The Neurobiology of Audiovisual Integration:

A Voxel-Based Lesion Symptom Mapping Study

by

Julia Cai

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2017 by the
Graduate Supervisory Committee:

Corianne Rogalsky, Chair
Tamiko Azuma
Julie Liss

ARIZONA STATE UNIVERSITY

May 2017

ABSTRACT

Audiovisual (AV) integration is a fundamental component of face-to-face communication. Visual cues generally aid auditory comprehension of communicative intent through our innate ability to "fuse" auditory and visual information. However, our ability for multisensory integration can be affected by damage to the brain. Previous neuroimaging studies have indicated the superior temporal sulcus (STS) as the center for AV integration, while others suggest inferior frontal and motor regions. However, few studies have analyzed the effect of stroke or other brain damage on multisensory integration in humans. The present study examines the effect of lesion location on auditory and AV speech perception through behavioral and structural imaging methodologies in 41 left-hemisphere participants with chronic focal cerebral damage. Participants completed two behavioral tasks of speech perception: an auditory speech perception task and a classic McGurk paradigm measuring congruent (auditory and visual stimuli match) and incongruent (auditory and visual stimuli do not match, creating a "fused" percept of a novel stimulus) AV speech perception. Overall, participants performed well above chance on both tasks. Voxel-based lesion symptom mapping (VLSM) across all 41 participants identified several regions as critical for speech perception depending on trial type. Heschl's gyrus and the supramarginal gyrus were identified as critical for auditory speech perception, the basal ganglia was critical for speech perception in AV congruent trials, and the middle temporal gyrus/STS were critical in AV incongruent trials. VLSM analyses of the AV incongruent trials were used to further clarify the origin of "errors", i.e. lack of fusion. Auditory capture (auditory stimulus) responses were attributed to visual processing deficits caused by lesions in the posterior temporal lobe, whereas visual capture (visual stimulus) responses were attributed to lesions in the anterior temporal cortex, including the temporal pole, which is widely considered to be an amodal semantic hub. The implication of anterior temporal regions in AV integration is novel and warrants further study. The behavioral and VLSM results are discussed in relation to previous neuroimaging and case-study evidence; broadly, our findings coincide with previous work indicating that multisensory superior temporal cortex, not frontal motor circuits, are critical for AV integration.

TABLE OF CONTENTS

Page

Page

APPENDIX

LIST OF FIGURES

The Neurobiology of Audiovisual Integration:

A Voxel-Based Lesion Symptom Mapping Study

I. Visual Aspects of Communication

Effective face-to-face communication often includes both auditory and visual information. Visual information can affect communicative intent and can be obtained from nonverbal gestures, such as from the speaker's head and face. Examples include determination of emphatic stress (Bernstein et al., 1998) and prosody (Munhall et al., 2004) from the movements of the speaker's head and eyebrows. While auditory-only communication is often sufficient, audiovisual (AV) integration is a fundamental component of daily conversation; speech is typically both seen and heard. Listeners integrate both the auditory and visual information provided during communication to create a representation of the message. AV integration can increase comprehension of the information being communicated (Calvert et al., 1998) and can improve intelligibility of clear and distorted speech (Neely, 1956; Arnold & Hill, 2001; Ross et al., 2007), as experienced in a noisy room, for example. Visual cues often aid the auditory signal by providing phonemic cues for the listener to predict what the speaker is saying and/or predict what the person is going to say (Sumby & Pollack, 1954). Listeners automatically process these cues, which may include place (tongue/teeth placement), manner (air emissions, i.e., blockage of air, continuous flow), and voicing (Shriberg & Kent, 2013), as well as onsets, offsets, and rate of change between phonemes, which allow the listener to focus on the expected acoustic speech signal (Callan et al., 2003). This phenomenon is especially prevalent in instances of sensory deprivation (Calvert et al., 1998), such as in individuals with early onset hearing loss (Bernstein et al., 2000; Ellis et al., 2001; Auer &

Bernstein, 2007) and cochlear implant users (Strelnikov et al. 2009), who must then rely on lipreading. Lipreading refers to recognition of speech based on visual cues as described above. Similarly, this would explain why older adults (who often experience some loss of hearing) are able to communicate well during face-to-face interactions, but often have difficulty communicating over the phone.

II. Mechanisms of Audiovisual Integration

Visual information can alter a listener's perception of auditory speech, as seen in the classic example of the McGurk Effect. In the McGurk paradigm, an audio recording of /pa/, which contains a voiceless, bilabial, stop consonant, is overlaid onto a video recording of the lip movements for /ka/, which contains a voiceless, velar, stop consonant. When listening to and watching this video, control participants report hearing /ta/, which contains a voiceless, lingua-alveolar, stop consonant, a mix (or "fusion") between the /p/ and /k/ consonants and is otherwise known as the McGurk Effect. When these same participants listen to only the audio recording or only an untreated video recording, they accurately report hearing either /pa/ or /ka/. This paradigm demonstrates how individuals can report perceiving the syllable /ta/ in the fused trial, despite the absence of auditory or visual input that is /ta/ (McGurk & MacDonald, 1976). This phenomenon is a result of the speech-processing system's ability to predict messages based on auditory and visual signals (van Wassenhove et al., 2005).

III. Neurobiology of Audiovisual Integration: Neuroimaging

Several functional magnetic resonance imaging (fMRI) studies have been performed on control participants to understand the neural basis of AV integration. Multisensory integration has been found to bilaterally activate the posterior superior temporal sulcus (pSTS) (Balk et al., 2010; Beauchamp et al., 2004a, 2004b; Callan et al., 2004) to a greater extent than individual modalities (auditory only or visual only) (Calvert et al., 2001). It has been hypothesized that the pSTS consists of auditory, visual, and auditory-visual neurons and is likewise organized. Electroencephalography studies in macaque monkeys have found activation for individual neurons to auditory, visual, and auditory-visual stimuli (Benevento et al., 1977). A similar distribution, with activation for AV perception lying between the activation for auditory-only and visual-only stimuli, has been identified in humans along the STS (Beauchamp et al., 2004b). Transcranial magnetic stimulation (TMS) has been used to test the role of the STS in AV integration by imitating a “lesion” at the site of stimulation and disrupting the function of that site. Using TMS with fMRI to more accurately localize the STS per subject, Beauchamp et al. (2010) stimulated the STS while having the participants identify stimuli consisting of both McGurk and non-McGurk stimuli. They found a significant decrease in the occurrence of the McGurk Effect when a temporary “lesion” site was stimulated at the STS, indicating the STS as the center for AV integration. Similarly, individuals demonstrating deficits in multisensory integration show no benefit to AV speech, which may negatively affect their ability for speech perception, such as individuals with autism (Iarocci & McDonald, 2006). Individuals with autism have difficulty both with lipreading (Smith & Bennetto, 2007) and blending auditory and visual speech (Williams et al.,

2004) and therefore present with a lower fusion rate than typical individuals during a McGurk task. This deficit may be attributed to the differing neurobiology of individuals with autism. The STS in individuals with autism has been found to activate differently during perceptual tasks, such as perception of faces (McCarthy et al., 1999) and detection of speech-sounds (Boddaert et al., 2004), in addition to having decreased connectivity with other centers for integration (Brambilla et al., 2004).

In addition to the pSTS being activated during AV integration tasks, some studies have implicated motor speech circuits, such as Broca's area, as a necessity for speech perception (Liberman & Mattingly, 1985; Ojanen et al., 2005). They suggest that a listener reconstructs the motoric programs for each speech sound based on visual and auditory cues as the speaker is saying each sound in order to perceive the information. It is well known that speech production can be affected by the auditory system, as seen in delayed or distorted auditory feedback paradigms (Houde & Jordan, 1995; Stuart et al., 2002; Katseff et al., 2011). The effects of auditory information on motor control are especially evident in individuals who experience adult-onset deafness and present with an obvious decline in articulation (Waldstein, 1990). More recently, studies have implicated that while the motor speech circuits may be active during moments of speech perception, it is only modulated during speech perception tasks that were more taxing (Matchin et al., 2014). For example, during tasks including decreased frame rate (Fridriksson et al., 2008) and degraded speech intelligibility (David et al., 2005) the left posterior inferior frontal gyrus became activated. In a behavioral and fMRI study, Matchin and colleagues (2014) examined the effect of the motor speech circuit on the McGurk Effect and likewise, AV integration. During the behavioral portion of the study, participants completed a classic McGurk task while modulating the motor speech circuit by articulatory suppression (continuously articulating a sequence of sounds without actually producing sound or continuously performing a finger tapping sequence), which should then modulate the McGurk Effect. However, Matchin et al. (2014) found no effect of modulating the motor speech circuit on the McGurk Effect. Then, participants were asked to complete the same task, though only given the option to continuously articulate the sequence of sounds (pa, ta, ka), during the fMRI portion of the study. Further supporting the notion that the motor speech circuit is not involved in simple AV integration tasks, the motor speech circuit showed significant activation in response to visual only speech stimuli, while the STS demonstrated significant activation in response to AV stimuli.

IV. Lower Cortical Representations of Audiovisual Integration

While it is well known that AV integration activates the STS bilaterally (a higher level of cortical processing), non-human and human species studies have been performed to identify which lower cortical stages are involved. Both eye position (Werner-Reiss et al., 2003) and somatosensory (Schroeder et al., 2001) input has been found to activate primary auditory cortex, indicating the role of the auditory cortex as an early stage for AV integration. For example, in a rhesus monkey study identifying the relationship between eye position and primary auditory cortex, Werner-Reiss and colleagues (2003) found eye position to activate approximately one-third of the neurons in primary auditory cortex and alter their response to audition approximately one-fourth of the time in both the light and the dark.

These same findings are well replicated in human studies. Several studies have found the auditory cortex to activate in response to visual-only stimuli (Calvert et al., 1997; Calvert & Campbell, 2003; Pekkola et al., 2005). When asked to silently lipread a given stimulus with no auditory input, fMRI and positron emission tomography (PET) studies have found bilateral activation in the primary auditory as well as auditory association cortices including, but not limited to, portions of the superior temporal gyrus, posterior STS, inferior frontal gyrus, and the premotor cortex (Calvert et al., 1997; Calvert & Campbell, 2003; MacSweeney et al., 2000; Okada et al., 2009). This activation may be an automatic response to seeing (i.e., phonemic cues) and anticipating acoustic features of speech and further indicates multisensory interactions in daily production and perception.

V. Audiovisual Integration After Stroke

A handful of authors have performed single patient studies with individuals with aphasia to identify their abilities for AV integration. Supporting the idea of the motor speech circuit in speech perception, Ramachandran et al. (1999) studied a single patient with Broca's aphasia, an expressive language disorder, who did not experience the McGurk Effect and concluded that Broca's area is necessary for AV speech. In contrast, Anderson and Starrfelt (2015) suggest that the motor speech circuit may only play a supplementary role. Anderson and Starrfelt (2015) studied two individuals suffering from stroke. The first patient suffered from a left middle cerebral artery (MCA) ischemic stroke affecting the inferior and middle frontal gyri with some superior temporal gyrus, insular, and basal ganglia damage. This patient originally presented with Global aphasia that resolved into Broca's aphasia. The second patient suffered from a left MCA hemorrhagic stroke affecting the frontotemporoparietal region with some basal ganglia and underlying white matter to the STS damage. This patient originally presented with Global aphasia that resolved into Wernicke's aphasia, a receptive language disorder. Both individuals were able to experience the McGurk Effect and Anderson and Starrfelt (2015) therefore concluded that Broca's area is not necessary for AV speech as Ramachandran et al. (1999) originally assumed. Additionally, Anderson and Starrfelt (2015) found both patients exhibited visual bias when responding to an incongruent trial, trials where the individual should be fusing, incorrectly. This was attributed to the "boost" visual presentation provided auditory presentation. Both patients performed significantly higher on congruent AV trials (the auditory and visual presentations matched) in comparison to auditory-only trials. Visual bias in the patient with Wernicke's aphasia may be a result of the nature of the receptive disorder.

AV speech has otherwise been incorporated into treatment post-stroke to help improve speech production. In therapy, patients are trained to focus on visual input (a speaker's face and lip movements) while listening to matching auditory stimuli. When compared to performance given auditory-only stimuli, patients demonstrated improved trained and untrained picture naming abilities (Fridriksson et al., 2009). Additionally, Fridriksson et al. (2012) studied the benefit of speech entrainment (mimicking AV speech provided by a video) given AV stimuli versus auditory-only stimuli in patients with Broca's aphasia. AV speech during speech entrainment was found to improve both speech output and fluency for each patient when compared to the auditory-only stimuli. This may indicate that individual's with Broca's aphasia can still benefit from visual information provided during speech (i.e., mouth movements, etc.). Likewise, AV speech has been found to benefit individuals with aphasia when utilizing computerized aphasia treatments targeting various modalities (i.e., verbal expression, writing, reading comprehension, etc.) (Choe & Stanton, 2011). Choe & Stanton (2011) examined performance differences in a confrontational naming task in a Broca's aphasic and Anomic aphasic given two types of cueing: AV cues (i.e., video recording where the participant both sees and hears a phonemic cue: "It starts with /k/. What is this?") and auditory-only cues (i.e., hearing an audio-recording of the phonemic cue) (Choe & Stanton, 2011, p. 989). Although the participants demonstrated increased performance with both cues, the participants required a decreased level of support when provided AV cues in comparison to auditory-only cues.

VI. Present Study

As discussed above, neuroimaging and lesion case studies to date have produced inconclusive results regarding the neuroanatomy critical for AV integration. The present study utilizes voxel-based lesion-symptom mapping (VLSM), a well-tested method used to identify which areas of the brain are crucial for specific tasks. This is performed by mapping behavioral measures onto specific voxels in the brain and compared across a large number of subjects (Bates et al., 2003). To our knowledge, the present study is the first VLSM study to examine the effect of stroke or other brain injury on the processing of AV speech or the resulting deficits based on lesion location. The following hypotheses are proposed:

1. Participants with damage to the left superior temporal sulcus and underlying white matter will be correlated with audiovisual integration impairments.
2. Damage to the left inferior frontal lobe will cause a significant increase in auditory capture (/pa/) when responding to a fusion trial.
3. Damage to auditory centers will cause a significant increase in visual capture (/ka/) when responding to a fusion trial.

Materials and Methods

Participants

Forty-one adults (12 female, 29 male) participated in the present study. All participants were recruited via the Multi-site Aphasia Research Consortium (MARC) as part of a larger ongoing research project. Participants were included in this study if they successfully completed the tasks of interest (described below) and met the following inclusion criteria: native English speaker, MRI exhibiting a chronic focal (6 months or more post-onset) lesion due to a stroke in the left hemisphere, no self-reported contraindications for MRI, and no self-reported or documented history of additional psychological or neurological disease. The vast majority of participants were strongly right-handed (83%) pre-stroke as determined by a modified Edinburgh Handedness Scale. Participants ranged in age from 31 to 86 years ($M = 58.78$, $sd = 11.99$). Written informed consent was obtained prior to participation in the study. All procedures were in compliance with the Code of Ethics of the World Medical Association and approved by the Institutional Review Boards of University of California Irvine, San Diego State University, University of Iowa, Medical College of Wisconsin and Arizona State University. The participants received monetary compensation for their time.

Materials

To address our goal of characterizing the effect of lesion location on the processing of AV speech, two tasks were administered: (1) an auditory speech perception task and (2) an AV speech perception task (i.e. a McGurk task; McGurk & MacDonald, 1976). Each participant was administered these tasks as part of an extensive psycholinguistic test battery to assess specific speech perception and production abilities. Individual tests within the battery were presented in a non-fixed pseudorandom order; items within each test were presented in a fixed random order.

Stimuli. Each trial consisted of the words "Get Ready", appearing for 1000ms, followed by a "X", appearing for 1200ms, and then the audio stimulus or AV stimulus, described below. After each stimulus, three printed response options were displayed horizontally across the computer screen: "Pa Ta Ka", with the serial positions of the three options presented in a fixed random order across trials for each subject (Figures 1 and 2). Participants were asked to point to the corresponding identity of the played acoustic stimuli on the computer screen. Responses were self-paced. A mouse click began the next trial. Both tasks were delivered through a laptop computer with PowerPoint software (Microsoft Office) placed at a distance comfortable to the participant.

All stimuli were recorded by a native English male speaker and were presented through supra-aural headphones at a volume level that was clearly audible and comfortable for each participant. To ensure the participant was able to perform the task and to familiarize the participant with each of the two tasks, one sample trial was presented prior to the start of both tasks.

Audiovisual Speech Perception. This task is based on the classic McGurk task (McGurk & MacDonald, 1976), and has previously been used in adult control subjects (Matchin et al., 2014). The task consisted of 30 trials in which participants were asked to indicate which of three sounds, /pa/ /ka/ or /ta/, was presented. The 30 trials consisted of 20 congruent trials (in ten trials the auditory and visual stimuli both correspond to /pa/ and in ten trials both correspond to /ka/) and ten incongruent trials (auditory stimulus was /pa/, visual stimulus was /ka/). These incongruent trials reliably generate a perception of /ta/ in control subjects (McGurk & MacDonald, 1976; Matchin et al., 2014).

For each AV stimulus, the participants were instructed to pay close attention to both the face in the video and the sound played. The total duration of each recording was 3000ms and the duration of each auditory stimulus was ~500ms. Each of the 20 congruent AV stimuli were generated by overlaying a corresponding auditory stimulus onto a visual stimulus and aligning the visual and auditory onset of the consonant burst, whereas each of the 10 incongruent AV stimuli were generated by overlying the auditory stimulus /pa/ onto the visual stimulus /ka/. Each syllable was played once. The AV speech perception task was consistently administered prior to the auditory speech perception task to prevent participants from guessing the incongruent nature of the AV stimuli.

Auditory Speech Perception. This task consisted of 30 trials in which participants were asked to indicate which of two sounds, /pa/ or /ka/, was presented. All three response options ("Pa Ta Ka") were included in the auditory speech perception trials to prevent the participants from guessing the nature of the AV fusion stimulus. In 20 trials the sound /pa/ was presented, and in ten trials the sound /ka/ was presented. Auditory stimuli were played while the participant looked at an image of a microphone on the screen. The total duration of each recording was 3000ms and the duration of each auditory stimulus was ~500ms; video recordings of the speaker had a frame rate of 30 fps. Each syllable was played once. This task serves as a baseline measure of auditory speech perception from which to compare performance on the AV speech perception, described above.

Imaging and Lesion Analyses. A high-resolution T1 MRI was collected for each participant. Lesion mapping was performed within Brainvox software (Frank et al., 1997) using MAP-3 lesion analysis methods (Damasio, 2000). In the MAP-3 lesion analysis method, each lesioned brain is transferred into the space of a template brain to appropriately compare the participant's lesion sites in a three-dimensional space. For each participant, anatomical markers such as sulci and gyri are first used to reslice the template brain to maximally orient each slice to the lesion's native space. Then, the lesion is manually demarcated onto the template using the same anatomical markers to identify lesion boundaries (see Figure 3 for an overlap map of all of the participant's lesions). All lesion mapping was completed by individuals with extensive training in this technique and supervised by an expert neuroanatomist. The above techniques within the MAP-3 lesion analysis method have been shown to have high reliability within and between raters and in some cases have demonstrated higher accuracy than automated methods (Fiez et al., 2000; Panatazis et al., 2010).

Voxel-Based Lesion Symptom Mapping. Voxel-based lesion symptom mapping (Bates et al., 2003) was used amongst all 41 participants to identify voxels in the left hemisphere where a participant with damage in that voxel performs significantly different than a participant with no damage in that voxel through a t-test. A voxel-wise threshold of p < .005 was used. In addition, an Analysis of Covariance (ANCOVA) was used to regress out variance due to lesion size and voxels were only included in the VLSM analysis if a minimum of 10% of participants (i.e. n=4) demonstrated damage in that voxel. Clusters were only reported if they met a minimum of 20 voxels.

Results

Behavioral Results: Auditory & Audiovisual Speech Perception

Overall accuracy on the speech perception tasks is presented in Figure 4 and performance was found to be significantly above chance through a one-sample t-test (Auditory speech perception task $M = .89$, $sd = .14$, $t(40) = 25.32$, $p < .001$; congruent trials in the AV speech perception task: $M = .91$, $sd = .15$, $t(40) = 24.52$, $p < .001$; incongruent trials in the AV speech perception task: $M = .81$; $sd = .28$, $t(40) = 10.94$, $p <$

.001). Incongruent AV perception trials were considered “correct” when participants fused the auditory and visual information and thus responded with /ta/. Performance on the auditory speech perception task was not significantly different than performance on AV congruent trials (*t*(40) = -.76, *p* = .450). However, performance on the AV congruent trials was significantly higher than performance on AV incongruent trials (*t*(40) = 2.20, *p* = .034).

A breakdown of performance on AV speech perception incongruent, or “fusion” trials is presented in Figure 5. To reiterate, on incongruent trials in the AV speech perception task, participants overall performed significantly above chance (81% correct). However, incorrect fusion trial responses can further be broken into two categories and were labeled either auditory or visual capture. Auditory capture indicates the participant responded to a fusion trial with the auditory stimulus, /pa/. Visual capture indicates the participant responded to a fusion trial with the visual stimulus, /ka/. Overall, the distribution of auditory (*M* = .10, *sd* = .20) and visual (*M* = .9; *sd* = .20) capture responses were similar.

VLSM Results: Speech Perception Tasks

We conducted voxel-based lesion symptom mapping (VLSM) across the sample of 41 left-hemisphere subjects to identify brain regions associated with auditory speech perception and congruent and incongruent AV speech perception performance. More specifically, we identified significant clusters of lesioned brain areas associated with lower performance on each task and within each error type.

The VLSM analysis for overall performance (proportion correct) on the auditory speech perception task identified a significant cluster (peak *t* at -49 42 19; number of voxels: 4679; $p < .005$) in left Heschl's gyrus extending into the left frontal operculum and left supramarginal gyrus (Figure 6).

Analysis of performance on the AV speech perception task yielded the following results: the congruent AV speech perception VLSM identified a significant cluster (peak *t* at -20 -20 -7; number of voxels: 2731; $p < .005$) in the basal ganglia of the left hemisphere, including the caudate nucleus, putamen and internal capsule (Figures 7). The incongruent AV speech perception VLSM identified a significant cluster (peak *t* -42 12 -21; number of voxels: 2934; $p < .005$) spanning the length of the left middle temporal gyrus, including the posterior STS and anterior temporal lobe (Figure 8). These results indicate that the left supramarginal gyrus and Heschl's gyrus as critical for auditory speech perception and the left basal ganglia and left middle temporal gyrus as critical for AV speech perception.

VLSM Results: Error Types

To further characterize the areas of lesion affecting performance on incongruent AV speech perception trials, VLSMs were used to analyze each type of error: auditory capture (incorrect response of /pa/, the auditory stimulus) and visual capture (incorrect response of /ka/, the visual stimulus). A larger number of auditory capture responses was found to be associated with damage in a significant cluster (peak *t* -52 75 11; number of voxels: 23140; *p* < .005) in the left middle temporal gyrus and left middle occipital gyrus (Figure 9). Conversely, a larger number of visual capture responses implicated a significant cluster (peak *t* -35 -3 -13; number of voxels: 8268; *p* < .005) in the left anterior temporal lobe (ATL), more specifically the anterior temporal pole, as well as the left post central gyrus (Figure 10). These results indicate that in AV speech perception, reliance on auditory information (auditory capture) is associated with lesions in the left posterior temporal lobe, whereas, reliance on visual information (auditory capture) is associated with lesions in the left anterior temporal lobe (approximately Brodmann's area 38).

In addition to whole brain analyses performed on all 41 participants, a lesion overlay map was created for (1) participants who made more auditory capture responses than visual capture responses and (2) participants who made more visual capture responses than auditory capture responses in AV incongruent trials. These lesion overlay maps allowed us to identify overlapping lesion locations for participants who make more auditory or visual capture responses. Participants who made at least one more auditory capture response in comparison to visual capture responses (n = 9) had maximum overlap in lesion location in the left posterior middle temporal gyrus/pSTS (Figure 11). On the other hand, participants who made at least one more visual capture response in comparison to auditory capture responses (n = 7) had the greatest overlap in the left insula, left inferior frontal gyrus, and left anterior superior temporal gyrus (Figure 12).

Discussion

The present study investigated the effect of lesion location on the processing of AV speech using voxel-based lesion symptom mapping (VLSM). Participants with chronic, focal left hemisphere brain lesions due to stroke completed two tasks of speech perception, an auditory speech perception task and an AV speech perception task (McGurk & MacDonald, 1976). Overall, we found high behavioral performance on both tasks (auditory only: 89%; AV congruent: 91%, AV incongruent: 81%), replicating findings in controls (McGurk & MacDonald, 1976; Matchin et al., 2014) with participant's performing most poorly on AV incongruent trials. Although most participants were overall able to fuse the auditory and visual information on the AV incongruent trials, it was expected that participants would demonstrate some "errors" on these trials. It is also the case that control participants do not always have a 100% fusion rate when completing a McGurk task and there are some control participants who are unable to fuse with no neurological indications for deficiency. In McGurk and MacDonald's (1976) paradigm, approximately 2% of adult control participants were unable to fuse when viewing an incongruent AV stimulus and this percentage varies in several studies replicating the classic McGurk paradigm (e.g., 36%, MacDonald & McGurk, 1978).

Participants' performance in the auditory and AV speech perception tasks were then related to lesion location. Whole brain VLSM analyses examining brain regions associated with performance in the auditory speech perception task coincides with previous literature (Morosan et al., 2011; Alexander et al., 1990; Amunts et al., 2012;

Sliwinska et al., 2012). Heschl’s gyrus (Brodmann’s area 41 & 42; Forstmann, Keuken, & Alkemade, 2015), the frontal operculum, and the supramarginal gyrus were identified as critical for auditory speech perception. These three regions are well supported in previous literature for language performance and auditory comprehension: Heschl’s gyrus is well known to function as the primary auditory cortex (Morosan et al., 2011), the frontal operculum has been found to be essential for a variety of language functions (Alexander et al., 1990; Amunts et al., 2012) including phoneme discrimination (Meister et al., 2007; D’Ausillio et al., 2009), and the supramarginal gyrus has previously been implicated in phonological processing (Sliwinska et al., 2012). It is likely that no significant regions were found at a corrected threshold because this is a fairly simple task with high reliability of performance (McGurk & MacDonald, 1976; Matchin et al., 2014) that can also be processed in the right hemisphere (Hickok et al., 2008).

Conversely, our findings regarding the brain regions associated with performance for congruent AV speech perception trials is novel in comparison to previous literature. As discussed in the introduction, previous studies have found activation of the left pSTS in response to multisensory stimuli consisting of visual and auditory input (Balk et al., 2010; Beauchamp et al., 2004a, 2004b; Callan et al., 2004). We expected our VLSM analyses to indicate regions in and adjacent to the pSTS in response to congruent AV speech perception trials. However, VLSM analyses indicated lesions to the caudate nucleus, putamen and internal capsule were associated with lower congruent AV performance. This is again a simple task, which should result in a higher performance than the auditory only speech perception task since the visual input should aid the auditory input for speech recognition (Calvert et al., 1998; Neely, 1956; Arnold & Hill, 2001). Furthermore, purely visual or auditory deficits would not decrease congruent performance because participants would be able to rely on the input they are able to process and respond appropriately. The function of the basal ganglia is a highly-debated topic. The basal ganglia is known to play a role in motor control of competing mechanisms for precise movements (Greybiel, 1995; Mink, 1996). However, the basal ganglia has also been frequently implicated in perceptual decision making tasks defined as deliberative, rather than reflexive, decision making processes (Ding & Gold, 2013). The basal ganglia has been identified to play equal role in cognition as motor control during selection between multiple responses (Redgrave et al., 1999). More specifically, the caudate nucleus has been indicated for perceptual judgment in adult rhesus monkeys when responding to a visual, motion discrimination task given multiple choices (Ding & Gold, 2010). Our results may support a higher-level task related effect, perhaps related to attention or the decision-making process; however, this is a prospective avenue for future studies.

As discussed in the introduction, the STS has been identified as crucial for AV integration in both humans (Balk et al., 2010; Beauchamp et al., 2004a, 2004b; Callan et al., 2004) and non-humans (Benevento et al., 1977). VLSMs for incongruent AV speech perception associated lower performance with the left middle temporal gyrus, a region in and adjacent to the left STS and underlying white matter. The lesion location spanned from the left anterior middle temporal gyrus to the left posterior middle temporal gyrus.

This finding supported our first hypothesis that AV integration impairments would be associated with lesions to the left posterior STS and underlying white matter.

There are several reasons why participants may incorrectly respond (i.e. not experience a fused percept) in the incongruent AV trials. Participants could have an auditory deficit that affects their ability to fuse the auditory information with visual information or a visual deficit that affects their ability to fuse visual information with auditory information. Or, participants could instead have intact visual and auditory representations, but these representations are not interacting with each other. This idea begs the question, what is driving these integration errors?

There are two types of errors the participants could make in the incongruent AV trials, auditory capture and visual capture. As a reminder, auditory capture is defined as an incorrect response of the auditory stimulus in an incongruent trial, whereas visual capture is defined as an incorrect response of the visual stimulus in an incongruent trial. The behavioral data indicates that overall participants made a similar number of auditory and visual capture responses in contrast to the notion of a visual “boost” gained in all AV trials (Anderson & Starrfelt, 2015). The VLSM analyses identified two very different regions associated with auditory and visual capture. VLSMs of increased auditory capture responses indicated a significant cluster in the left posterior temporal lobe, more specifically, the left posterior middle temporal gyrus. A second cluster was identified in the left middle occipital gyrus (Brodmann’s area 19). Brodmann’s area 19 is a part of the visual association cortex and plays a role in visual perception and processing (Forstmann, Keuken, & Alkemade, 2015), specifically visuospatial and motion processing (Colligon et al., 2011). To further support our finding, an overlay of the participants' lesions who responded with more auditory capture responses in comparison to visual capture responses was created (n = 9). Participants who responded with more auditory capture responses than visual capture responses mirror the VLSM results in that the maximum overlap (n = 6) was identified in the left posterior middle temporal gyrus/pSTS (Figure 11). These findings did not support our hypothesis that damage to the inferior frontal lobe would cause a significant increase in auditory capture responses. Although they did not support our hypothesis, these findings are logical as one would expect participants with a lesion located in the visual cortex to demonstrate difficulty processing visual input and instead rely on auditory input.

On the other hand, VLSMs for increased visual capture responses identified a significant cluster in the left anterior temporal lobe (Brodmann's area 38, Forstmann, Keuken, & Alkemade, 2015) closer to the left anterior temporal pole, and post-central gyrus (Brodmann's area 1, Forstmann, Keuken, & Alkemade, 2015). We again created an overlay of the participants' lesions who responded with more visual capture responses in comparison to auditory capture responses (n = 7). Participants who responded with more visual capture responses than auditory capture responses differ from the VLSM results in that the maximal overlap (n = 7) was identified in the left insula and left inferior frontal gyrus, as well as some anterior temporal lobe (Figure 12). The anterior temporal lobe has been indicated in several language processing functions. First, the anterior temporal lobe has been indicated in lexical processing and retrieval, evidenced by reduced ability for naming (Damasio et al., 1996, 2004). The anterior temporal lobe has additionally been indicated as the core center for semantic processing of words and objects (Patterson, Nestor, & Rogers, 2007; Rogers et al., 2004, 2006), as evidenced in semantic dementia, a neurodegenerative disease resulting in atrophy of the anterior and lateral temporal lobes hallmarked by progressive loss of semantic, or conceptual, knowledge. Finally, disorders of speech recognition, such as auditory processing disorder (Han et al., 2015), have been associated with anterior temporal resections (Boatman et al., 2006; Han et al., 2015) and anterior temporal lobe epilepsy can result in impaired temporal perception (Lavasani et al., 2016). We hypothesized that increased visual capture responses would be associated with damage to primary auditory centers (e.g., Heschl's gyrus, Morosan et al., 2011), however, our hypothesis was unsupported. Our results indicate that reliance on visual input may not be caused by damage to these basic auditory perception regions. It also is likely that the intact basic auditory speech perception centers in the right hemisphere are sufficient in the presence of left hemisphere primary auditory cortex damage. The anterior temporal lobe result was quite surprising, and future studies of more patients with anterior temporal lobe damage are needed to better characterize the relationship between the anterior temporal lobe and AV integration.

Overall, our whole brain findings in lesion patients are highly consistent with previous literature indicating primary auditory cortex for auditory speech perception and left STS/middle temporal gyrus for AV integration (Balk et al., 2010; Beauchamp et al., 2004a, 2004b; Callan et al., 2004). Although some studies have implicated motor speech circuits, specifically Broca's area, for speech perception (Liberman & Mattingly, 1985;

Ojanen et al., 2005), our VLSM analyses did not associate lower performance on AV integration tasks with lesions to the left posterior inferior frontal gyrus. This further supports more recent beliefs that the motor speech circuit is not involved in AV integration (Matchin et al., 2014). However, our results implicated lesions to the left basal ganglia in association with congruent AV integration trials. The basal ganglia has not previously been implicated in processing of multisensory integration and may support higher-level task related effects. We also identified unique lesion patterns associated with reliance of the auditory or visual stimulus in the presence of AV incongruent information. Increased auditory reliance resulted in a significant cluster in the left middle temporal gyrus and left middle occipital gyrus, whereas increased visual reliance resulted in the left anterior temporal lobe. Both of the results for auditory and visual reliance suggest visual or auditory deficits, rather than deficits in AV integration given intact auditory and visual representations. Additionally, an overlay of the VLSM results from all AV incongruent trials, increased auditory capture responses and increased visual capture responses was created (Figure 13). Although small, we found one area of overlap in the STS/middle temporal gyrus, indicated in a sea green color. Previous neuroimaging studies implicate a similar region in AV integration (Balk et al., 2010; Beauchamp et al., 2004a, 2004b; Callan et al., 2004) and may indicate a possible integration region since damage to this region was similarly associated with both error types.

Future Studies

One remaining question that arises from our results is why was the basal ganglia implicated in congruent AV integration performance? This area has not otherwise been indicated for multisensory processing. However, the basal ganglia has been implicated in perceptual decision making (Redgrave et al., 1999; Ding & Gold, 2010, 2013). Our findings may support this non-motor cognitive function of the basal ganglia resulting in higher-level task effects driving this result. For example, participants demonstrating basal ganglia deficits may have reduced attention or sequential processing abilities and therefore decreased performance even though both auditory and visual information could be utilized.

Similarly, the left anterior temporal lobe has not otherwise been indicated for reliance on visual information. Previous literature has implicated the anterior temporal in language processing and retrieval (Damasio et al., 1996; 2004), semantic processing of words and objects (Patterson, Nestor, & Rogers, 2007; Rogers et al., 2004, 2006), as well as auditory processing disorder (Boatman et al., 2006; Han et al., 2015). Future studies are needed to better characterize this relationship in a larger group of participants with anterior temporal lobe damage.

Possible Clinical Implications

Another avenue relates to clinical implications of lesion locations associated with lower performance on AV integration tasks, auditory or visual reliance, and how these findings can benefit patients during rehabilitation. A portion of speech therapy already utilizes the benefit of visual information throughout aphasia treatment, as previously discussed in the introduction (Choe & Stanton, 2011; Fridriksson et al., 2009). For instance, patients with aphasia have been found to demonstrate increased performance measured by decreased need for cueing when provided AV cues during computerized naming treatment in comparison to auditory only cues (Choe & Stanton, 2011). It is important to provide appropriate and individualized support throughout rehabilitation to support maximal language outcomes. However, our findings may indicate that providing auditory and visual input perhaps does not benefit all patients. If the patient is relying on auditory or visual information and demonstrates deficits in AV integration, therapy may want to focus on directing the participants to attend to the preferred information rather than providing excess stimulation.

Conclusion

Results from the present study support the idea that the STS is crucial for AV integration, and do not implicate frontal motor regions in AV integration. Lesions to the STS result in integration deficits resulting in a combination of auditory and visual capture responses. On the other hand, consistent auditory or visual capture responses are more likely caused by distinct lesion patterns in the left middle temporal gyrus/middle occipital gyrus and the left anterior temporal lobe, respectively. Future studies are needed to characterize the roles of the basal ganglia and the temporal pole in AV integration.

References

Alexander, M.P., Naeser, M.A., & Palumbo, C. (1990). Broca's area aphasias: Aphasia after lesions including the frontal operculum. *Neurology, 40*(2), 353-362.

Amunts, K. & Zilles, K. (2012). Architecture and organizational principles of Broca's region. Trends in *Cognitive Sciences, 16*(8), 418-426.

Andersen, T.S. & Starrfelt, R. (2015). Audiovisual integration of speech in a patient with Broca's aphasia. *Frontiers in Psychology, 6*(435), 1-10.

Arnold, P. & Hill, F. (2001). Bisensory augmentation: a speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology*, *92*, 339–355.

Auer E.T. & Bernstein, L.E. (2007) Enhanced visual speech perception in individuals with early-onset hearing impairment. *Journal of Speech, Language, and Hearing Research, 50*, 1157-1165.

Balk MH, Ojanen V, Pekkola J, Autti T, Sams M et al. (2010) Synchrony of audio-visual speech stimuli modulates left superior temporal sulcus. Neuroreport 21: 822-826.

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.

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

Beauchamp, M.S., Argall, B.D., Bodurka, J., Duyn, J.H., & Martin, A. (2004b). Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. *Nature Neuroscience, 7*, 1190–1192.

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.

Benevento, L.A., Fallon, J., Davis, B.J., & Rezak, M. (1977). Auditory-visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. *Experimental Neurology, 57*(3), 849-872.

Bernstein, L.E., Demorest, M.E., & Tucker, P.E. (2000). Speech perception without hearing. *Perception & Psychophysics, 62*(2), 233–252.

Bernstein, L.E., Everhardt, S.P., & Demorest, M.E. (1989). Single-channel vibrotactile supplements to visual perception of intonation and stress. *Journal of Acoustical Society of America, 85*(1), 397-405.

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*, 2225-2234.

Boatman, D.F., Lesser, R.P., Crone, N.E., Krauss, G., Lenz, F.A., & Miglioretti, D.L. (2006). Speech recognition impairments in patients with intractable right temporal lobe epilepsy. *Epilepsia, 47*(8), 1397-1401.

Boddaert, N., Chabane, N., Belin, P., Bourgeois, M., Royer, V., Barthelemy, C., … & Zilbovicius, M. (2004). Perception of complex sounds in autism: Abnormal auditory cortical processing in children. *American Journal of Psychiatry, 161*, 2117-2120.

Brambilla, P., Hardan, A.Y., di Nemi, S.U., Caverzasi, E., Soares, J.C., Perez, J. & Barala, F. (2004). The functional Neuroanatomy of autism. *Functional Neurology, 19*, 9-17.

Callan, D.E., Jones, J.A., Munhall, K., Callan, A.M., Kroos, C., Vatikiotis-Bateson, E. (2003). Neural processes underlying perceptual enhancement by visual speech gestures. *Neuroreport, 14*, 2213–2218.

Callan, D.E., Jones, J.A., Munhall, K., Kroos, C., Callan, A.M., & Vatikiotis-Bateson, E. (2004). Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. Journal of Cognitive Neuroscience, 16, 805–816.

Calvert, G.A., Brammer, M.J., & Iverson, S.D. (1998). Crossmodal identification. *Trends in Cognitive Sciences, 2*, 247–253.

Calvert, G.A., Bullmore, E.T, Brammer, M.J., Campbell, R., Williams, S.C.R., McGuire, P.W.R. … & David, A.S. (1997) Activation of auditory cortex during silent lipreading. *Science, 276*, 593-596.

Calvert, G.A. & Campbell, R. (2003). Reading speech from still and moving faces: The neural substrates of visible speech. *Journal of Cognitive Neuroscience, 15*, 57-70.

Choe, Y. & Stanton, K. (2011). The effect of visual cues provided by computerised aphasia treatment. *Psychology Press, 25*(9), 983-997.

Colligon, O., Vandewalle, G., Voss, P., Albouy, G., Charbonneau, G., Lassonde, M., & Lepore, F. (2011). Functional specialization for auditory-spatial processing in the occipital cortex of congenitally blind humans. *Processings of the National Academy of Sciences in the United States of America, 108*(11), 4435-4400.

Damasio, H. (2000). The lesion method in cognitive neuroscience. In Boller, F., Graffman, J., & Rizzolati, G. (Eds.), *Handbook of Neuropsychology, 2nd edition* (77-102). Philadelphia: Elsevier.

Damasio, H., Grabowski, T.J., Tranel, D., Hichwa, R.D., & Damasio, A.R. (1996). A neural basis for lexical retrieval. *Nature, 380*, 499-505.

Damasio, H. Tranel, D., Gradowsky, T., Adolphs, R., & Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition, 92*, 179-229.

D'Ausillo, A., Pulvermuller, F., Salmas, P., Bufalari, I., Begliomini, C. & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology, 19*, 381-385.

Ding, L. & Gold, J.I. (2010). Caudate encodes multiple computations for perceptual decisions. *The Journal of Neuroscience, 30*(47), 15747-15759.

Ding, L. & Gold, J.I. (2013). The basal ganglia's contributions to perceptual decision making. *Neuron, 79*, 640-649.

Ellis, T., MacSweeney, M., & Campbell, R. (2001). TAS, a new test of adult speechreading: Deaf people really can be better speechreaders. Paper presented at the International Conference on Auditory-Visual Speech Processing (AVSP 2001), Aalborg, Denmark.

Fiez, J.A., Damasio, H., & Grabowski, T.J. (2000). Lesion segmentation and manual warping to a reference brain: intra- and interobserver reliability. *Human Brain Mapping, 9*(4), 192-211.

Forstmann, B.U., Keuken, M.C., & Alkemade, A. (2015). An introduction to human brain anatomy. In B.U. Forstmann & E. Wagenmakers (Eds.), *An introduction to model-based cognitive neuroscience* (71-99). New York: Springer.

Frank, R.J., Damasio, H., & Grabowski, T.J. (1997). Brainvox: An interactive, multimodal visualization and analysis system for neuroanatomical imaging. *Neuroimage, 5*(1), 13-30.

Fridriksson, J., Baker, J.M., Whiteside, J., Eoute, D., Moser, D., Vesselinov, R., & Rorden, C. (2009). Treating visual speech perception to improve speech production in non-fluent aphasia. *Stroke, 40*(3), 853–858.

Fridriksson, J., Hubbard, H.I., Hudspeth, S.G., Holland, A.L., Bonilha, L., Fromm, D., & Rorden, C. (2012). Speech entrainment enables patients with Broca's aphasia to produce fluent speech. *Brain, 135*, 3815–3829.

Graybiel, A.M. (1995). Building action repertoires: memory and learning functions of the basal ganglia. *Current Opinion in Neurobiology, 5*, 733-741.

Han, S., Lee, E.M., Choi, E.J., Ryu, H.U., Kang, J.K., & Chung, J.W. (2015). Changes in central auditory processing in patients with mesial temporal lobe epilepsy after anterior temporal lobectomy with amygdalohippocampectomy. *Journal of Clinical Neurology, 12*(2), 151-159.

Hickok, G., Okada, K., Barr, W., Pa, J., Rogalsky, C., Donnelly, K. … Grant, A. (2008). Bilateral capacity for speech sound processing in auditory comprehension: Evidence from Wada procedures. *Brain and Language, 107*(3), 179-184.

Houde, J.F. & Jordan, M.L. (1995). Adaptation in speech production to transformed auditory feedback. *Journal of the Acoustical Society of America, 97*(5), 3243.

Iarocci, G. & McDonald, J. (2006). Sensory integration and the perceptual experience of persons with autism. *Journal of Autism and Developmental Disorders, 36*, 77-90.

Katseff, S., Houde, J., & Johnson, K. (2011). Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback? *Language and Speech, 55*(2), 295-308.

Lavasani, A.N., Mohammadkhani, G., Motamedi, M., Karimi, L.J., Jalaei, S., Shojaei, F.S., Danesh, A., & Azimi, H. (2016). Auditory temporal processing in patients with temporal lobe epilepsy. *Epilepsy & Behavior, 60*, 81-85.

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

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

MacSweeney, A., Amaro, E., Calvert, G.A., Campbell, R., David, A.S., McGuire, P. … & Brammer, M.J. (2000). Silent speechreading in the absence of scanner noise: An event-related fMRI study. *Neuroreport, 11*(8), 1729-1733.

Marchand, W.R., Lee, J.N., Thatcher, J.W., Hsu, E.W., Rashkin, E., Suchy, Y., … Barbera, S.S. (2008). Putamen coactivation during motor task execution. *Neuroreport, 19*(9), 957-960.

Matchin, W., Groulx, K., & Hickok, G. (2014). Audiovisual speech integration does not rely on the motor system: Evidence from articulatory suppression, the McGurk effect, and fMRI. *Journal of Cognitive Neuroscience, 26*(3), 606-620.

McCarthy, G., Puce, A., Belger, A., & Allison, T. (1999). Electrophysiological studies of human face perception II: Response properties of face-specific potentials generated in occipitotemporal cortex. *Cerebral Cortex, 9*, 431-444.

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

Meister, I.G., Wilson, S.M. Deblieck, C., Wu, A.D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology, 17*, 1692-1696.

Mink. J.W. (1996). The basal ganglia: Focused selection and inhibition of competing motor programs. *Progress in Neurobiology, 50*, 381-425.

Morosan, P., Rademacher, J. Schleicher, A., Amunts, K., Schormann, T. & Zilles, K. (2011). Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage, 13*, 684-701.

Munhall, K.G., Jones, J.A., Callan, D.E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological Science, 15*(2), 133-137.

Neely, K.K. (1956) Effect of visual factors on the intelligibility of speech. *Journal of the Acoustical Society of America, 28,* 1275–1277.

Ojanen, V., Mottonen, R., Pekkola, J., Jaaskelainen, I.P., Joensuu, R., Autti, T., & Sams, M. (2005). Processing of audiovisual speech in Broca's area. *NeuroImage, 25*, 333–338.

Okada, K. & Hickok, G. (2009). Two cortical mechanisms support the integration of visual and auditory speech: A hypothesis and preliminary data. *Neuroscience Letters, 452,* 219-223.

Pantazis, D., Joshi, A., Jiang, J., Shattuck, D.W., Bernstein, L.E., Damasio, H., & Leahy, R.N. (2010). Comparison of landmark-based and automatic methods for cortical surface registration. *Neuroimage, 49*(3), 2479-93.

Patterson, K., Nestor, P.J., & Rogers, T.T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, 8*, 976-987.

Pekkola, J., Ojanen, V., Autti, T., Jääskeläinen, I.P., Möttönen, R., Tarkiainen, A., & Sams, M. (2005). Primary auditory cortex activation by visual speech: an fMRI study at 3T. *Neuroreport, 16*, 125-128.

Ramachandran, V.S., Rogers-Ramachandran, D., Altschuler, E.L., & Wisdom, S.B. (1999). A test of Rizzolatti's theory of language evolution; McGurk effect and lip reading in Broca's aphasia? Society of Neuroscience Abstract, 25, 1636. Program No. 654.11.

Redgrave, P., Prescott, T.J. & Gurney, K. (1999). The basal ganglia: A vertebrate solution to the selection problem? *Neuroscience, 89*(4), 1009-1023.

Rogers, T.T., Hocking, J., Noppeney, U. Mechelli, A., Gorno-Tempini, M.L., Patterson, K., & Price, C.J. (2006). Anterior temporal cortex and semantic memory: Reconciling findings from neuropsychology and functional imaging. *Cognitive Affective and Behavioral Neuroscience, 6*, 201-213.

Rogers, T.T., Lambon Ralph, M.A., Garrad, P., Bozeat, S., McClelland, J.L., Hodges, J.R., & Patterson, K. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review, 111*(1), 205-235.

Ross, L.A., Saint-Amour, D., Leavitt, V.M., Javitt, D.C., & Foxe, J.J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex, 17*, 1147–1153.

Schroeder, C.E., Lindsley, R.W., Colleen, S., Marcovici, A., Smiley, J.F., & Javitt, D.C. (2000). Somatosensory input to auditory association cortex in the macaque monkey. *Journal of Neurophysiology, 85*, 1322-1327.

Shriberg, L.D. & Kent, R.D. (2013). *Clinical Phonetics* (4th ed.). Boston, MA: Pearson Education, Inc.

Sliwinska, M.W., Khadilkar, M., Campbell-Ratcliffe, J., Quevenco, F., & Devlin, J.T. (2012). Early and sustained supramarginal gyrus contributions to phonological processing. *Frontiers in Psychology, 3*(161), 1-10.

Smith, E.G. & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry, 48*(8), 813-821.

Strelnikov, K., Rouger, J., Barone, P., Deguine, O. (2009). Role of speechreading in audiovisual interactions during the recovery of speech comprehension in deaf adults with cochlear implants. *Scandinavian Journal of Psychology, 50*, 437-444.

Stuart, A., Kalinowski, J., Rastatter, M.P., & Lynch, K. (2002). Effect of delayed auditory feedback on normal speakers at two speech rates. *Journal of the Acoustical Society of America, 111*, 2237–2241.

Sumby, W.H. & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*, 212–215.

van Wassenhove, V., Grant, K.W., Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences, 102*, 1181-1186.

Waldstein, R.S. (1990). Effects of postlingual deafness on speech production: Implications for the role of auditory feedback. *Journal of the Acoustical Society of America, 88*, 2099– 2114.

Werner-Reiss, U., Kelly, K.A., Trause, A.S., Underhill, A.M., & Groh, J.M. (2003). Eye position affects activity in primary auditory cortex of primates. *Current Biology, 13,* 554-562.

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

APPENDIX A

EXAMPLE TRIAL OF SPEECH PERCEPTION TASKS

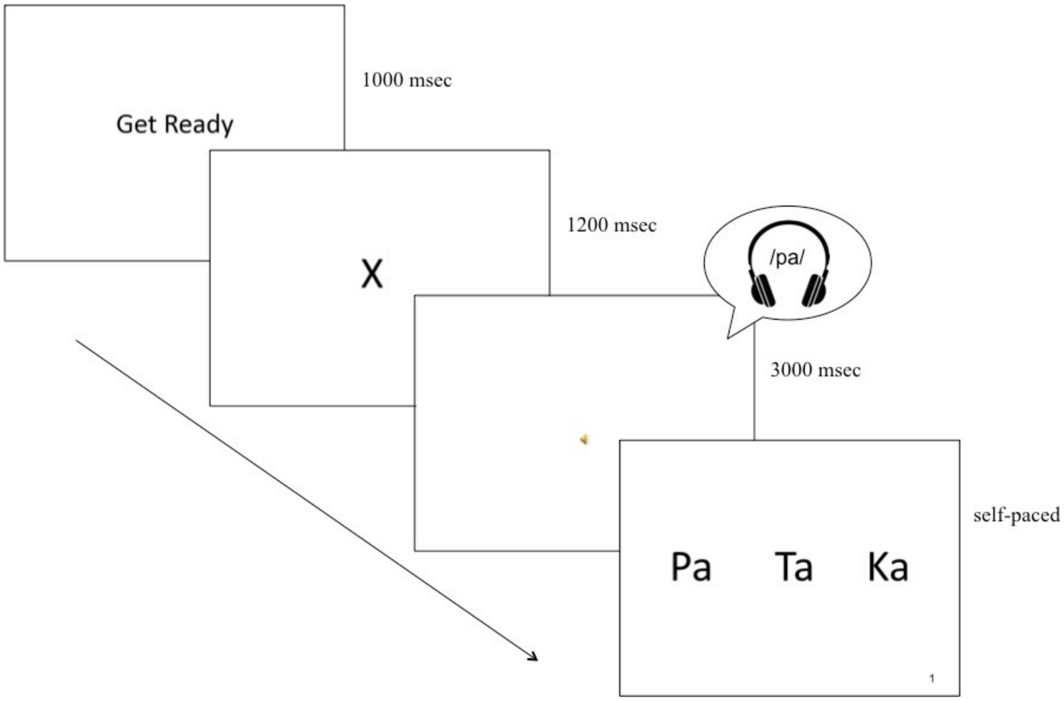

Figure 1. Example auditory speech perception stimulus.

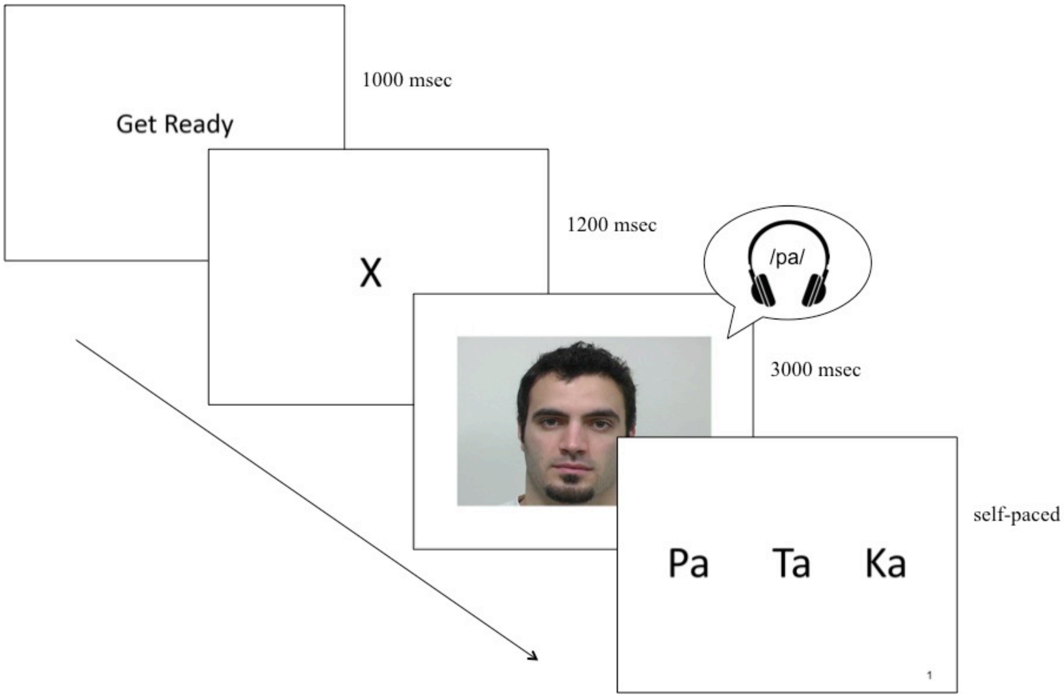

Figure 2. Example audiovisual speech perception stimulus.

APPENDIX B

BEHAVIORAL & VLSM RESULTS

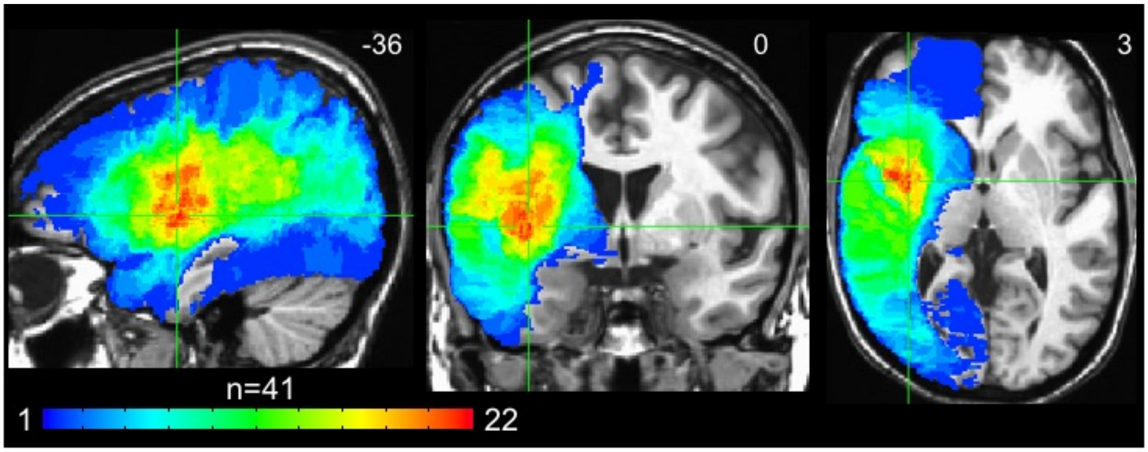

Figure 3. Overlap of all the patients' lesions included in the VLSM analyses (-36 0 3, max overlap = 22).

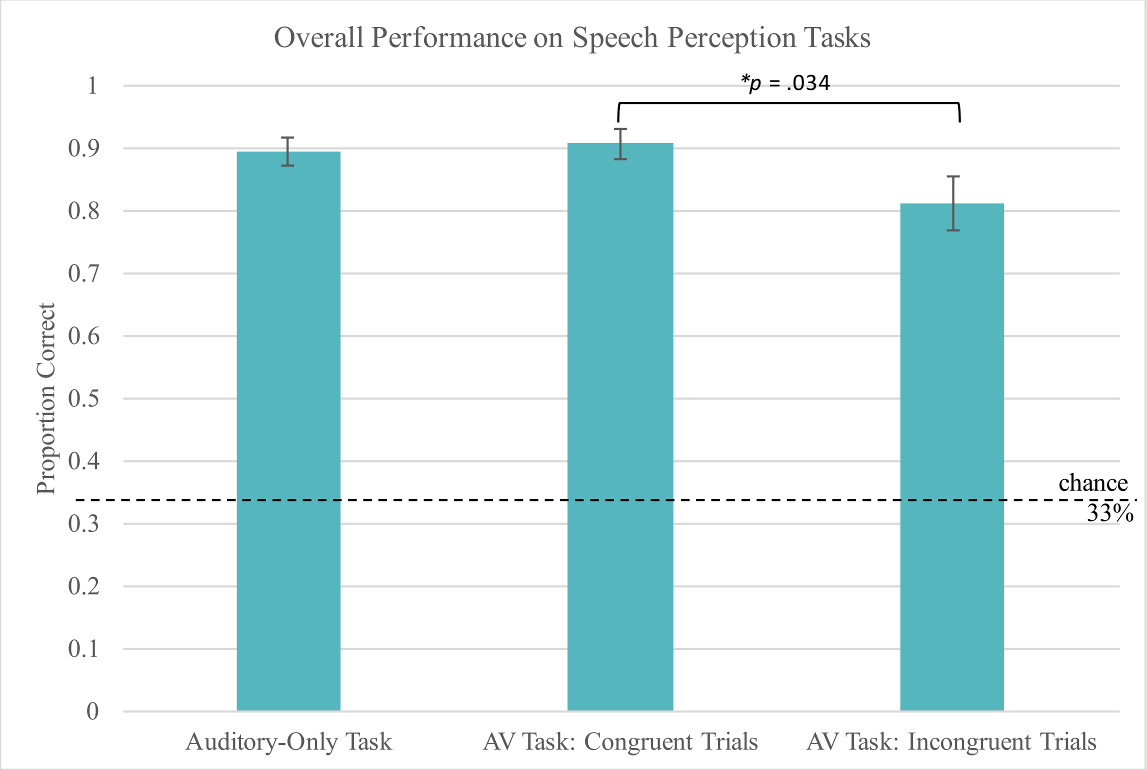

Figure 4. Overall average performance on syllable identification tasks.

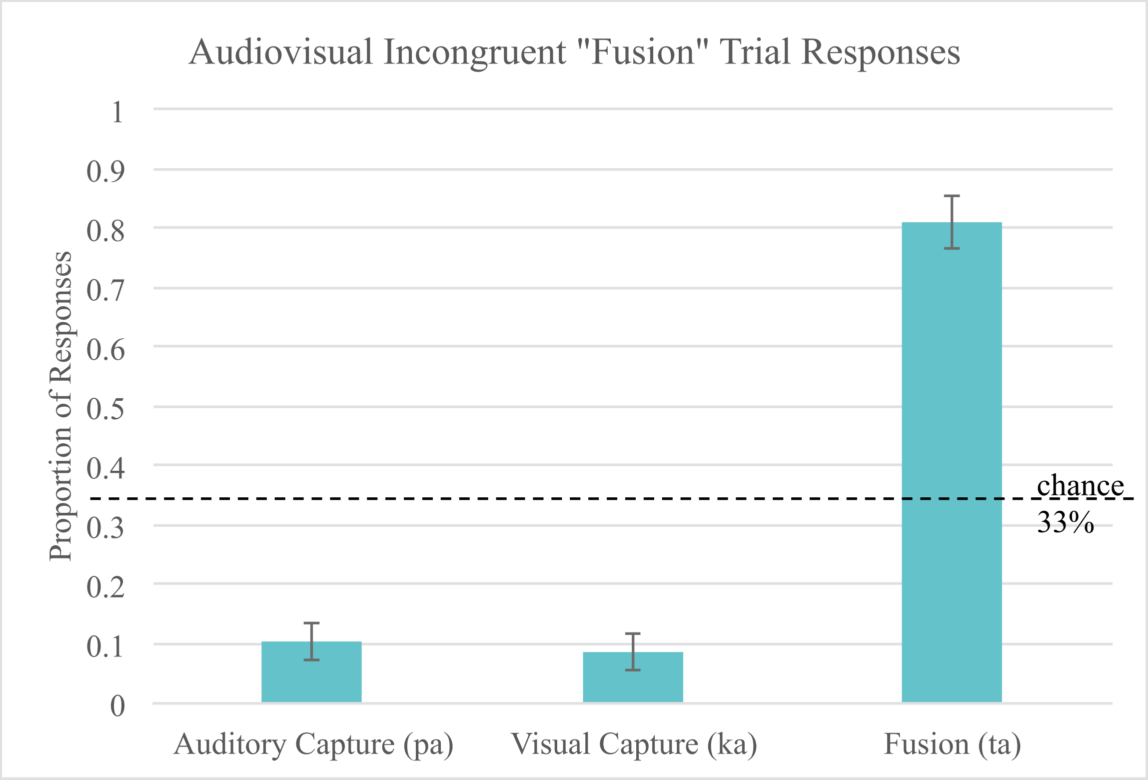

Figure 5. Average responses on audiovisual incongruent, or “fusion”, trials.

Note. Auditory capture indicates the participant responded to a fusion trial with the auditory stimulus, /pa/. Visual capture indicates the participant responded to a fusion trial with the visual stimulus, /ka/.

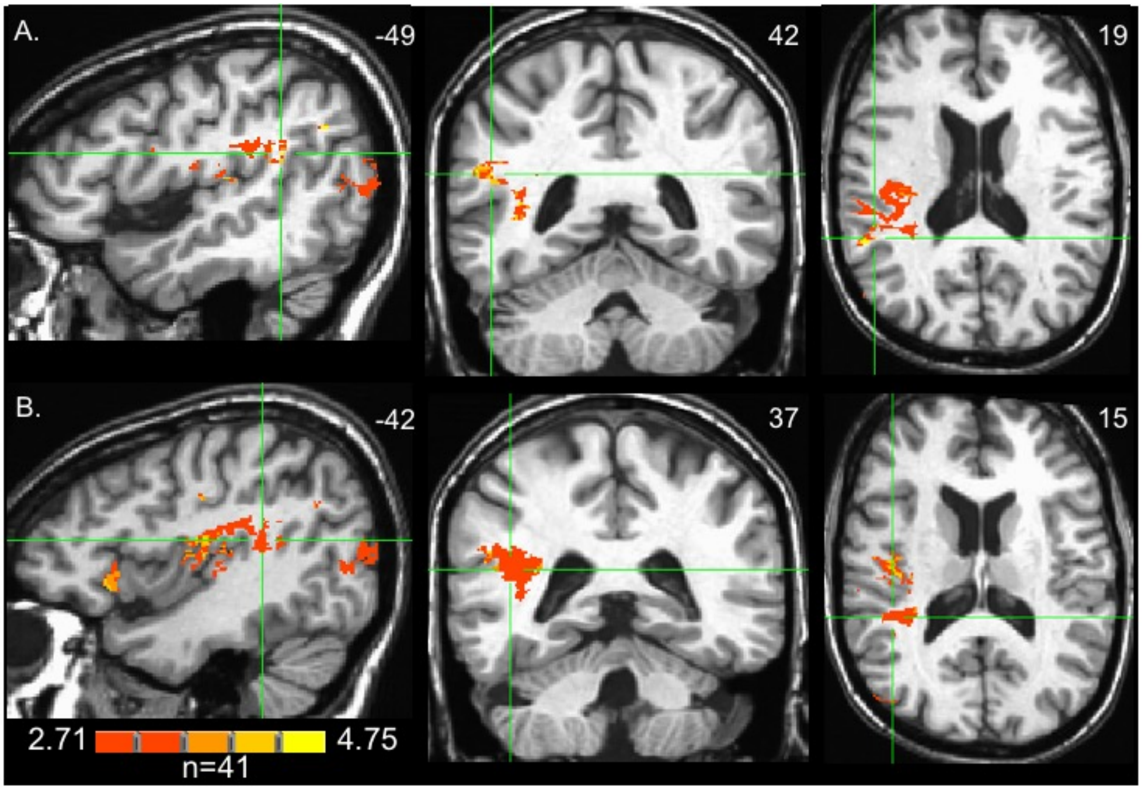

Figure 6. Orthogonal views of the VLSM analysis for auditory speech perception trials (p < .005). (A) Crosshairs are on the peak *t* value (-49 42 19). (B) Crosshairs are on Heschl's gyrus (*t* = -42 37 15).

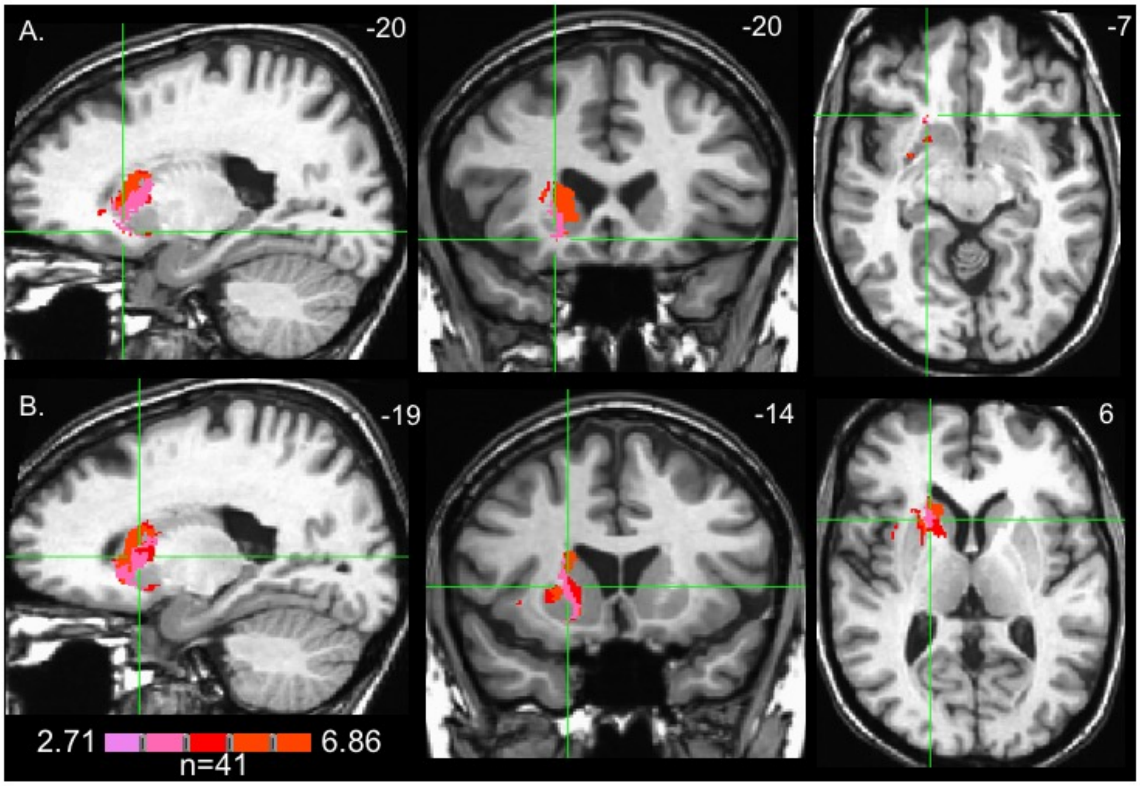

Figure 7. Orthogonal views of the VLSM analysis for congruent audiovisual speech perception trials (p < .005). (A) Crosshairs are on the peak *t* value (-20 -20 -7). (B) Another view of the lesions significant in the basal ganglia. Crosshairs are on *t* = -19 -14 6.

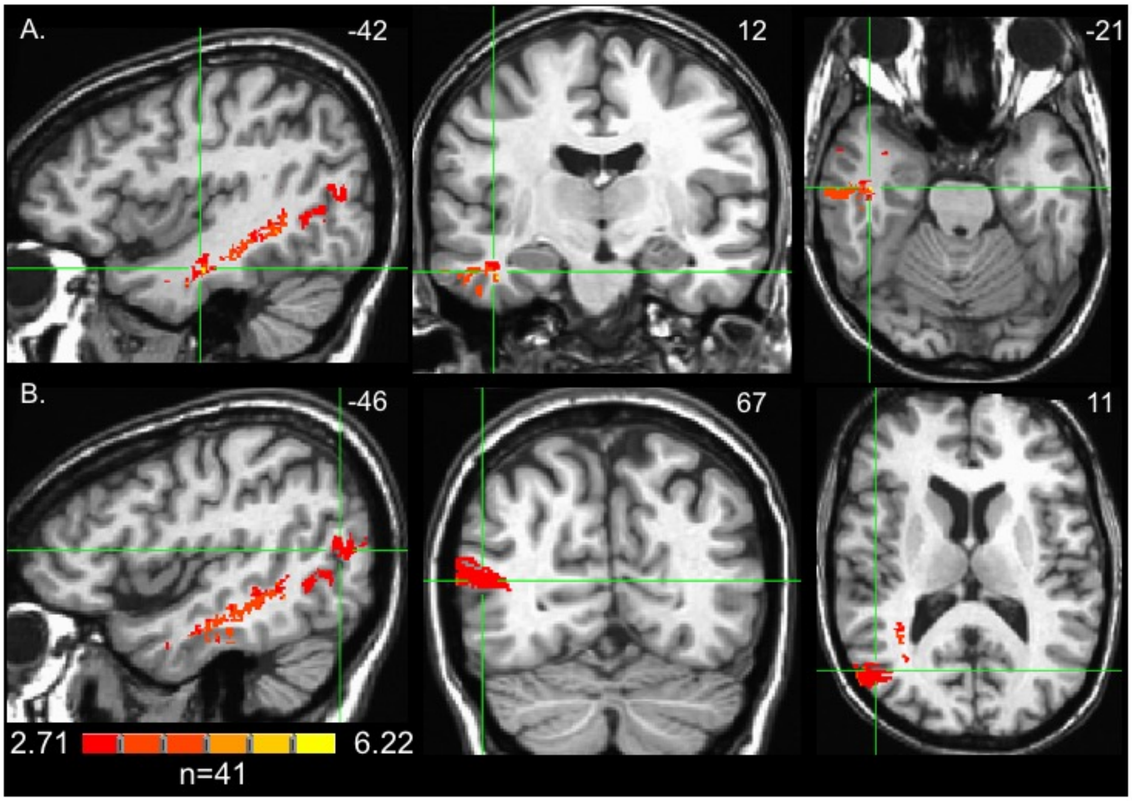

Figure 8. Orthogonal views of the VLSM analysis for incongruent audiovisual speech perception trials ($p < .005$). Crosshairs are on the peak *t* value (-42 12 -21).

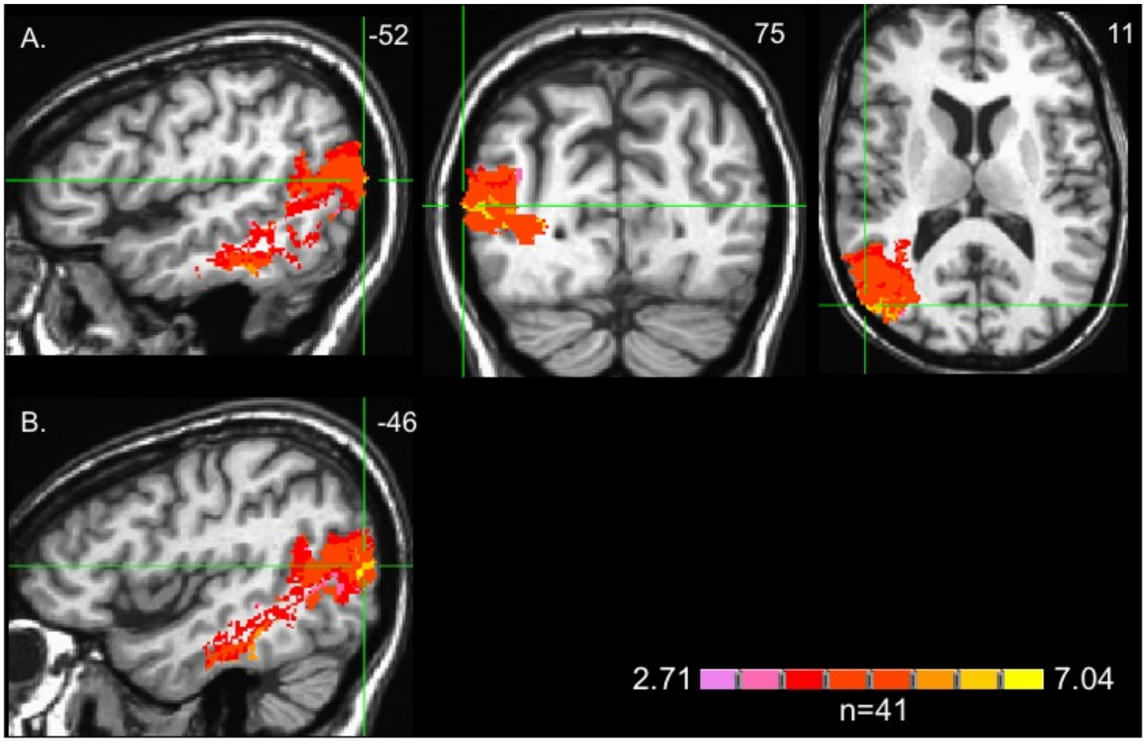

Figure 9. Orthogonal views of the VLSM analysis for auditory capture responses on incorrect incongruent audiovisual speech perception trials (p < .005). (A) Crosshairs are on the peak *t* value (-52, 75, 11). (B) Another sagittal slice displaying middle occipital gyrus as significant for increased auditory capture responses (*t* = -46 75 11).

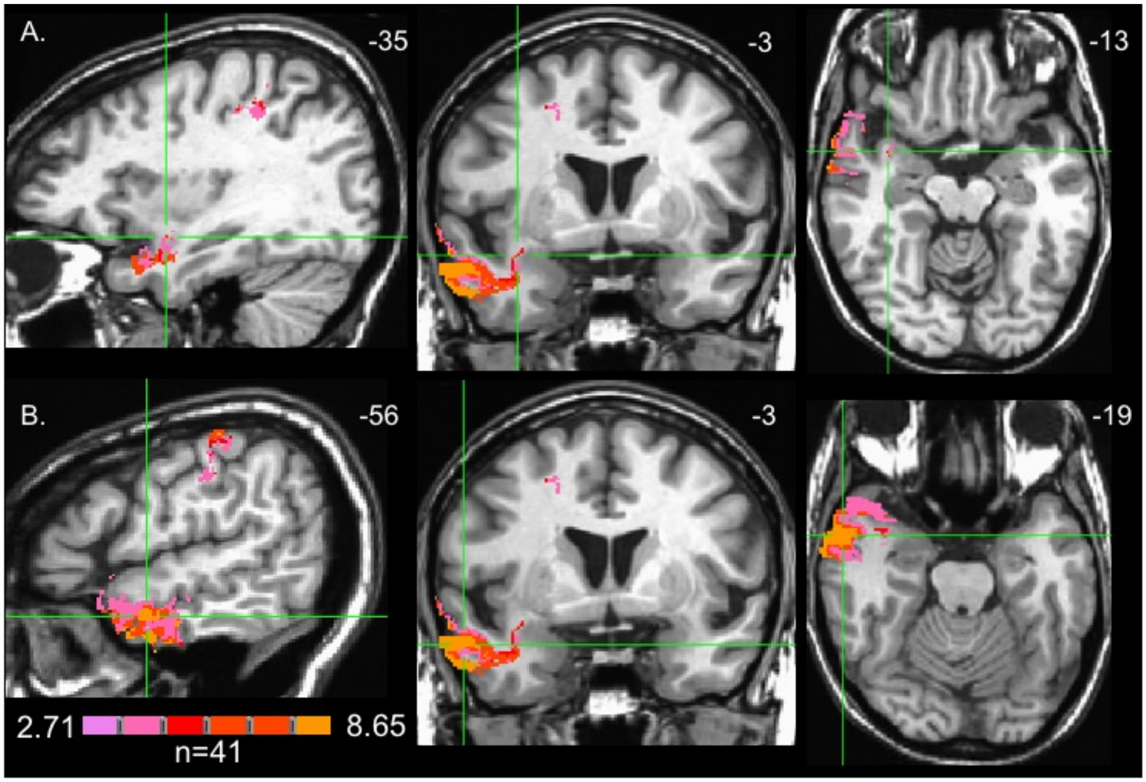

Figure 10. Orthogonal views of the VLSM analysis for visual capture responses on incorrect incongruent audiovisual speech perception trials (p < .005). (A) Crosshairs are on the peak *t* value (-35 -3 -13). (B) Another view of lesion locations significant for visual capture within the anterior temporal lobe (*t* = -56 -3 -19).

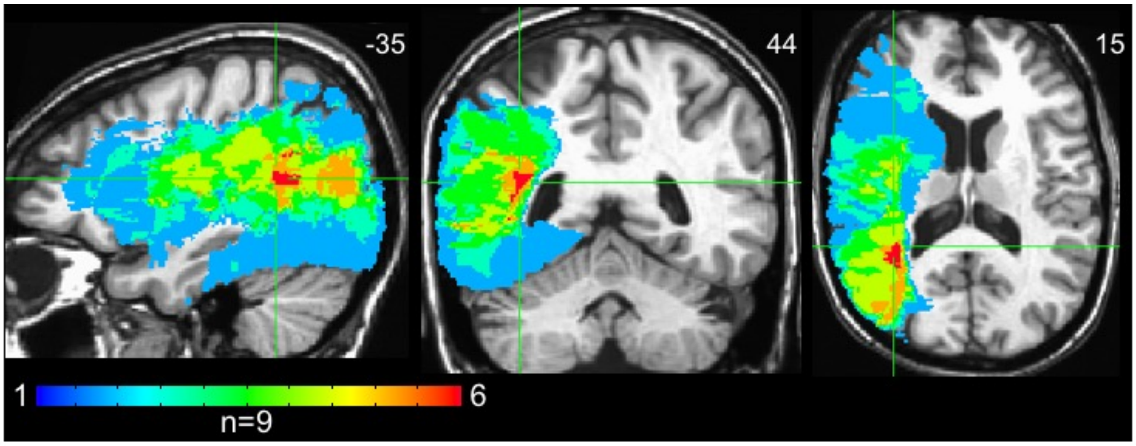

Figure 11. Overlay of the participant's lesions who responded with more auditory than visual capture responses in incorrect incongruent trials (n = 9). Crosshairs are on the greatest overlap of lesions (-35 44 15, max overlap = 6).

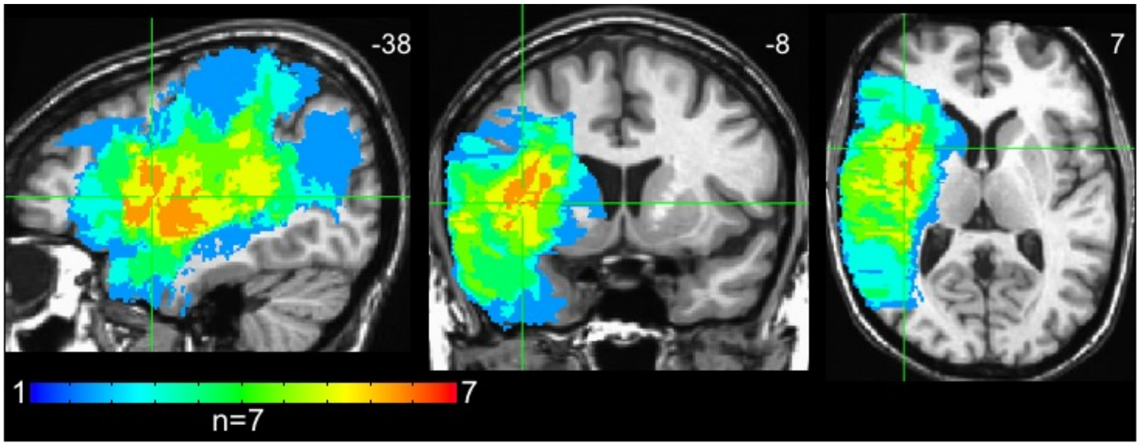

Figure 12. Overlay of the participant's lesions who responded with more visual than auditory capture responses in incorrect incongruent trials (n = 7). Crosshairs are on the greatest overlap of lesions (38 -8 7, max overlap = 7).

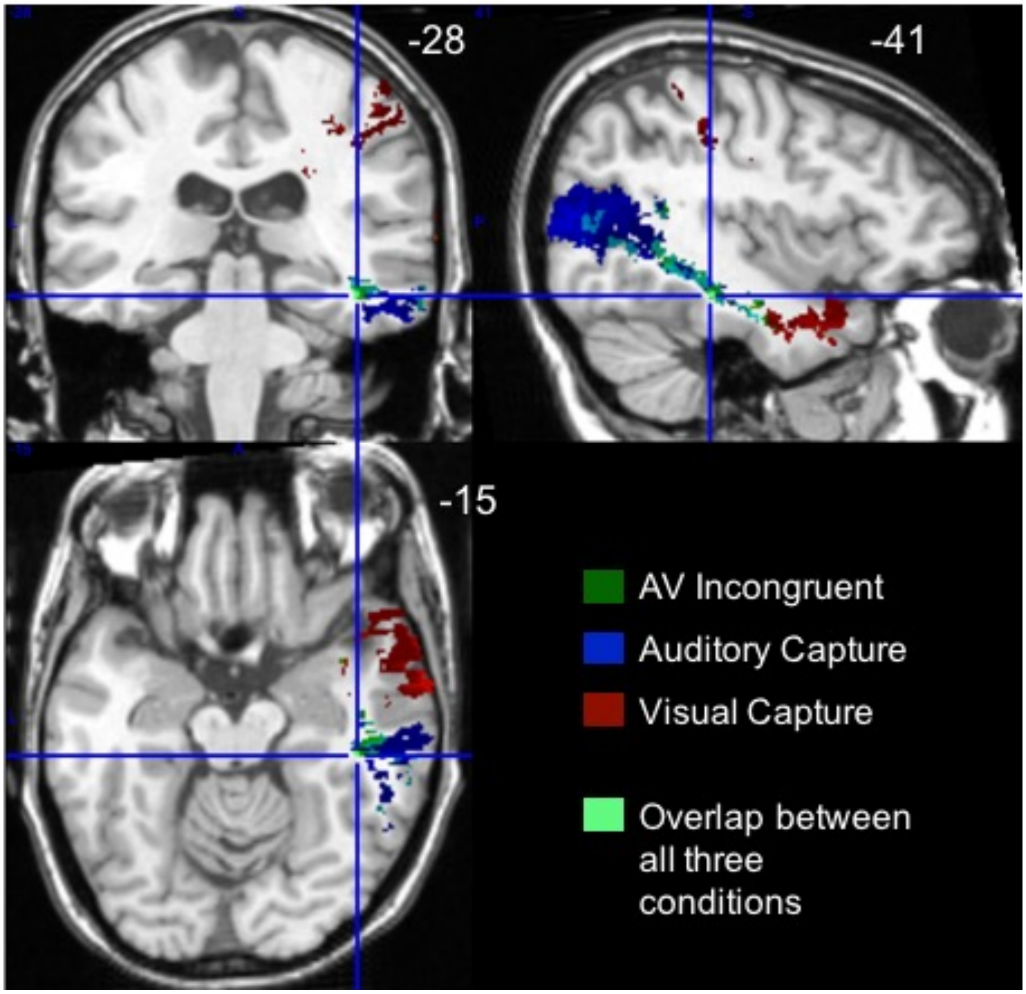

Figure 13. Overlay of VLSM for incongruent AV speech perception (green), auditory capture (blue), and visual capture (red) ($p < .005$). Crosshairs are on overlapping lesion locations for all three conditions (sea green, -41 -28 -15).