Neural dynamics of the motor system in speech perception

Margaret Clarke

A dissertation

Submitted in partial fulfillment of the

Requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Patricia K. Kuhl, Chair

Samu Taulu

Adrian KC Lee

Ludo Max

Program Authorized to Offer Degree:

Speech and Hearing Sciences

University of Washington

Abstract

Neural Dynamics of the motor system in speech perception

Margaret Clarke

Chair of the Supervisory Committee:

Patricia K. Kuhl

Department of Speech and Hearing Sciences

The role of the motor system in speech perception has provoked considerable debate over the past 50 years. Motor and sensory cortices are traditionally thought to be functionally separate systems. However, many studies have shown their roles in both action and perception to be highly integrated. In particular, this has been observed in regard to speech, where listening to speech sounds elicits neural activity in motor regions of the brain in both adults and infants. The primary goal of the present study was to investigate the role of motor and motor planning brain regions in speech perception throughout various stages of language development. Three datasets containing magnetoencephalography (MEG) data from 2-, 6-, 7-, and 11-month-old infants and adults were used to address four experimental questions related to the role that motor brain systems play in the auditory perception of speech. The four experimental questions examine the relationship between activation in auditory sensory and motor regions of the brain with respect to: 1) the *temporal structure* of activation in sensory as opposed to motor brain regions, 2) the *development* of neural responses with increasing age, 3) the role of *language experience*, and 4) potential differences between *speech as opposed to nonspeech* auditory signals. Results showed that motor and motor planning regions are activated during speech perception across all ages. At 2 months of age, infants show activity in both motor and motor planning regions in response to speech, but not to nonspeech acoustic stimuli. This provides evidence that infants' activation of sensory and motor brain regions in response to speech does not require experience producing speech.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest appreciation to my supervisor and committee chair, Dr. Patricia K. Kuhl. I thank Dr. Kuhl for allowing me to explore my research interests and providing me with invaluable teaching throughout this journey. I'd also like to thank her for her confidence and trust in me, her support, and her empathy. It has been a great privilege and honor to work and study under her guidance, and I have grown both professionally and personally because of her.

I would also like to thank the members of my Supervisory and Reading Committees, Drs. Adrian KC Lee, Samu Taulu, Ludo Max and Steve Perlmutter. I thank my committee members for their continued guidance and support through each stage of this process. I am incredibly grateful to have had the opportunity to learn and draw inspiration from such a scientifically diverse, bright and passionate group of scientists.

My dissertation work would not be possible without the help of my friends and colleges at I-LABS. It has been an incredible privilege to work alongside such an amazing group of individuals. Thank you to Dr. Christina Zhao and Dr. Kambiz Tavabi for allowing me to use their datasets. I'd also like to thank Dr. Samu Taulu and Dr. Eric Larson for their mentorship on MEG physics and methodology. I have learned an incredible amount from them both. I'd also like to extend my gratitude to Dr. Alexis Bosseler and Julia Mizrahi. Words cannot express how grateful I am for their friendship, support, and advice they have provided me through this journey.

I cannot express the appreciation I have for my amazing friends and family, who have been my greatest source of strength and support throughout this process. Thank you to my wonderful parents for everything they have done for me, always believing in

me, and always making me laugh when I was about to cry. I love you both so much. Thank you to Margot, Curt, and the boys for opening their home to me and their continued encouragement, love, and support. Thank you to Jason, Erica, and Brandon, for always supporting me and being my biggest fans.

I. Introduction

Language is a fundamentally important skill in humans. It is how we communicate with one another and how we express and understand feelings. Language supports thinking and problem solving and helps us to develop and maintain relationships. Learning to understand and use language is the critical first step in literacy, which provides a tool for learning and innovation.

The complex nature of human language makes it an active field of research which is arguably among one of the most challenging in neuroscience. Infants learn language effortlessly in a remarkably short time. By 12 months, infants can discriminate speech sounds native to their language better than nonnative speech sounds (Kuhl, et al, 2006; Werker & Tees, 1984), show evidence of knowing the legal versus illegal sound combinations in that language (Jusczyk, Hirsch-Pasek, Kemler Nelson, Kennedy, Woodward et al., 1992; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993), track statistical properties of speech input (Grieser & Kuhl, 1989; Kuhl ,Williams, Lacerda, Stevens, & Lindblom, 1992; Maye, Werker, & Gerken, 2002), and produce speech patterns that can be interpreted by their parents (Kuhl, 2004).

Neurophysiological and behavioral research studies show that exposure to language in the first year of life influences the brain's neural circuitry even before infants speak their first words (Kuhl, 2010), a process that has been shown to begin in utero (Mehler, 1999; Moon & Fifer 2000; Moon, Lagercrantz & Kuhl, 2013). Understanding the underlying neural circuitry involved in encoding language in the developing brain presents a difficult challenge. However, studies of the developing brain provide invaluable insight into the mechanisms behind humans' unique ability to acquire language so proficiently, which is critical for treatment of language and speech disorders.

Like other forms of complex information, language has to be transformed from raw input into meaningful representations. One of the first stages of this transformation is perceptual. Speech perception refers to the earliest level of processing involved in the mapping of acoustics to meaning. Despite the effortlessness with which adults perceive speech, there are a number of complex perceptual and cognitive processes involved in achieving this mapping (Holt & Lotto, 2010). It is unquestionable that speech perception recruits various multisensory cortical resources (Sams, 1991; Skipper, Van Wassenhove, Nusbaum, & Small, 2007; Bernstein & Liebenthal 2014; Nath & Beauchamp 2012), including auditory, visual, somatosensory and motor systems. What role each of these systems play, how they interact, and how they develop is not yet fully understood.

A. Is the motor system involved in speech perception?

The role of the motor system in speech perception has provoked considerable debate over the past 50 years (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Our motor and sensory cortices are traditionally thought to be functionally separate systems. However, many studies have shown their roles in both action and perception to be highly integrated. In particular, several studies have shown that both motor and sensory cortices are active during perception (Skipper et al, 2007). This has also been observed in regard to speech, where listening to speech sounds elicits neural activity in motor regions of the brain (Wilson et al., 2004; Pulvermüller, Huss, Kherif, Moscoso del Prado Martin, Hauk, et al., 2006).

One interpretation of the observed motor activity during speech perception is that "the objects of speech perception are the intended phonetic gestures of the speaker"- as posited by Liberman's motor theory of speech perception (Liberman et al., 1967; Liberman and Mattingly, 1985). Liberman's theory argues that people perceive spoken words by identifying the intended vocal tract gestures with which they are pronounced rather than by identifying the auditory sound patterns. Additionally, it is important to highlight that the theory claims that it is the neuro-motor commands to the articulators (e.g., tongue, lips, and vocal folds), or "intended articulatory gestures," rather than actual articulatory movements or gestures that are the basis of speech perception. The theory assumes that there is a close relationship between speech production and perception; the role of the speech motor system is to produce speech articulations as well as to detect them.

The motor theory is supported by the phenomenon of "categorical perception" (CP) of speech sounds (Liberman, Harris, Hoffman, & Griffith, 1957). The study of categorical perception often uses experiments involving discrimination and identification tasks in order to categorize participants' perceptions of sounds. Liberman and colleagues reported that when people listen to sounds that vary along the voicing continuum, for example, from /ba/ to /pa/, they categorize the sounds as belonging to one category or the other, rather than something in between. This effect, in which a perceived sound jumps abruptly from one category to another at a certain point along a continuum, instead of changing gradually, defines categorical perception. Liberman et al. (1967) proposed that CP was unique to speech, and to humans, and that CP provided evidence that speech was processed in a unique manner.

Classic models of speech perception that focus on auditory and general perceptual processing have been proposed as alternatives to the motor theory (Diehl, 1989; Massaro, 1998; Kuhl, 2000). These models claim that speech is not directly mediated through articulatory gestures and that speech is processed by the same mechanisms as nonspeech sounds. Support for this approach is shown in studies demonstrating that both speech and nonspeech sounds can be perceived categorically (Pisoni, 1977), can be induced from learning alone (Lane, 1965; Goldstone, 1994), and that the critical discrimination component (which is the component of CP shown in human infants, see below) can also be shown in nonhuman animals (Kuhl & Miller 1975, 1978). These studies provided evidence that CP was not unique to humans or to speech, and tempered enthusiasm for Liberman’s idea that CP provided the critical support for the theory that speech requires a special process, one unique to humans and unique to speech.

If motor production mediates speech perception, then one would assume that categorical perception of speech sounds is a result of learning to produce speech.

However, studies by Peter Eimas with 1- and 4-month-old infants demonstrated that infants perceive speech sounds categorically before they are able to speak (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). The CP studies conducted on animals are consistent with the idea that CP in infants does not require learning, and raise the possibility that, instead, it could be a function of general auditory processing, and further that the evolution of language may have involved exploiting characteristics of general mammalian auditory processing in selecting the sounds employed in the world's languages (Kuhl, 2000).

B. Multimodal approach

A unique property of speech compared to other auditory signals is that it is multisensory (Guellai, Mersad, & Streri, 2015). Speech involves not only auditory processing, but also visual, motor, as well as somatosensory processing. Moreover, studies have begun to examine brain responses to speech to understand how multimodal inputs are integrated in speech perception.

1. Audio-visual effects on speech perception

a. Adults

It is well established that visual information greatly influences speech perception (McGurk & MacDonald 1976; 1978). Evidence supporting this largely comes from two lines of research: The first line involves studies in which auditory speech stimuli are presented in noise (Sumby & Pollack, 1954; MacLeod & Summerfield, 1987; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007). For example, Ross et al. (2007) found that when auditory speech stimuli are degraded, seeing the talker visually can improve perceptual accuracy. Additionally, when auditory stimuli are intact, visual speech stimuli have an effect on speech perception and comprehension (Reisberg, McLean, & Goldfield, 1987). Perception of a nonnative second language is more accurate with audio-visual stimuli than with auditory-only stimuli (Hazan, Sennema, Faulkner, Ortega-Llebaria, Iba, & Chung, 2006) and comprehension of difficult verbal materials can be made easier under audio-visual conditions (Reisberg et al., 1987). The second line of research comes from the presentation of mismatched auditory and visual speech stimuli. McGurk & MacDonald (1976) reported a compelling illusion occurring with audiovisual speech. Subjects listened to a speaker articulating a consonant while simultaneously watching a face articulating a different consonant. Although the acoustic speech signal was well recognized on its own, it was perceived as a separate consonant after dubbing with incongruent visual speech. The illusion has been termed the McGurk effect. Several studies of the McGurk effect have demonstrated the perceptual link between the auditory and visual system during speech perception. When auditory and visual speech consonants are mismatched, perceivers often hear a third consonant that is different from both the auditory and visual stimuli. For example, auditory /ba/ presented with visual /ga/ is often interpreted as /da/. In this demonstration, perceptual information about the observable place of articulation integrates with acoustic speech.

b. Infants

The influence of visual speech information on speech perception has also been demonstrated in infancy (Kuhl & Meltzoff, 1982; Rosenblum, Schmuckler, & Johnson, 1997; Patterson & Werker, 1999). For example, Kuhl and Meltzoff (1982) tested 4.5-month-old infants' abilities to detect auditory-visual correspondences for speech. In a first experiment infants were presented with two side-by-side visual images of a speaker producing two different vowel sounds: /a/ (as in pop) and /i/ (as in peep), while listening to an auditory sound matching one of the two vowels. In a second experiment, the process was repeated with altered auditory stimuli where the spectral information of each vowel was removed, but the temporal characteristics remained. Results from the first experiment showed that infants fixated longer on the face which matched the auditory vowel sound. This finding was not shown in the second experiment. These results show that infants are able to recognize the correspondence between visual and auditory speech information and that this correspondence is likely based on spectral information. This effect is shown in older infants, however by 9–10 months is observed only for native speech sounds (Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009).

Additionally, typical speech and language development takes place in an audio-visual context, which promotes native language acquisition (Kuhl & Meltzoff, 1996; Lewkowicz, & Ghazanfar, 2006; Irwin, Avery, Brancazio, Turcios, Ryherd, & Landi, 2018). This can be demonstrated in studies which report deficits in audio-visual processing in populations which struggle with speech perception, such as children with autism spectrum disorders (ASD). For example, Irwin et al. (2011) tested children with ASD on audio-visual speech perception tasks, including a McGurk task and an audio-visual speech in noise task. Children with ASD were less influenced by visible articulatory information than typically developing controls, in speech reading (visual-only), speech in noise and with mismatched audio-visual (McGurk) stimuli (see Bosseler & Massaro, 2003).

2. Somatosensory

a. Adults

While the contribution of visual inputs on auditory perception of speech is well documented, the somatosensory effects on speech production is a much smaller and more recent body of research. Several studies show evidence for a role of somatosensory systems in speech perception (Möttönen, Järveläinen, Sams, & Hari, 2005; Skipper et al., 2007; Ito, Tiede, & Ostry, 2009; Correia, Jansma, & Bonte, 2015; Bartoli, Maffongelli, Campus, & D'ausilio, 2016). The most compelling results thus far are studies done with skin stretching or jaw displacement during speech perception, which selectively alter somatosensory feedback. Robotic devices are used to deform the skin on the face and mouth or move the jaw to configurations that would normally accompany speech production. When the subjects' skin is stretched during speech perception, the sounds they perceive are altered (Ito et al., 2009). These results indicate that somatosensory inputs affect the processing of speech sounds and that the somatosensory system is involved in the perceptual processing of speech.

In addition, studies like this have been used to demonstrate that plasticity in sensory and motor systems may be linked. Nasir and colleagues (2009) used a robotic device which displaced subjects' jaws and selectively altered somatosensory feedback during speech. With practice during this “forcefield training” speakers progressively corrected for the perturbation. After this “motor learning” occurred they also showed systematic changes in their perceptual classification of speech sounds. For example, on a speech continuum between the words “head” and “had,” the boundary at which subjects identified one word as one or another, shifted after the force field training. Furthermore, individuals that showed greater amounts of motor learning also showed greater perceptual change in classifying speech sounds, suggesting this perceptual shift is tied to motor learning. This effect was not seen in control subjects who did not experience jaw displacement. The results from this study further support the involvement of the somatosensory system in the neural processing of speech sounds and suggest that speech motor learning results in changes to the way we perceive speech.

More recent studies show evidence of a relationship between somatosensory information and the mechanisms that underlie perceptual learning of speech. A study by Borrie et al. (2018) compared perception of dysarthric speech before and after training. The control group received no training, but 4 other groups received training with dysarthric speech under several conditions; 1) auditory targets, 2) auditory targets and written feedback, 3) auditory targets and a vocal imitation task, 4) auditory targets, written feedback and vocal imitation.

Training improved intelligibility of dysarthric speech in all conditions, with the largest improvements observed when the auditory targets were accompanied by both written feedback and an imitation task. These results suggest that somatosensory information can strengthen the activation of speech sound movements in degraded speech. The findings, therefore, suggest a relationship between somatosensory information and the mechanisms that underlie perceptual learning of speech.

b. Infants

The importance of the somatosensory system for speech perception has been shown in early infancy. Studies demonstrating alterations to articulator movements were shown to affect the discrimination of speech sounds even in the first year (Yeung & Werker, 2013; Bruderer, Danielson, Kandhadai, & Werker, 2015). For example, Bruderer et al. (2015) studied pre-verbal 6-month-old infants during a nonnative (Hindi /d̪/–/ɖ/) speech sound discrimination task, which requires a distinction in tongue placement during production. During three experimental studies, teething toys were used to control the position and movement of the tongue tip while the infants listened to the speech sounds. Experiment 1 was conducted without oral-motor manipulation to verify the successful discrimination of the two sounds. Experiment 2 used a flat teether that selectively blocked tongue tip movement, and experiment 3 used a teether which did not interfere with tongue tip movement. Ultrasound imaging was used to determine infants' tongues were appropriately placed during Experiments 2 and 3. A looking-time procedure was then used to compare between trials where a single speech sound repeated, or one with alternating speech sounds. If infants spent more time attending during the trial with alternating sounds, this was interpreted as an ability to discriminate between them. The results showed that infants were able to discriminate between speech sounds in experiment 1, as well as in experiment 3, but not experiment 2. Temporarily restraining infants' articulators impeded their discrimination of a nonnative speech sound contrast but only when the relevant articulator was selectively restrained to prevent the movements associated with producing those sounds. These results provide evidence that sensorimotor information from the articulators influences speech perception even before infants are able to speak.

Additionally, it is well established with behavioral studies that newborns are able to imitate facial movements such as tongue protrusion and mouth opening, suggesting early integration of vision and proprioception (Meltzoff & Moore, 1977, 1989, 1994). Studies using speech stimuli support these results. For example, Chen, Striano, & Rakoczy (2004) found that newborns will produce more mouth openings when listening to /a/ than /m/ sounds and produce more mouth closing when listening to /m/ than /a/. Supporting data show that when 20-week-old infants were presented with auditory-visual representations of the vowels /a/, /i/, and /u/ in the laboratory, they were more likely to produce the target vowels they heard more than they produced the alternative two vowels (Kuhl & Meltzoff, 1996).

C. Evidence from Neuro-imaging: Adults

The perception of speech signals has traditionally been attributed to the auditory cortices only (Scott & Johnsrude, 1993). Results from neuroimaging studies have consistently shown that activation in left superior temporal regions (superior temporal gyrus, STG, and superior temporal sulcus, STS) is notably greater for speech sounds than that of nonspeech sounds (Dewitt & Rauschecker, 2012). In contrast, processing of nonspeech sounds has been attributed mostly to the primary auditory cortex (A1) and dorsolateral portions of STG. These findings suggest that whereas processing the acoustic features of nonspeech sounds occurs at early levels in the A1, the phonetic processing of speech sounds occurs at a later stage in the auditory processing stream in the secondary auditory cortex (in the surrounding regions of STG and STS; (Woods, Herron, Cate, Kang, & Yund, 2011). However, it is difficult to draw conclusions about speech-specific mechanisms from comparing cortical activations evoked by speech and nonspeech sounds with distinctly different acoustic profiles/features. That is, any observed differences in response may simply be due to differences in the acoustic features of the signals rather than the distinction of whether a signal is speech or not. Thus, it may be that the left STG/STS region is not involved in the processing of speech sounds, but rather contributes more generally in the processing of the kinds of complex acoustic features that are also characteristic of speech sounds.

In addition to temporal areas, frontal and sensorimotor regions seem to be activated in speech processing (Schomers & Pulvermüller, 2016). Many neuroimaging studies have demonstrated activation in the left inferior frontal cortex during passive speech perception (Benson, Whalen, Richardson, Swainson, Clark et al., 2001). Premotor and motor regions have also been implicated in speech perception tasks.

TMS studies have shown that motor-evoked potentials (MEPs) in the articulatory motor cortex are enhanced during speech perception (Fadiga, Craighero, & D'Ausilio, 2009). In addition, many TMS studies inhibiting motor regions show disruption of subjects' ability to perform a phonetic discrimination task.

Sensorimotor integration (SMI) has been said to provide an interface for speech perception and production that is essential for verbal communication (Tourville & Guenther, 2011; Guenther & Vladusich, 2012; Jenson et al., 2014). However, the nature and timing of SMI is not fully understood and relatively few studies attempt to address it (Wilson, Saygin, Sereno, & Iacoboni, 2004; Hickok, Houde, & Rong, 2011). Neuroimaging studies have identified the auditory dorsal pathway (posterior temporal lobe, inferior parietal lobe, premotor cortex; PMC, primary motor cortex M1) as playing a role in both speech perception and production. In perception, SMI is explained through independent yet convergent "dual streams" of neural activity (Scott & Johnsrude, 2003; Hickok & Poeppel, 2004; Hickok, 2009; Dewitt, & Rauschecker, 2012; Specht, 2013; Jensen et al., 2014). The ventral stream (mainly within auditory brain regions) provides speech decoding and comprehension. The dorsal stream (including sensorimotor regions) is thought to provide an audio-motor interface linking auditory input to articulatory goals in speech perception. Though dorsal stream activation has previously been reported to be left-lateralized, there is recent evidence of bilateral organization (Cogan et al., 2014; Simmonds, Leech, Collins, Redjep, & Wise, 2014).

1. fMRI

Functional magnetic resonance imaging (fMRI) is a neuroimaging technique which measures brain activity by detecting changes associated with blood flow. The technique relies on the fact that cerebral blood flow and neuronal activation are coupled. fMRI studies have been used to assess which regions of the brain are active during speech related tasks and whether motor brain areas play a role in speech perception. Early adult studies using fMRI were primarily used to investigate active brain regions during speech perception. These studies used phonetic tasks to show activation in motor regions implicated in speech production, including Inferior Temporal Gyrus (IFG) or Broca's area, the cerebellum, PMC, M1, as well as auditory regions, such as the STG (Démonet et al., 1992; Benson et al., 2001). Because these early studies established motor activation during speech perception and speech production, subsequent studies were then needed to examine whether the observed motor activation was due to general auditory input or was specific to speech stimuli (Watkins, Strafella, & Paus, 2003; Wilson et al., 2004). For example, an fMRI experiment by Wilson et al. (2004) compared activation under three conditions: listening to speech syllables, production of speech syllables, and listening to non- speech. While the authors found activation in the PMC to all stimuli, they observed increased activation in the ventral PMC that occurred during both speech conditions (perception and production) and was greatly reduced in response to perceiving nonspeech stimuli. The results of this study indicate that there are differences in motor region activation to speech vs non- speech stimuli. Building from this finding, other fMRI studies investigated whether during speech perception, activity of articulatory motor areas reflected phonological information (Pulvermüller et al., 2006; Raizada & Poldrack, 2007). Pulvermüller et al. (2006) had subjects listen to syllables starting with a lip-related bilabial /p/ or a tongue-related alveolar phoneme /t/, presented in the absence of any overt motor task. A localizer task was used where subjects produced nonspeech related lip and tongue movements to define regions of interest (ROIs) in the sensorimotor cortex. When examining these ROIs, and other areas in the precentral cortex, the researchers found that areas of the sensorimotor cortex corresponding to the regions which controlled the articulators, were more active. For example, brain areas representative of the lips were active during lip-related bilabial phonemes and tongue areas were active during tongue-related alveolar phonemes. Results from this study suggest motor regions reflect phonological processing and are consistent with the idea that experience producing speech enables the generation of internal models of motor commands required to produce speech. These internal models of motor commands are then compared with incoming sensory data and used during perception.

Although these findings are consistent with the view that speech perception involves the motor system in a process of auditory-to-motor mapping, the role the motor system plays is not clear. One way to explore this is to examine situations where subjects are presented with native and nonnative speech. Because classic brain regions involved in speech perception (i.e., auditory cortex, STG) show differences in activation depending on whether the input is native or nonnative speech, examining whether there are differences in activation in motor regions could determine what role these regions serve.

Several fMRI studies suggest a compensatory role for motor activation in speech perception by looking at motor region activation differences in native and nonnative speech perception. Activation of inferior frontal areas, including IFG were found to be associated with nonnative speech perception and learning, while frontal regions were less active during native language perception (Callan, Callan, & Jones 2014; Golestani & Zatorre, 2004). Callan et al. (2014) used the English /l/-/r/ contrast to compare brain activity in native English and native Japanese speaking adults. The authors reported greater activity in auditory areas (STG) to the native phonetic contrast, and greater activity in motor brain areas (IFG and M1) to the nonnative contrast. Parallel to these findings, Skipper et al. (2007) demonstrated that audio-visual tests of the McGurk Illusion activate motor areas, including the cerebellum and cortical motor areas involved in planning and executing speech movements.

Although these studies do indicate a functional connection between auditory and motor speech systems, and a possible role of the motor system in speech perception, they provide no direct evidence that motor system activation facilitates or is necessary for speech perception.

2. TMS

Transcranial magnetic stimulation (TMS) is a neuromodulatory technique which applies magnetic pulses to the brain via a “coil.” An electric current is delivered to the coil, which generates a magnetic field. The generated magnetic field then induces an electrical current in the brain to excite or inhibit specific regions. TMS has widely been used in cognitive neuroscience, including studies on speech (Murakami & Ziemann, 2013).

Although early fMRI studies have suggested some level of functional connection between auditory and motor-speech systems, they cannot provide direct evidence that activation from motor regions facilitates speech perception without the ability to inhibit motor regions of the brain. In contrast, TMS studies have provided such causal evidence, reporting a reduction in the ability to perform syllable perception tasks when repetitive TMS is applied in order to inhibit the premotor or motor areas (D’ausilio, Bufalari, Salmas, Busan, & Fadiga, 2011; Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Möttönen & Watkins, 2009). For example, Meister et al. (2007) applied repetitive TMS (rTMS) stimulation to the premotor cortex in order to disrupt subjects’ ability to perform a phonetic discrimination task. Subjects performed worse on the task when the premotor cortex was disrupted compared to the same task without disruption and compared to a nonspeech control task which was matched for difficulty, and structure. In addition, Möttönen and Watkins (2009) used rTMS to temporarily disrupt the lip representation in the left primary motor cortex and had subjects perform a categorical perception task with speech sounds that involve the lips (/ba/-/da/ and /pa/-/ta/), and speech sounds which do not (/ka/-/ga/ and /da/-/ga/). This disruption impaired CP of stimuli ranging between two speech sounds that differed in place of articulation (i.e., /ba/-/da/ and /pa/-/ta/). In contrast, it did not impair CP of stimuli ranging between speech sounds that do not involve the lips in their articulation (i.e., /ka/-/ga/ and /da/-/ga/). Furthermore, an rTMS-induced disruption of the hand representation had no effect on categorical perception of either set of speech stimuli. In addition, studies have shown that combining TMS with electromyography can be used as a method to measure excitability in motor areas during speech perception (Adank, Nuttall, & Kennedy-Higgins, 2017; Möttönen & Watkins, 2012; Möttönen, Van de Ven, & Watkins, 2014). For example, TMS pulses over the representations of the articulators (i.e., tongue and lips) in the primary motor cortex elicit motor evoked potentials (MEPs) in the corresponding muscles. Changes in the size of MEPs reflect changes in the excitability of the motor pathways connecting muscles with their cortical representations. Using this technique, several studies have shown that the excitability of the primary motor cortex is enhanced while listening to speech (Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Watkins et al., 2003; Nuttall, Kennedy- Higgins, Hogan, Devlin, & Adank, 2016; Nuttall, Kennedy-Higgins, Devlin, & Adank, 2017).

Although these studies do provide some preliminary evidence that motor regions may have a faciliatory role in speech perception, the precise contribution of motor system activation and whether it is essential for speech perception and comprehension is still under debate.

3. MEG/EEG

Magnetoencephalography (MEG) and Electroencephalography (EEG) are functional imaging methods used to measure brain activity directly and non-invasively (from outside the head). EEG detects synchronized electrical activity from large groups of neurons, whereas MEG detects the changes in magnetic fields that are produced from this electrical activity. While TMS and fMRI are powerful tools to study the brain, MEG and EEG offer an alternative means for measuring neural activity during cognitive processing. MEG/EEG's advantage over most other non-invasive neuro-imaging methods is excellent temporal resolution (~1 ms). Current models of speech perception hypothesize the dorsal pathway is involved in mapping sensory or phonological information onto articulatory-motor representations. The functional organization of the pathway, in terms of the interactions between auditory, somatosensory and motor regions is still largely unknown. In order to gain a better understanding of the functional contributions of dorsal stream activity in speech perception, it is necessary to address the time- course of such activity.

Many MEG and EEG studies that investigate sensorimotor integration during speech perception focus on the mu (*μ*) rhythm, which is characterized by peaks of activity in the (α) alpha (~10 Hz) and (β) beta (~20 Hz) frequency bands. This rhythm is commonly localized to sensorimotor regions (Hari, 2006) including the premotor (Jenson et al., 2014, Schnitzler & Vary, 2000) and primary motor (Schnitzler & Vary, 2000) cortices, within the anterior dorsal stream. Suppression of power is thought to indicate increased cortical activity in the alpha band. This effect has been reported to be found when subjects are performing a motor task, watching performance of a motor task (Muthukumaraswamy & Johnson, 2004; Perry & Bentin, 2010), viewing visually presented speech (Crawcour, Bowers, Harkrider, & Saltuklaroglu, 2009), identification of auditory speech in noise (Cuellar, Bowers, Harkrider, Wilson, & Saltuklaroglu, 2012), and listening to action-based sounds (Pineda et al., 2013).

A study by Jenson et al. (2014) using EEG employed Independent Component Analysis (ICA) to isolate the μ rhythm during discrimination of /ba/-/da/ syllable pairs in both quiet and noisy conditions. Mean event-related changes in spectral power across time, or "event-related spectral perturbations (ERSP)" revealed distinct patterns of activity across the time course of perceptual events. Activity prior to and during stimulus presentation was characterized by β event-related desynchronization (ERD) and α event-related synchronization (ERS) in both quiet and noisy backgrounds. Remarkably, α ERS (i.e., stronger sensorimotor inhibition) became more robust when discriminations were made in the presence of noise. Alpha ERS is thought to reflect cortical inhibition (Jenson & Mazaheri, 2010), which can serve to enhance attention to task-relevant information by filtering out competing information (Zion Golumbic, Poeppel, & Schroeder, 2012). The presence of α ERS prior to stimulus onset was therefore considered "evidence of attentional tuning to enhance upcoming stimulus processing" (Jenson et al., 2014). The strongest pattern of μ activity, which consisted of both α and β ERD, was found following stimulus offset. This pattern was seen in both quiet and noisy conditions and has been observed in a number of speech discrimination studies (Bowers, Saltuklaroglu, Harkrider, Cuellar, & Sinigaglia, 2013; Bowers, Saltuklaroglu, Harkrider, Wilson, & Toner, 2014; Jenson et al., 2014; Saltuklaroglu et al., 2017). This pattern can also be observed during speech production tasks (Gunji, Ishii, Chau, Kakigi,

& Pantev, 2007; Tamura et al., 2012) and in working memory tasks (Behmer & Fournier, 2014). This has led authors to conclude that, in speech discrimination tasks, α and β ERD characterizes sensorimotor processing for the purpose of retaining stimuli within phonological working memory (Hickok, Buchsbaum, Humphries, & Muftuler, 2003).

It is noteworthy that covert speech (i.e., internal speaking) has been previously used to explain motor activity observed in speech perception tasks (Burton, Small, & Blumstein, 2000; Jenson et al., 2014; Roa Romero, Senkowski, & Keil, 2015). However, this claim lacks support without temporal data showing when activity occurred relative to the speech stimulus. By demonstrating anterior sensorimotor disinhibition aligned with auditory inhibition following stimulus offset in the absence of EMG activity, these data support the theory that covert speech can account for some of the motor activity observed during accurate speech discrimination tasks such as these (Burton et al., 2000; Wilson et al., 2004; Callan, Callan, Tajima, & Akahane-Yamada, 2006; Bowers, Saltuklaroglu, Harkrider, Cuellar, & Sinigaglia, 2013; Jenson et al., 2014).

It should also be noted that the source localization of auditory α clusters should be interpreted with caution based on the inherent uncertainty of source localization when performing EEG with only 68 electrodes. The temporal alignment of sensorimotor μ ERD and temporal α ERS strongly suggests the presence of a sensorimotor feedback loop for online monitoring and hypothesis testing and warrants further investigation with methods able to establish better spatial resolution and connectivity between regions.

D. Evidence from Neuro-imaging: Infants

Perception of speech begins as a universal phenomenon where infants are able to discriminate phonetic contrasts in all languages (Kuhl, Ramírez, Bosseler, Lin, & Imada, 2014). It is known that by the end of the first year of life, perception begins to become more specialized to the infant's native language. Discrimination of nonnative contrasts diminishes, while speech perception in the native language improves. What is not known is the role of the perceptual-motor link during speech perception early in life. To date, few neuro-imaging studies exist which examine this topic.

An early study by Imada et al. (2006) demonstrated the existence of motor activity during speech perception in early life. The authors used MEG to compare speech syllables, harmonic chords (to resemble speech) and pure tones (as a nonspeech contrast) between newborns, 6 and 12-month-old infants. Consistent with previous adult research, the authors found activation for all sounds for all ages in superior temporal (ST) regions. As a novel finding, 6- and 12-month-old infants showed activation to speech-like stimuli (i.e., syllables and harmonic chords, but not tones) in inferior frontal (IF) regions. This effect was not seen in newborns. The study provided early evidence for IF contribution to speech perception as early as six months of age. This result approximates the timeline of the onset of canonical babbling, which typically develops between six and eight months and is thought to rely on sensory-motor connections. Although this study suggests the existence of auditory-motor connections in the brain, it does not provide insight into the role that the speech-motor system plays in early speech perception.

A more recent MEG study by Kuhl et al. (2014) examined both auditory and motor brain areas during discrimination of native and nonnative syllables in infants at 7 and 12 months. Results showed 7-month-old infants respond equivalently to native and nonnative contrasts, showing activation in both auditory and motor brain regions. Thus, activation in motor brain regions to speech at 7 months is not limited to sounds infants most commonly hear in their environment and appears to be reacting to all speech-like sounds. By 12 months, infants show higher activation in auditory (i.e., ST) regions for native compared to nonnative stimuli, and higher activation in motor regions (i.e., IF and cerebellum) for nonnative compared to native stimuli. These results suggest that by the end of the first year of life infants' auditory and motor regions change in different directions for native versus nonnative stimuli, suggesting these two brain regions are coding different types of information in response to speech. The authors interpret the data by noting that experience with speech throughout the first year of life increases infants' auditory sensitivity to differences in speech sounds within their native language. Incoming auditory representations of native speech are linked to internally generated speech motor models based on infants' prior production experience, resulting in strengthened sensory-motor pairings. Language experience strengthens both perceptual and motor connections for native language speech. By 12 months the perceptual-motor link relating speech movements to sounds would describe native speech better than nonnative speech, where it would be more difficult to generate internal models. Thus, nonnative speech elicits greater activation in motor regions (e.g., IF and cerebellum) which indicates greater effort, compared to native speech.

I. The Current Study

Three datasets were used in four experiments to explore how sensory and motor representations of speech are connected, and aim to investigate the following;

Experiment 1: Temporal structure (all datasets): (*i*) What is the temporal structure of activation in auditory and motor regions in response to hearing native speech? Hypotheses: a) Auditory and motor regions will be simultaneously activated across all ages. b) The latency of the response in all regions will decrease as age increases c) the magnitude of the response in both auditory and motor regions will increase as age increases.

Experiment 2: Relationship between auditory and motor activation (all datasets): (*ii*) How do neural responses from auditory and motor regions change in relation to another across *age*? *Hypothesis:* The ratio between auditory and motor regions will change throughout development. Specifically, we should see changes around the onset of canonical babbling (7 months) and when production is mastered (adulthood).

Experiment 3: Perceptual experience (dataset 2 - 7 & 11 months): (*iii*) How does the activity in motor regions in response to speech change with perceptual experience? *Hypothesis*: Activation for native and nonnative contrasts will be equivalent at 7 months, but by 11 months we expected native and nonnative speech to activate auditory and motor brain areas differentially, with native activation greater than nonnative in auditory brain areas, and nonnative activation greater than native in motor brain areas.

Experiment 4: Speech vs nonspeech (dataset 1 - 2 & 6 months): (*iv*) How does the activity in auditory and motor regions differ between speech and nonspeech auditory stimuli? *Hypothesis:* Activation in auditory and motor regions in response to speech should differ from nonspeech, based on auditory and motor experience.

A. Methods

A. *Participants*

All participants were recruited through the University of Washington subject pool. All experimental procedures were approved by the Institutional Review Board of the University of Washington, and all informed consent was obtained from the adult participants and the parents of the infant participants.

Dataset 1 (2 & 6 months):

Thirty-two typically developing infants from monolingual American-English households participated in a longitudinal MEG study, where they were recorded at 2 and 6 months of age. The inclusion criteria included the following: (a) no regular exposure to any languages other than English since birth, (b) full term and born within 14 days of due date, (c) no known health problems and no more than three ear infections.

Dataset 2 (7 & 11 months):

Forty-nine 7-month-old infants and 39 11-month-old infants participated in the study. The inclusion criteria included the following: (a) no regular exposure to any languages other than English since birth, (b) full term and born within 14 days of due date, (c) no known health problems and no more than three ear infections and (d) birth weight ranging from 6 lb to 10 lb.

Dataset 3 (adults):

Monolingual English speakers (*N* = 15, male = 9, age = 21.48 ± 2.15) were recruited. All participants were healthy right-handed adults with no reported speech, hearing or language disorders.

B. MEG recordings

All MEG data were recorded using a whole-head Elekta-Neuromag Vectorview system (Elekta, Oy, Helsinki, Finland) with 306 superconducting quantum interference device sensors (SQUIDS), including 204 planar gradiometers and 102 axial magnetometers. To reduce external magnetic interference, the MEG system is housed in a two-layer magnetically shielded room (IMEDCO, Noblesville, IN) at the University of Washington's Institute for Learning & Brain Sciences. To prepare the participants for the recording, five head position estimation (HPI) coils were attached to adults' heads, and attached to a nylon cap and placed on the infants' heads. The locations of the HPI coils with respect to the head as well as 3 cardinal points and at least 150 additional points along the head surface were digitized with the Fastrak 3D stylus digitizer (Polhemus, Colchester, VT, USA). During recordings, HPI coils were activated continuously to generate alternating magnetic fields at frequencies between 83 and 321 Hz and were used to track participant's head positions throughout the measurement. In adults, two electrooculogram (EOG) electrodes were attached above and below the left eye to record eye blinks and movements that may occur during data collection, and two electrocardiogram (ECG) electrodes were attached to the chest and back to record heartbeat artifacts for both adult and infant subjects. Adults were placed in an upright chair under the MEG sensor array, while infants were placed in a custom-made adjustable chair with cushions that made it easy to adjust their height for an optimal position within the helmet for MEG recording. During infant recordings, an assistant waved toys silently in front of the infants to entertain them. As the task requires only passive listening to auditory stimuli, a silent video played for all participants in the background throughout the session.

C. MEG stimuli and procedures

The sound stimuli from all datasets were delivered from a TDT RP 2.7 (Tucker-Davis Technologies Real-Time Processor), controlled by custom Python software, to either insert earphones (for adults) or a flat speaker (for infants). All audio stimuli were presented at 65 dB SPL (infants) or to a comfortable hearing level (adults) and resampled to 24,414 Hz for the TDT.

Dataset 1 (2 & 6 months):

A double oddball paradigm was used to elicit the mismatch response (MMR) response in infants during MEG recordings. Oddball stimuli were voice onset time (VOT) token syllables (/ba/, /wa/) pseudo randomly presented alongside the standard stimulus conveying the mid-point consonantal VOT duration between VOT in deviant syllables. The stimuli were pseudo-randomly arranged using a 4:1 ratio of standard-to-deviant stimulus, such that the same deviant never occurred consecutively and was always separated by at least 3 occurrences of the standard. To establish a strong auditory memory trace, 80 consecutive trials of the standard stimulus were presented at the onset of each MEG recording session. Each participant was presented with a sequence of 680 auditory syllabic trials separated by an onset asynchrony of 1000 ± 500 ms. Stimuli were presented in a single block at a sampling rate of 1800 Hz. Only the deviant syllable /ba/ was analyzed for experiments. In a second run, a 1000 ms amplitude modulated (40 Hz) chirp tone was presented 120 times. The stimuli were presented in a single block at a sampling rate of 1800 Hz.

Datasets 2 & 3 (adults, 7 & 11 months):

A double oddball paradigm was used with a bilabial stop consonant with +10 ms VOT common to English and Spanish (perceived as /ba/ in English and /pa/ in Spanish) served as the standard sound. A bilabial stop consonant with -40 ms prevoiced VOT exclusive to Spanish (perceived as /ba/) served as one deviant sound, and a bilabial stop consonant with +40ms VOT exclusive to English (perceived as /pa/) served as the second deviant. The duration of the standard stimuli was 100 ms, and the duration of all deviant stimuli were 130 ms. The inter-stimulus intervals were 150 ms with jitters within a ±10-ms window. The stimuli were presented in two blocks (3000 sounds per block) at a sampling rate of 5000 Hz.

D. Data Simulation

In order to test whether our analysis pipeline would be effective to differentiate sources related to speech perception in infant brains, we first simulated ‘realistic’ MEG data. Three spatially distinct signals with different amplitudes, latencies and durations at three seed locations were simulated within a 3-month-old infant head model. The first source was intended to represent activity in the left motor/somatosensory cortices in response to listening to speech. We simulated a time series where the latency was 0.05 s with a duration of 0.4 s and an amplitude of 5 nAm. The two remaining sources were intended to represent the bilateral auditory processing within the superior temporal gyri (STG) in response to speech perception. The latency was 0.1 s for the left auditory source and 0.12 s for the right auditory source. Both sources had a duration of 0.2 s and amplitude of 8 nAm. Simulated sources are shown below.

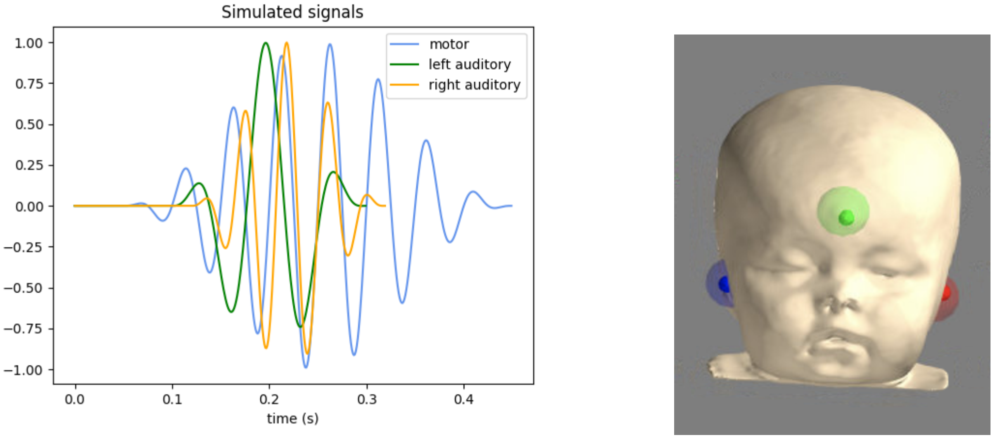

Figure 1. Time Courses of simulated signals (left). 3-month-old infant head model surface (right).

Source waveforms were simulated in each seed location; the center of mass vertex of ROI labels which were defined using the anatomical cortical parcellation proposed by Glasser and colleagues (2016). Data structure and information were used from a real recording of an 11-month-old infant during a speech perception task. Realistic head movements and noise calculated from the noise covariance were added from the same recording. Cardiac artifacts simulated from real signals were also added to the simulated data. *Processing:* The simulated data was preprocessed using Temporal Signal Space Separation (tSSS) (Taulu & Simola, 2006) to suppress external interference artifacts and perform movement compensation from HPI coil data. The MEG signals were transformed into the subject's time-weighted averaged head position. Principal component analysis (PCA) was used to identify the cardiac artifact components from the simulated ECG channel, and Signal Space Projection (SSP) (Uusitalo & Ilmoniemi, 1997) was used to project these signals out of the data. The data was epoched around the onset of the stimulus (-0.1 - 0.5 s) and averaged together to create event-related fields (ERFs). A noise covariance was computed from the baseline signals preceding the onset of the stimuli as an estimate of the noise of the data and used to whiten the data (i.e., improve the signal-to-noise ratio [SNR]). A template source space and conductor model was produced from an average of MRIs from 3-month-old infants. The MRI was aligned and warped to optimally fit the subject's digitized head points, and the resulting affine transformation was applied to the template source space and inner skull surface. A Boundary Element Model (BEM) was used for forward modeling (i.e., to calculate the expected signal at each sensor for each dipole location). Sources were reconstructed with both a linear inverse solver based on $\ell 2$ minimum norm, and $\ell 1/\ell 2$ mixed-norm solver. The mixed-norm approach has the ability to incorporate some additional assumptions about the sources compared to minimum norm estimates. Since we are interested in measuring independent brain areas that are relatively close in the brain, within a small head, the mixed-norm solver should best suit our purposes. Reconstructed source time courses were then extracted from each ROI and compared.

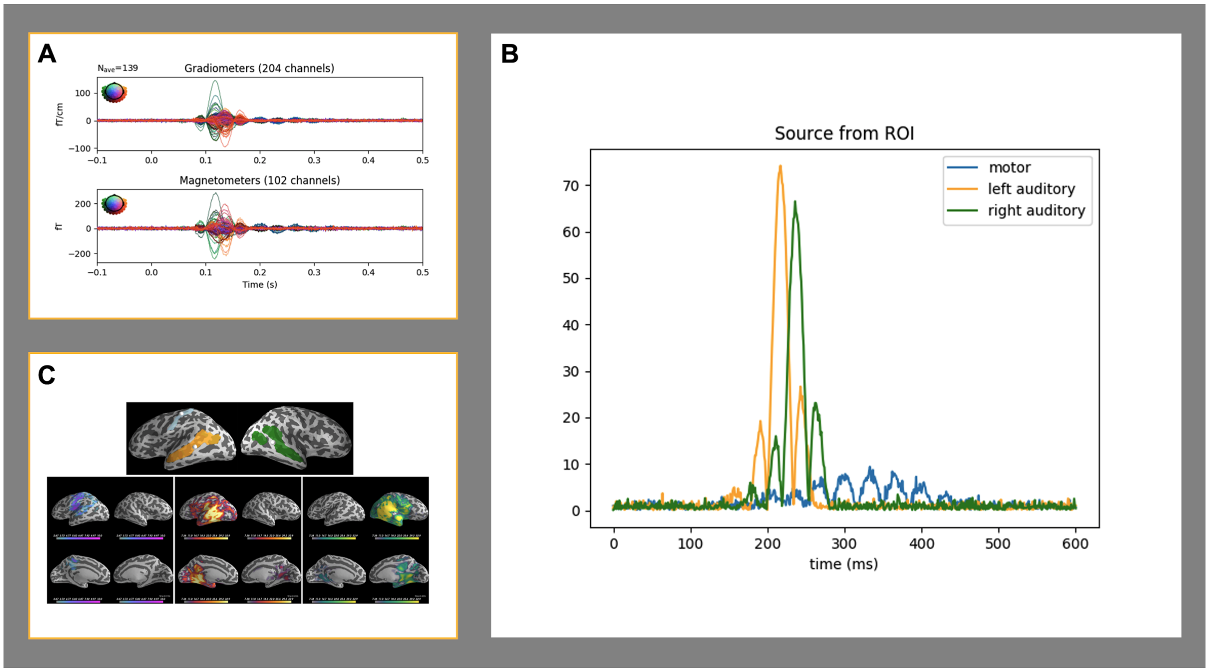

Figure 2. A). Sensor plot of simulated data with three temporally overlapping sources. **B)**. Source time courses of temporally overlapping sources from each ROI, reconstructed with mixed-norm solver. **C)**. Top row: colored labels show ROIs from each hemisphere, and white spheres show center of mass for each label. Bottom row: Source distribution reconstructed at the peak times of each simulated signal (0.11, 0.26 and 0.42 s). These temporally non-overlapping signals were simulated in order to show the spatial distribution patterns of each source.

Results: Although the simulation demonstrates point-spread in each region, when the waveforms were simulated within the three original seed regions (motor/somatosensory, left and right auditory), the source results revealed intact and separate sources from each using both ℓ2 minimum norm, and ℓ1/ℓ2 mixed-norm solver.

E. Analysis

All data were analyzed using MNE Python software. Data were preprocessed using tSSS (Taulu & Simola, 2006) to suppress external interference artifacts and perform movement compensation from HPI coil data. Some tSSS parameters were adjusted to optimize external noise suppression for age and head size; 1) The tSSS duration length was 6 s for ages 2 and 6 months, 4 s for ages 7 and 11 months and 10 s for adults. 2) The tSSS correlation limit was .95 for 2 months and .98 for all other age groups. 3) The internal subspace value was 6 for 2, 6 and 7 months, and 8 for 11 months and adults. MEG signals were transformed into the time-weighted averaged position for each individual subject. PCA was used to identify any eye and cardiac artifact components and Signal Space Projection (SSP) (Uusitalo & Ilmoniemi, 1997) was used to project these signals out of the data. EOG and ECG electrodes were used as representative channels in adults, 7- and 11-month-old infants. The 2- and 6-month-old infant datasets did not have ECG or EOG electrodes, so a representative channel showing visible ECG artifacts was chosen as an alternative. All data was epoched around the onset of the stimulus (-0.1 - 0.7 s) and averaged together to create event-related fields (ERFs). Epochs were rejected if there were large artifacts which were not removed by preprocessing methods, or if the SNR of less than 3 HPI coils is unsatisfactory (since accurate movement compensation requires 3 or more coils). Noise covariances were calculated from empty room recordings as an estimate of the noise and used to whiten the data. Evoked data as well as removed ECG & EOG projection components were visually inspected to ensure each dataset was of good quality.

Source Estimation: A template source space and conductor model was produced for each infant age group from an averaged brain template of a similar age (3, 6, 12 months and adult). The MRIs were aligned and warped to optimally fit each subject's digitized head points, and the resulting affine transformation was applied to the template source space and inner skull surface. A Boundary Element Model (BEM) was used for forward modeling.

Distributed source modeling: To calculate spatiotemporal source distribution, the present study employed dynamic statistical parametric mapping (dSPM), a noise-normalized, anatomically constrained linear inverse technique (Dale et al., 2000) that creates statistical maps estimating dipole strength at each voxel and time point in 5 ms increments.

Regions of interest (ROIs): ROI labels were defined using the anatomical cortical parcellation proposed by Glasser and colleagues (2016). The average activation time-course across vertices within each label was extracted for all subjects. Time windows of interest (TOI) were chosen based on the average peak latency for each age group (s); 0.2-0.35 (2-months), 0.15-0.25 (7 & 11months) and 0.05-0.15 (adults). Mean values were calculated over each TOI for each ROI in both hemispheres. In order to compare the ratio between auditory and motor regions across age, mean values for each motor ROI were subtracted from the mean STG value for the same hemisphere then normalized to the geometric mean. This resulted in normalized difference values: $S - M/\sqrt{(S \cdot M}$ where S represents activity from the STG region and M represents activity from each motor region.

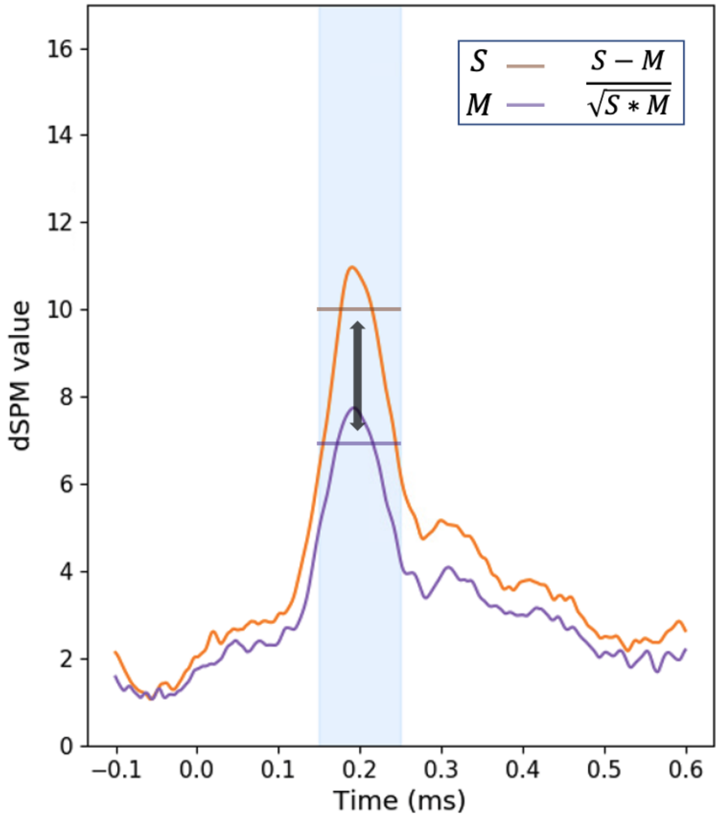

Figure 3. Normalized difference values are calculated to examine the ratio between auditory and motor brain regions. Source waveforms are shown for auditory regions (orange) and motor regions (purple). The shaded blue region represents the TOI. For each subject the mean value within the TOI is calculated for both STG (S) and motor region (M). These values are subtracted and normalized to the geometric mean.

Point spread modelling

In order to ensure any differences in activation over age are not explained by point spread, we modelled the point spread between each ROI for each age to compare to our results. We computed all-to-all pairwise leakage among our 6 ROIs in source space based on MNE inverse solutions and the FreeSurfer (Fischl, 2012) cortical parcellation labels. Label-to-label leakage was estimated as the correlation among the labels' point-spread functions (PSFs) then compared to brain data to check for any correlation.

II. Results

Three datasets containing data from 2-, 6-, 7-, and 11-month-old infants and adults were used to address four experimental questions related to the role that motor brain systems play in the auditory perception of speech. The four experimental questions examine the relationship between activation in auditory sensory and motor regions of the brain with respect to: 1) the *temporal structure* of activation in sensory as opposed to motor brain regions, 2) the *development* of neural responses with increasing age, 3) the role of *language experience*, and 4) potential differences between *speech as opposed to nonspeech* auditory signals.

Experiment 1 - Temporal Structure

In order to investigate the time course of brain activation in both auditory (superior temporal gyrus - STG) and motor regions (primary motor cortex - M1 and inferior frontal gyrus - IFG) we examined both the magnitude and latency of the response to the speech across age groups. All subject groups were tested using the CV syllable /ba/ (see Methods for details). A linear mixed effects model (LME) was performed to evaluate the effects of age, ROI and hemisphere on both latency and mean magnitude.

Latency effect

As anticipated, the latency of response decreased as a function of age in both auditory and motor regions. A display of the latency data as a function of age is shown in Figure

4. As shown, the average latency of response in the left STG region was 0.296, 0.223, 0.211 and 0.105 s and 0.296, 0.216, .200 and .117 s in the right hemisphere for 2-,7-,11-month and adults, respectively. This general pattern of results obtained for the STG analysis was also shown in the M1 and the IFG.

There were no identified extreme outliers, and the data was normally distributed, as assessed by the Shapiro-Wilk's test of normality ($p > 0.05$). There was homogeneity of variances ($p > 0.05$) as assessed by Levene's test of homogeneity of variances. Mauchly's test of sphericity indicated that the assumption of sphericity was met ($p > 0.05$).

There was no statistically significant three-way interaction between Age, ROI and hemisphere: [$F(6, 348) = 0.826$, $p = 0.55$]. For the two-way interactions and main effects, a Bonferroni adjustment was applied leading to statistical significance being accepted at the $p < 0.025$ level.

There was no statistically significant two-way interaction between age and ROI: [$F(6, 348) = 1.599$, $p = 0.15$], age and hemisphere: [$F(3, 348) = 1.944$, $p = 0.12$] or ROI and hemisphere: [$F(2, 348) = 0.344$, $p=0.71$].

There was a significantly significant main effect of age: [$F(3, 348) = 423.602$, $p < 0.0001$], however no significant main effects for ROI: [$F(2, 348) = 1.533$, $p = 0.22$], or hemisphere: [$F(1, 348) = 2.488$, $p = 0.12$].

All pairwise comparisons were run between age groups for each ROI. A Bonferroni adjustment was applied where the p-value was adjusted to 0.25. The mean latency was significantly different between all age groups in all ROIs ($p < 0.01$).

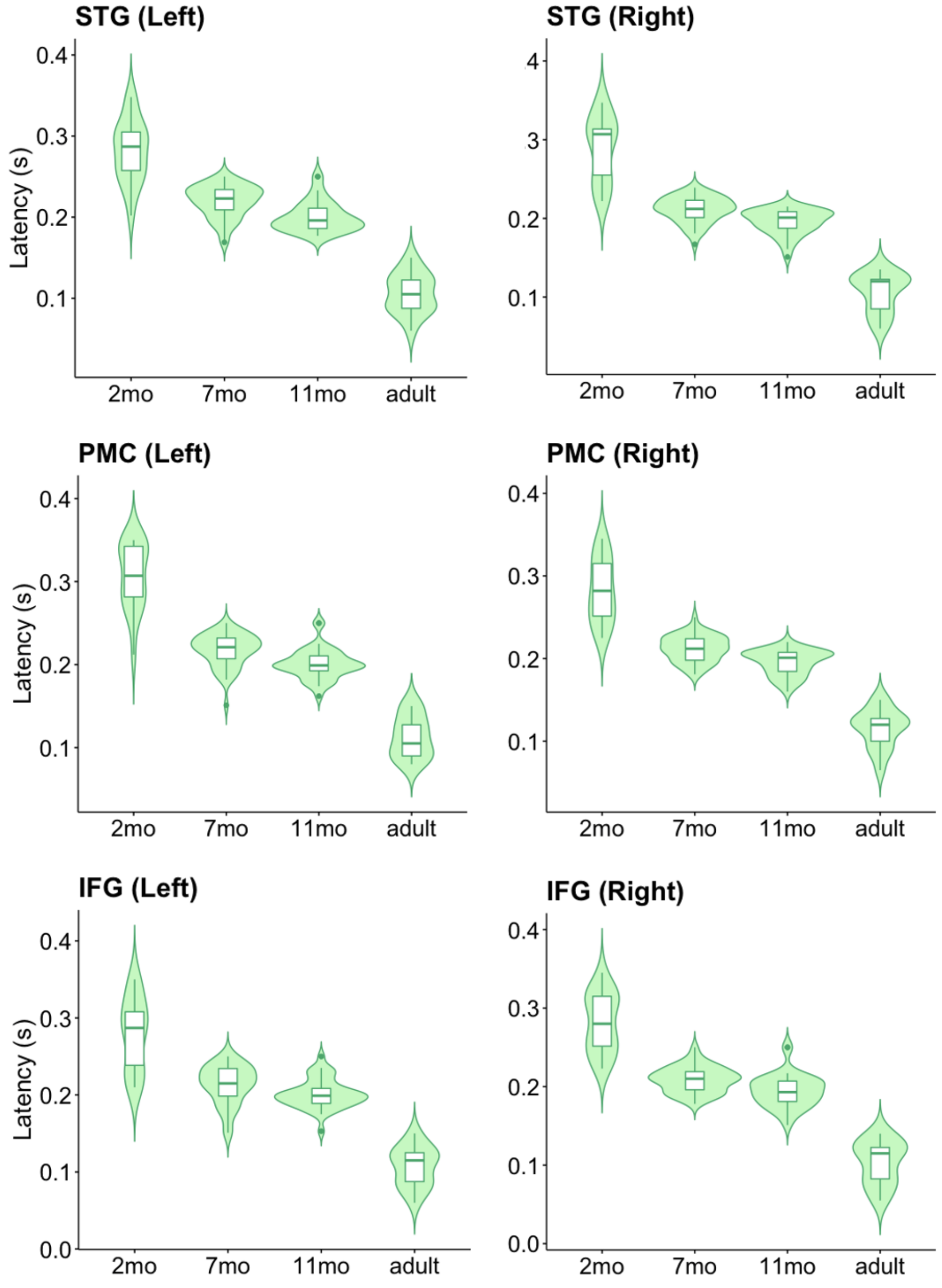

Figure 4. *Younger children exhibited later latencies across all ROIs.* Violin plots of response latencies for all age groups in superior temporal gyrus (top row), primary motor cortex (middle row) and inferior frontal gyrus (bottom row). Significant differences were found between all age groups.

Magnitude effect

As anticipated, all age groups showed activity in both STG and primary motor cortices in response to the speech. As depicted in Figure 5, the mean magnitude of neural response to the syllable /ba/ generally increased between 2-month and adults across both sensory and motor areas.

There were no extreme outliers, and the data was normally distributed, as assessed by the Shapiro-Wilk's test of normality ($p > 0.05$). There was homogeneity of variances ($p > 0.05$) as assessed by Levene's test of homogeneity of variances. For the three-way interaction effect, Mauchly's test of sphericity indicated that the assumption of sphericity was met ($p > 0.05$).

There was no statistically significant three-way interaction between Age, ROI and hemisphere: [$F(6, 348) = 1.264$, $p=0.27$]. For the two-way interactions and main effects, a Bonferroni adjustment was applied leading to statistical significance being accepted at the $p < 0.025$ level.

There was a statistically significant two-way interaction between age and ROI: [F(6, 348 = 4.628, $p < 0.001$], but not age and hemisphere: [$F(3, 348) = 1.991$, $p = 0.11$] or ROI and hemisphere: [$F(2, 348) = 0.525$, $p = 0.59$].

There was a statistically significant main effect of age: [$F(3, 348) = 23.507$, $p < 0.0001$], and ROI: [$F(2, 348) = 91.539$, $p < 0.0001$]. However, there was no significant main effect of hemisphere: [$F(1, 348) = 1.268$, $p = 0.26$].

All pairwise comparisons were run between age groups for each ROI. A Bonferroni adjustment was applied. The mean value was significantly different between 2-months and other age groups In the STG and IFG ($p < 0.01$), but not the M1.

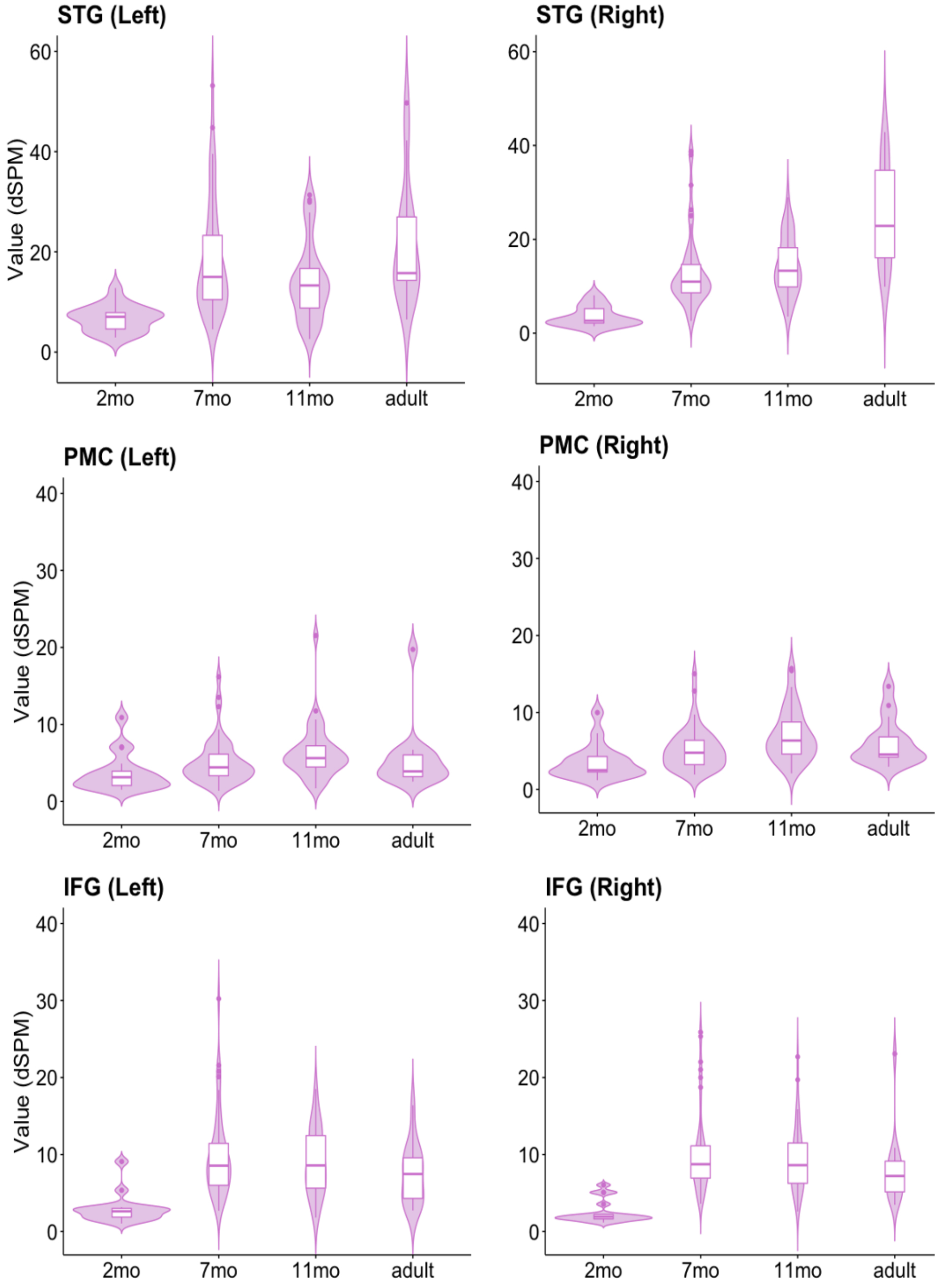

Figure 5. *Younger infants show lower magnitude responses to speech in STG and IFG.* Violin

plots of peak magnitude for all age groups in superior temporal gyrus (top row), primary motor cortex (middle row) and inferior frontal gyrus (bottom row).

Experiment 2 - Development

Ratio effect

In order to examine the interaction between auditory and motor brain activation with age as the variable of interest, the ratio between auditory (STG) and motor regions (M1 or IFG) was examined. An LME model was performed to evaluate the effects of age, ROI and hemisphere on the normalized difference-values for each motor ROI (i.e., STG-M1 and STG-IFG). As displayed in Figures 6 & 7, the relationship between the magnitude of auditory and motor brain activation in response to speech changed with age (Figure 6A & B) and hemisphere (Figure 7B).

There was no statistically significant three-way interaction between Age, ROI and hemisphere: [$F(3, 236) = 2.126$, $p = 0.09$]. For the two-way interactions and main effects, a Bonferroni adjustment was applied leading to statistical significance being accepted at the $p < 0.025$ level.

There was a statistically significant simple two-way interaction between age and ROI: [$F(3, 236) = 7.575$, $p < 0.0001$], and ROI and hemisphere [$F(1, 236) = 6.609$, $p = 0.01$], but not age and hemisphere: [$F(3, 236) = 0.208$, $p = 0.89$]

There was a statistically significant main effect of age: [$F(3, 236) = 14.184$, $p < 0.0001$], however, no main effect of ROI: [$F(1, 236) = 0.375$, $p = 0.54$] and hemisphere: [$F(1, 236) = 0.015$, $p = 0.91$].

All pairwise comparisons were run between age groups for each ROI. A Bonferroni adjustment was applied. The mean value was significantly different between 2-months and all other age groups In the STG and IFG ($p < 0.01$), but not the M1.

In order to ensure that point spread between auditory and motor regions is not responsible for our effects, we estimated the label-to-label leakage (point-spread) correlations which we then compared to the normalized difference values for each label to label comparison. No significant correlation was found between point-spread and ratio values across sensory auditory and motor ROIs ($R = -0.019$).

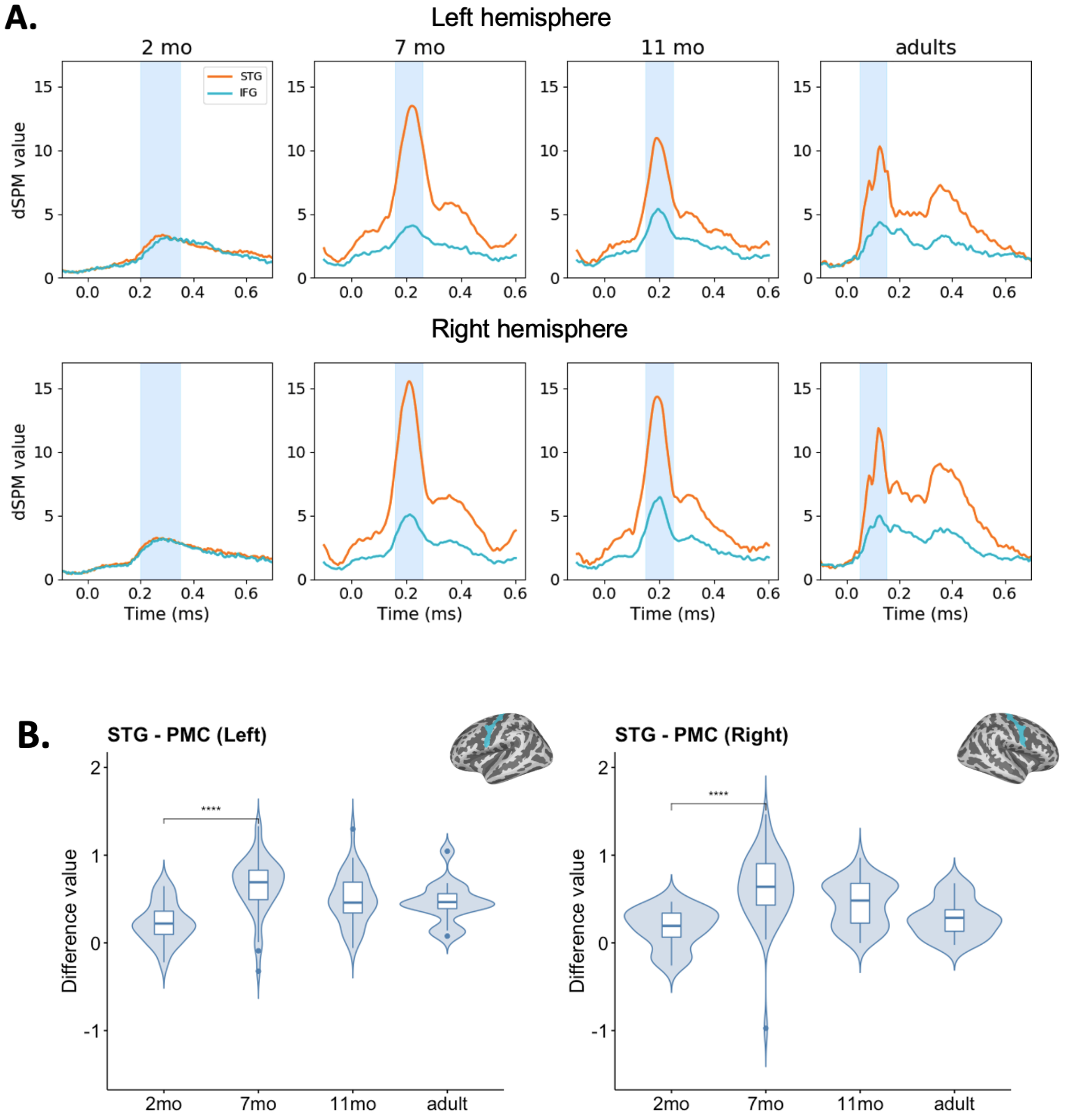

Figure 6. *Ratio between STG and M1 activation changes between 2 and 7 months*. Grand averaged source waveforms **(A)** for STG (orange traces) and M1 (teal traces) are shown for each age group. Time-windows of interest are shown as blue shaded regions. Violin plots in **(B)** show the distribution of M1 difference-values for each age group. Asterisk represents significant differences between age groups which represent main developmental change.

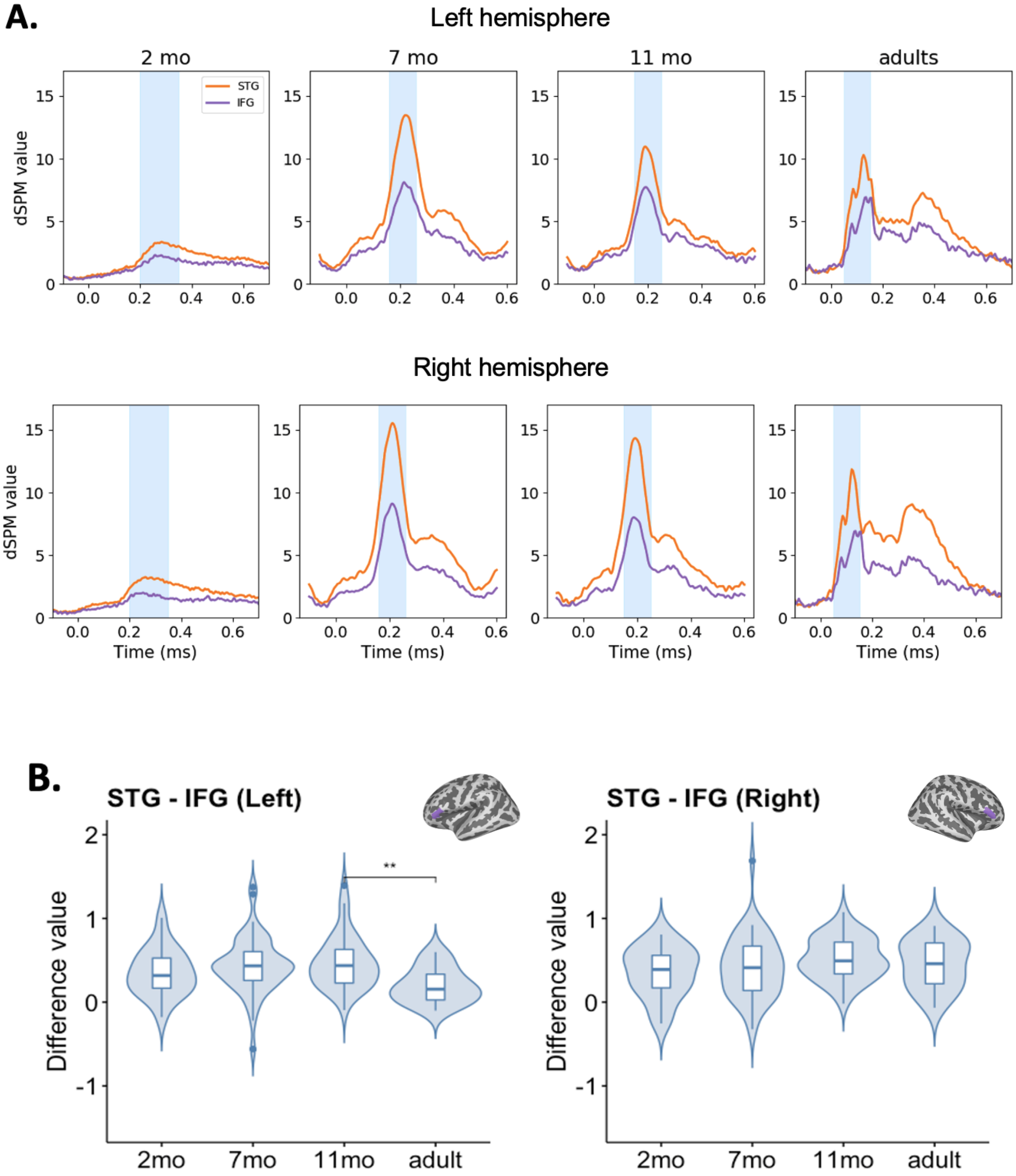

Figure 7. *Ratio between STG and IFG activation changes between 11-months and adults*. Grand averaged source waveforms **(A)** for STG (orange traces) and IFG (purple traces) are shown for each age group. Time-windows of interest are shown as blue shaded regions. Violin plots in **(B)** show the distribution of IFG difference-values for each age group. Asterisk represents significant differences between age groups which represent main developmental

change.

Point-spread

Label-to-label leakage between all ROIs was estimated as the correlation among the labels' point-spread functions (PSFs) then compared to brain data to check for any correlation. No significant correlation was found between PSF and ratio between auditory and motor results ($R = -0.91$)

Experiment 3 - Experience

In order to investigate the contribution of auditory and motor brain regions on the developmental transition where native discrimination improves and nonnative declines between 7 and 11-months we compared peak magnitudes across three conditions: Age (7 & 11mo), language condition (native, nonnative) and ROI (STG, M1, and IFG). A LME model examined the effects of age, language condition and ROI.

Magnitude effect

There were no extreme outliers and the data was normally distributed, as assessed by the Shapiro-Wilk's test of normality ($p > 0.05$). There was homogeneity of variances ($p > 0.05$) as assessed by Levene's test of homogeneity of variances, and the assumption of sphericity was met ($p > 0.05$), determined by Mauchly's test of sphericity.

The results revealed two significant main effects: age: [$F(2,87) = 7.6066$, $p < 0.01$], and ROI: [$F(5,87) = 135.03$, $p < 0.0001$], as well as a significant interaction, age × ROI: [$F(5,87) = 2.44$, $p = 0.033$]. There was no significant effect of language condition: [$F(1,87) = 0.22$, $p = 0.63$] (Figure 8).

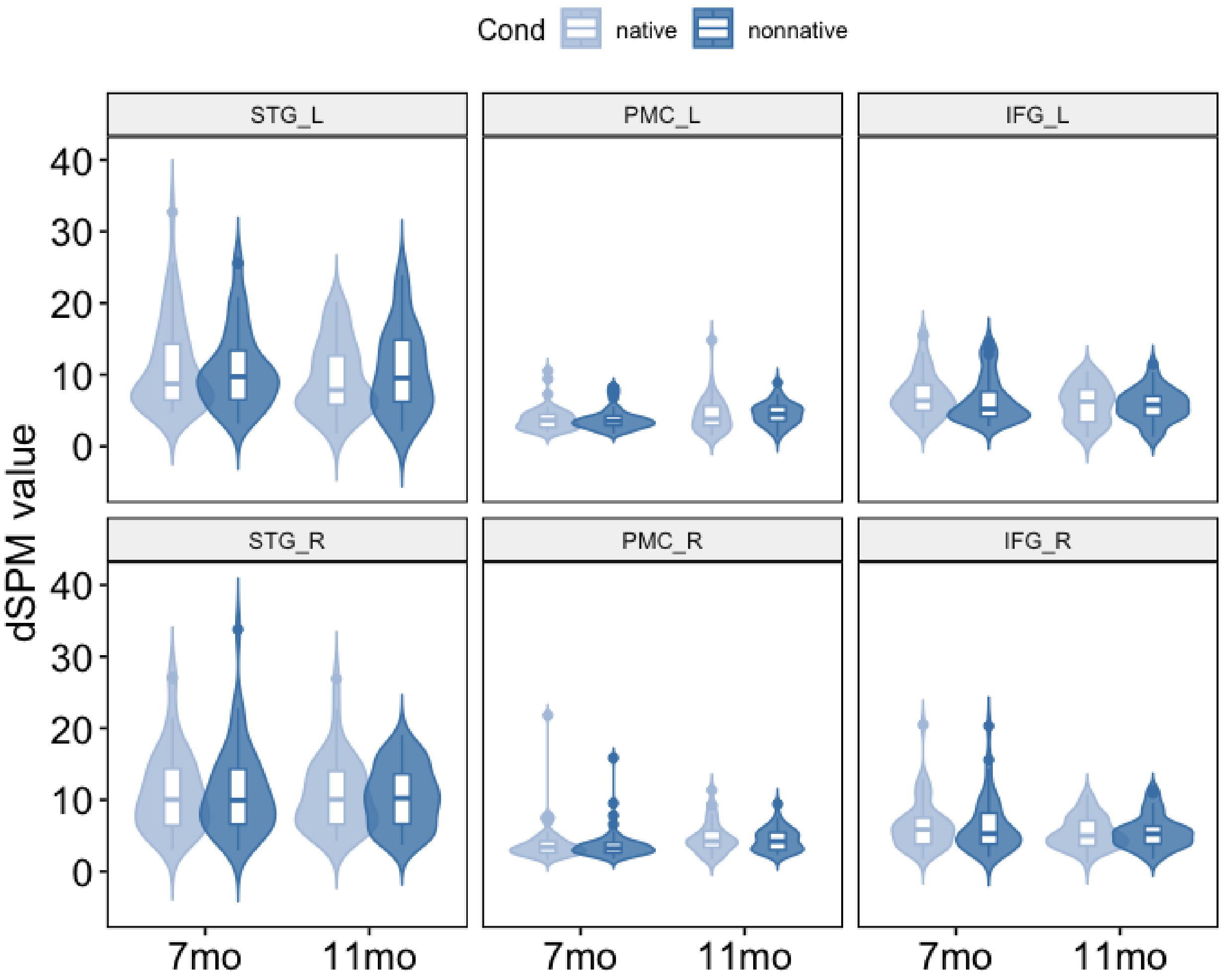

Figure 8. Violin plots of response magnitude for 7- and 11-month-old infants. No significant differences were found between native and nonnative speech sounds across ROIs. Left hemisphere (top row), right hemisphere (bottom row).

Experiment 4 - Speech vs Nonspeech

Ratio effect

In order to examine the interaction between sensory and motor brain activation with stimulus type as the variable of interest, the ratio between auditory (STG) and motor regions (M1 or IFG) was examined. Two-way repeated measures analysis of variance (ANOVA) was performed to evaluate the effects of stimuli type (speech, nonspeech) and ROI (STG, M1 and IFG) on the normalized difference-values for each motor ROI (i.e., STG-M1 and STG-IFG).

2-months

As displayed in Figure 9 & 10, the ratio of the magnitude between auditory and motor brain activation changed across stimuli type, where the ratio was larger for nonspeech stimuli across all brain regions.

There were no extreme outliers, and the data was normally distributed, as assessed by the Shapiro-Wilk's test of normality ($p > 0.05$). There was homogeneity of variances ($p > 0.05$) as assessed by Levene's test of homogeneity of variances. For the two-way interaction effect, Mauchly's test of sphericity indicated that the assumption of sphericity was met ($p > 0.05$).

There was no statistically significant two-way interaction between stimulus type and ROI: [$F(1.98, 31.64) = 0.962$, $p = 0.392$]. The ANOVA revealed a statistically significant effect of stimulus type: [$F(1, 16) = 16.210$, $p = 0.0009$], but no significant effect of ROI: [$F(1.86, 29.77) = 2.977$, $p = 0.07$].

Post hoc pairwise comparisons were run between stimulus types for each ROI. A Bonferroni adjustment was applied. The pairwise comparisons revealed a significant difference between speech and nonspeech for every ROI; Left IFG: $p = 0.03$, Left M1: $p = 0.018$, Right IFG: $p = 0.028$ and Right M1: $p = 0.021$.

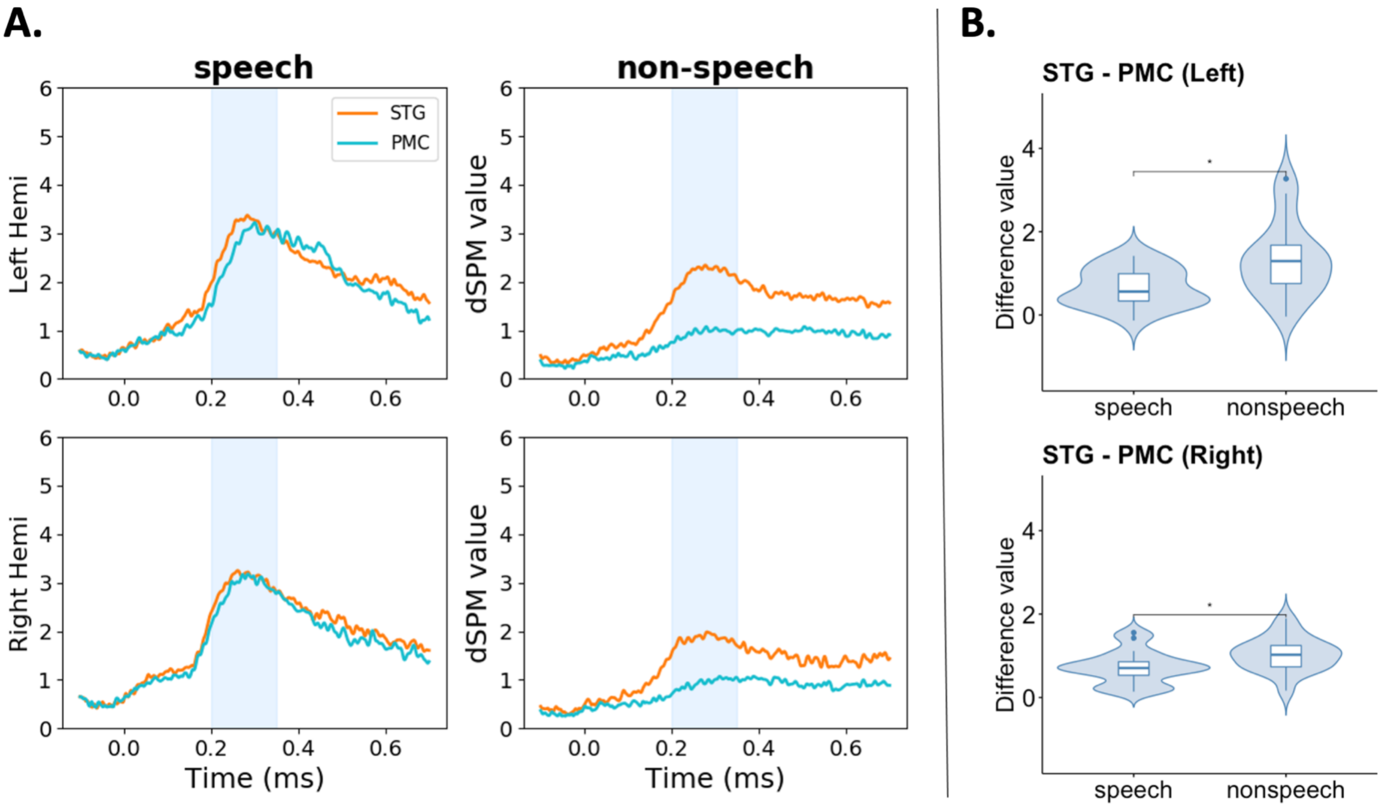

Figure 9. *2-month-old infants show differences between speech and nonspeech stimuli in the M1***.** Grand averaged source waveforms of 2-month-old infants **(A)** for STG (orange traces) and M1 (teal traces) are shown for each stimulus type (column) and hemisphere (row). Time-windows of interest are shown as blue shaded regions. Violin plots in **(B)** show the distribution of M1 difference-values for each hemisphere and stimulus type. Asterisk represents significant differences between stimulus types.

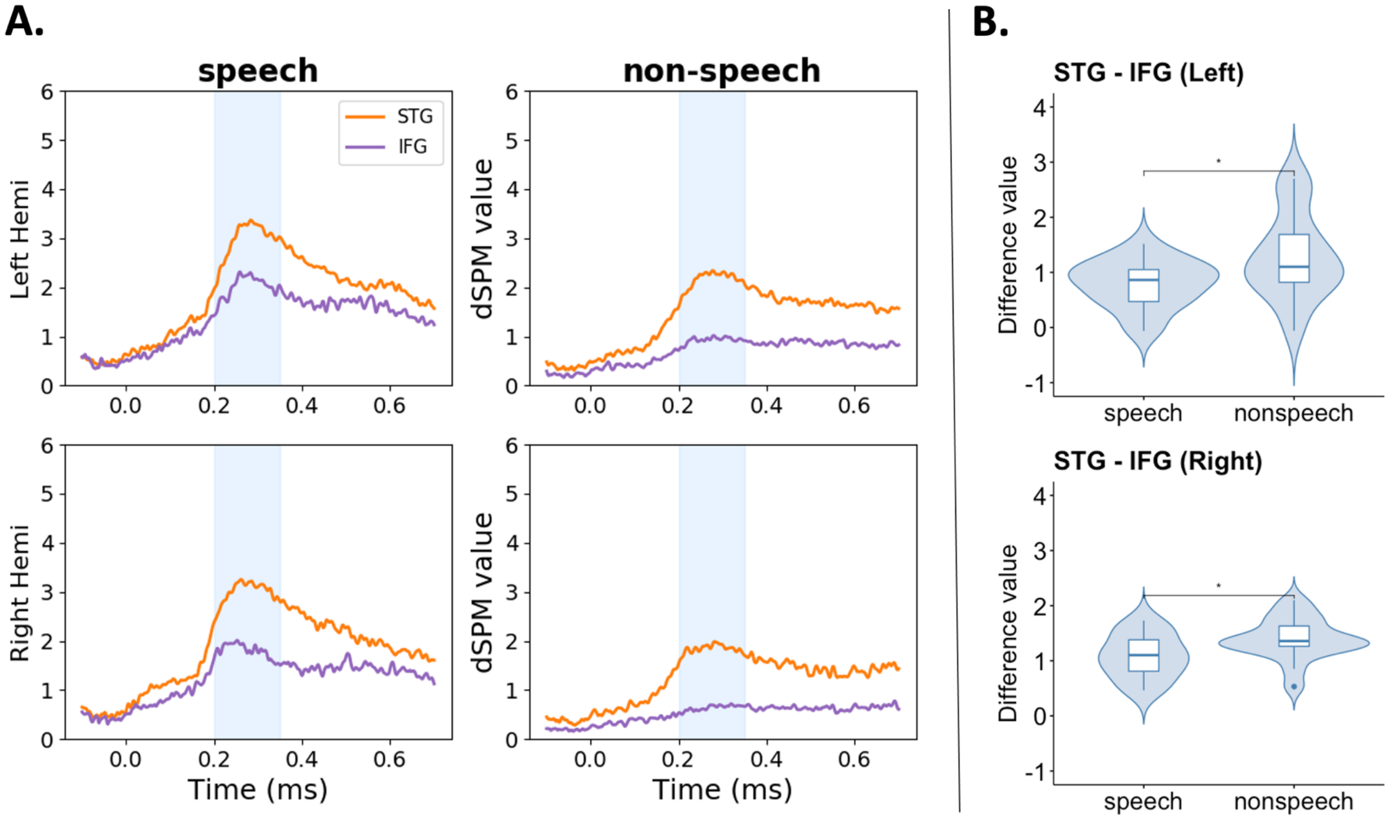

Figure 10. *2-month-old infants show differences between speech and nonspeech stimuli in the IFG***.** Grand averaged source waveforms of 2-month-old infants **(A)** for STG (orange traces) and IFG (purple traces) are shown for each stimulus type (column) and hemisphere (row). Time-windows of interest are shown as blue shaded regions. Violin plots in **(B)** show the distribution of IFG difference-values for each hemisphere and stimulus type. Asterisk represents significant differences between stimulus types.

6 mo - Ratio effect

As displayed in Figure 11 & 12., the ratio of the magnitude between auditory and motor brain activation changed across stimuli type, where the ratio was larger for speech stimuli across all brain regions.

There were no extreme outliers, and the data was normally distributed, as assessed by the Shapiro-Wilk's test of normality ($p > 0.05$). There was homogeneity of variances ($p > 0.05$) as assessed by Levene's test of homogeneity of variances. For the two-way interaction effect, Mauchly's test of sphericity indicated that the assumption of sphericity was met ($p > 0.05$).

There was a statistically significant two-way interaction between stimulus type and ROI: [$F(3, 30) = 6.217$, $p < 0.01$], as well as a statistically significant effect of stimulus type: [$F(1, 10) = 17.788$, $p < 0.01$], and ROI: [$F(1.75, 17.47) = 14.377$, $p < 0.001$].

Post hoc pairwise comparisons were run between stimulus types for each ROI. A Bonferroni adjustment was applied. The pairwise comparisons revealed a significant difference between speech and nonspeech for the right hemisphere ROIs; right IFG: $p < 0.001$ and Right M1: $p < 0.001$.

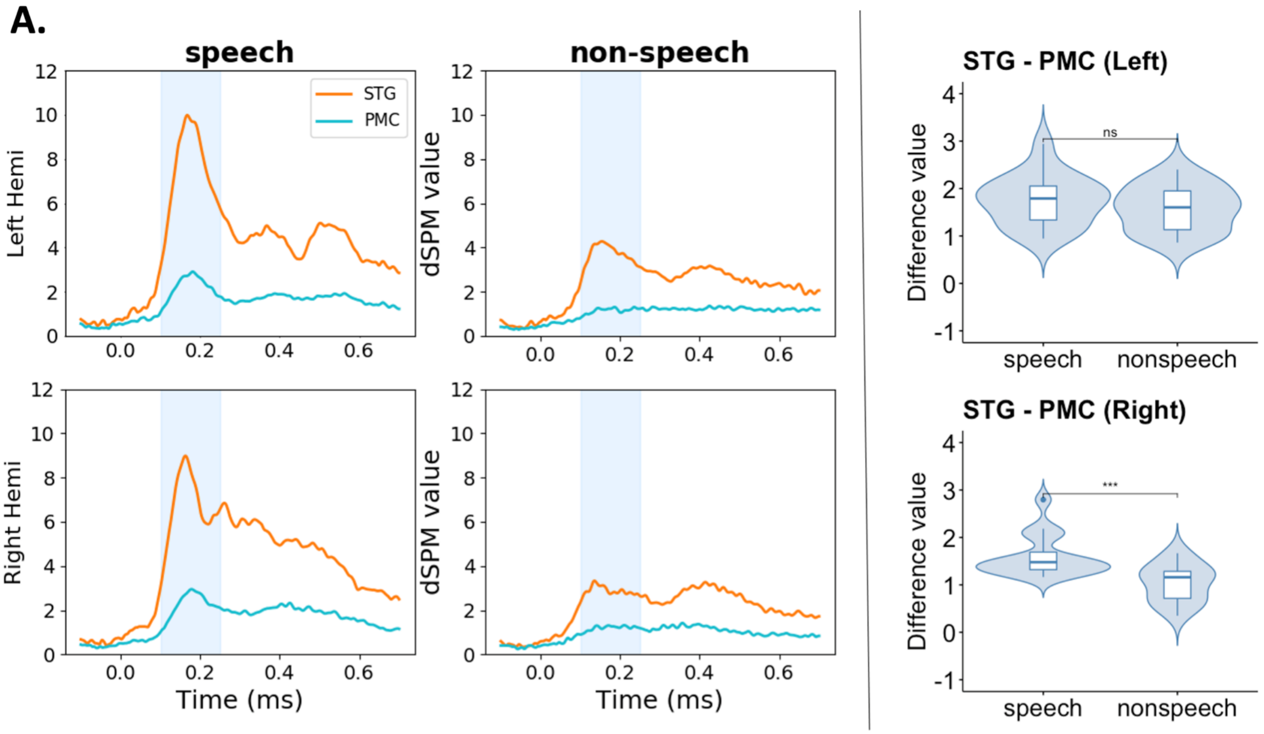

Figure 11. *6-month-old infants show differences between speech and nonspeech stimuli in the M1***.** Grand averaged source waveforms of 6-month-old infants **(A)** for STG (orange traces) and M1 (teal traces) are shown for each stimulus type (column) and hemisphere (row). Time-windows of interest are shown as blue shaded regions. Violin plots in **(B)** show the distribution of M1 difference-values for each hemisphere and stimulus type. Asterisk represents significant differences between stimulus types.

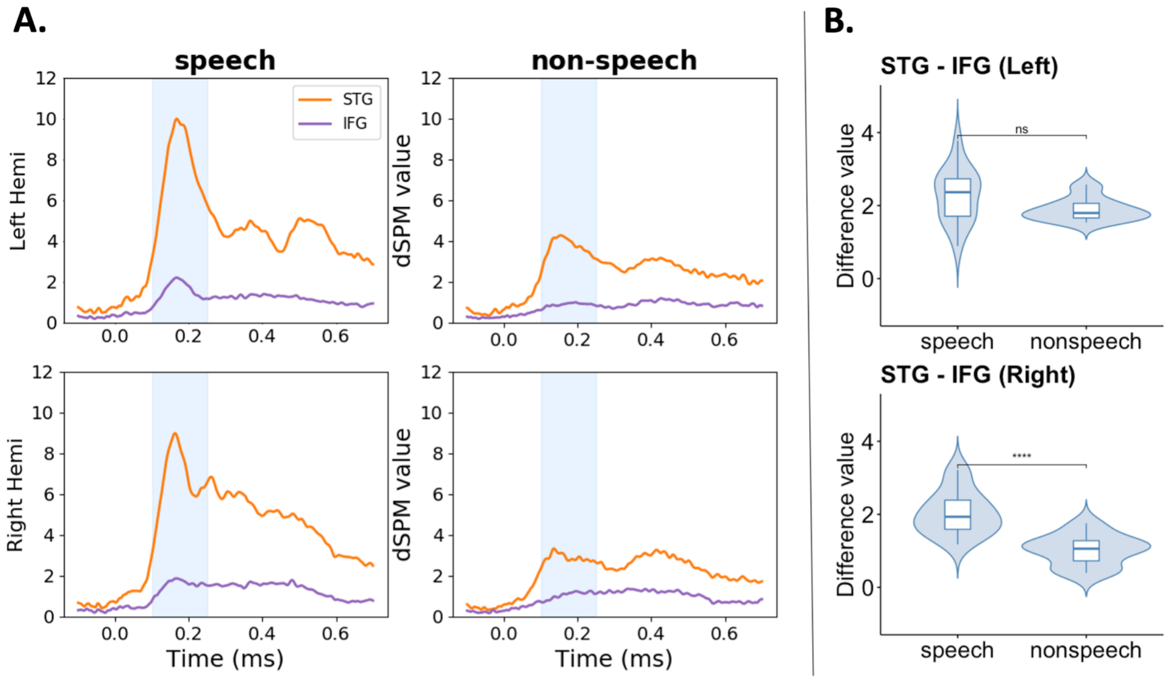

Figure 12. *6-month-old infants show differences between speech and nonspeech stimuli in the IFG***.** Grand averaged source waveforms of 6-month-old infants **(A)** for STG (orange traces) and IFG (purple traces) are shown for each stimulus type (column) and hemisphere (row). Time-windows of interest are shown as blue shaded regions. Violin plots in **(B)** show the distribution of IFG difference-values for each hemisphere and stimulus type. Asterisk represents significant differences between stimulus types.

III. Discussion

The primary goal of the present study was to investigate the role of motor and motor planning brain regions in speech perception throughout various stages of language development. Specifically, this thesis sought to determine 1) what is the *temporal structure* of activation in auditory sensory, motor, and motor planning brain regions?, 2) how do neural responses from *auditory, motor, and motor planning brain regions change in relation to another across age?*, 3) How does the activity in motor and motor planning brain regions change with *perceptual experience* in response to speech*?,* and

4) How does the activity in motor and motor planning brain regions differ between *speech and nonspeech* auditory stimuli?

1) Experiment 1: Temporal structure

Experiment 1 examined MEG-recorded cortical activation in response to the English syllable /ba/ in 2-, 7-, 11-month-old infants and adult subjects. The goal of the experiment was twofold: first, we tested the hypothesis that early in development speech activates not only cortical areas related to auditory perception, but also motor and motor planning areas. Second, we wanted to investigate any changes in the latency and magnitude of the neural response to speech in both auditory, motor and motor planning brain regions.

As expected from previous studies which show language-specific patterns emerge in speech perception prior to speech production (Dehaene-Lambertz, 2006 & Imada et al, 2006), activity was observed in both auditory and motor, and motor planning regions across all ages, and the time courses of activation in motor regions were temporally correlated with activation from auditory processing areas. In our current analysis, we observed speech-specific activity in the IFG in 2-month-old infants. These findings indicate speech specific IFG activation occurs younger than what has been reported in previous studies. For example, Imada et al (2006) reported speech-specific IFG activity in 6-month-old infants, but not newborns. Dehaene-Lamberz et al (2003) reported activity in frontal and right premotor regions at 3 months of age to forward, but not backward presented speech (Dehaene-Lambertz, 2003). These findings suggest that sensory-motor connections develop with increased motor experience. Our findings show activity in M1 in response to speech across all ages. Our results add to this literature and suggest that infants' sensory-motor connections develop as early as 2-months of age, well before the onset of speech production.

In addition, we observed systematic age-related changes in both the peak magnitude and peak latency of the neural response to speech. Consistent with the ERP literature (Wunderlitch et al., 2006), there was a significant decrease in peak latency in all ROIs that occurred as a function of age. We also observed an increase in peak magnitude from 2-months to all other age groups in both the STG and the IFG, but not the M1. Age-related changes in magnitude and latency are assumed to reflect maturation of the neural processes generating the response. For example, peak amplitude is thought to reflect changes in synaptic density (Eggermont, 1988), which doubles over the first 3 months of life in auditory processing regions (Huttenlocher & Dabholkar, 1997). The decrease in latency as a function of age has been assumed to indicate an increase in myelination and synaptic efficiency (Eggermont, 1992). Increasing neural synchrony may also result in the averaged response occurring earlier and with larger magnitude (Thomas et al., 1997). Changes in the magnitude and latency of the response over time may also be influenced by changes in the location and/or orientation of the neural substrates.

Considering the STG and IFG are involved in more complex speech processing, (i.e., STG is critical for extracting meaningful linguistic auditory features of speech, whereas the IFG is involved in more complex sequencing and motor planning**),** it is reasonable that we see an increase in magnitude in both regions at 7 months**.** By this age, infants have more experience hearing speech in their environments and are beginning to produce speech sounds. In other words, the increase in response magnitude that we observed across age in both the STG and IFG are consistent with the view that activity in these regions change as a function of both listening to and producing speech.

In contrast, the absence of age-related changes in magnitude in M1 suggests that the M1 may provide a different function in speech perception than the IFG. Theories surrounding the role of the M1 in speech perception largely come from two lines of research: The first suggests that the M1 reflects phonological processing and is consistent with the idea that experience producing speech enables the generation of internal models of motor commands required to produce speech (Kuhl et al., 2014; Halle & Stevens, 1959, 1962). The second line suggests that the M1 plays a compensatory role in speech perception (Du et al, 2014).

Experiment 2: Relationship between auditory and motor

Experiment 2 aimed to evaluate the relationship between activation in auditory and motor regions in response to native speech sounds in 2-, 7-, 11-month-old infants and adult subjects.

In the IFG, a significant reduction in the ratio between auditory and motor regions was found between 11 months and adults, reflecting an increase of activity in the IFG compared to the STG. This effect was seen in the left hemisphere only, which is in line with previous studies reporting higher left hemisphere dominant IFG activation in adults while listening to speech. Left inferior frontal regions have been argued to support motor speech planning and imitation, serving as the link between speech perception and production (Hickok & Poeppel, 2007).

In the M1, a significant increase in the ratio was found between 2 & 7 months, reflecting an increase in activity in the STG compared to the M1 at 7 months. As in Experiment 1, this difference in where the developmental change occurs could suggest that the M1 and IFG have different roles in speech perception. Although younger infants appear to recruit M1 more than older infants and adults for speech processing, adults appear to recruit IFG to a greater degree than infants.

2) *Experiment 3: Perceptual experience*

Experiment 3 aimed to examine brain activity to native and nonnative speech syllables in 7- and 11-month-old infants. We expected equivalent activation for native and nonnative contrasts at 7-months, but by 11 months we expected native and nonnative speech to activate auditory and motor brain areas differentially. Specifically, we hypothesized that activation to native speech would be greater than activation to nonnative speech in auditory brain areas. In contrast, we hypothesized that activation to nonnative speech would be greater than activation to native speech in motor brain areas, a pattern that would be consistent with findings from previous infant and adult studies (Kuhl et al., 2014). Kuhl et al (2014) argued that the developmental increase in left IFG for the nonnative speech contrasts reflects increased experience both hearing and producing native speech, which effectively reduces the ability to generate internal models for nonnative speech, thus, increasing left IFG activation.

Our results did not provide support for the hypothesis. At both ages, auditory and motor areas were activated in response to speech. However, we did not observe any differences between native and nonnative speech in any ROI, at either age. This inconsistency is likely due to some differences in our paradigm and stimuli from that of previous studies. In the current study, both native and nonnative speech sounds were acoustically different lengths (130 ms) when compared to the standard sound (100 ms). This difference in stimuli length likely resulted in discrimination of all speech contrasts, resulting in no difference in response magnitudes between the native and nonnative condition. Further research is needed to investigate whether the M1 follows the same developmental pattern reported for the IFG (more active in response to nonnative speech than native speech as a function of language experience), which would provide better insight into the role of the M1 in speech perception.

Experiment 4: Speech vs Nonspeech

Experiment 4 examined cortical activation in response to native speech and a nonspeech contrast in 2- and 6-month-old infants. It was expected that the magnitude of the activity in both auditory and motor ROIs would be reduced in the nonspeech condition (Wunderlitch et al., 2006). We found that the magnitude of activity in both auditory and motor regions was reduced to nonspeech in comparison to speech, but to different degrees between regions and across ages. At 2 months infants appear to use auditory and motor regions equally to process speech. We see a reduction in the magnitude of activation in both auditory and motor regions to nonspeech stimuli, but to a larger degree in motor regions (particularly the M1). By 6 months infants are using the STG to a greater degree than motor regions to process speech sounds, and less so for nonspeech sounds.

As previously mentioned, results from the current study strongly suggest that the M1 and the IFG provide different contributions to speech perception. Earlier studies theorize that motor regions are involved in phonological processing. This view is consistent with the idea that experience producing speech enables the generation of internal models of motor commands that are required to produce speech (Fadiga et al, 2009). If this is true, we might expect to see changes in M1 activation to occur as a function of experience with producing speech, for example from 11-months to adulthood. Our working hypothesis is that the M1 may play a compensatory role in speech perception at least during the early stages in language learning and is active when the brain is “working harder” to process speech stimuli. At 2 months, infants can recognize speech sounds as important stimuli from their environments (compared to nonspeech stimuli) and recruit M1 when trying to process these more “relevant” stimuli. Similar findings have been shown in adults as they learn a new language (Stein, Winkler, Kaiser, & Dierks, 2014) or when learning morse code (Kujala, Huotilainen, Uther & Shtyrovl 2003; Schlaffke, Leemans, Schweizer, Ocklenburg, & Schmidt-Wilcke, 2017). By 6 months of age, infants are more familiar with speech sounds and are thus more sophisticated at processing speech as compared to 2-month-old infants. We argue that the developmental increase in SGT activity across age reflects a stronger ability to extract meaningful features from speech, and more generally, better speech perception abilities.

Additionally, results from the nonspeech condition at 2 months provide evidence that our motor regions are distinct from auditory processing regions and our results cannot be attributed to point spread from auditory to motor regions.

Study Limitations

There are limitations of the present study that should be recognized and taken into consideration. First, since the current study compared previously collected data from different protocols, some parameters were not completely controlled for across datasets. For example, although the stimuli compared across age groups were acoustically similar, they were not identical. In addition, the number of trials collected varied across different age groups. Using previously collected data, however, allowed us to encompass several groups of subjects who are at different stages in their language development which is difficult to achieve in a single study. Another limitation to consider is that when doing MEG source-modelling in infants the cortical surface maps derived from structural MRIs to generate locations of potential sources from the infant brain can be difficult to characterize as infant brains have poor differentiation along the grey matter / white matter boundary. As a result, automated brain segmentation methods frequently yield poor results from individual infant MRIs. In the current study we employed templates from similarly aged infant MRIs and attempted to carefully inspect each stage of the source modelling process to reduce risk of error. We do acknowledge that, in general, source modelling is developed for adult subjects and advances in this area for infants would greatly increase reliability of source localization

V. Conclusion and future directions

Some key points related to these findings advance our understanding of infant speech perception and raise additional questions. First, both the M1 and IFG seem to be active during speech perception, even well before speech production begins. A novel finding was that both the M1 and IFG are active in infants as young as 2 months of age in response to speech, as previous research showed activity in the IFG as young as 6 months and frontal and right premotor regions active in 3-month-olds. Additionally, the current data show that both the IFG and M1 are not activated to the same extent when infants listen to nonspeech auditory stimuli. Moreover, the current data suggest that the roles of the M1 and IFG are different. Based on the results presented here, we hypothesize that the M1 plays a compensatory role in speech perception, while the IFG is involved in more sophisticated processing and motor planning, consistent with findings from previous research (e.g., Imada et al., 2006; Kuhl et al., 2014). Further research with this data investigating the patterns of association between brain data and outcome measures on these same children (which were obtained in the previous studies) could provide information on the importance of both the M1 and IFG in early development. In addition, studies involving musical stimuli and specific kinds of training could strengthen knowledge of sensorimotor integration. Music is a complex auditory stimulus, and playing music adds the motor component. Carefully designed, studies have the potential to reveal whether the sensorimotor activation observed in the current study for speech are also shown with increasing auditory and motor experience with music.

VII. References

Adank, P., Nuttall, H., & Kennedy-Higgins, D. (2017). Transcranial magnetic stimulation and motor evoked potentials in speech perception research. *Language, Cognition and Neuroscience*, 32(7), 900–909.

Bartoli, E., Maffongelli, L., Campus, C., & D’Ausilio, A. (2016). Beta rhythm modulation by speech sounds: Somatotopic mapping in somatosensory cortex. *Scientific Reports*, 6(1), 31182.

Benson, R., Whalen, D., Richardson, M., Swainson, B., Clark, V., Lai, S., & Liberman, A. (2001). Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain and Language*, 78(3), 364–396.

Behmer, L., & Fournier, L. (2014). Working memory modulates neural efficiency over motor components during a novel action planning task: An EEG study. Behavioural *Brain Research*, 260, 1.

Bernstein, L., & Liebenthal, E. (2014). Neural pathways for visual speech perception. *Frontiers in Neuroscience*, 8, 386.

Borrie, S., Lansford, K., & Barrett, T. (2018). Understanding dysrhythmic speech: When rhythm does not matter and learning does not happen. *Journal Of The Acoustical Society Of America*, 143(5), 379–385.

Bosseler, A., & Massaro, D. (2003). Development and evaluation of a computer-animated tutor for vocabulary and language learning in children with autism. *Journal of Autism and Developmental Disorders*, 33(6), 653–672.

Bowers, A., Saltuklaroglu, T., Harkrider, A., Cuellar, M., & Sinigaglia, C. (2013). Suppression of the μ rhythm during speech and nonspeech discrimination revealed by independent component analysis: Implications for sensorimotor integration in speech processing. *PLoS ONE*, 8(8), 72024.

Bowers, A., Saltuklaroglu, T., Harkrider, A., Wilson, M., & Toner, M. (2014). Dynamic modulation of shared sensory and motor cortical rhythms mediates speech and nonspeech discrimination performance. *Frontiers in Psychology*, 5, 366.

Bruderer, A., Danielson, D., Kandhadai, P., & Werker, J. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of*

Sciences of the United States of America, 112(44), 13531–13536.

Burton, M., Small, S., & Blumstein, S. (2000). The role of segmentation in phonological processing: An fMRI investigation. *Journal of Cognitive Neuroscience*, 12, 679–690.

Callan, D., Callan, A., & Jones, J. (2014). Speech motor brain regions are differentially recruited during perception of native and foreign-accented phonemes for first and second language listeners. *Frontiers in Neuroscience*, 8, 275.

Callan, A., Callan, D., Tajima, K., & Akahane-Yamada, R. (2006). Neural processes involved with perception of non-native durational contrasts. *Neuroreport*, 17, 1353–1357

Cuellar, M., Bowers, A., Harkrider, A., Wilson, M., & Saltuklaroglu, T. (2012). Mu suppression as an index of sensorimotor contributions to speech processing: Evidence from continuous EEG signals. *International Journal of Psychophysiology*, 85(2), 242–248.36

Chen, X., Striano, T., & Rakoczy, H. (2004). Auditory-oral matching behavior in newborns. *Developmental Science*, 7(1), 42–47.

Cogan, G., Thesen, T., Carlson, C., Doyle, W., Devinsky, O., & Pesaran, B. (2014). Sensory-motor transformations for speech occur bilaterally. *Nature*, 507(7490), 94–98.

Correia, J., Jansma, B., & Bonte, M. (2015). Decoding articulatory features from fMRI responses in dorsal speech regions. *The Journal of Neuroscience*: The Official Journal of the Society for Neuroscience, 35(45), 15015–15025.

Crawcour, S., Bowers, A., Harkrider, A., & Saltuklaroglu, T. (2009). Mu wave suppression during the perception of meaningless syllables: EEG evidence of motor recruitment. *Neuropsychologia*, 47(12), 2558–2563.

D'Ausilio, A., Bufalari, I., Salmas, P., Busan, P., & Fadiga, L. (2011). Vocal pitch discrimination in the motor system. *Brain and Language*, 118(1–2), 9–14.

D'Ausilio, A., Bufalari, I., Salmas, P., & Fadiga, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex*, 48, 882–887.

Dale, A., Liu, A., Fischl, B., Buckner, R., Belliveau, J., Lewine, J., & Halgren, E. (2000). Dynamic Statistical Parametric Mapping: Combining fMRI and MEG for High-Resolution Imaging of Cortical Activity. *Neuron,* 26(1), 55-67.

Dehaene-Lambertz, G., Dehaene, S & Hertz-Pannier, G. (2003). Functional neuroimaging of speech perception in infants. *Science*, 298(5600), 2013-2015.

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Mériaux, S., Roche, A., Sigman, M., & Dehaene, S. (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proceedings of the National Academy of Sciences*, 103(38), 14240-14245

Démonet, J., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J., Wise, R., . . . Frackowiak, R. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain: A Journal of Neurology*, 115, 1753–1768.

Dewitt, I., & Rauschecker, J. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences of the United States of America*, 109(8), E505–E514.

Diehl, R. (1989). Remarks on Stevens' quantal theory of speech. *Journal of Phonetics*, 17(1–2), 71–78.

Eggermont J. J. (1988). On the rate of maturation of sensory evoked potentials. *Electroencephalography and clinical neurophysiology*, *70*(4), 293–305.

Eggermont J. J. (1992). Development of auditory evoked potentials. *Acta otolaryngologica*, *112*(2), 197–200.

Eimas, P., Siqueland, E., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968), 303–306.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, 15(2), 399–402.

Fadiga, L., Craighero, L., & D'Ausilio, A. (2009). Broca's area in language, action, and music. *Annals of the New York Academy of Sciences*, 1169(1), 448–458.

Fischl, B. (2012). FreeSurfer. *Neuroimage*, 62(2), 774–781.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., & Van Essen,

D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178.

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology-General*, 123(2), 178–200.

Golestani, N., & Zatorre, R. (2004). Learning new sounds of speech: Reallocation of neural substrates. *Neuroimage*, 21(2), 494–506.

Goodsitt, J., Morgan, J., & Kuhl, P. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language*, 20(2), 229–252.

Guenther F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, 25, 408–422.

Guellai, B., Mersad, K., & Streri, A. (2015). Suprasegmental information affects processing of talking faces at birth. *Infant Behavior and Development*, 38, 11–19.

Gunji, A., Ishii, R., Chau, W., Kakigi, R., & Pantev, C. (2007). Rhythmic brain activities related to singing in humans. *Neuroimage*, 34, 426–434.

Halle, M., & Stevens, K. M. (1959). Analysis by synthesis. In W. Wathen-Dunn & L. E. Woods (Eds.), *Proceedings of the Seminar of Speech Compression and Processing* (Vol. 2).

Halle, M., & Stevens, K. (1962). Speech recognition: A model and a program for research. *IRE Transactions of the PGIT*, IT-8, 155-159.

Hazan, V., Sennema, A.,Faulkner, A., & Ortega-Llebaria, M.(2006). The use of visual cues in the perception of non-native consonant contrasts. *Journal of the Acoustical Society of America*, 119(3), 1740–1751.

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience*, 21(7), 1229–1243.

Hickok, G. (2012). The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. *Journal of Communication Disorders*, 45(6), 393–402.

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*, 15(5), 673–682.

Hickok, G., Costanzo, M., Capasso, R., & Miceli, G. (2011). The role of Broca’s area in speech perception: Evidence from aphasia revisited. *Brain and Language*, 119(3), 214–220.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, 69(3), 407–422.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4(4), 131–138.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1–2), 67–99.

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

Holt, L., & Lotto, L. (2010). Speech perception as categorization. *Attention, Perception, & Psychophysics*, 72(5), 1218–1227.

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P. (2006). Infant speech perception activates Broca’s area: A developmental magnetoencephalography study. *Neuroreport*, 17(10), 957–962.

Irwin, J., Avery, T., Brancazio, L., Turcios, J., Ryherd, K., & Landi, N. (2018). Electrophysiological indices of audiovisual speech perception: Beyond the McGurk Effect and Speech in Noise. *Multisensory Research*, 31(1–2), 39–56.

Irwin, J., Tornatore, L., Brancazio, L., & Whalen, D. (2011). Can children with autism spectrum disorders “hear” a speaking face? *Child Development*, 82(5), 1397–1403.

Ito, T., Tiede, M., & Ostry, D. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences of the United States of America*, 106(4), 1245–1248.

Jenson, D., Bowers, A., Harkrider, A., Thornton, D., Cuellar, M., & Saltuklaroglu, T. (2014). Temporal dynamics of sensorimotor integration in speech perception and production: Independent component analysis of EEG data. *Frontiers in Psychology,* 5, 656.

Jusczyk, P., Hirsh-Pasek, K., Kemler Nelson, D., Kennedy, L., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants *Cognitive Psychology*, 24(5), 252-293. https://doi.org/10.1016/0010-0285(92)90009-Q.

Jusczyk, P., Friederici, A., Wessels, J., Svenkerud, V., & Jusczyk, A. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32*(3), 402–420. https://doi.org/10.1006/jmla.1993.1022

Kuhl, P. (2000). Theoretical approaches to speech perception. *The Journal of the Acoustical Society of America*, 108(5), 2627.

Kuhl, P. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5(11), 831–843.

Kuhl, P. (2010). Brain mechanisms in early language acquisition. *Neuron*, 67(5), 713–727.

Kuhl, P., & Miller, J. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, 190(4209), 69–72.

Kuhl, P., & Meltzoff, A. (1982). The bimodal perception of speech in infancy. *Science*, 218, 1138.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255(5044), 606–608. https://doi.org/10.1126/science.1736364

Kuhl, P., & Meltzoff, A. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, 100(4), 2425–2438.

Kuhl, P., Tsao, F., & Liu, H. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America*, 100(15), 9096–9101.

Kuhl, P., Conboy, B., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the “critical period”. *Language*

Learning and Development, 1(3–4), 237–264.

Kuhl, P., & Rivera-Gaxiola, M. (2008). Neural substrates of language acquisition. *Annual Review of Neuroscience*, 31, 511–534.

Kuhl, P., Ramírez, R., Bosseler, A., Lin, J., & Imada, T. (2014). Infants' brain responses to speech suggest analysis by synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 111(31), 11238–11245.

Kujala, A., Huotilainen, M., Uther, M., & Shtyrov, Y. (2003). Plastic cortical changes induced by learning to communicate with non-speech sounds. *Neuroreport*, 14(13), 1683-1687.

Lane, H. (1965). The motor theory of speech perception: A critical review. *Psychological Review,* 72(4), 275–309.

Lewkowicz, D., & Ghazanfar, A. (2006). The decline of cross-species intersensory perception in human infants. *Proceedings of the National Academy of Sciences of the United States of America*, 103(17), 6771–6774.

Liberman, A., Harris, K., Hoffman, H., & Griffith, B. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54(5), 358–368.

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

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

Lively, S., Logan, J., & Pisoni, D. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, 94(3), 1242–1255.

Lively, S., Pisoni, D., Yamada, R., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America*, 96(4), 2076–2087.

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

Massaro, D. (1998). Integrating cues in speech perception. *Behavioral and Brain Sciences*, 21(2), 275.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111. https://doi.org/10.1016/s0010-0277(01)00157-3

Mcgurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746–748.

Mcgurk, H., & Macdonald, J. (1978). Auditory-visual coordination in the first year of life. *International Journal of Behavioral Development*, 1(3), 229–239.

Meister, I., Wilson, S., Deblieck, C., Wu, A., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, 17(19), 1692–1696.

Meltzoff, A., & Moore, M. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198(4312), 75–78.

Meltzoff, A., & Moore, M. (1989). Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology*, 25(6), 954–962.

Meltzoff, A., & Moore, M. (1994). Imitation, memory, and the representation of persons. *Infant Behavior and Development*, 17(1), 83–99.

Mehler, J. (1999). Cortical representation of language in monolinguals and bilinguals. *The Journal of the Acoustical Society of America*, 105(2), 1034.

Moon, C., & Fifer, W. (2000). Evidence of transnatal auditory learning. *Journal of Perinatology*, 20(S8), S37–S44.

Moon, C., Lagercrantz, H., & Kuhl, P. (2013). Language experienced in utero affects vowel perception after birth: A two-country study. *Acta Paediatrica*, 102(2), 156–160.

Möttönen, R., J.rvel.inen, J., Sams, M., & Hari, R. (2005). Viewing speech modulates activity in the left SI mouth cortex. *Neuroimage*, 24(3), 731–737.

Möttönen, R., & Watkins, K. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 29(31), 9819–9825.

Möttönen, R., & Watkins K. (2012). Using TMS to study the role of the articulatory motor system in speech perception. *Aphasiology*, (26), 1103–1118.

Möttönen, R., van de Ven, G. M., & Watkins, K. (2014). Attention fine-tunes auditory-motor processing of speech sounds. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *34*(11), 4064–4069. https://doi.org/10.1523/JNEUROSCI.2214-13.2014

Murakami, T., Ugawa, Y., & Ziemann, U. (2013). Utility of TMS to understand the neurobiology of speech. *Frontiers in Psychology*, 4, 446.

Muthukumaraswamy, S., & Johnson, B. (2004). Primary motor cortex activation during action observation revealed by wavelet analysis of the EEG. *Clinical Neurophysiology*, 115(8), 1760–1766.

Nasir, S., & Ostry, D. (2009). Auditory plasticity and speech motor learning. *Proceedings of the National Academy of Sciences of the United States of America*, 106(48), 20470–20475.

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

Nuttall, H., Kennedy-Higgins, D., Hogan, J., Devlin, J., & Adank, P. (2016). The effect of speech distortion on the excitability of articulatory motor cortex. *NeuroImage*, 128, 218–226.

Nuttall, H., Kennedy-Higgins, D., Devlin, J., & Adank, P. (2017). The role of hearing ability and speech distortion in the facilitation of articulatory motor cortex. *Neuropsychologia*, 94, 13–22.

Patterson, M., & Werker, J. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, 22(2), 237–247.

Perry, A., & Bentin, S. (2010). Does focusing on hand-grasping intentions modulate electroencephalogram μ and α suppressions? *Neuroreport*, 21(16), 1050–1054.

Pineda, J., Grichanik, M., Williams, V., Trieu, M., Chang, H., & Keysers, C. (2013). EEG sensorimotor correlates of translating sounds into actions. *Frontiers in Neuroscience*, 7(203), 203.

Pisoni, D. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *The Journal of the Acoustical Society of America*, 61(5), 1352–1361. 121–166). Cambridge, MA: MIT Press.

Pons, F., Lewkowicz, D., Soto-Faraco, S., & Sebasti.n-Gall.s, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, 106(26), 10598–10602.

Poeppel, D. (2014). The neuroanatomic and neurophysiological infrastructure for speech and language. Current Opinion in *Neurobiology*, 28, 142.

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

Purves, D., Augustine, G. J., Fitzpatrick, D., Lawrence, C. K., LaMantia, A., O McNamara, J., & Williams, M. S. (Eds.). (2001). *Neuroscience* (2nd ed.). Sunderland, MA: Sinauer Associates.

Rauschecker, J. (2012). Ventral and dorsal streams in the evolution of speech and language. *Frontiers in Evolutionary Neuroscience*, 4, 7.

Raizada, R., & Poldrack, R. (2007). Selective amplification of stimulus differences during categorical processing of speech. *Neuron*, 56(4), 726–740.

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A speech reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology lip-reading* (pp. 97–113). London: Erlbaum.

Romero, Y. R., 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.

Rogalsky C., Love T., Driscoll D., Anderson S. W., & Hickok G. (2011). Are mirror neurons the basis of speech perception? Evidence from five cases with damage to the purported human mirror system. *Neurocase*, 17, 178–187.

Rosenblum, L., Schmuckler, D., & Johnson, M. (1997). The McGurk effect in infants. *Perception & Psychophysics*, 59(3), 347–357.

Ross, L., Saint-Amour, D., Leavitt, V., Javitt, D., & Foxe, J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, 17(5), 1147–1153.

Saltuklaroglu, T., Harkrider, A. W., Thornton, D., Jenson, D., & Kittilstved, T. (2017). EEG Mu (µ) rhythm spectra and oscillatory activity differentiate stuttering from non-stuttering adults. *NeuroImage*, 153, 232–245.

Sams, M. (1991). Cortical responses to changes in auditory stimuli: Magnetoencephalographic studies. Acta Oto-Laryngologica, 111(Sup 491), 124–131.

Schlaffke, L., Leemans, A., Schweizer, L. M., Ocklenburg, S., & Schmidt-Wilcke, T. (2017). Learning Morse Code Alters Microstructural Properties in the Inferior Longitudinal Fasciculus: A DTI Study. *Frontiers in human neuroscience*, *11*, 383.

Schnitzler, J., & Vary, P. (2000). Trends and perspectives in wideband speech coding. *Signal Processing*, 80(11), 2267–2281.

Schomers, M., & Pulvermüller, F. (2016). Is the sensorimotor cortex relevant for speech perception and understanding? An integrative review. *Frontiers in Human Neuroscience*, 10, 435.

Scott, S., & Johnsrude, I. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences,* 26(2), 100–107.

Simmonds, A., Leech, R., Collins, C., Redjep, O., & Wise, R. (2014). Sensory-motor integration during speech production localizes to both left and right plana temporale. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 34(39), 12963–12972.

Skipper, J., Van Wassenhove, V., Nusbaum, H., & Small, S. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, 17(10), 2387–2399.

Specht, K. (2013). Mapping a lateralization gradient within the ventral stream for auditory speech perception. *Frontiers in Human Neuroscience*, 7, 629.

Stein, M., Winkler, C., Kaiser, A., & Dierks, T. (2014). Structural brain changes related to bilingualism: does immersion make a difference?. *Frontiers in psychology*, *5*, 1116.

Sumby, W., & Pollack, I. (1954). Erratum: Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26(4), 583.

Tamura, T., Gunji, A., Takeichi, H., Shigemasu, H., Inagaki, M., Kaga, M., & Kitazaki., M. (2012). Audiovocal monitoring system revealed by mu-rhythm activity. *Frontiers in Psychology*, 3, 225.

Taulu, S., & Simola, J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Physics in medicine and biology*. 51, 1759-68. 10.1088/0031-9155/51/7/008.

Tourville J. A., & Guenther F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26, 952–981.

Trudeau-Fisette, P., Ito, T., & Menard, L. (2019). Auditory and somatosensory interaction in speech perception in children and adults. *Frontiers in Human Neuroscience*, 13, 344.

Uusitalo, M., & Ilmoniemi, R. (1997). Signal-space projection method for separating MEG or EEG into components. *Medical & biological engineering & computing*. 35, 135-40.

Watkins, K., Strafella, A., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41(8), 989–994.

West, M., & King, A. (1988). Female visual displays affect the development of male song in the cowbird. *Nature,* 334(6179), 244–246.

Werker, J., & Hensch, T. (2015). Critical periods in speech perception: New directions. *Annual Review of Psychology*, 66(1), 173–196.

Werker, J., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. I*nfant Behavior and Development*, 7(1), 49–63.

Wilson, S., Saygin, A., Sereno, M., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7(7), 701–702.

Woods, D. L., Herron, T. J., Cate, A. D., Kang, X., & Yund, E. W. (2011). Phonological processing in human auditory cortical fields. *Frontiers in Human Neuroscience*, 5, 42.

Wunderlich, J., Cone, B., & Shepherd, R. (2006). Maturation of the cortical auditory evoked potential in infants and young children. *Hearing Research*, 212, 185-202.

Yeung, H., & Werker, J. (2013). Lip movements affect infants' audiovisual speech perception. *Psychological Science*, 24(5), 603–612.

Zion Golumbic, E., Poeppel, D., & Schroeder, C. (2012). Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective. *Brain and Language*, 122(3), 151–161.