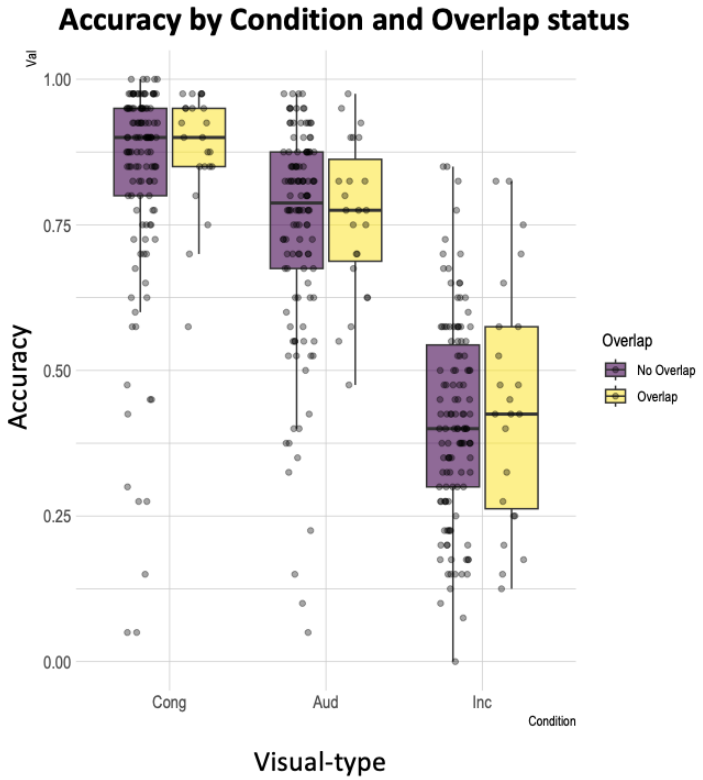
3.4.3 Region-of-Interest Analyses

Given the limitations of our lesion coverage to conduct a whole brain analysis, we continued with a targeted approach to specifically examine the impact of pSTS lesions on audiovisual processes. We hypothesized that those with lesions overlapping the left pSTS ROI would show a significantly reduced McGurk effect. On the other hand, we hypothesized that other audiovisual measures including congruity and the incongruity effect of audiovisual speech would show no or less of a difference between those with and without overlap in the ROI.

Again, going back to our prior finding from Chapter 2 that the location of tumor cores is the main driver of behavioral impairments in brain tumor patients, we restricted our overlap analysis to only the core masks. Out of 182 total subjects, 143 subjects had a tumor core (the remaining only had FLAIR hyperintensity regions which encompass peritumoral edema). And of these 143 tumor cores, 23 cores fell within the left pSTS ROI with a 20 mm radius. The performance on auditory-only, congruent audiovisual, and incongruent audiovisual conditions across all noise conditions was compared between subjects whose core overlapped with the ROI versus subjects whose core did not overlap with the ROI. The same comparison was made for the frequency of McGurk responses. Boxplots of the accuracies across the 3 conditions (audio, congruent audiovisual, and incongruent audiovisual) for the overlap group versus the non-overlap group are displayed in Figure 13. Similarly, a boxplot of the frequency of McGurk responses is shown in Figure 14. As hypothesized, the comparison revealed no significant differences in accuracies between the two groups in the auditory-only condition ($t(40.49) = 0.73$, $p = 0.47$) nor the incongruent audiovisual condition ($t(27.88) = 0.73$, $p = 0.47$). However, in contrast to our hypothesis, we found 1.) a significant difference in the two groups for the congruent audiovisual condition ($t(58.23) = 2.05$, $p = 0.044$) with subjects with lesions in the

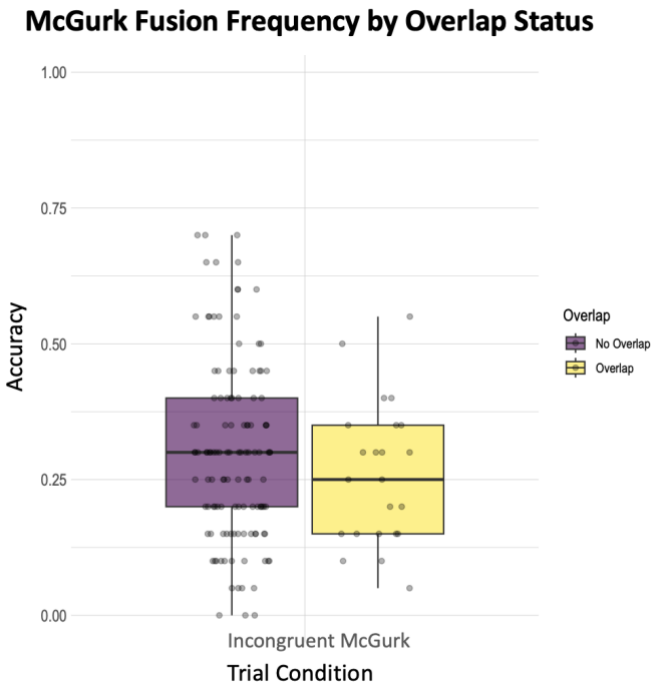
pSTS having higher accuracies in these trials; and 2.) a non-significant difference in the frequency of reported fusion responses ($t(36.6635) = 1.2207$, $p = 0.23$) between the two patient groups. Upon closer examination of the data, we noticed presence of few subjects whose performance was near or below the chance accuracy level at 25% across all conditions, suggesting that these subjects were not properly paying attention or simply not responding for the majority of the trials. Upon removal of these outliers (n = 6) identified using median absolute deviation, the significant difference between the two overlap groups disappeared for congruent audiovisual condition ($t(38.57) = 0.99$, $p = 0.328$), in line with our *a priori* hypothesis. However, the pattern of results of all other t-tests and ANOVAs, however, remained the same even after the removal of these 6 subjects.

Figure 13. *Boxplot of accuracies by trial conditions.*



Note. Overlap group denotes whether the group consists of subjects who had core masks overlapping the 20 mm left pSTS ROI.

Figure 14. *Boxplot of McGurk fusion response frequencies.*



Note. Overlap group denotes whether the group consists of subjects who had core masks overlapping the 20 mm left pSTS ROI.

Next, we examined how the results of the ROI analyses varied as a function of different noise levels added to the audio (low vs. high). A two-way ANOVA was performed to analyze the effect of noise level and ROI overlap status on McGurk fusion frequency. As expected, our data (Fig. 15) showed the main effects of noise level [$F(1,286) = 12.65$, $p = 4.41\text{E-}4$, $\eta_{p^2} = .042$] but neither a main effect of overlap status [$F(1,286) = 1.542$, $p = 0.215$, $\eta_{p^2} = .005$] nor a significant interaction between noise level and overlap status [$F(1,286) = 0.238$, $p = 0.626$, $\eta_{p^2} = 8.30\text{E-}4$]. Post-hoc analyses revealed that trials with higher noise levels were associated with greater fusion reports from subjects [$t(288) = 3.56$, $p = 4.38\text{E-}4$, $d = .418$]. This result is consistent with numerous prior findings that listeners rely on and benefit more from visual information when the auditory stimuli are compromised (Ma et al., 2009, Stevensen and James, 2009; Tye-Murray et

al., 2010). We then focused on the accuracies of the 3 trial conditions (audio-only, congruent audiovisual, and incongruent audiovisual) across the noise levels and ROI overlap status (Figure 16). Running a 3-way ANOVA, our data revealed the main effects of trial conditions [$F(2,858) = 377.496$, $p < 2E\text{-}16$] and noise levels [$F(1,838) = 153.500$, $p < 2E\text{-}16$], as well as a significant interaction between trial conditions and noise levels [$F(2,858) = 22.02$, $p = 4.73E\text{-}10$].

Figure 15. *Boxplot showing McGurk response frequency by audio noise level and overlap status with the 20 mm spherical ROI.*

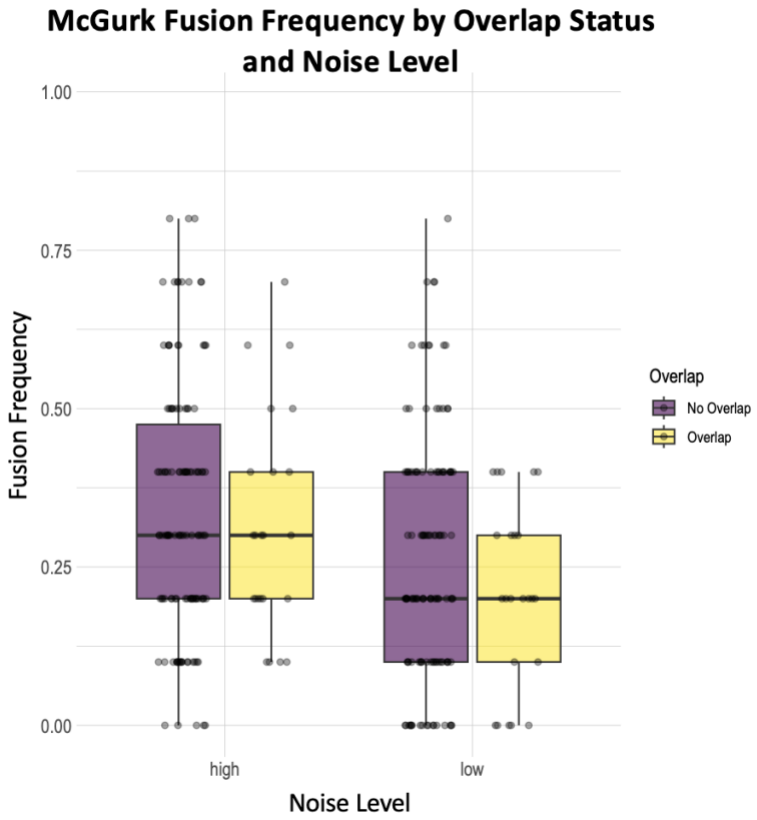
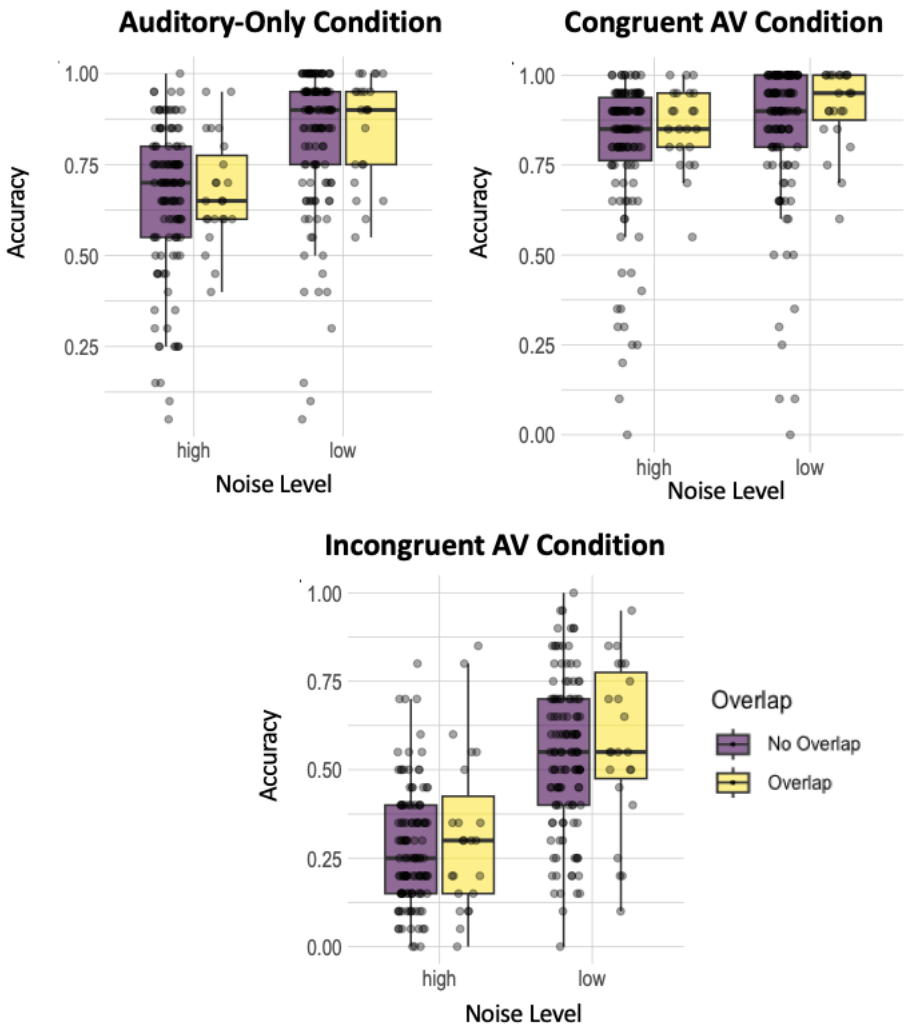


Figure 16. *Boxplot of accuracy across conditions by noise level and overlap status.*



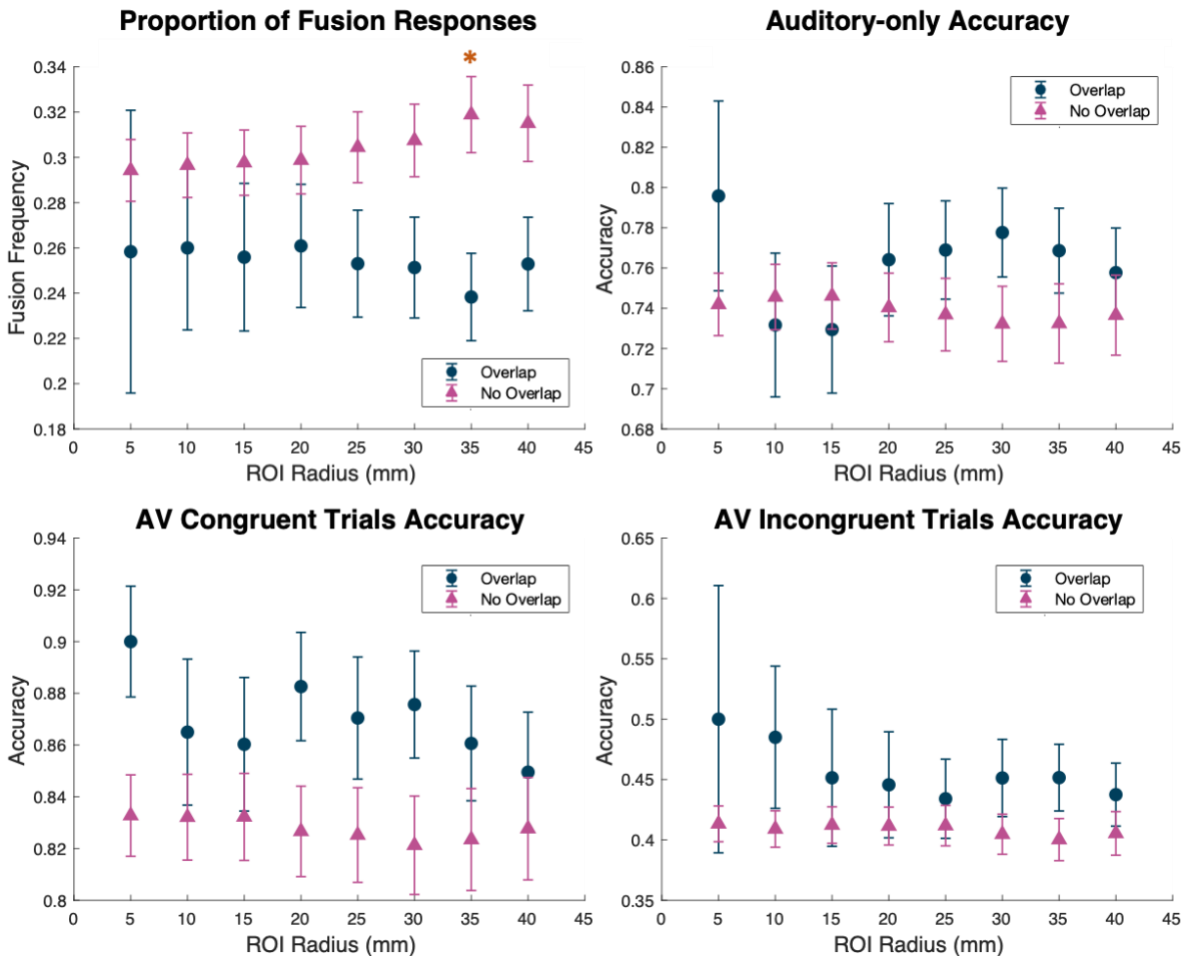
Finally, to understand the specificity of our findings using the 20 mm ROI, we investigated the dependency of our obtained results on our selected ROI size. This was done by performing the same analyses for the 20 mm radius ROI across the 8 different ROI sizes (from 5 mm to 40 mm radius using 5 mm increments). For each behavioral measure of interest (fusion frequency, auditory accuracy, congruent audiovisual accuracy, and incongruent audiovisual accuracy), 8 comparisons were made between the scores of those with an ROI overlap and those without an ROI overlap. Table 3 shows the number of subjects who had their cores overlapping the ROI for the values of radii used. The results of these comparisons are shown in Figure 17.

Threshold correction was performed using False Discovery Rate (FDR) with a threshold of p-value at $p < 0.05$. Of all the comparisons within each behavioral measure, the only difference to survive FDR corrections was in the McGurk fusion frequency using the 35 mm ROI radius. This specific comparison showed that those with lesions outside of the ROI more frequently report McGurk fusion percepts compared to those with lesions in this ROI ($p_{adj} = .017$).

Table 3. *Number of subjects whose core overlaps with the ROI of different sizes.*

ROI Radius	5mm	10mm	15mm	20mm	25mm	30mm	35mm	40mm
# subjects overlapping ROI (of 143 cores)	6	15	17	23	33	38	47	52

Figure 17. *Behavioral measures comparison between those with overlap in the ROI with those who do not across different radii of ROIs.*



Note. After the false discovery rate correction for multiple comparisons, the only significant difference between the group with ROI overlap and without ROI overlap occurred using 35mm radius ROI for McGurk fusion frequency (denoted with an asterisk). The error bars denote the standard error of the mean.

While the results of this analysis revealed patterns in line with our predictions, we wanted to ensure that our results were not due to random sampling given the relatively smaller number of subjects that had lesions overlapping the ROI compared to those with lesions outside of the ROI. Therefore, we permuted the subjects' ROI overlap status (overlap versus non-overlap), compared their behavioral measures for each ROI radius, and evaluated statistical significance based on an alpha of 0.05. This permutation test revealed similar patterns of results as our

parametric t-test (both resulting in the same conclusion of significance for alpha = 0.05) and confirmed that our results were not due to random sampling chance.

Similarly, to assess the specificity of our selected ROI location (MNI coordinates -53, -48, 10) and its associations with McGurk fusion perceptibility, we compared the test statistics of our 35 mm radius ROI with randomly selected ROIs of the same size within the boundaries of our total core overlay map in 2000 permutations. We required that the randomly generated ROIs have at minimum the same number of voxels overlapping with the core overlay map. The results showed that only 1 out of 2000 randomly generated ROIs had t-statistics as large as our left pSTS ROI, deriving a p-value of .0005.

3.5 Discussion

In Study 2, we examined the influence of visual speech information on various behavioral measures of speech perception using patients with intrinsic brain tumors. Unlike the majority of audiovisual speech perception studies, our study used naturalistic stimuli using real instead of phonemes appearing in isolation from other sounds. Using such stimuli, our data showed expected patterns of results: congruent audiovisual information boosted the accuracy of speech perception while incongruent audiovisual reduced the accuracy of speech perception in non-McGurk stimuli. Similarly, our data demonstrated that listeners' reliance on visual information increases when auditory information becomes compromised (with high noise) with greater benefit experienced when the audiovisual information was congruent and larger reduction in accuracy when the audiovisual information was incongruent.

Through TMS, Beauchamp et al. (2010) showed that the left pSTS is causally involved in the maintenance of the McGurk perception. While this work was highly influential, this study left open the question whether this area is also causally involved in the general audiovisual

speech perception. While many studies in the past regarded McGurk processing as being synonymous with general audiovisual integration, more recent evidence started to question this underlying assumption. More recently, researchers have begun to find evidence that McGurk processing, where listeners fuse together mismatching audiovisual information, likely relies on a different neural mechanism than congruent audiovisual processing, where listener experience enhanced perception from the matching information. For example, the general audiovisual speech enhancement may result from the temporal information (Besle et al., 2008; Chandrasekaran et al., 2009) or speaker identity (Brang, 2019; Vatakis & Spence, 2007) that visual stimuli relays to the auditory system, which McGurk perception alone cannot capture. McGurk perception, may additional recruit additional processes to resolve a conflict between the two modalities, something that's uncharacteristic of naturally congruent speech perception.

With the goal of addressing this distinction between the two processes, our study found small support that a tumor in the left pSTS leads to weak audiovisual speech integration processes including a weaker McGurk effect (in terms of the number of fusions experienced) compared to patients with a tumor elsewhere in the brain. While there was evidence that lesions in pSTS is associated with reduced McGurk illusions in the ROI analyses using ROI of varying sizes, this finding was not robust. Specifically, we were surprised that there was no significant difference in the frequency of McGurk responses between patients with lesions in the pre-selected ROI and patients with lesions outside of the ROI. Yet, we did not interpret our findings as evidence against our original hypotheses. Indeed, prior literature provides compelling, robust evidence that left pSTS is responsible for facilitating McGurk illusions (Beauchamp et al., 2009; Hickok et al., 2018) and that superior temporal regions are heavily involved in maintaining both phonological processing and speech representation (Boatman, 2004; Hickok and Poeppel, 2000).

Therefore, we took this contradictory finding as evidence that our analyses were not well-powered. For example, using the ROI with a 20 mm radius, we only had 23 patients with cores within the ROI compared to the 120 who did not. Another explanation for why our study failed to replicate Hickok et al. (2018)'s finding that lesions in pSTS impair McGurk perception may be the difference in our patient population. Compared to the acute nature of stroke and the resulting lesion, tumors tend to develop much more gradually. Additionally, whereas stroke causes immediate cell death, cells in tumor lesions can still retain some of their functionality (Aabedi et al., 2021). These two differences can translate into a more moderate behavioral impairment observed in tumor patients compared to stroke patients (Cipolotti et al, 2015; Vaidya et al., 2019). And with more moderate behavioral impairment, this implies that tumor lesion studies can require greater sample sizes to detect the same size effects.

If our contradictory findings reflect the true effects of lesions in the left pSTS, we expected to see this effect maintained across the different ROI sizes used to test the reliability of our results. Unsurprisingly, the difference in the auditory accuracies between overlap and non-overlap groups disappears across the different ROI sizes used once the multiple comparisons corrections are applied. The only significant difference that does survive the multiple comparisons corrections across testing of all behavioral measures is the difference in McGurk fusion reports between non-overlap and overlap group of subjects with an ROI radius of 35mm. Here, we find patients with lesions in ROI report fewer McGurk illusions than those with lesions elsewhere. However, given that this finding was not robust across the other ROI sizes, we interpret it as evidence, albeit weak, that left pSTS is responsible for maintaining the McGurk percept based on these specific data. In the context of prior TMS and stroke VLSM research

however, these data add some support to the causal relevance of the left pSTS in the generation of the McGurk effect.

Hickok et al. (2018)'s lesion mapping work used stroke aphasia patients to identify causal regions critical for audiovisual speech integration. Using VLSM across 100 patients with left hemisphere stroke lesions, the researchers identified that lesions in the superior temporal lobe and temporal-parietal junction were significantly associated with McGurk susceptibility in patients. On the other hand, they failed to find any significant brain regions associated with auditory facilitation measures (benefit of congruent visual information) covarying the McGurk fusion frequency; notably, their measures only used 2 phonemes, /ka/ and /pa/. In the original conception of the present study, we intended on extending this study in tumor patients with wider lesion coverage, a larger number of subjects, and more naturalistic stimuli (compared to their use of phonemic stimuli). With more lesions and wider lesion coverage, we hoped to identify brain regions that were differentially involved in audiovisual facilitation versus McGurk fusion percepts. Unfortunately, our analyses were significantly restricted by the limited core lesion coverage we observed in our pool of subjects. Despite our much larger sample size, we did not impose any inclusion criteria for patients based on where their lesions were. While this was done to maximize the number of voxels that could be included in the VLSM, it resulted in sparse coverage of the left temporal area where we expected to find significance. In the initial proposal of the study, given our very large sample size for a lesion study, we were not concerned with this possibility. Consequently, this limitation translated to a low statistically powered study, limiting our inferences even with the use of a more targeted approach like the region of interest analysis.

Despite the shortcomings of our study, we believe our data will be valuable in dissociating the brain regions involved in the facilitatory effects versus the modulatory effects

(resolving conflicting information) of audiovisual speech in a future study. One potential route to achieve this goal is to set a specific number of subjects that present with tumor cores in the left hemisphere before completing a parallel set of analyses that are presented here. This will ensure that the study has sufficient lesion coverage in areas where we expect to find a strong relationship with the audiovisual impairments. Such a study could help determine whether auditory enhancements from congruent visual speech (e.g., better detection and faster RTs) and visual modulations of what is heard (i.e., the McGurk effect) are supported by two distinct mechanisms and brain regions.

Chapter 4 Posterior Superior Temporal Sulcus Serves as Cortical Locus for McGurk Effect but Not for Congruent Audiovisual Processing

4.1 Abstract

Speech perception benefits from congruent multisensory speech information. Moreover, mismatched visual speech can alter the processing of the accompanying acoustic speech, leading to an illusory percept known as the McGurk effect. This illusion has been widely used to study audiovisual speech integration, illustrating that auditory and visual information is combined in the brain to create a single coherent percept. While prior transcranial magnetic stimulation (TMS) and neuroimaging studies have identified left posterior superior temporal sulcus (pSTS) as a causal region involved in the generation of the McGurk effect, it remains unclear whether this region is critical only for this illusion or also for the more general benefits of congruent visual speech. For example, recent work suggests that the audiovisual facilitation effect (a measure of benefit gained from audiovisual congruent stimuli compared to audio-only stimuli) and the audiovisual McGurk fusion rate (a measure of how frequently McGurk stimuli are fused) reflect distinct audiovisual processing mechanisms. To better understand how these different indices and dimensions of audiovisual integration are causally associated with the left pSTS, we used single-pulse TMS to disrupt the cortical excitability in this region during trials with either incongruent McGurk combinations or congruent audiovisual combinations. Consistent with the prior TMS study, we observed that single-pulse TMS to the left pSTS significantly reduced the strength of the McGurk effect. Importantly, however, left pSTS stimulation did not affect the positive benefits of congruent audiovisual speech (higher accuracy and faster reaction times),

demonstrating a causal dissociation between the two processes. Our results are consistent with models proposing that the pSTS is only one of the multiple critical areas supporting audiovisual speech interactions. Moreover, these data add to a growing body of evidence suggesting that the McGurk effect should not be employed as a surrogate measure for more general and ecologically valid audiovisual speech behaviors.

4.2 Introduction

While speech perception is regarded largely as an auditory process, our perception heavily relies on concurrent visual cues to provide both complementary and redundant information about an auditory stimulus, particularly in noisy environments (Campbell, 2008; MacLeod & Summerfield, 1987; Sumby & Pollack, 1954). For example, listeners can derive both timing and phonemic information about spoken words simply by watching a speaker's mouth movements (Luo et al., 2010; Plass et al., 2020; Schroeder et al., 2008). This visual augmentation can significantly improve speech perception accuracy, especially when acoustics are compromised. Given the inherent multisensory nature of speech, we examine how the brain enables vision to support language processes to truly understand speech processing in a naturalistic context.

The present study aims to identify the causal brain structures necessary for audiovisual speech integration using transcranial magnetic stimulation. TMS is a noninvasive brain stimulation method that involves the application of magnetic pulses to targeted brain areas that, depending on stimulation parameters, can temporarily excite or suppress activity at local neuronal populations (Hallett, 2000). Applying single pulse TMS (also referred to as virtual lesioning) at specific time points during a task, researchers have used this technique to study the causal mechanisms underlying numerous cognitive and perceptual processes (Hallett, 2000;

Rossini & Rossi, 2007; Walsh & Cowey, 2000). By inhibiting specific brain regions believed to be involved in facilitating audiovisual speech, TMS can clarify whether the targeted region is causally necessary for the tested audiovisual behavior.

Traditionally, studies of audiovisual processing have relied on the use of the McGurk effect, in which an auditory phoneme (e.g., /ba/) is paired with the visual movie from a different phoneme (e.g., /ga/), resulting in the perception of a fused or unique sound (e.g., /da/) (McGurk & MacDonald, 1976). Using McGurk stimuli, prior research has identified the left posterior superior temporal sulcus (pSTS) as a crucial region that facilitates the integration of auditory and visual information (Sekiyama et al., 2003; Bernstein et al., 2008; Benoit et al., 2010; Irwin et al., 2011; Nath et al., 2011; Szycik et al., 2012). For example, individual differences in the strength of the McGurk effect are correlated with fMRI activity in the pSTS during the experience of the illusion (Nath & Beauchamp, 2012), and both inhibitory transcranial magnetic stimulation (TMS) and damage following a stroke in this region are associated with reduced McGurk effect percepts (Beauchamp et al., 2010; Hickok et al., 2018). However, researchers have recently questioned whether the findings from research using McGurk combinations can generalize to more natural audiovisual integration processes (for review see Alsius et al., 2018). For example, numerous studies (Brown & Braver, 2005; Hickok et al., 2018; Van Engen et al., 2017) have reported weak correlations between the McGurk effect and other measures of audiovisual speech processing in individuals, raising doubts about whether the mechanisms that enable the McGurk effect are the same as those that process natural audiovisual speech.

Whether the McGurk effect and more general audiovisual processes rely on the same mechanisms is a significant issue in the field because the McGurk effect had long been accepted as a standard measure for quantifying audiovisual speech integration (Alsius et al., 2018; Van

Engen et al., 2022). This holds especially true for clinical populations including those with autism spectrum disorder (ASD). Individuals with ASD often exhibit difficulties in communication and social interaction and many researchers believe that this could be in part attributed to impairments in multisensory processing. To study this, researchers have repeatedly used the likelihood of McGurk perception to demonstrate that individuals with ASD show altered multisensory processing (Gelder, et al., 1991; Feldman et al, 2022; Stevenson et al, 2014; Zhang et al., 2019; Williams et al., 2004). These studies have consistently shown that individuals with ASD experience significantly weaker McGurk effects than the neurotypical population. However, it is important to note that the McGurk effect focuses on the cognitive cost of processing conflicting auditory and visual information. In contrast, in normal speech contexts, listeners avoid integrating conflicting audiovisual speech information (Brang, 2019; Seijdel et al., 2023). Specifically, this incongruent pairing that is necessary to elicit McGurk fusion responses has been regarded as highly artificial and unnatural, showing limited features present in everyday speech (Van Engen et al., 2022) because face-to-face conversations yield congruent combinations of auditory and visual speech. Moreover, McGurk studies tend to examine audiovisual processing using isolated phonemes rather than complete words, casting further doubt on their applicability to natural speech. Consequently, tasks that use more naturalistic stimuli, like complete words and congruent audiovisual pairings, may be better able to clarify the role of visual information in everyday speech perception, thus advancing beyond the limited context of the McGurk effect.

While strong correlative research has identified the left pSTS as a region associated with the generation of the McGurk effect, only two studies to date have used causal methods (Beauchamp et al., 2010; Hickok et al., 2018). In a 2010 study, Beauchamp et al. reported that

single pulse TMS applied to the left pSTS significantly reduced perception of the McGurk effect, providing strong evidence for the role of the pSTS in the generation of this illusion. In their study, the authors conducted two separate experiments, each with sample sizes of 9, using similar task designs but two different speakers (experiment 1 using a female speaker and experiment 2 using a male speaker) and two different phonemes (experiment 1 used auditory /BA/ with visual /GA/ and experiment 2 used auditory PA with visual /NA/ or /KA/). Two separate experiments using different phoneme and speaker were to ensure the robustness of their findings across different speakers and stimuli combinations. Results showed that single-pulse stimulation to the left pSTS reduced the average frequency of fused percepts in McGurk conditions by 54% (experiment 1) and 21% (experiment 2) compared to stimulation applied to the control site. The authors also showed that only stimulation applied within 100 ms of the onset of the auditory stimulus reduced the frequency of McGurk percepts, while stimulation applied outside of this time range did not influence frequency. Extending this research to clinical populations, Hickok et al. (2018) tested patients with a recent stroke using a McGurk effect paradigm with the goal of identifying lesioned areas of the brain that reduced McGurk effect percepts. Partially replicating the TMS work, Hickok et al. (2018) showed that stroke lesions in the broad superior temporal lobe (as well as in auditory and visual areas in the superior temporal and lateral occipital regions) resulted in the greatest deficits in McGurk perception, adding support to the model that the left pSTS enables the generation of the McGurk effect.

While both prior TMS and stroke lesion mapping studies identified the left pSTS as being causally relevant to the generation of the McGurk effect, those studies were not designed to test the relevance of this structure on more natural, congruent audiovisual speech perception behaviors. Building upon Beauchamp et al. (2010)’s findings, here we sought to address the

involvement of left pSTS in other aspects of audiovisual speech processing beyond the generation of the McGurk percept and further investigate whether the McGurk effect is a good proxy for audiovisual processing. Based on past literature, two clear predictions emerged: 1) transient damage to the left pSTS will impair both the McGurk effect and the normal benefits from audiovisual speech. Such findings would indicate that left pSTS is a critical hub for audiovisual speech generation in general and that the McGurk effect is a good proxy for natural audiovisual speech behaviors. Or 2) transient damage to the left pSTS impairs the McGurk effect with minimal impact on the normal benefits from audiovisual speech. Such findings would indicate that this region is likely only one of many critical structures necessary for audiovisual speech generation in general, reflecting only a subset of the information relayed from visual to auditory speech regions and laying the groundwork for research to understand which information is relayed through this hub.

To test these predictions using TMS, we examined both the frequency of subjects' McGurk effect percepts (which captures how visual information can change, or modulate, the perception of auditory information) along with a measure of audiovisual facilitation (which focuses on how visual information aids and facilitates the processing of the auditory information). By distinguishing between the two features, we can better understand how mismatched visual information can modulate auditory perception and whether this is dissociable from the ability of congruent visual information to improve and facilitate the processing of concurrent auditory perception. Towards this goal, we used an audiovisual task with full word stimuli, rather than phonemes that are typically utilized in most McGurk-type studies, as well as a larger sample (n = 21) than the prior TMS study. We hypothesized that pSTS stimulation would affect the McGurk effect more than congruent audiovisual benefits, providing evidence

that the visual modulation of auditory speech, relies on different neural mechanisms and consequently brain regions from audiovisual facilitation.

4.3 Methods

4.3.1 Subjects

Twenty-four healthy subjects (10 males, $\overline{age} = 25$, 4 left-handed) with self-reported normal hearing and vision without a history of neurological disorder participated in the study. Three of the 24 total subjects either voluntarily withdrew from the study or experienced data errors resulting in 21 total subjects who completed the study. Beauchamp et al. (2010) prescreened their subjects to include only those who reported strong McGurk effects. They justified this pre-selection process because there are large individual differences in susceptibility to the McGurk effect (Nath & Beauchamp, 2012) and variable reliance on lip movements during speech perception (Gurler et al., 2015). However, we did not exclude any subjects based on their McGurk susceptibility. While previous research has shown that the specific audiovisual stimulus used affects the strength and frequency of McGurk effects experienced by subjects (Basu Mallick et al., 2015), the stimuli used in the current study successfully evoked McGurk percepts in the majority of individuals tested (Brang et al., 2020) and the same stimulus set has been shown to produce robust congruent audiovisual benefits during speech recognition (Ross et al., 2007). Therefore, to maximize the generalizability of our study to everyday speech perception, we did not exclude subjects based on their susceptibility to the McGurk effect.

To estimate the necessary sample size, we conducted an a priori power analysis using G*Power (Faul et al., 2007) based on Beauchamp et al. (2010)'s data. In comparing the frequency of reported fusion responses during pSTS stimulation versus control site stimulation, their results yielded Cohen's D values of 3.22 and 8.43 across two experiments. Given the effect

size of 3.22 (the smaller of the two estimates), considered extremely large using Cohen's criteria (1988), we would need a minimum sample size of 4 to replicate their effects with a significance criterion of alpha = .05 and power = .95. However, as our goal was to examine whether congruent audiovisual behaviors were affected as well, we assumed conservatively that if present, TMS effects on congruent audiovisual behaviors would be at least 25% the magnitude of effect of TMS on McGurk percepts. Repeating the power analysis with an alpha of 0.05, power of .80, and effect size of .805 in a two-tailed paired t-test design yielded a minimum sample size of 15. We sought to exceed this number and collect as much data as possible before the end of April 2023.

All participants gave informed consent prior to the experiment. This study was approved by the institutional review board at the University of Michigan.

4.3.2 Task

The task for the present study was very similar to the audiovisual task used in Experiment 2 (refer to the task description in Study 2 for detail). In each trial, a female speaker produced a high frequency monosyllabic word starting with the consonants 'b', 'f', 'g', or 'd'. The trials were either audio-only, visual-only, audiovisual incongruent, or audiovisual congruent. Pink noise was applied to all auditory stimuli. The noise level was set based on a piloting of approximately 70% accuracy in the auditory-only condition. In trials with a visual stimulus, the speaker's video appeared 500 ms prior to the onset of the auditory stimulus. In audio-only trials, a gray box appeared 500 ms prior to the auditory onset to provide an equivalent temporal cue. Again, visual stimuli were recorded at 29.97 frames per second, trimmed to 1100 ms in length, and adjusted so that the first consonantal burst of sound occurred at 500 ms during each video. In Beauchamp et al. (2010), the authors found maximally diminished fusion responses when the

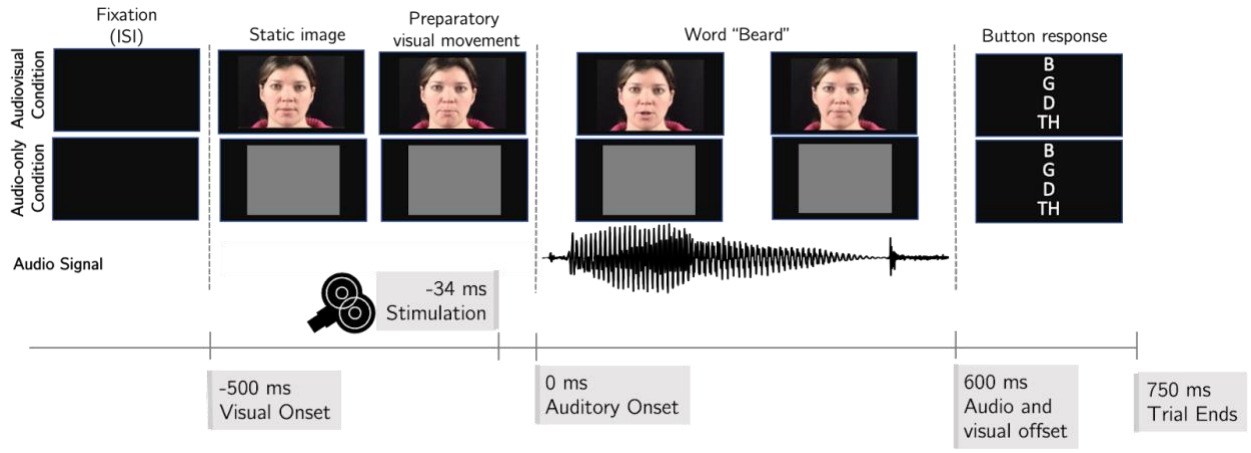
pSTS was stimulated just before auditory onset. Accordingly, here we applied a single TMS pulse 34 ms prior to the onset of the audio. Six hundred milliseconds following auditory onset, subjects were prompted to report the initial consonant of the seen (in visual trials) or heard word using a gamepad (Logitech F310) from the 4 options displayed on the screen. The task was completed while seated upright on a desktop computer using Psychtoolbox-3 (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) with participants seated approximately 60 cm away from the screen at eye level.

The task consisted of 3 blocks in total. Two blocks included stimulation of one anatomical region per block (the left pSTS or vertex) and one block included no stimulation. The order of stimulation was counterbalanced across participants. In total, the study consisted of 384 trials, with 128 trials in each block with 4 audiovisual conditions (audio-only, visual-only, audiovisual congruent, audiovisual incongruent) divided equally within each block (32 trials per condition per block). To ensure that all auditory and visual components of each word stimuli were presented the same number of times, half of the incongruent audiovisual trials had their incongruent auditory and visual stimuli flipped. These flipped conditions were considered to be non-McGurk incongruent conditions as the audiovisual combinations are not expected to generate fused responses, although they were still expected to cause lower accuracy and slower reaction time. Each block took approximately 7 minutes to complete, and participants were given the option to take breaks between each block. The schematic of the trials is shown in Figure 18.

Unlike the task used in Study 2, the current experiment included visual-only conditions and all auditory stimuli included the same level of pink noise (SNR of -4.6 dB) to increase the relative difficulty of the audio-only condition (to avoid ceiling effects in any condition) and because the addition of noise tends to increase the reliance on visual speech information (Ross et

al., 2007). While visual-only conditions were excluded from the Study 2 task design due to the timing constraint of testing patients, the condition was added back to test the effect of pSTS stimulation on lipreading performance.

Figure 18. *TMS Task Schematic.*



Note. Schematic of audiovisual (AV) and auditory-only trial for the word 'beard'. All trials started out with a blank black screen with an inter-stimulus interval lasting between 125 - 375 ms. 500 ms prior to the auditory onset, either a gray box (audio-only condition) or the video of the speaker (AV condition) appeared. For blocks in which TMS was applied, stimulation was applied 34 ms prior to the onset of the audio. Following the audio/visual offset, participants were given 1.5 seconds to respond via button press which initial letter the word they heard started with.

4.3.3 TMS

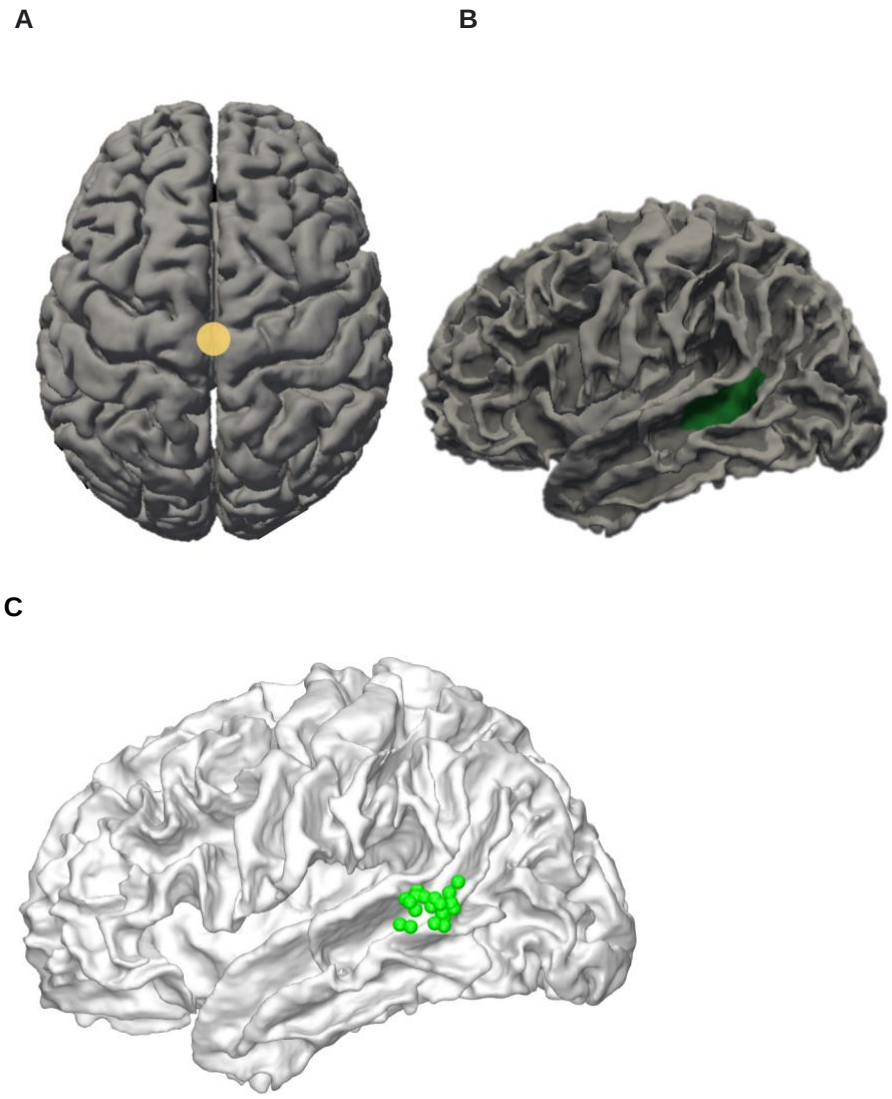
Prior to the TMS session, all participants underwent a structural MRI scan using a 3T GE MR 750 scanner to acquire T1-weighted images to be used for TMS guidance; three of 21 participants' MRI data was acquired as part of a different experimental paradigm from our lab (Ganesan et al., in review). Individual participants' T1 scans were processed using Freesurfer (http://surfer.nmr.mgh.harvard.edu/) for cortical reconstruction and volumetric segmentation. The Freesurfer-generated pial and white matter reconstructions were then used to localize the left pSTS target coordinates according to the labels of the automatic cortical parcellation and automatic segmentation volumes. Specifically, we used the individual subject coordinates from the center of 'lh-bankssts', which is the bank of the left hemisphere superior temporal sulcus, as our pSTS target.

TMS was applied through a MagPro X100 using a 70 mm figure-8 shaped TMS coil (MCF-B70, MagVenture Inc.). For each subject, their respective stimulation intensity was determined by obtaining their resting motor threshold and multiplying it by 1.1 (110%) as is the standard for single-pulse TMS thresholding (Kallioniemi and Julkunen, 2016; Sondergaard et al., 2021). The resting motor threshold is the lowest stimulation intensity necessary to evoke a consistent motor response while targeting the motor cortex. Specifically, this is the threshold at which an electromyographic motor response that is greater than 50 μV from the first dorsal interosseus muscle measuring is observed 5 out of 10 times (Mills and Nithi, 1997; Rossini & Rossi, 2007). In our motor thresholding, our study targeted the right first dorsal interosseous muscle. The mean resting motor threshold used in our study was 51.7% of the maximum stimulator output. Once the target intensity was determined for each subject, the same intensity (110% of the resting motor threshold) was used for the entirety of their TMS session.

Following motor thresholding, participants completed the audiovisual task with each block targeting a different TMS site: pSTS, vertex, and no stimulation. The vertex stimulation block served as a control condition (Jung et al., 2016) to mimic a similar TMS experience (including the placement of the coil against the head and the sound generated for each stimulation pulse) as the pSTS stimulation block. Brainsight's neuro-navigation system (Brainsight; Rogue Research) was used to target the stimulation sites by registering participants' facial landmarks to participants' structural T1s using headbands containing trackers. The target coordinates for the pSTS that were obtained via Freesurfer reconstructions were used to guide the stimulation. Figure 19B shows the location of the left pSTS stimulation site extracted from Freesurfer on a standard Montreal Neurological Institute (MNI)-152 brain as well as the vertex (Fig. 19A), the control stimulation site. Specific target sites used for each subject is shown on the

MNI brain in Figure 19C. The average coordinates of the left STS stimulation site across all subjects were (x = -65.1 ± 3.3, y = -48.5 ± 4.8, z = 7.3 ± 3.5) on a standard MNI-152 brain.

Figure 19. *Stimulation Sites on MNI brain.*



Note. (A) Vertex stimulation site (B) Freesurfer's segmentation denoting banks of STS was used as a guide to localize left pSTS stimulation site. (C) Stimulation sites used for the 21 subjects transformed to standard MNI space – each sphere denotes a specific STS site used for one subject.

To ensure precise delivery of the stimulation pulse relative to the onset of the auditory stimuli, the TMS pulse was auto-triggered through our MATLAB script using the MAGIC

toolbox (Saatlou et al., 2018). Due to the position of the coil required to target the pSTS, earbuds, instead of headphones, were used to deliver the audio. Headband positions were validated both before and after the motor thresholding prior to the audiovisual task. If significant deviations from the original facial landmarks were observed, the landmarks were re-registered.

4.3.4 Preliminary data acquired with continuous theta burst stimulation (cTBS)

Beauchamp et al. (2010) utilized a single pulse TMS paradigm to investigate the role of pSTS in facilitating the McGurk perception. However, in the initial design of our study, we aimed to use continuous theta burst stimulation (cTBS), a form of repetitive TMS (rTMS). cTBS consists of applying a set of three 50Hz pulses every 200 ms for a total of 40 seconds and is shown to inhibit cortical excitability of the targeted region for approximately 30 - 50 minutes (Huang et al., 2005; Wischnewski & Schutter, 2015). Therefore, the stimulation is applied right before the subject begins the task without needing to apply stimulation throughout the task. Due to the relatively long-lasting effects of cTBS, this preliminary design required TMS to occur over two sessions (one session per day) with each session targeting a different stimulation site. For the present speech perception task, we believed that cTBS would be ideal as participants may be distracted by any TMS clicks during the task itself and the cTBS approach can minimize the interference of the stimulation on the task performance.

While we assumed that the behavioral effect of cTBS would be more subtle than that of single-pulse TMS, we hypothesized that we will still be able to detect the effect of pSTS stimulation. Given the large effect size (Cohen's d between 3.22 and 8.43) reported in the original study (Beauchamp et al., 2010), we still expected to detect the effect of pSTS stimulation on fusion response frequency using our pilot data (n = 4) with cTBS. However, we found no trend towards a reduction in fusion response with cTBS applied to the pSTS. As we

were unable to replicate the main finding in the original study that we were hoping to extend, we, therefore, decided to switch to a single-pulse TMS design to more closely replicate the original study. To our knowledge, there has been limited work investigating the difference between the effect of cTBS and single-pulse TMS on behavioral measures.

4.3.5 Behavioral Measures of Interest

Our main behavioral measures of interest included audiovisual facilitation, audiovisual conflict, McGurk fusion frequency, and lip reading accuracy (Table 4). These measures were selected to be comparable to prior causal audiovisual research such as Beauchamp et al. (2010) and Hickok et al. (2018). Audiovisual facilitation refers to the benefit in comprehension from having congruent visual information accompanying the auditory stimuli. Quantitatively, this is represented as the accuracy (ACC) or reaction time (RT) of congruent audiovisual trials subtracted by the auditory-only trials. The audiovisual conflict measure captures the cognitive cost of processing incongruent auditory and visual information. This is numerically represented as the ACC or RT of incongruent audiovisual trials subtracted by the auditory-only trials. McGurk fusion frequency is the proportion of McGurk stimuli (a combination of auditory and visual stimuli that are specifically intended to result in a fused percept) in which the listener reports hearing the fused percept of the audiovisual stimuli. Finally, lip reading accuracy simply refers to the proportion of visual-only trials that were correctly identified.

As noted above, the incongruent audiovisual condition included 50% of stimuli combinations whose phonemic pairings typically evoke McGurk fusions (e.g., auditory ‘bet’ and visual ‘get’ yield the percept of ‘debt’) and the reversed pairings (e.g., auditory ‘get’ and visual ‘bet’) that do not typically result in McGurk fusions for the purpose of counterbalancing purposes. Because the McGurk Frequency measure quantifies McGurk fusions it is calculated

using from 50% of the trials in the incongruent audiovisual condition that lead to fusion effects. To ensure comparability with the McGurk Frequency measure, AV Facilitation, and AV Conflict were similarly calculated using only the trials in which the auditory stimulus was included in the McGurk Frequency measure. For example, the auditory word 'bet' was included in all three analyses but the auditory word 'get' was excluded. Conversely, the lip reading accuracy measure included all stimuli in that condition though effects were similar when restricted to the smaller set of stimuli.

Table 4. *Behavioral Measures of Interest.*

Measurement	Equation	Description
McGurk Frequency	*P*(fusion response | McGurk trials)	Proportion of reported fusion responses for McGurk trials
AV Facilitation	*Accuracy of* $AV_{congruent} - A$	Comprehension benefit received from congruent visual stimuli accompanying auditory stimuli. Larger value is associated with greater enhancement in comprehension.
AV Conflict	*Accuracy of* $AV_{incongruent} - A$	The amount that comprehension is impaired by having incongruent visual stimuli accompanying auditory stimuli. Larger value is associated with less perceptual impairment.
Lip Reading Accuracy	*P*(Vis = correct)	Proportion of visual-only trials that were correctly identified

4.3.6 Analysis

We first conducted a two-way ANOVA to test the impact of stimulation conditions (pSTS, vertex, or no stimulation) on the audiovisual measures. We followed this omnibus analysis with planned two-sample t-test comparisons between the main comparisons of interest (AV Facilitation, AV Conflict, McGurk Frequency, and Lip Reading accuracy) between vertex and pSTS stimulation conditions. Analyses were conducted both on reaction time and accuracy

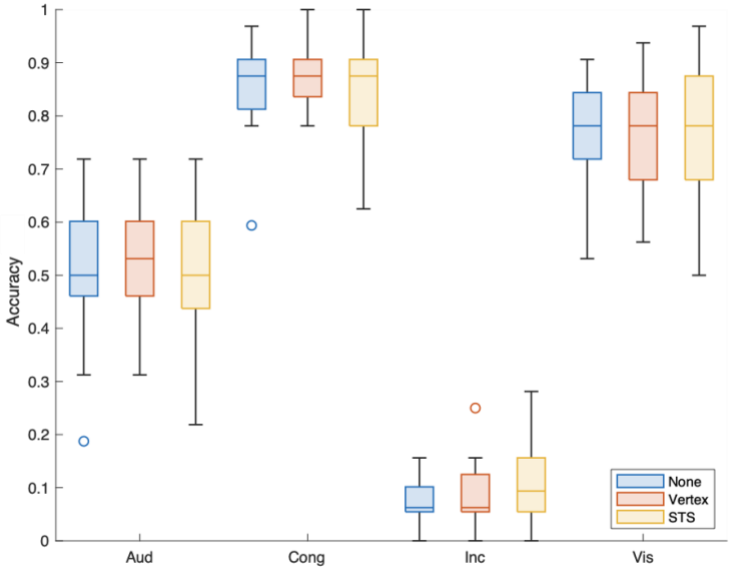
data. Although the original degrees of freedom are reported here for clarity, p values were subjected to Greenhouse–Geisser correction where appropriate (Greenhouse & Geisser, 1959).

4.4 Results

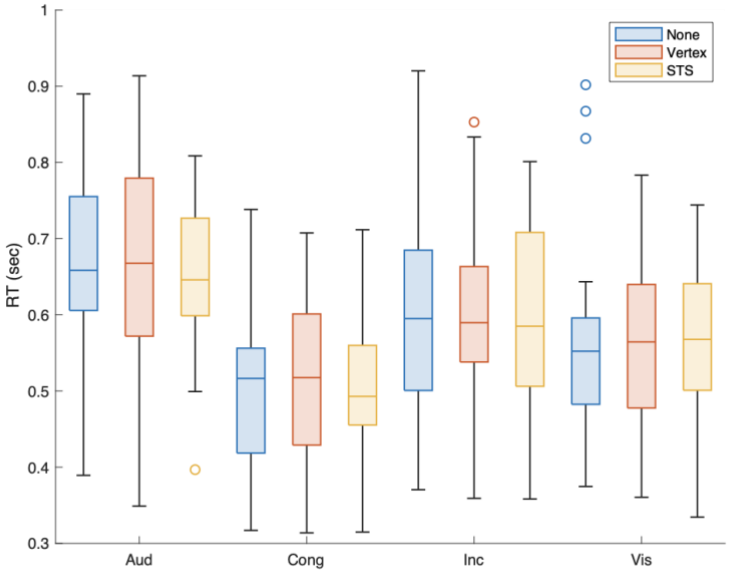
Boxplots showing the distribution of accuracy and reaction time measures are shown in Figures 20a and 20b respectively. A repeated measures ANOVA {Note, this includes all 4 visual types} applied to accuracy data (Fig 20a) revealed a main effect of visual type [$F(3,60) = 561.9$, $p = 7.4E\text{-}39$, $\eta_{p^2} = 0.913$], but no effect of stimulation site [$F(2,40) = 0.242$, $p = .737$, $\eta_{p^2} = 0.002$], nor an interaction between visual type and stimulation site [$F(6,120)$, $p = .725$, $\eta_{p^2} = 0.007$]. RT data (Figure 20b) mirrored those of the accuracy data with a main effect of visual type [$F(2.3,46.08) = 54.379$, $p = 1.19E\text{-}13$, $\eta_{p^2} = 0.185$], but no effect of stimulation site [$F(2,40) = 0.152$, $p = 0.860$, $\eta_{p^2} = 0.001$], nor an interaction between visual type and stimulation site [$F(6,120) = 0.820$, $p = 0.556$, $\eta_{p^2} = 0.002$]. These results demonstrated a strong influence of visual content on task performance (such that congruent visual speech improves speech recognition and incongruent visual speech impairs it) and that there was no systematic effect of stimulation across conditions.

Figure 20. *Boxplot of task accuracy (A) and reaction time (B) by conditions.*

A) Accuracy by Stimulation and Behavioral Measure



B) Reaction Time by Stimulation and Behavioral Measure



Note. For auditory, congruent, and incongruent conditions, responses that matched the auditory stimuli were counted as the correct answer. For visual-only trials, responses that matched the visual stimuli were counted as the correct answer.

To more directly compare the influence of visual information on speech perception, we conducted the same repeated measures ANOVAs, excluding the lip reading (visual-alone) condition, which revealed a similar pattern of results: a main effect of visual type [$F(2,40) = 741.229$, $p = 2.46E\text{-}32$, $\eta_{p^2} = 0.930$], but no effect of stimulation site [$F(2,40) = 0.516$, $p = 0.601$, $\eta_{p^2} = 0.003$], nor an interaction between visual type and stimulation site [$F(4,80) = 0.834$, $p = 0.507$, $\eta_{p^2} = 0.010$]. RT data mirrored those of the accuracy data with a main effect of visual type [$F(2,40) = 67.953$, $p = 1.37E\text{-}13$, $\eta_{p^2} = 0.233$], but no effect of stimulation site [$F(2,40) = 0.152$, $p = 0.212$, $\eta_{p^2} = 0.001$], nor an interaction between visual type and stimulation site [$F(4,80) = 1.091$, $p = 0.367$, $\eta_{p^2} = 0.002$].

Table 5. *Group-averaged behavioral measures of interest across different stimulation conditions.*

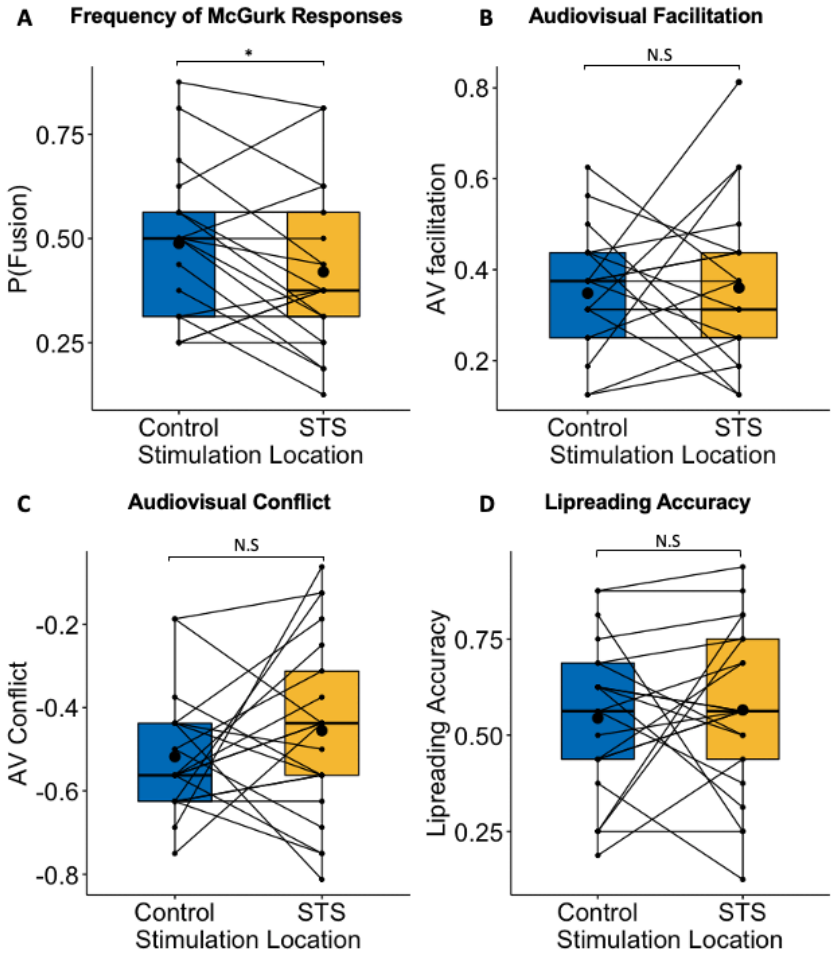
Behavioral Measure	No Stimulation	Vertex (control)	Left STS
Frequency of Fusion Responses	0.446 ± 0.156	0.488 ± 0.175	0.420 ± 0.192
Audiovisual Facilitation	0.408 ± 0.195	0.348 ± 0.093	0.360 ± 0.174
Audiovisual Conflict	-0.455 ± 0.145	-0.445 ± 0.132	-0.455 ± 0.218
Lipreading Accuracy	0.768 ± 0.098	0.768 ± 0.109	0.769 ± 0.124
Congruent AV Accuracy	0.866 ± 0.082	0.874 ± 0.061	0.853 ± 0.087
Incongruent AV Accuracy	0.074 ± 0.043	0.080 ± 0.064	0.106 ± 0.079
Auditory-only Accuracy	0.510 ± 0.128	0.525 ± 0.107	0.524 ± 0.120

Because the repeated measures ANOVAs showed neither the main effects of the stimulation site nor interactions with the stimulation site, we did not conduct follow-up tests on the individual conditions. Instead, we next examined the four planned comparisons based on a priori hypotheses to replicate and extend the findings from Beauchamp et al. (2010).

First, we sought to directly replicate the comparison made by Beauchamp et al. (2010) by examining the proportion of fusion responses made by subjects. Consistent with their data, single-pulse TMS to the left pSTS significantly reduced the subjects' likelihood of perceiving a McGurk fusion percept compared to the TMS to the vertex ($t(20) = 2.16$, $p = 0.043$, $d = 0.373$; Fig 4a).

Next, we compared whether accuracy in the incongruent audiovisual condition was affected by the stimulation site; this measure used the same trials as above but was only concerned with whether the subject identified the auditory stimulus correctly (disregarding the type of incorrect response made). Here we found a trend towards pSTS stimulation improving auditory accuracy in the presence of incongruent visual speech (subjects were better able to identify the heard stimulus) relative to vertex stimulation ($t(20) = 1.70$, $p = 0.105$, $d = 0.371$). Importantly, however, we observed no influence of pSTS stimulation (relative to vertex stimulation) on performance in congruent audiovisual trials ($t(20) = 0.00$, $p = 1.00$, $d = 0.00$). Comparing McGurk fusion rates and congruent audiovisual accuracy in a 2x2 repeated measure ANOVA additionally demonstrated a significant interaction between the two [$F(1,20) = 5.151$, $p = .034$, $\eta p^2 = 0.205$]. Lastly, there was no influence of stimulation on lipreading ability between pSTS and vertex stimulation ($t(20) = 0.372$, $p = 0.714$, $d = 0.098$). Similarly, we examined the above comparisons using RT (excluding fusion responses as it refers to specific responses made by subjects) as the dependent measure, and no differences were found (Table 6).

Figure 21. *Boxplot showing the difference between various audiovisual measures across TMS stimulation sites.*



Note. (A) The proportion of trials in which participants reported McGurk fusion responses in incongruent McGurk trials. (B) The change in accuracy between congruent audiovisual trials and auditory-only trials. Larger value in audiovisual facilitation denotes that participants benefit more from having congruent visual cues. (C) The change in accuracy between incongruent audiovisual trials and auditory-only trials. Larger audiovisual conflict measure denotes that there is smaller cognitive cost to processing incongruent visual stimuli. (D) The proportion of trials in which participants correctly identify the viseme in visual-only trials.

Table 6. *Reaction time comparisons for pSTS stimulation versus vertex stimulation.*

Comparison Type	Trial Condition	*df*	*t-statistics*	*p-value*	*Cohen's D*
Reaction Time for pSTS stimulation vs. vertex stimulation	Auditory-only	20	0.238	0.814	0.040
	Congruent AV	20	1.280	0.215	0.224
	Incongruent AV	20	0.815	0.425	0.158
	Visual only	20	0.041	0.968	0.008

Finally, to examine the relationship between the various audiovisual measures we computed Spearman's rank correlation between the relevant conditions. The correlation matrix between these conditions is shown in Table 6. Two significant effects emerged. First, audiovisual facilitation effect in individual subjects is positively correlated with audiovisual conflict effects (r_s = 0.688, p = 4.67E-10); this strong positive correlation suggests that listeners who experience greater perceptual benefit from congruent audiovisual speech experience less perceptual impairment from having incongruent audiovisual speech. Importantly, this is the opposite pattern of what would be predicted if the McGurk effect and congruent audiovisual effects are based on a single unified mechanism. Second, the frequency of listener's fusion reports is negatively correlated with their lip reading performance (r_s = -0.462, p = 1.39E-4); counterintuitively, this moderate correlation showed that those with good lip reading performance tended to experience fewer McGurk percepts.

Table 7. *Spearman's Correlation Matrix of the behavioral measures.*

	Audiovisual Facilitation ($AV_{cong} - A$)	**Audiovisual Conflict ($AV_{incong} - A$)**	**Frequency of Fusion Response**	**Lip Reading**
Audiovisual Facilitation				
Audiovisual Conflict	0.688 (*4.647E-10*)			
Frequency of Fusion Response	0.197 (*0.122)*	-0.063 *(0.626)*		
Lip Reading	-0.084 *(0.514)*	-0.149 *(0.243)*	-0.462 (*1.392E-4*)	

4.5 Discussion

The current study used real word stimuli to investigate the role of left pSTS on various aspects of audiovisual speech processes including McGurk perception, audiovisual facilitation, audiovisual conflict processing, and lip reading. Using TMS in healthy subjects, our data demonstrate that stimulation of the left pSTS significantly disrupts the experience of the McGurk effect, reducing the frequency of reported fusion percepts, while leaving the other audiovisual processes intact. Our study extends the work of Beauchamp et al. (2010) demonstrating that inhibitory TMS applied over the left pSTS, but not a control site, reduces the strength of the McGurk effect. However, as their behavioral task could not robustly capture the benefit of congruent visual stimuli on auditory speech perception, it remained unclear whether this region contributes specifically to McGurk processes, or to the general audiovisual speech integration processes at large. Our study extends Beauchamp et al. (2010)'s finding by confirming that the neural mechanism facilitating the McGurk perception, which many consider unnatural and artificial, can be dissociated from the other audiovisual processes like audiovisual facilitation and

lipreading, which occur in everyday speech. Similarly, our results are consistent with models proposing that the pSTS is only one of the multiple critical areas supporting audiovisual speech interactions. It also adds to the growing evidence that McGurk processing relies on additional neural mechanisms beyond our everyday audiovisual speech.

One way to explain the involvement of pSTS in McGurk processing but not in audiovisual facilitation is that direct projections from visual motion area MT/V5 to the auditory cortex (Besle et al., 2008) allow the visual facilitation of auditory process to bypass the pSTS. Similarly, it is also possible that the neural processes underlying audiovisual facilitation alternatively recruit frontal structures to recover speech information through sensorimotor integration (Du et al., 2014; Du et al., 2016; Hickok & Poeppel, 2007) without involving the pSTS. Along with Beauchamp et al. (2010)'s findings of reduced McGurk effect when targeting the left pSTS we replicated here, a similar report of a weakened McGurk effect has been reported in patients following strokes near the left pSTS (Hickok et al., 2018). Adding onto these findings, our work provides compelling evidence that McGurk processing is a specific form of audiovisual speech integration that's independent of other audiovisual speech enhancement and even general conflict processing (when audiovisual stimuli are not combined). That is to say, while McGurk processing may share some naturalistic components of normal, everyday audiovisual speech processing, it also contains additional properties, i.e., having to combine mismatched auditory and visual information, that's uncharacteristic of natural speech. Consequently, our findings suggest that the left STS is more specifically responsible for detecting minor incongruities across modalities and re-evaluating or directly transforming visemes to phonemes.

The works of Hickok et al. (2018) and Van Engen et al. (2017) have aimed to similarly investigate the relationship between McGurk susceptibility and the use of visual information to facilitate speech perception. Indeed, both studies report minimal correlations between the two measures, providing compelling evidence against the widespread use of McGurk susceptibility as an index for audiovisual speech integration (Alsius et al., 2007; Jones & Callan, 2003; Paré et al., 2003; Van Wassenhove et al., 2007). Replicating Hickok et al. (2018) and Van Engen et al. (2017), our results show no significant correlation between audiovisual facilitation, the ability to make use of visual information to facilitate speech perception, and the frequency of reported McGurk fusion percept. However, our correlation matrix indicates a moderate correlation of r_s = 0.688 (p = 4.647E-10) between the audiovisual facilitation measure and the audiovisual conflict measure, suggesting that the processing of both congruent and incongruent audiovisual speech relies on some shared mechanisms utilized by both processes.

We also find a significant negative correlation between lip reading performance and the frequency of McGurk responses. This tells us that those with better lip-reading abilities experience more McGurk percepts more infrequently. Greater accuracy in lip reading (performance in visual only trials) implies that the listener is able to extract more speech information from lip movements. Therefore, we were surprised by the direction of the correlation between lip reading performance and McGurk fusion frequency. The negative correlation between lip reading performance and McGurk fusion frequency suggested that good lip readers are less susceptible to McGurk effects. Using eye-tracking, Gurler et al. (2015) found that those who fixate longer on the speaker's mouth during speech tend to perceive McGurk effects more frequently. Based on this finding, we expected the relationship between lip reading performance and the McGurk susceptibility to be positive. However, given our findings, we postulate that

good lip readers are less likely to integrate the incongruent audiovisual McGurk stimuli due to the high fidelity of their visual speech information. Further research will be needed to confirm whether individuals with greater reliance on visual cues are less prone to the McGurk illusion.

While we replicated Beauchamp et al. (2010)'s main finding that single pulse TMS to left pSTS diminishes the McGurk effect, we observed very large differences between the two studies in the magnitude of the effect size for the difference in McGurk percept frequency between pSTS and control site stimulation. In comparison to the large effect size reported by Beauchamp et al. (2010), with a Cohen's D of 3.22 and 8.43 (across two separate experiments using different speakers and phonemes), our effect size for the difference in the frequency of fusion responses was much more moderate with a Cohen's D of 0.2. Such disparity in the effect size may be accounted for by the differences in the two study designs.

Our study differed from Beauchamp et al. (2010) in six notable ways. First, Beauchamp et al. (2010) only had 2 conditions: congruent and incongruent McGurk conditions. All trials used the same single auditory phoneme either matched with its congruent viseme or an incongruent viseme that is known to create a fusion percept. However, to capture the measures of audiovisual facilitation, audiovisual conflict, and lip reading, our study also included auditory-only conditions, visual-only conditions, and incongruent audiovisual conditions that were not expected to produce McGurk percept. Second, we used real monosyllabic words rather than phonemes for more naturalistic speech stimuli. This was done to address a common criticism of McGurk studies which argues that phonemes are highly artificial and do not reflect natural speech. Third, Beauchamp et al. (2010) determined the location of left STS using both anatomical (5 subjects based on landmarks) and functional (7 subjects based on individual subjects' fMRI activation patterns) approaches. However, our study relied only on anatomical

landmarks to determine the intended stimulation site. Fourth, we used the vertex as a control site rather than "a control TMS site dorsal and posterior to the STS" as reported by Beauchamp et al. (2010). This was to ensure we were using a consistent control site across subjects. Fifth, we did not exclude subjects based on whether they experienced strong McGurk effects. Our prior work (Brang et al., 2020) using similar stimuli set showed that most individuals report some level of fusion responses when presented with our word stimuli. Therefore, we wanted to ascertain that the results of our TMS were generalizable and not restricted to only those who experience strong McGurk percepts. Indeed, in the no-stimulation condition for the current dataset, all participants experienced a decrease in accuracy in the McGurk audiovisual incongruent condition relative to the auditory-alone condition (range 12.5 - 75% decrease in accuracy, mean = 45.5%). Fifth, we added pink noise to all of our auditory stimuli whereas no noises were added to Beauchamp et al. (2010)'s auditory stimuli. Our decision to add noise was based on the prior literature showing that dependence on visual speech information increases with introduction of noise (Alsius et al., 2016; Buchan et al., 2008; Stacey et al., 2020). The addition of pink noise may in part have aided in generating McGurk percepts in all of our participants. Lastly, the two studies differed slightly in the TMS threshold used to apply single pulse stimulation. Whereas Beauchamp et al. (2010) used motor threshold intensity, or resting motor threshold (RMT) for pulses, we used a slightly higher threshold at 110% of the RMT. We opted for the higher threshold as taking 110% or 120% of RMT is the more widely reported approach found in TMS literature (Cuypers et al., 2014; Kallioniemi and Julkunen, 2016; Sondergaard et al., 2021) which would naturally be expected to produce greater inhibition of the involved region.

Given these differences in the study designs, it's possible that the disparity in the effect sizes of the pSTS stimulation on the frequency of McGurk effect may have been driven by a few

of these factors. Specifically, we speculate that the largest driver of the disparity is the difference in the exclusion criteria (which were accompanied by other task designs to ensure that subjects still get fusion percepts). By broadening the subject pool, we similarly broaden our inference beyond the individuals who almost always experience the McGurk effect with a particular stimulus pairing. Because our pool of subjects is less likely to experience the McGurk percept compared to those from Beauchamp et al. (2010), it's possible the effect of STS stimulation is less pronounced as the integration of the incongruent stimuli does not always occur as is the case for Beauchamp et al. (2010)'s subjects. In addition, it is also widely accepted that the likelihood of the McGurk effect varies largely across the stimuli used (Beauchamp et al., 2010; Basu Mallick et al., 2015). It is plausible that the audiovisual phoneme pairing used in Beauchamp et al. (2010) elicits a stronger McGurk percept compared to the audiovisual word pairings used in our study.

Taken together, our data points to evidence that audiovisual speech integration is not exclusively dependent on a single major hub in the left pSTS; instead, the left pSTS is more important for the generation of McGurk perception, resolving the conflict between auditory and visual information so that the information can be perceived as a single percept, rather than two mismatching percepts. While pSTS has been dubbed the multisensory hub of the brain and is indeed necessary for certain facets of multisensory perception, the importance of this region has likely been inflated due to the field's heavy reliance on McGurk stimuli in the study of audiovisual integration. Indeed, our data provides converging and complementary evidence with the growing number of both behavioral and electrophysiological studies (Arnal et al., 2009; Arnal et al., 2011; Eskelund et al., 2011; Faivre et al., 2015; Fingelkurts et al., 2003; Lange et al., 2013; Palmer & Ramsey, 2012; Roa Romero et al., 2015) that points to the existence of multiple

processing pathways and question the generalizability of McGurk perception to more naturalistic audiovisual speech processing.

Collectively, these converging evidence suggests that congruent audiovisual processing responsible for speech enhancements and incongruent audiovisual processing responsible for McGurk processing rely on two distinct neural pathways: 1.) one that enhances the initial encoding of auditory information based on the information passed by the visual cues and 2.) one that modifies auditory representation based on the integrated audiovisual information. The first early feedforward process occurs early in the processing stream and aligns the auditory encoding with the temporal and acoustic features of the accompanying visual input (Arnal et al., 2009; Arnal et al., 2011; van Wassenhove et al., 2005). The later feedback process is engaged following the detection of mismatch between the auditory and visual cues, with the brain subsequently altering and adjusting the processing of the unisensory speech based on the combined audiovisual information (Arnal et al., 2011; Kayser & Logothetis, 2009; Olasagasti, Bouton, & Giraud, 2015). Given that this late feedback process is facilitated by higher order areas like pSTS, the limited reliance of this pathway during congruent audiovisual processing can explain why inhibitory stimulation to pSTS shows limited disruption on the audiovisual speech enhancement benefits.

In summary, our data demonstrate that while TMS to the left pSTS can limit audiovisual speech integration and result in a weaker McGurk effect, it does not universally reduce the ecologically important benefits of congruent visual information on speech perception. This suggests a dissociation in neural mechanisms such that the pSTS reflects only one of multiple critical areas necessary for audiovisual speech interactions.

Chapter 5 General Discussion

In this dissertation, I investigated the causal brain regions that facilitate two important components of everyday language: semantic naming processes and audiovisual speech integration. Given the critical role of these processes in healthy social function, numerous studies aimed to identify brain areas and networks subserving speech production and perception function. However, as majority of these studies rely on non-invasive, correlative neuroimaging methods such as fMRI, EEG, MEG, or PET, it remains unclear which brain regions that show activations through these means are necessary to maintain the function and which regions are merely associated with the process. For example, correlative methods can reveal involvement of brain regions that are non-specific to the cognitive task; this consequently limits us from having a precise mapping between a brain region and cognitive function. To address this shortcoming and expand our understanding of how the brain enables language and speech perception, my dissertation examined the unique and causal contributions of different brain regions utilizing causal approaches such as lesion mapping and TMS.

In study 1, we established that lesions in the middle temporal region causally facilitate semantic naming function using brain tumor patients demonstrating that tumor lesion mapping successfully replicates the widely used stroke lesion mapping approaches. Traditionally, in comparison to the sheer number of lesion mapping using post-stroke aphasic lesions, there have been a significantly smaller number of studies that investigate the relationship between tumor locations and the associated behavioral deficits. Study 1 provides evidence that tumor lesions mapping allows valid conclusions to be drawn about the neural origins of cognitive functions.,

This is not to say that the differences in the etiologies producing stroke versus tumor lesion should be overlooked (see van Grinsven et al., 2023 for review). Instead, lesion mapping literature across these two patient populations can be used as a tool to review the diverging and converging findings to further expand our understanding of the brain function.

In Study 1, we also demonstrated that different sub-regions of the tumor can better explain the relationship between lesion and behavioral symptom. Compared to the more homogenous makeup of stroke lesions, different regions within a single tumor lesion demonstrate unique MRI-characteristics that can help break down the lesions into smaller regions. These sub-regions can capture different ratios of healthy to tumor cells with a greater ratio of normal to tumor cells the further we move away from the center of the tumor mass (Ji et al., 2015). Consequently, these sub-regions can reflect the degree of remaining functionality within the lesion and reveal whether the area can still retain its normal cognitive functions. As hypothesized, our results revealed that the tumor core, the main tumor mass, is a better predictor for the degree and location of the behavioral deficit compared to the FLAIR hyperintensity which encompasses the edema. Our findings provided both methodological and clinical implications for future studies. Using an underutilized clinical population within the lesion mapping literature, our finding highlights the need for future studies to consider the distinction between different areas within a tumor lesion such that the identified relationship between lesion and behavior can be more precise and accurate. Our findings similarly reinforce the need for consideration of these different areas in the resection of gliomas which balances the need to maximize tumor tissue resected while minimizing the behavioral impairments.

In Study 2, we extended our findings from Study 1 by investigating the brain regions involved in audiovisual processing through tumor lesion mapping to address whether the same

mechanism subserves both the congruent audiovisual processing and McGurk processing. Since the publication of McGurk illusion in 1976 (McGurk and MacDonald), McGurk perception has become an extremely popular way of demonstrating and studying audiovisual speech integration. In most of these studies, McGurk perceptibility was used as a broad measure of audiovisual integration without much consideration for its generalizability to natural audiovisual speech. Inevitably, researchers raised valid and important concerns that McGurk perception may comprise of distinct mechanisms that make it different from conversational audiovisual speech. Therefore, it became important to understand whether natural speech perception and McGurk speech perception are in fact the same processes and if not, distinguish which of the findings established using McGurk stimuli can be generalized to everyday speech processing and which were unique to processing incongruent fused audiovisual information.

To this goal, Study 2 found support, albeit weak, for the view that a lesion in pSTS, the area found to be critical for McGurk processing, disrupts McGurk perception without disrupting congruent audiovisual speech processing. Specifically, while we were unable to conduct voxel lesion symptom mapping approach due to the limitations in lesion coverage, our region of interest analyses using various ROI sizes demonstrated that those with lesions in the left pSTS were less likely to report McGurk fusion perceptions with no impact on congruent audiovisual processing. Future work with a broader coverage of the tumor core throughout the brain will be necessary to conduct voxel level mapping and identify distinction between brain areas necessary for each process. However, even without voxel level lesion symptom mapping, our TMS work in Study 3 was able to extend and strengthen our Study 2's finding by reporting similar findings in healthy adults.

The two different causal approaches to brain mapping used in this dissertation, voxel lesion symptom mapping and TMS, like any other approach, face certain limitations. Fortunately, the advantages of each methods can complement the limitations of the other. For example, voxel lesion symptom mapping, with enough lesions, allows causal mapping of the whole brain. However, clinical subjects with specific conditions are inevitably difficult to recruit and since researchers do not dictate the location of lesions, it's hard to control for the number of subjects who present with lesions in a particular area to ensure sufficient power in the analysis. In addition, using any type of clinical population raises the question of whether the findings will generalize to healthy individuals.

On the other hand, TMS has the advantage of being a non-invasive procedure with minimally risky and temporary effects that can be performed on healthy individuals. At the same time, TMS faces the limitation in that this method requires researchers to identify a pre-defined area of interest for stimulation. As such, TMS can only reveal information about the specific brain region that was previously identified and targeted. Therefore, TMS is ideal for answering questions that already have strong priors as to which brain regions are involved. By utilizing both lesion mapping and TMS in addressing our research questions, we're able to take advantage of the larger areas of the brain for investigation through lesion mapping and the ability to then replicate this causal association in healthy individuals through TMS.

In Study 3, we stimulated the left pSTS with single-pulse TMS and found that electrical stimulation of this region reduces the likelihood of McGurk perception while having no effect on congruent audiovisual perception. This demonstrates that left pSTS is not critical for processing ecologically valid congruent audiovisual information while it is necessary for combining

information from mismatching auditory and visual information to create a percept that captures information from both modalities.

To contextualize our findings within existing literature, we consider the growing evidence in multisensory literature that suggest audiovisual information is integrated along multiple stages in speech perception (Arnal et al., 2011, Eskelund et al., 2011; Gao et al., 2023; O'Sullivan et al., 2021; Peelle and Sommers, 2015). According to the multistage integration model, the integration of cross-modal information occurs both at the level of the primary sensory cortices between the auditory and visual cortex and at the level of higher order areas such as STS, inferior parietal lobule (IPL) and prefrontal cortex. Based on neuroarchitecture in nonhuman primates, researchers discovered that STS receives convergent inputs from the primary auditory and visual cortices (Seltzer and Pandya, 1994; Lewis and Van Essen, 2000) and consists of neurons that preferentially respond to auditory, visual, and audiovisual stimuli (Beauchamp et al., 2004; Bruce et al., 1981; Dahl et al., 2009). STS, therefore, has been widely considered the multisensory hub where unimodal information converges. The multistage integration model specifies that in the early stage of integration, the visual speech can reset the ongoing phase of cortical oscillations in the auditory cortex, thereby conferring temporal information to the auditory cortex (Luo and Poeppel, 2010; Megevand et al., 2020; Schroeder et al., 2008; Zion-Golumbic et al., 2013) and during the late stage of integration, higher order regions such as pSTS integrate the outputs of the lower-level processing (articulatory representation of visual speech and acoustic information of auditory speech) by weighing the reliability of each modality to create a unified percept of speech. It has been proposed that in the early stage of integration, visual information provides correlated, or redundant cues to the auditory signal (known as the *correlated mode*), whereas during the late stage of integration,

visual information provides complementary information about the auditory signal (known as the *complementary mode*), supporting a compensatory role as needed (Campbell, 2008; Peelle and Sommers, 2015; O'Sullivan et al., 2021).

Placing this into the context of the dissertation, our data suggest that congruent audiovisual speech perception relies more heavily on the *correlated mode* during the early-stage integration, whereas incongruent McGurk perception relies more heavily on the *complementary mode* during the late-stage integration that takes place in the pSTS. Specifically, the correlated mode may be the process of optimizing the feedforward encoding of auditory information whereas the complementary mode is the process of optimizing the later feedback processes that can alter auditory representation. It is possible that in cases where the visual system cannot contribute new information to the auditory system, there's little need to engage the complementary mode or the feedback process. This may result in speech perception process bypassing or relying less heavily on higher order areas such as the pSTS. However, in cases where the visual system can in fact recover information about the acoustic speech that cannot be extracted by the auditory system alone, i.e., in noisy environments or when there isn't a 100% match with auditory and visual information, complementary mode, which relies on the left pSTS, may become activated. In support of this view, an fMRI study by Nath and Beauchamp (2011) showed that there is an increase in the functional connectivity between STS and visual cortex during noisy speech perception compared to non-noisy speech. This suggests that STS may be recruited in more challenging or complex audiovisual contexts where pSTS is needed to evaluate the reliability of stimuli from each modality to create the most statistically likely percept of the stimuli.

Despite the converging evidence from my dissertation suggesting the two independent mechanisms subserve congruent audiovisual perception and McGurk perception, our work still leaves open the important question of whether there exist brain areas that are causally involved in maintaining congruent audiovisual processing that are not involved in McGurk processing. This open question may benefit from future studies that make use of sufficiently powered VLSM study to investigate the double dissociation between congruent audiovisual processing and McGurk processing. In addition, a large-scale longitudinal study of tumor patients with lesions within the pSTS may be used to investigate how different audiovisual processing recovers over time, focusing on whether the recovery of McGurk processing and congruent audiovisual processing is closely associated. These future studies can help bring more compelling evidence that visual facilitation in congruent speech and visual modulations in McGurk percept, are supported by two distinct mechanisms. This future works that can identify brain regions that are causally involved in congruent processing while remaining uninvolved in McGurk processing will bring significant contributions to the mapping of human brain function.

In summary, my dissertation work demonstrated that tumor lesions may be used to conduct lesion symptom mapping and utilized this specific approach, in conjunction with another causal approach, TMS, to demonstrate that while pSTS is responsible for facilitating the illusory McGurk perception, it is not responsible for facilitating the benefits of congruent visual information on speech perception. Our work adds to the growing body of evidence suggesting that McGurk perceptibility should not be considered a general assay of audiovisual integration as there are multiple mechanisms that support what we generalize as “audiovisual integration.” Importantly, future studies need to make a clear distinction on which process they’re investigating so that inferences drawn from those studies can apply to the correct mechanism. By

clarifying this distinction between modulatory audiovisual processing and facilitatory audiovisual processing, researchers can better understand the different components that contribute simultaneously to general audiovisual integration processes.

Bibliography

Aabedi, A. A., Kakaizada, S., Young, J. S., Ahn, E., Weissman, D. H., Berger, M. S., Brang, D., & Hervey-Jumper, S. L. (2021). Balancing task sensitivity with reliability for multimodal language assessments. *Journal of Neurosurgery, 135*(6), 1817-1824.

Aabedi, A. A., Lipkin, B., Kaur, J., Kakaizada, S., Valdivia, C., Reihl, S., Young, J. S., Lee, A. T., Krishna, S., & Berger, M. S. (2021). Functional alterations in cortical processing of speech in glioma-infiltrated cortex. *Proceedings of the National Academy of Sciences, 118*(46), e2108959118.

Adam, G., Ferrier, M., Patsoura, S., Gramada, R., Meluchova, Z., Cazzola, V., Darcourt, J., Cognard, C., Viguier, A., & Bonneville, F. (2018). Magnetic resonance imaging of arterial stroke mimics: a pictorial review. *Insights into Imaging, 9*(5), 815-831.

Alsius, A., Navarra, J., & Soto-Faraco, S. (2007). Attention to touch weakens audiovisual speech integration. *Experimental Brain Research, 183*, 399-404.

Alsius, A., Paré, M., & Munhall, K. G. (2018). Forty years after hearing lips and seeing voices: the McGurk effect revisited. *Multisensory Research, 31*(1-2), 111-144.

Alsius, A., Wayne, R. V., Paré, M., & Munhall, K. G. (2016). High visual resolution matters in audiovisual speech perception, but only for some. *Attention, Perception, & Psychophysics, 78*, 1472-1487.

Anderson, S. W., Damasio, H., & Tranel, D. (1990). Neuropsychological impairments associated with lesions caused by tumor or stroke. *Archives of Neurology, 47*(4), 397-405.

Andersen, T. S., & Starrfelt, R. (2015). Audiovisual integration of speech in a patient with Broca's Aphasia. *Frontiers in Psychology, 6*, 435.

Arnal, L. H., Morillon, B., Kell, C. A., & Giraud, A. L. (2009). Dual neural routing of visual facilitation in speech processing. *The Journal of Neuroscience, 29*(43), 13445-13453.

Arnal, L. H., Wyart, V., & Giraud, A. L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nature Neuroscience, 14*(6), 797.

Baart, M., Stekelenburg, J. J., & Vroomen, J. (2014). Electrophysiological evidence for speech-specific audiovisual integration. *Neuropsychologia, 53*, 115-121.

Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R. T., Berger, C., Ha, S. M., & Rozycki, M. (2018). Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv preprint arXiv:1811.02629.

Baldo, J. V., Arévalo, A., Patterson, J. P., & Dronkers, N. F. (2013). Grey and white matter correlates of picture naming: evidence from a voxel-based lesion analysis of the Boston Naming Test. *Cortex, 49*(3), 658-667.

Banerjee, S., & Mitra, S. (2020). Novel volumetric sub-region segmentation in brain tumors. *Frontiers in Computational Neuroscience, 14*, 3.

Barajas Jr, R. F., Hodgson, J. G., Chang, J. S., Vandenberg, S. R., Yeh, R.-F., Parsa, A. T., McDermott, M. W., Berger, M. S., Dillon, W. P., & Cha, S. (2010). Glioblastoma multiforme regional genetic and cellular expression patterns: influence on anatomic and physiologic MR imaging. *Radiology, 254*(2), 564-576.

Barajas Jr, R. F., Phillips, J. J., Parvataneni, R., Molinaro, A., Essock-Burns, E., Bourne, G., ... & Nelson, S. J. (2012). Regional variation in histopathologic features of tumor specimens from treatment-naive glioblastoma correlates with anatomic and physiologic MR Imaging. Neuro-oncology, 14(7), 942-954.

Basu Mallick, D., F Magnotti, J., & S Beauchamp, M. (2015). Variability and stability in the McGurk effect: contributions of participants, stimuli, time, and response type. *Psychonomic Bulletin & Review*, *22*, 1299-1307.

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronkers, N. F. (2003). Voxel-based lesion–symptom mapping. *Nature Neuroscience*, *6*(5), 448-450.

Baum, S. H., Martin, R. C., Hamilton, A. C., & Beauchamp, M. S. (2012). Multisensory speech perception without the left superior temporal sulcus. *NeuroImage*, *62*(3), 1825-1832.

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. Neuron, 41(5), 809-823.

Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *The Journal of Neuroscience*, *30*(7), 2414-2417.

Benoit, M. M., Raij, T., Lin, F. H., Jääskeläinen, I. P., & Stufflebeam, S. (2010). Primary and multisensory cortical activity is correlated with audiovisual percepts. *Human Brain Mapping*, *31*(4), 526-538.

Bernstein, L. E., Lu, Z. L., & Jiang, J. (2008). Quantified acoustic–optical speech signal incongruity identifies cortical sites of audiovisual speech processing. *Brain Research*, *1242*, 172-184.

Besle, J., Fischer, C., Bidet-Caulet, A., Lecaignard, F., Bertrand, O., & Giard, M. H. (2008). Visual activation and audiovisual interactions in the auditory cortex during speech perception: intracranial recordings in humans. *The Journal of Neuroscience*, *28*(52), 14301-14310.

Besle, J., Fort, A., Delpuech, C., & Giard, M. H. (2004). Bimodal speech: early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, *20*(8), 2225-2234.

Binder, J. R., Tong, J. Q., Pillay, S. B., Conant, L. L., Humphries, C. J., Raghavan, M., Mueller, W. M., Busch, R. M., Allen, L., & Gross, W. L. (2020). Temporal lobe regions essential for preserved picture naming after left temporal epilepsy surgery. *Epilepsia*, *61*(9), 1939-1948.

Blumstein, S. E. (1994). Impairments of speech production and speech perception in aphasia. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *346*(1315), 29-36.

Boatman, D. (2004). Cortical bases of speech perception: evidence from functional lesion studies. Cognition, 92(1-2), 47-65.

Boss, S. M., Moustafa, R. R., Moustafa, M. A., El Sadek, A., Mostafa, M. M., & Aref, H. M. (2019). Lesion homogeneity on diffusion-weighted imaging is a marker of outcome in acute ischemic stroke. *The Egyptian Journal of Neurology, Psychiatry and Neurosurgery*, *55*, 1-4.

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433-436.

Brang, D. (2019). The Stolen Voice Illusion. *Perception*, *48*(8), 649-667.

Brang, D., Plass, J., Kakaizada, S., & Hervey-Jumper, S. L. (2020). Auditory-Visual Speech Behaviors are Resilient to Left pSTS Damage. bioRxiv, 2020.2009. 2026.314799.
Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science*, *307*(5712), 1118-1121.
Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. Journal of neurophysiology, 46(2), 369-384.
Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior research methods, 41(4), 977-990.
Buchan, J. N., Paré, M., & Munhall, K. G. (2008). The effect of varying talker identity and listening conditions on gaze behavior during audiovisual speech perception. *Brain Research*, *1242*, 162-171.
Buchan, J. N., & Munhall, K. G. (2011). The influence of selective attention to auditory and visual speech on the integration of audiovisual speech information. *Perception, 40*(10), 1164-1182.
Campbell, R. (2008). The processing of audio-visual speech: empirical and neural bases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1001-1010.
Cella, D., Lai, J.-S., Nowinski, C., Victorson, D., Peterman, A., Miller, D., Bethoux, F., Heinemann, A., Rubin, S., & Cavazos, J. (2012). Neuro-QOL: brief measures of health-related quality of life for clinical research in neurology. *Neurology, 78*(23), 1860-1867.
Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology, 5*(7), e1000436. doi:10.1371/journal.pcbi.1000436
Cipolotti, L., Healy, C., Chan, E., Bolsover, F., Lecce, F., White, M., Spanò, B., Shallice, T., & Bozzali, M. (2015). The impact of different aetiologies on the cognitive performance of frontal patients. Neuropsychologia, 68, 21-30.
Cohen, J. (1988). The effect size. Statistical power analysis for the behavioral sciences, 77-83.
Cuypers, K., Thijs, H., & Meesen, R. L. (2014). Optimization of the transcranial magnetic stimulation protocol by defining a reliable estimate for corticospinal excitability. *PloS One, 9*(1), e86380.
Dahl, C. D., Logothetis, N. K., & Kayser, C. (2009). Spatial organization of multisensory responses in temporal association cortex. Journal of Neuroscience, 29(38), 11924-11932.
Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., & Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition, 92*(1-2), 179-229.
DeMarco, A. T., & Turkeltaub, P. E. (2018). A multivariate lesion symptom mapping toolbox and examination of lesion-volume biases and correction methods in lesion-symptom mapping (1065-9471).
Desmurget, M., Bonnetblanc, F., & Duffau, H. (2007). Contrasting acute and slow-growing lesions: a new door to brain plasticity. *Brain, 130*(4), 898-914.
Dronkers, N. F., Wilkins, D. P., Van Valin Jr, R. D., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition, 92*(1-2), 145-177.
Duffau, H. (2005). Lessons from brain mapping in surgery for low-grade glioma: insights into associations between tumour and brain plasticity. *The Lancet Neurology, 4*(8), 476-486.

Duffau, H. (2011). Do brain tumours allow valid conclusions on the localisation of human brain functions? *Cortex*, *47*(8), 1016-1017.

Eidel, O., Burth, S., Neumann, J.-O., Kieslich, P. J., Sahm, F., Jungk, C., Kickingereder, P., Bickelhaupt, S., Mundiyanapurath, S., & Bäumer, P. (2017). Tumor infiltration in enhancing and non-enhancing parts of glioblastoma: a correlation with histopathology. *PloS One*, *12*(1), e0169292.

Erickson, L. C., Zielinski, B. A., Zielinski, J. E., Liu, G., Turkeltaub, P. E., Leaver, A. M., & Rauschecker, J. P. (2014). Distinct cortical locations for integration of audiovisual speech and the McGurk effect. Frontiers in psychology, 5, 534.

Eskelund, K., Tuomainen, J., & Andersen, T. S. (2011). Multistage audiovisual integration of speech: dissociating identification and detection. *Experimental Brain Research*, *208*(3), 447-457.

Faivre, N., Mudrik, L., Schwartz, N., & Koch, C. (2014). Multisensory integration in complete unawareness: Evidence from audiovisual congruency priming. Psychological Science, 25(11), 2006-2016.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.

Fekonja, L. S., Wang, Z., Doppelbauer, L., Vajkoczy, P., Picht, T., Pulvermüller, F., & Dreyer, F. R. (2021). Lesion-symptom mapping of language impairments in patients suffering from left perisylvian gliomas. cortex, 144, 1-14.

Feldman, J. I., Conrad, J. G., Kuang, W., Tu, A., Liu, Y., Simon, D. M., Wallace, M. T., & Woynaroski, T. G. (2022). Relations between the McGurk effect, social and communication skill, and autistic features in children with and without autism. *Journal of Autism and Developmental Disorders*, *52*(5), 1920-1928.

Fernández, L. M., Visser, M., Ventura-Campos, N., Ávila, C., & Soto-Faraco, S. (2015). Top-down attention regulates the neural expression of audiovisual integration. *NeuroImage*, *119*, 272-285.

Fingelkurts, A. A., Fingelkurts, A. A., Krause, C. M., Möttönen, R., & Sams, M. (2003). Cortical operational synchrony during audio–visual speech integration. Brain and language, 85(2), 297-312.

Fussen, S., De Boeck, B. W., Zellweger, M. J., Bremerich, J., Goetschalckx, K., Zuber, M., & Buser, P. T. (2011). Cardiovascular magnetic resonance imaging for diagnosis and clinical management of suspected cardiac masses and tumours. *European Heart Journal*, *32*(12), 1551-1560.

Ganesan, K., Plass, J., Beltz, A. M., Liu, Z., Grabowecky, M., Suzuki, S., . . . Tao, J. X. (2020). Visual speech differentially modulates beta, theta, and high gamma bands in auditory cortex. *bioRxiv*.

Ganesan G., Cao C., Demidenko M., Jahn A., Stacey W., Wasade V., Brang D., (in review). Auditory cortex encodes lipreading information through spatially distributed activity. *bioRxiv*.

Gao, C., Green, J. J., Yang, X., Oh, S., Kim, J., & Shinkareva, S. V. (2023). Audiovisual integration in the human brain: a coordinate-based meta-analysis. Cerebral Cortex, 33(9), 5574-5584.

Gelder, B. d., Vroomen, J., & Van der Heide, L. (1991). Face recognition and lip-reading in autism. *European Journal of Cognitive Psychology*, *3*(1), 69-86.

Gill, B. J., Pisapia, D. J., Malone, H. R., Goldstein, H., Lei, L., Sonabend, A., Yun, J., Samanamud, J., Sims, J. S., & Banu, M. (2014). MRI-localized biopsies reveal subtype-specific differences in molecular and cellular composition at the margins of glioblastoma. *Proceedings of the National Academy of Sciences*, *111*(34), 12550-12555.

Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *Journal of the Acoustical Society of America*, *108*(3), 1197-1208. doi:Doi 10.1121/1.1288668

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95-112.

Gurler, D., Doyle, N., Walker, E., Magnotti, J., & Beauchamp, M. (2015). A link between individual differences in multisensory speech perception and eye movements. *Attention, Perception, & Psychophysics*, *77*(4), 1333-1341.

Habets, E. J., Hendriks, E. J., Taphoorn, M. J., Douw, L., Zwinderman, A. H., Vandertop, W. P., Barkhof F., De Witt Hamer P.C., & Klein, M. (2019). Association between tumor location and neurocognitive functioning using tumor localization maps. Journal of neuro-oncology, 144, 573-582.

Hallett, M. (2000). Transcranial magnetic stimulation and the human brain. *Nature*, *406*(6792), 147-150.

Hawkins-Daarud, A., Rockne, R. C., Anderson, A. R., & Swanson, K. R. (2013). Modeling tumor-associated edema in gliomas during anti-angiogenic therapy and its impact on imageable tumor. Frontiers in oncology, 3, 66.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393-402.

Hickok, G., Rogalsky, C., Matchin, W., Basilakos, A., Cai, J., Pillay, S., Ferrill, M., Mickelsen, S., Anderson, S. W., & Love, T. (2018). Neural networks supporting audiovisual integration for speech: A large-scale lesion study. *Cortex*, *103*, 360-371.

Herbet, G., & Duffau, H. (2022). Contribution of the medial eye field network to the voluntary deployment of visuospatial attention. *Nature Communications*, *13*(1), 328.

Hervey-Jumper, S. L., Zhang, Y., Phillips, J. J., Morshed, R. A., Young, J. S., McCoy, L., Lafontaine, M., Luks, T., Ammanuel, S., & Kakaizada, S. (2023). Interactive effects of molecular, therapeutic, and patient factors on outcome of diffuse low-grade glioma. *Journal of Clinical Oncology*, *41*(11), 2029-2042.

Hickok, G., Rogalsky, C., Matchin, W., Basilakos, A., Cai, J., Pillay, S., . . . Love, T. (2018). Neural networks supporting audiovisual integration for speech: A large-scale lesion study. *Cortex*, *103*, 360-371.

Hillis, A. E., Wityk, R. J., Tuffiash, E., Beauchamp, N. J., Jacobs, M. A., Barker, P. B., & Selnes, O. A. (2001). Hypoperfusion of Wernicke's area predicts severity of semantic deficit in acute stroke. *Annals of Neurology*, *50*(5), 561-566.

Hoffman, P., & Morcom, A. M. (2018). Age-related changes in the neural networks supporting semantic cognition: A meta-analysis of 47 functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, *84*, 134-150.

Huang, Y. Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron*, *45*(2), 201-206.

Husstedt, H. W., Sickert, M., Köstler, H., Haubitz, B., & Becker, H. (2000). Diagnostic value of the fast-FLAIR sequence in MR imaging of intracranial tumors. European radiology, 10, 745-752.

Irwin, J., & DiBlasi, L. (2017). Audiovisual speech perception: A new approach and implications for clinical populations. *Language and Linguistics Compass, 11*(3), 77-91.
Irwin, J. R., Frost, S. J., Mencl, W. E., Chen, H., & Fowler, C. A. (2011). Functional activation for imitation of seen and heard speech. *Journal of Neurolinguistics*, *24*(6), 611-618.
Ji, M., Lewis, S., Camelo-Piragua, S., Ramkissoon, S. H., Snuderl, M., Venneti, S., Fisher-Hubbard, A., Garrard, M., Fu, D., & Wang, A. C. (2015). Detection of human brain tumor infiltration with quantitative stimulated Raman scattering microscopy. *Science Translational Medicine*, *7*(309), 309ra163-309ra163.
Jones, J. A., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport*, *14*(8), 1129-1133.
Jung, J., Bungert, A., Bowtell, R., & Jackson, S. R. (2016). Vertex stimulation as a control site for transcranial magnetic stimulation: a concurrent TMS/fMRI study. *Brain Stimulation*, *9*(1), 58-64.
Kallioniemi, E., & Julkunen, P. (2016). Alternative stimulation intensities for mapping cortical motor area with navigated TMS. *Brain Topography*, *29*, 395-404.
Kaplan, E., Goodglass, H., & Weintraub, S. (2001). Boston naming test.
Karnath, H.O., & Steinbach, J. P. (2010). Do brain tumours allow valid conclusions on the localisation of human brain functions?--Objections. *Cortex*, *47*(8), 1004-1006.
Kayser, C., & Logothetis, N. (2009). Directed interactions between auditory and superior temporal cortices and their role in sensory integration. *Frontiers in Integrative Neuroscience*, *3*, 7.
Keil, J., & Senkowski, D. (2018). Neural oscillations orchestrate multisensory processing. *The Neuroscientist*, *24*(6), 609-626.
Kissela, B. M., Khoury, J. C., Alwell, K., Moomaw, C. J., Woo, D., Adeoye, O., Flaherty, M. L., Khatri, P., Ferioli, S., La Rosa, F. D., Broderick, J. P., & Kleindorfer, D. O. (2012). Age at stroke: temporal trends in stroke incidence in a large, biracial population. Neurology, 79(17), 1781-1787.
Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3?
Krieger-Redwood, K., & Jefferies, E. (2014). TMS interferes with lexical-semantic retrieval in left inferior frontal gyrus and posterior middle temporal gyrus: Evidence from cyclical picture naming. *Neuropsychologia*, *64*, 24-32.
Krishna, S., Choudhury, A., Keough, M. B., Seo, K., Ni, L., Kakaizada, S., Lee, A., Aabedi, A., Popova, G., & Lipkin, B. (2023). Glioblastoma remodelling of human neural circuits decreases survival. *Nature*, 1-9.
Lange, J., Christian, N., & Schnitzler, A. (2013). Audio–visual congruency alters power and coherence of oscillatory activity within and between cortical areas. Neuroimage, 79, 111-120.
Larsson, H. B. W., Stubgaard, M., Frederiksen, J. L., Jensen, M., Henriksen, O., & Paulson, O. B. (1990). Quantitation of blood-brain barrier defect by magnetic resonance imaging and gadolinium-DTPA in patients with multiple sclerosis and brain tumors. Magnetic Resonance in Medicine, 16(1), 117-131.
Levelt, W. J., Praamstra, P., Meyer, A. S., Helenius, P., & Salmelin, R. (1998). An MEG study of picture naming. Journal of cognitive neuroscience, 10(5), 553-567.
Lewis, J. W., & Van Essen, D. C. (2000). Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. Journal of Comparative Neurology, 428(1), 112-137.

Lin, Z., Yang, R., Li, K., Yi, G., Li, Z., Guo, J., Zhang, Z., Junxiang, P., Liu, Y., & Qi, S. (2020). Establishment of age group classification for risk stratification in glioma patients. *BMC Neurology*, *20*, 1-11.

Luo, H., Liu, Z., & Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biology*, *8*(8), e1000445.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics*, *24*(3), 253-257.

MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, *21*(2), 131-141.

Magnotti, J. F., Ma, W. J., & Beauchamp, M. S. (2013). Causal inference of asynchronous audiovisual speech. *Frontiers in Psychology, 4,* 798. doi:10.3389/fpsyg.2013.00798

Mah, Y. H., Husain, M., Rees, G., & Nachev, P. (2014). Human brain lesion-deficit inference remapped. *Brain*, *137*(9), 2522-2531.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746-748.

Mégevand, P., Mercier, M. R., Groppe, D. M., Golumbic, E. Z., Mesgarani, N., Beauchamp, M. S., Schroeder C. E., & Mehta, A. D. (2020). Crossmodal phase reset and evoked responses provide complementary mechanisms for the influence of visual speech in auditory cortex. Journal of Neuroscience, 40(44), 8530-8542.

Members, W. G., Roger, V. L., Go, A. S., Lloyd-Jones, D. M., Benjamin, E. J., Berry, J. D., Borden, W. B., Bravata, D. M., Dai, S., & Ford, E. S. (2012). Heart disease and stroke statistics—2012 update: a report from the American Heart Association. *Circulation*, *125*(1), e2-e220.

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., & Wiest, R. (2014). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*, *34*(10), 1993-2024.

Menze, B. H., Van Leemput, K., Lashkari, D., Riklin-Raviv, T., Geremia, E., Alberts, E., Gruber, P., Wegener, S., Weber, M.-A., & Székely, G. (2015). A generative probabilistic model and discriminative extensions for brain lesion segmentation—with application to tumor and stroke. *IEEE Transactions on Medical Imaging*, *35*(4), 933-946.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746-748.

Mesulam, M. M. (1981). A cortical network for directed attention and unilateral neglect. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *10*(4), 309-325.

Mills, K. R., & Nithi, K. A. (1997). Corticomotor threshold to magnetic stimulation: normal values and repeatability. *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine*, *20*(5), 570-576.

Molinaro, A. M., Hervey-Jumper, S., Morshed, R. A., Young, J., Han, S. J., Chunduru, P., Zhang, Y., Phillips, J. J., Shai, A., & Lafontaine, M. (2020). Association of maximal extent of resection of contrast-enhanced and non–contrast-enhanced tumor with survival within molecular subgroups of patients with newly diagnosed glioblastoma. *JAMA Oncology*, *6*(4), 495-503.

Munhall, K., MacDonald, E., Byrne, S., & Johnsrude, I. (2009). Talkers alter vowel production in response to real-time format perturbation even when instructed not to compensate. *The Journal of the Acoustical Society of America, 125*(1), 384-390.

Nath, A. R., & Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *NeuroImage, 59*(1), 781-787.

Nath, A. R., Fava, E. E., & Beauchamp, M. S. (2011). Neural correlates of interindividual differences in children's audiovisual speech perception. *Journal of Neuroscience, 31*(39), 13963-13971.

Olasagasti, I., Bouton, S., & Giraud, A. (2015). Prediction across sensory modalities: A neurocomputational model of the McGurk effect. *Cortex, 68*, 61-65.

O'Sullivan, A. E., Crosse, M. J., Di Liberto, G. M., de Cheveigné, A., & Lalor, E. C. (2021). Neurophysiological indices of audiovisual speech processing reveal a hierarchy of multisensory integration effects. Journal of Neuroscience, 41(23), 4991-5003.

Palmer, T., & Ramsey, A. (2012). The function of consciousness in multisensory integration. *Cognition, 125*(3), 353-364.

Paré, M., Richler, R. C., ten Hove, M., & Munhall, K. (2003). Gaze behavior in audiovisual speech perception: The influence of ocular fixations on the McGurk effect. *Perception & Psychophysics*, *65*, 553-567.

Parrish-Novak, J., Holland, E. C., & Olson, J. M. (2015). Image guided tumor resection. Cancer journal (Sudbury, Mass.), 21(3), 206.

Pedersen, P. M., Stig Jørgensen, H., Nakayama, H., Raaschou, H. O., & Olsen, T. S. (1995). Aphasia in acute stroke: incidence, determinants, and recovery. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *38*(4), 659-666.

Peelle, J. E., & Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. Cortex, 68, 169-181.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437-442.

Plass, J., Brang, D., Suzuki, S., & Grabowecky, M. (2020). Vision perceptually restores auditory spectral dynamics in speech. *Proceedings of the National Academy of Sciences, 117*(29), 16920-16927. doi:10.1073/pnas.2002887117

Plass, J., Guzman-Martinez, E., Ortega, L., Grabowecky, M., & Suzuki, S. (2014). Lip reading without awareness. *Psychological Science, 25*(9), 1835-1837.

Puschmann, S., Daeglau, M., Stropahl, M., Mirkovic, B., Rosemann, S., Thiel, C. M., & Debener, S. (2019). Hearing-impaired listeners show increased audiovisual benefit when listening to speech in noise. *NeuroImage, 196,* 261-268.

Pustina, D., Avants, B., Faseyitan, O. K., Medaglia, J. D., & Coslett, H. B. (2018). Improved accuracy of lesion to symptom mapping with multivariate sparse canonical correlations. *Neuropsychologia, 115*, 154-166.

Qin, X., Liu, R., Akter, F., Qin, L., Xie, Q., Li, Y., Qiao, H., Zhao, W., Jian, Z., & Liu, R. (2021). Peri-tumoral brain edema associated with glioblastoma correlates with tumor recurrence. *Journal of Cancer, 12*(7), 2073.

Rathore, S., Akbari, H., Rozycki, M., Abdullah, K. G., Nasrallah, M. P., Binder, Z. A., Davuluri, R. V., Lustig, R. A., Dahmane, N., & Bilello, M. (2018). Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1. *Scientific Reports, 8*(1), 5087.

Roa Romero, Y., Senkowski, D., & Keil, J. (2015). Early and late beta-band power reflect audiovisual perception in the McGurk illusion. Journal of neurophysiology, 113(7), 2342-2350.

Ruben, R. J. (2000). Redefining the survival of the fittest: communication disorders in the 21st century. *The Laryngoscope*, *110*(2), 241-241.

Rennig, J., Wegner-Clemens, K., & Beauchamp, M. S. (2020). Face viewing behavior predicts multisensory gain during speech perception. *Psychonomic Bulletin & Review*, *27*(1), 70-77.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Molholm, S., Javitt, D. C., & Foxe, J. J. (2007). Impaired multisensory processing in schizophrenia: deficits in the visual enhancement of speech comprehension under noisy environmental conditions. *Schizophrenia Research*, *97*(1), 173-183.

Rossini, P. M., & Rossi, S. (2007). Transcranial magnetic stimulation: diagnostic, therapeutic, and research potential. *Neurology*, *68*(7), 484-488.

Saatlou, F. H., Rogasch, N. C., McNair, N. A., Biabani, M., Pillen, S. D., Marshall, T. R., & Bergmann, T. O. (2018). MAGIC: An open-source MATLAB toolbox for external control of transcranial magnetic stimulation devices. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation*, *11*(5), 1189-1191.

Schneider, F., Marcotte, K., Brisebois, A., Townsend, S. A. M., Smidarle, A. D., Loureiro, F., da Rosa Franco, A., Soder, R. B., Nikolaev, A., & Marrone, L. C. P. (2021). Neuroanatomical correlates of macrolinguistic aspects in narrative discourse in unilateral left and right hemisphere stroke: A voxel-based morphometry study. *Journal of Speech, Language, and Hearing Research*, *64*(5), 1650-1665.

Schoenegger, K., Oberndorfer, S., Wuschitz, B., Struhal, W., Hainfellner, J., Prayer, D., Heinzl, H., Lahrmann, H., Marosi, C., & Grisold, W. (2009). Peritumoral edema on MRI at initial diagnosis: an independent prognostic factor for glioblastoma? *European Journal of Neurology*, *16*(7), 874-878.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences*, *12*(3), 106-113.

Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Faseyitan, O., Brecher, A., Dell, G. S., & Coslett, H. B. (2009). Anterior temporal involvement in semantic word retrieval: voxel-based lesion-symptom mapping evidence from aphasia. *Brain*, *132*(12), 3411-3427.

Seijdel, N., Schoffelen, J.-M., Hagoort, P., & Drijvers, L. (2023). Attention drives visual processing and audiovisual integration during multimodal communication. *bioRxiv*, doi: https://doi.org/10.1101/2023.05.11.540320

Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neuroscience Research*, *47*(3), 277-287.

Seltzer, B., & Pandya, D. N. (1994). Parietal, temporal, and occipita projections to cortex of the superior temporal sulcus in the rhesus monkey: A retrograde tracer study. Journal of Comparative Neurology, 343(3), 445-463.

Shallice, T., Mussoni, A., D'Agostino, S., & Skrap, M. (2010). Right posterior cortical functions in a tumour patient series. *Cortex*, *46*(9), 1178-1188.

Snyder, K. M., Forseth, K. J., Donos, C., Rollo, P. S., Fischer-Baum, S., Breier, J., & Tandon, N. (2023). Critical role of the ventral temporal lobe in naming. Epilepsia.

Sondergaard, R. E., Martino, D., Kiss, Z. H., & Condliffe, E. G. (2021). TMS motor mapping methodology and reliability: a structured review. *Frontiers in Neuroscience*, *15*, 709368.

Stacey, J. E., Howard, C. J., Mitra, S., & Stacey, P. C. (2020). Audio-visual integration in noise: Influence of auditory and visual stimulus degradation on eye movements and perception of the McGurk effect. *Attention, Perception, & Psychophysics*, *82*, 3544-3557.

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience, 19*(12), 1964-1973.

Stevenson, R. A., Segers, M., Ferber, S., Barense, M. D., & Wallace, M. T. (2014). The impact of multisensory integration deficits on speech perception in children with autism spectrum disorders. Frontiers in psychology, 5, 379.

Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *Journal of the Acoustical Society of America, 26*(2), 212-215. doi:Doi 10.1121/1.1907309

Szycik, G. R., Stadler, J., Tempelmann, C., & Münte, T. F. (2012). Examining the McGurk illusion using high-field 7 Tesla functional MRI. *Frontiers in Human Neuroscience*, *6*, 95.

Tiippana, K. (2014). What is the McGurk effect? *Frontiers in Psychology, 5*, 725.

Tofts, P. S., & Kermode, A. G. (1991). Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts. Magnetic resonance in medicine, 17(2), 357-367.

Tuomainen, J., Andersen, T. S., Tiippana, K., & Sams, M. (2005). Audio–visual speech perception is special. Cognition, 96(1), B13-B22.

Vaidya, A. R., Pujara, M. S., Petrides, M., Murray, E. A., & Fellows, L. K. (2019). Lesion studies in contemporary neuroscience. Trends in cognitive sciences, 23(8), 653-671.

Van Engen, K. J., Dey, A., Sommers, M. S., & Peelle, J. E. (2022). Audiovisual speech perception: Moving beyond McGurk. *The Journal of the Acoustical Society of America*, 152(6), 3216-3225.

Van Engen, K. J., Xie, Z., & Chandrasekaran, B. (2017). Audiovisual sentence recognition not predicted by susceptibility to the McGurk effect. *Attention, Perception, & Psychophysics*, *79*(2), 396-403.

van Grinsven, E., Smits, A., van Kessel, E., Raemaekers, M., de Haan, E., Wajer, I. H., Ruijters, V., Philippens, M., Verhoeff, J., & Ramsey, N. (2023). The impact of etiology in lesion-symptom mapping–A direct comparison between tumor and stroke. *NeuroImage: Clinical*, *37*, 103305.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *PNAS*, *102*(4), 1181-1186.

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, *45*(3), 598-607.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Attention, Perception, & Psychophysics, 69*(5), 744-756.

Vigneau, M., Beaucousin, V., Hervé, P.-Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., & Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage*, *30*(4), 1414-1432.

Visser, M., Embleton, K. V., Jefferies, E., Parker, G., & Ralph, M. L. (2010). The inferior, anterior temporal lobes and semantic memory clarified: novel evidence from distortion-corrected fMRI. *Neuropsychologia*, *48*(6), 1689-1696.

Vroomen, J., & Stekelenburg, J. (2009). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *Journal of Cognitive Neuroscience, 22*(7), 1583-1596.

Walker, G. M., Schwartz, M. F., Kimberg, D. Y., Faseyitan, O., Brecher, A., Dell, G. S., & Coslett, H. B. (2011). Support for anterior temporal involvement in semantic error production in aphasia: new evidence from VLSM. *Brain and Language, 117*(3), 110-122.

Walsh, V., & Cowey, A. (2000). Transcranial magnetic stimulation and cognitive neuroscience. *Nature Reviews Neuroscience, 1*(1), 73-80.

Wang, Y., Rudd, A. G., & Wolfe, C. D. (2013). Age and ethnic disparities in incidence of stroke over time: the South London Stroke Register. *Stroke, 44*(12), 3298-3304.

Watanabe, M., Tanaka, R., & Takeda, N. (1992). Magnetic resonance imaging and histopathology of cerebral gliomas. Neuroradiology, 34, 463-469.

Williams, J. H., Massaro, D. W., Peel, N. J., Bosseler, A., & Suddendorf, T. (2004). Visual–auditory integration during speech imitation in autism. *Research in Developmental Disabilities, 25*(6), 559-575.

Wilson, S. M., Eriksson, D. K., Schneck, S. M., & Lucanie, J. M. (2018). A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. *PloS One, 13*(2), e0192773.

Wischnewski, M., & Schutter, D. J. (2015). Efficacy and time course of theta burst stimulation in healthy humans. *Brain Stimulation, 8*(4), 685-692.

Wu, C.-X., Lin, G.-S., Lin, Z.-X., Zhang, J.-D., Liu, S.-Y., & Zhou, C.-F. (2015). Peritumoral edema shown by MRI predicts poor clinical outcome in glioblastoma. *World Journal of Surgical Oncology, 13*(1), 1-9.

Xing, S., Lacey, E. H., Skipper-Kallal, L. M., Jiang, X., Harris-Love, M. L., Zeng, J., & Turkeltaub, P. E. (2016). Right hemisphere grey matter structure and language outcomes in chronic left hemisphere stroke. *Brain, 139*(1), 227-241.

Xu, Z., Shen, B., Taji, W., Sun, P., & Naya, Y. (2020). Convergence of distinct functional networks supporting naming and semantic recognition in the left inferior frontal gyrus. Human Brain Mapping, 41(9), 2389-2405.

Ye, Y., Cai, Z., Huang, B., He, Y., Zeng, P., Zou, G., Deng, W., Chen, H., & Huang, B. (2020). Fully-automated segmentation of nasopharyngeal carcinoma on dual-sequence MRI using convolutional neural networks. *Frontiers in Oncology, 10*, 166.

Zion-Golumbic, E. M. Z., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., Goodman R. R., Emerson R., Mehta A. D., Simon J. Z.,& Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. Neuron, 77(5), 980-991.

Zhang, J., Meng, Y., He, J., Xiang, Y., Wu, C., Wang, S., & Yuan, Z. (2019). McGurk effect by individuals with autism spectrum disorder and typically developing controls: A systematic review and meta-analysis. *Journal of Autism and Developmental Disorders, 49*(1), 34-43.

Zhang, Y., Kimberg, D. Y., Coslett, H. B., Schwartz, M. F., & Wang, Z. (2014). Multivariate lesion-symptom mapping using support vector regression. *Human Brain Mapping, 35*(12), 5861-5876.

Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F.-X., & Lorenzi, C. (2005). Deficits in speech perception predict language learning impairment. *Proceedings of the National Academy of Sciences, 102*(39), 14110-14115.