Activation of word-level speech production regions during suprasegmental speech perception differs by modality and task

Benjamin Elgie, Integrated Program in Neuroscience
McGill University, Montreal
Submitted: 09/2011

A thesis submitted to McGill University in partial submission of the requirements of the degree of Master of Science

Table of Contents

Acknowledgments

I would firstly like to thank my joint supervisors, Dr. Vince Gracco and Dr. Shari Baum. They took me into their labs knowing I had very little knowledge of the field, but willing to give me the chance to learn. My experience here has been everything I wanted from graduate studies in neuroscience. I would particularly like to thank Dr. Baum for her extraordinarily quick turnaround time when it came to reviewing my work, and for pushing me to be clearer in my writing. I would also like to particularly thank Dr. Gracco for his discussions of theory, which have helped me look at this research from new perspectives, and to think harder about what goes on at both the perceptual and neural levels. Thanks also to Drs. Robert Zatorre and Denise Klein, for serving on my thesis committee and providing helpful feedback on my thesis proposal and my thesis itself (respectively).

Several fellow students have also contributed to this thesis, directly and indirectly. Pascale Tremblay, a former lab member, originally conducted Experiment #1, and provided valuable expertise in fMRI analysis whenever I needed advice. She was crucial in helping to form some of my earliest ideas, and caused me to look at the literature with more critical eyes. Thanks as well to Laura Copeland, another former lab member, for originally conceiving and conducting Experiment #2. Isabelle Deschamps, a graduate student in Dr. Baum's lab, and Melanie Segado, a summer student in Dr. Gracco's lab, both assisted me in reconstructing the protocols for both experiments, and provided useful discussion regarding fMRI analysis. Krystyna Grabski, a visiting graduate student, provided interesting theoretical discussions as well as several key papers that contributed to my work. Other lab members, including Francois-Xavier Bradjot and Thomas Gisiger, provided support, a good working environment, and helpful advice whenever I needed to think out loud.

I would like to add a special note of thanks to Dr. Josephine Nalbantoglu, the director of the IPN, for assisting me in changing labs and providing essential help in securing Drs. Baum and Gracco as supervisors.

Abstract

This study addresses recent ideas regarding the contribution of motor and frontal brain regions, traditionally engaged during speech production, to speech perception. Using an fMRI experiment concerned with word-level speech production and perception, the overlap between perception and production was investigated using a functional mask derived from a conjunction analysis of that experiment's word production tasks. This same mask was used to analyse activity during multi-modal sentence-level speech perception in another experiment. Common activity was found between word production and word perception, but not between word production and the more complex sentence-level speech perception tasks. Contrary to certain claims, visual speech perception did not lead to increased activation of speech production regions. Whole-brain analyses of the sentence-level experiment revealed complex differences between modality- and task-specific regions in frontal, temporal, and occipital regions. Activation in this experiment was clearly influenced by inherent demands of the speech level, task, and modality. Results are discussed in light of the task demands of both experiments, as well as their implication for current understanding of motor/frontal contributions to speech perception.

Résumé

Cette étude s'appuie sur de récentes hypothèses concernant la contribution en perception de la parole des aires cérébrales motrices et frontales, traditionnellement recrutées lors de la production de la parole. La création d'un masque fonctionnel calculé à partir des données d'une étude en imagerie par résonance magnétique fonctionelle (IRMf) portant sur la perception et la production de mots nous a permis de rechercher une éventuelle superposition entre la perception et la production de la parole. Ce masque a été à nouveau utilisé pour analyser d'éventuelles activations pendant une tâche de perception de phrases multi-modales issue d'une autre expérience d'IRMf. Des activités communes à la production et à la perception de mots, mais pas entre la production de mots et la production plus complexe de phrases, ont été mises en évidence. Contrairement à certaines affirmations, la perception visuelle de la parole n'a pas entraîné d'augmentation des activations dans les régions dédiées à la production de la parole. Des analyses de l'ensemble du cerveau lors de la perception et de la production des phrases ont révélé des différences complexes entre les régions spécifiques de la tâche ou de la modalité dans des aires frontales, temporales et occipitales. La modalité, la tâche et le niveau de complexité de la parole ont clairement influencé les activations observées lors de cette expérience. Les résultats obtenus sont discutés en regard des demandes spécifiques dues aux tâches et aux expériences menées ainsi que de la compréhension actuelle des contributions motrices/frontales lors de la perception de la parole.

Activation of word-level speech production regions during suprasegmental speech perception differs by modality and task

Overview

Our understanding of speech perception is incomplete. Firstly, the speech environment renders accurate perception difficult. Artificial speech recognition devices experience difficulty with speaker variability in age, gender, dialect, and idiosyncrasies, emotion, fatigue, or stress, in recognizing words across varying speech rates, and in separating speech from background noise (Pallett 1985). Humans are faced with the same problems, but our ability to accurately perceive and follow speech is still superior to that of automated speech recognition programs for spontaneous sentence-level speech (Lippmann 1997). Given the acoustic variability of speech, our speech perception abilities are impressive in themselves.

The speeds at which humans are able to perceive speech is again impressive. Our ability to accurately perceive speech at increasing word rates declines slowly until around 300 words per minute (wpm), and then declines rapidly thereafter; but even at 400 wpm, we are still capable of perceiving individual words (Foulke 1968). Humans easily order sequences of the 70-80ms long speech sounds used in normal conversation, and only need 50ms long notes to distinguish the order of a sequence of musical notes (Warren *et al* 1969). But the order of a repeated sequence of arbitrary sounds (ex. hisses, buzzes, and tones) cannot be reliably reported if the sounds in the sequence are each 200ms long; reliable reports are possible only when the sounds are at least 700ms long (Warren *et al* 1969). At 700ms per phoneme, and 4-5 phonemes per word, we would only

be able to perceive around 170 words per minute, if we perceived speech as made up of simple, discrete auditory events (Liberman *et al* 1967). Thus, we must perceive speech in some other way, or using a different mechanism than for auditory perception of non-speech sounds.

The fundamental segment of speech is variously identified as the phoneme (ex. /d/ or /i/) or the syllable (/di/); the argument is complicated by forward and backward *coarticulation* of phonemes, whereby the abstract idea of a speech sound (the phoneme) is contextually altered by the preceding and succeeding phonemes (McNeil & Lindig 1973, Oden & Massaro 1978, Fowler 1984). Thus, "definable segments of sound do not correspond to segments at the phoneme level" (Liberman *et al* 1967). Consonant-vowel (CV) syllables such as /di/ cannot be divided acoustically into /d/ or /i/, although we can segment them mentally. The role of coarticulation is itself under debate. Speech is popularly conceived of as a series of discrete phonological segments which are blurred together to produce a continuous speech signal. Some classical theories of coarticulation proceed largely on this basis, assuming speech is a series of idealized segments which are blurred together by the biomechanical interactions of articulators; these generally assume that coarticulation degrades the speech signal (Tatham & Morton 2006, pp.24,41). Others suggest that coarticulation gives contextual information about segments in the series, and thus enhances the speech signal (Cooper *et al* 1952). Still others suggest that coarticulation is involuntary but can be influenced cognitively, and that different sorts of coarticulation can be intentionally manipulated for communicative purposes (Tatham & Morton 2006, pp.18, Ch. 6).

Because of the acoustic variability and coarticulation of speech, listeners cannot rely directly on the acoustic signal to provide an invariant and segmented speech signal. A number of theories of speech perception have therefore suggested that a strong coupling exists between speech perception and speech production, and that this is what allows for competent, real-time communication between speakers and listeners. These theories include gestural theories, such as the motor theory of speech perception (MTSP), which claim that speech

production regions in the brain facilitate perception of the complex auditory signals which make up normal speech (Galantucci *et al* 2006). By contrast, purely auditory theories do not posit a speech motor contribution to perception; they suggest speech perception relies on translation of featural aspects of acoustic signals (such as voice-onset time and spectral analysis of speech frequencies) into phonetic codes (Diehl *et al* 2004).

The research described herein examines the potential involvement of speech production regions in both low- and high-level speech perception. For the purposes of this work, low-level speech refers both to sub-lexical speech, such as phonemes or syllables, which are recognizable as speech but which have no semantic content, as well as to single words. High-level speech refers to phrases, sentences, and discourse, and must account for everyday knowledge (pragmatics), prosody, and the intended meaning of a perceived utterance. Gestural theories have largely been studied using low-level units of speech, and their findings extrapolated to higher speech levels. Several recent studies have looked at motor activation during the perception of action verbs and sentences related to motor gestures, as a part of the theory that action execution regions of the brain are also activated by semantic representations of those actions (Hauk *et al* 2004, Tremblay & Small 2010). Otherwise, though, there have been few attempts to directly study whether the motor system is active during higher-level speech perception tasks. Perception of speech at different levels could involve different mental processes and mechanisms; discerning the initial phoneme of a syllable is quite different from determining if the speaker is angry or happy, and whether they are conveying this information semantically or prosodically. Given the different requirements of different speech levels and tasks, it is not clear if one can generalize theories of perception across levels and tasks. If the motor system *is* active during higher-level speech perception, its involvement can then be examined under various tasks and conditions to reveal what specific roles it might play at higher levels.

For this purpose, I am using data drawn from two functional magnetic resonance imaging (fMRI) experiments to investigate networks common to both

speech production and perception. The first experiment addresses the brain regions involved in speech production and perception of single words; the second focuses on perception of suprasegmental speech, presented via different modalities. The overall findings are interpreted in the context of the specific perceptual demands of each task and experimental paradigm, and consider the continuity between word-level speech and higher-order, sentence-level speech, as well as the explanatory power of gestural theories in accounting for brain activation patterns in higher-order speech perception.

Background

Motor theory of speech perception

In the 1950s, researchers at Haskins Laboratories were studying synthetic speech using a spectrograph, which converts sound into a graphical display of sound frequencies over time, and a pattern playback machine, which converts spectrographs (natural or hand-drawn) back into sound (Cooper *et al* 1952). They noted that their participants had difficulty perceiving words when phonemes were presented simply in sequence, the way they appear symbolically in writing. In investigating this issue, Liberman and colleagues (1952, 1954) found that vocal gestures for consonants and vowels overlap in time, a phenomenon known as *coarticulation*; this came out of their investigation of spectrographic formant patterns as a possible basis for discovering acoustic invariants in the speech signal (Cooper *et al* 1952). Formants are concentrations of acoustic energy within a restricted frequency region caused by resonant properties of the vocal tract, and are commonly represented as dark focal bands on a spectrogram, showing frequency variation over time during speech (Liberman *et al* 1967). Typical speech shows three to four such formants, labelled F1-F4. The work undertaken by the Haskins group focused on the use of idealized, artificial formant shapes, created by drawing them by hand onto acetate, and playing them using a machine which converted the visible spectrograms into sound, known as a 'pattern playback' machine (Cooper *et al* 1951).

The Haskins researchers followed up on their findings and investigated the energy burst (stop burst) during the release of unvoiced stop consonants (/p/ /t/ /k/), the contextual effect of succeeding vowels, and the formant transitions from voiced (/d/ /b/ /g/) and unvoiced stop consonants to the following vowel (Liberman *et al* 1952, 1954, 1967). Prior investigations had shown that the primary energy concentration for /p/ was centred at a low or intermediate frequency, that for /t/ at a high frequency, and that for /k/ depended on the succeeding vowel, starting high for front vowels like /i/ and lowering as the succeeding vowel was articulated further back, like /u/ (Potter *et al* 1947, Cooper *et al* 1952). When testing their idealized formant shapes, the Haskins group found that when the stop burst frequency was centred at 1440 Hz, the sound that played back tended to be perceived as /k/ when followed by /a/, but tended to be perceived as /p/ when followed by /i/ (Cooper *et al* 1952, Liberman *et al* 1952). Because perception changed while the stop burst was held at a fixed frequency, but followed by different vowels (necessarily the product of different articulations), they rejected the stop burst as an invariant acoustic signal for consonants, and suggested that perception tracked articulation rather than the acoustic signal, based on the listener's implicit knowledge of the production of /k/ and /p/ (Liberman *et al* 1952).

In CV syllables, like the ones used in the study above, the formants of the vowel show variations based on the preceding consonant. These variations, which occur at the start of the formant band before the steady-state shape of the vowel proper, are known as formant transitions. The Haskins group drew on earlier research showing a falling F2 transition for /k/ and /g/, a level transition for /t/ and /d/, and a rising transition for /p/ and /b/ (Potter *et al* 1947, Joos 1948). To investigate the role of formant transitions in consonant perception, they presented voiced and unvoiced stop consonants across a range of rising -> falling F2 transitions, with each of a representative range of succeeding vowels approximating the same front -> back vowels used previously (Liberman *et al* 1954). They found /b/ and /p/, the labial consonants, were strictly associated with rising transitions, largely independent of the place of vowel articulation. /d/ and

/t/, the alveolar consonants, were mostly identified in association with flat transitions and front vowels (/e/ and /e/), and falling transitions with back vowels (/a/, /ɔ/, /o/, /u/). /g/ and /k/, the velar consonants, were associated with flat-falling transitions with back vowels (/o/, /u/), or rising transitions with front vowels (/i/, /e/, /e/). See Figure 1 to view their results. The study showed that the connection between acoustic cues and consonant perception is fairly loose, so that widely varying transitions and following vowels can still result in perception of a similar phoneme (Liberman *et al* 1967). Again, the Haskins group concluded that phoneme perception tracked the implied articulatory gestures of the phonemes instead of the acoustic signals generated (Liberman *et al* 1954).

The full MTSP was put forth in 1967, and updated in 1985 (Liberman & Mattingly 1985) and 2000 (Liberman & Whalen 2000), restating the three principal claims of the theory as described below. Firstly, the original MTSP claims that speech is a "special and especially efficient code" which requires a specific decoder in the human brain, separate from the mechanisms used for normal auditory perception (Liberman *et al* 1967, pp.431). Among the evidence given for this is the difficulty of reading visible speech in a spectrogram, where it is quite difficult to segment or even identify individual phonemes; too, they note that while it had been relatively simple to build machines which recognize print, the same was not true for ones which recognize speech, whereas humans recognize speech quite easily but often experience difficulty in reading. If speech is a special code, and humans possess a dedicated decoder for speech, this would explain why it is difficult for non-humans to recognize speech, even though machines, for example, can easily recognize visual symbols. This 'speech is special' claim breaks down into smaller ones: that speech is special with regards to auditory perception, because it involves a specialized code which recruits speech output knowledge, and that because speech draws on specialized neural circuitry, it is particular to humans (Galantucci *et al* 2006). In particular, Liberman and colleagues drew on their discovery of 'categorical perception'; when presented with a range of stop consonants, consisting of a gradient of F2 transitions ranging from /ba/ to /da/ to /ga/ (ie with the place of articulation moving from front to

back), listeners' perception of the stimuli switched abruptly from one initial phoneme to the other, with no variation within a phonemic "category" (Liberman *et al* 1962). Thus, the first steps along the continuum were perceived as /b/ for all until suddenly the percept shifted to /d/ after a certain *crossover boundary*, with no gradual uncertainty or slow transition on the part of the participants. This, they claim, is because perception tracks articulation rather than acoustics; there is less articulatory variation within a consonant group than between them (Liberman *et al* 1962).

This first claim for the uniqueness of speech has largely been discarded in modern conceptualizations of motor theories, as it is not generally supported by evidence from other species or from other aspects of audition (Galantucci *et al* 2006). Chinchillas have been trained to distinguish between various /t/ and /d/ CV pairs, and are able to identify /t/ and /d/ after training when paired with novel vowels, produced by new talkers, and between /ta/ and /da/ when produced using synthetic speech (Kuhl & Miller 1975). Furthermore, the same chinchillas displayed categorical perception indistinguishable from an adult human's for a synthetic continuum ranging from /da/ to /ta/ without further training. The authors argued that because the difference between /ta/ and /da/ is one of voicing (vibration of the vocal folds begins during or just after the stop in /da/ and after the stop in /t/a), categorical perception is a psychophysical property of the auditory system, rather than reflecting articulatory knowledge (Kuhl & Miller 1975). Macaques, too, can discriminate between the acoustic markers of consonants /b,d,g/ at the same phonetic boundary points as humans (Kuhl & Padden 1983). These experiments show that implicit knowledge of speech production, or a special decoder, is not necessary to perceive all aspects of speech. Chinchillas and macaques do not produce different stop consonants, and thus must be utilizing some general aspects of the acoustic signal instead. While this does not necessarily mean humans do not draw on their articulatory knowledge, it weakens the case for speech perception being special, or particular to humans.

The second primary claim of the MTSP is that in perceiving speech, humans are perceiving gestures of the vocal tract (Liberman & Mattingly 1985).

In making this claim, proponents draw not only on the original research done by the Haskins group described above, but also on non-auditory perception of vocal gestures. If one views a speaker making a well-defined vocal gesture, such as a labial /pa/, while the sound /ka/ is played through headphones, the two percepts can 'fuse' to produce /ta/. This is known as the McGurk-MacDonald effect, after the original researchers (McGurk & MacDonald 1976). If the hearer perceives auditory /ka/ as the product of a velar constriction of the vocal tract, and visual /pa/ as a labial constriction, it would make sense that the resulting percept is an articulation intermediate between the two, namely the alveolar constriction which produces /ta/. This effect persists during other perceptual modalities. Individuals are able to experience McGurk-MacDonald effects while hearing a phoneme and using their fingers to feel a speaker articulating a different phoneme (Fowler & Dekle 1991). A variant of the MTSP, Direct Realism, pays particular attention to this second claim, and takes as its central hypothesis the idea that the objects of perception tend to be the distal cause of a percept rather than the proximal signal which conveys the perceptual information (Fowler 1986). When experimental participants are given a designated response (/pa/) and told to repeat it after a model produces a random CV syllable, their response is faster when the model's production matches their set response /pa/ than when it conflicts (Fowler *et al* 2003). During a similar task participants shadowed VCV syllables, and when the model switched to a CV syllable, participants produced either the same CV as the model or a set CV pair. Responses were faster during imitation than during a fixed response, and the difference between the first simple response task and the second choice response task was only 26ms, much shorter than the canonical 100-150ms difference between simple and choice response times, suggesting the element of choice had somehow been reduced in the second task (Fowler *et al* 2003). This is taken in support of the theory that participants actually perceive gestures. An acoustic percept would not prime a production response, since brain activation would just be in auditory regions. But if the percept were articulatory, it would activate those brain regions and reduce the time required for response

selection by priming the production of matching, but not different, responses (Galantucci *et al* 2006).

More recently, this second claim has been modified to state that listeners perceive *intended* gestures of the vocal tract, rather than the physical gestures themselves (Liberman & Mattingly 1985). This modification arose, in part, because coarticulation makes it theoretically nearly as difficult to reconstruct the physical articulation as to acoustically segment and identify phonemes (though see Fowler 1986 for a contrasting view). Another reason for the modification was the results of studies which found a lack of articulatory as well as acoustic invariants; that is, while the relationship between phoneme-to-acoustics is one-to-many (ex. /di/ and /du/ have different F2 transitions but are perceived the same), so too can articulations-to-phonemes exhibit a many-to-one relationship (eg. Lindblom *et al* 1977). For example, in an experiment by Lindblom *et al* (1977), participants were asked to produce vowels with and without their jaw being fixed in place. Despite the abnormal and novel articulation required, participants managed to produce perceptually recognizable vowels with similar acoustic properties to those of normally articulated vowels. This suggests that while physical articulation can vary, the intended goal is shifted to maintain a constant percept for the listener. Similar studies using unexpected jaw perturbation during production of /baeb/ and /baez/ showed rapid but partial compensation, which maintained undistorted speech (Kelso *et al* 1984). Still, studies of unperturbed speech show that articulatory coordination for single sounds (as opposed to connected speech) is highly constrained (Gracco & Löfqvist 1994), and so despite minor or abnormal variations in vowel articulation, it may still be theoretically possible for listeners to recover intended articulatory gestures.

The third claim of the MTSP holds that speech perception recruits motor regions of the brain. This claim has been made through each version of the MTSP, from a discussion of the "neural commands...in the central nervous system" providing "the reference system in terms of which the decoding [of speech] is carried out" (Liberman *et al* 1962, pp. 8) in early papers to the specialized neural architecture for mapping speech to articulation discussed in

later work; the claim has essentially remained that there is a functional overlap between neural networks processing articulatory output, and those receiving auditory inputs (Liberman *et al* 1967, Liberman & Mattingly 1985). This is a natural extension of the first two claims; special processing for speech perception requires accompanying specialization in body or brain or both, and given the few major differences between humans and animals in the relevant systems, it is logical to look for uniqueness primarily within the neural architecture. And in claiming that perceiving speech is perceiving gestures, motor theorists invoke an implicit knowledge of articulatory -> acoustic mapping, instantiated within the same regions of the brain used to produce speech. Early evidence for this came from studies using *selective adaptation* (a form of repetition suppression). Participants were presented with a voiced -> unvoiced range of /ba/-/pa/ or /da/-/ta/ and asked to identify the phoneme (Eimas & Corbit 1973). Before each identification trial they listened for one minute to a repeated series of an exemplar of a voiced or an unvoiced syllable (in separate runs). Their categorical perception shifted in the direction of the adapted syllable (more of the test syllables were identified as the unadapted syllable), suggesting that whatever neural mechanism was involved in voicing identification became exhausted when activated repeatedly (Eimas & Corbit 1973). However, until recently this last claim has proven difficult to investigate directly. With the discovery of mirror neurons (di Pellagrino *et al* 1992) and the rise of neuroimaging techniques, most of the recent work concerning motor processes in speech perception has focused on this claim in particular.

Acoustic theories of speech

In contrast to the MTSP, other theories of speech perception have arisen. Several fall under the general label of gestural theories, such as analysis-by-synthesis, in which incoming auditory signals are compared to a hypothetical articulatory template, or direct realism, as mentioned above (Stevens & Halle 1967, Fowler 1986). Other theories include those known as general acoustic theories of speech perception. This collection of theories typically considers the

acoustic signals produced by speech as the percept of speech (Diehl *et al* 2004; Liberman & Whalen 2000). They do not assign a prominent role in perception to the articulatory gestures used in speech production; instead, speech perception is considered to be a well-practiced activity which relies on general auditory perception mechanisms to function, such as perceiving contrasts in timing or spectral analysis. Acoustic theories generally imply a matching of acoustic cues to phonetic exemplars contained in long-term memory, which are then sent on to other cognitive mechanisms (Diehl *et al* 2004). Despite the popularity of these auditory theories, motor theories have persisted, largely due to continued findings which link speech perception and production, as well as new data from neuroimaging of speech perception (Galantucci *et al* 2006).

Neuroimaging of speech perception and production

The rise of neuroimaging techniques like positron emission tomography (PET) and fMRI has opened up new avenues of inquiry in speech research, and allows researchers to better address questions of specific neural activation patterns which have been raised since the earliest days of speech research.

While there is still little agreement on details, there is a growing consensus on which regions are involved in various speech perception and production functions, according to a review by Price (2010). The transverse temporal gyrus (or Heschl's gyrus) is known as the primary auditory cortex, involved in low-level processing of sounds. Processing pseudowords additionally recruits the superior temporal gyrus; processing words extends this activity to the middle temporal gyrus, angular gyrus, and potentially parts of the inferior parietal lobe and ventrolateral prefrontal cortex. Similar regions are involved in sentence comprehension. Resolving syntactic and semantic ambiguities recruits parts of the inferior frontal gyrus, as does word retrieval. Beyond these regions, though, there is wide disagreement regarding whether certain regions' activity is general or task and/or modality-specific, and which tasks or modalities are associated with which patterns of activity (Price 2010). Speech production research has yielded a greater consensus, with activity in primary and pre-motor cortex, somatosensory

cortex, supplementary motor area and pre-supplementary motor area, the cerebellum, and potentially the inferior frontal gyrus (Price 2010). All of the activation patterns described above are typically left-lateralized, with the exception of purely motor-output activity.

Functional neuranatomical theories of speech perception

Some theories of speech production and perception have arisen which specifically address perception in terms of neural patterns of activity without considering broader questions of the object of perception or the precise coupling of perception and production. The Gradient Order Directions Into Velocities of Articulators (GODIVA) model of speech production incorporates perception while describing auditory and sensorimotor feedback and online (real time) speech correction and compensation (Guenther 1995, Bohland *et al* 2010). The theory's proponents claim that during speech production, efferent copies of expected feedback are sent from a sound map region in the inferior frontal gyrus and ventral premotor cortex to somatosensory and auditory cortices, to be matched against incoming sensorimotor and auditory feedback. This shows obvious parallels with analysis-by-synthesis (Stephens & Halle 1967). Another theory of neural organization of speech, espoused by Hickok & Poeppel (2004), suggests that from primary auditory areas, speech perception divides into two processing streams: a ventral stream which proceeds to the inferior-posterior temporal lobe and maps sound onto meaning, and a dorsal stream which proceeds dorso-posteriorly to an auditory-motor interface area at the posterior part of the Sylvian fissue, and then dorso-anteriorly to the inferior frontal gyrus, ventral premotor cortex, and supplementary motor area. This second pathway is considered non-critical for speech perception and comprehension under normal conditions. In explaining the general leftward lateralization of speech perception, these researchers and others have proposed this to be reflective of a general difference between the two hemispheres in the time windows over which they process information, a theory often referred to as the "asymmetric sampling in time" theory (Poeppel 2003, Boemio *et al* 2005, Giraud *et al* 2007.) Thus, the left

hemisphere processes information over short time windows relevant for speech, and the right hemisphere processes over longer time windows which are only relevant for suprasegmental aspects of speech, such as prosody. A similar theory, specific to audition, suggests that there is a trade-off between processing the temporal and spectral (frequency) domains of audition, with the left hemisphere being specialized for temporal information and the right hemisphere for spectral (Zatorre *et al* 2002, Hyde *et al* 2007, Warrier *et al* 2009). Again, this theory allows for a distinction between temporally relevant speech and spectrally relevant intonation. In both cases, the needs of computationally distinct tasks are met through specialized mechanisms in the two hemispheres.

Mirror neurons

The discovery by Rizzolatti and colleagues of neurons in the F5 region of macaque monkey brains, which seemed to respond both to internally generated actions as well as while observing the execution of the same action, sparked new interest in motor contributions to speech perception (di Pellegrino *et al* 1992, Rizzolatti *et al* 1996). Importantly for speech research, some of these *mirror neurons* responded to other, non-visual stimuli specific to the gesture; for example, the sound of nuts being cracked evoked activity similar to that evoked while seeing nuts cracked, or while cracking them oneself (Köhler *et al* 2002). Further experiments revealed that the responses to observation were not species-specific, but did not occur for gestures which are not part of an individual's motor repertoire; during functional imaging of humans while viewing various oral gestures made by humans, monkeys, and dogs, similar patterns of activity were evoked for biting and lip movements by all species, but not while viewing dogs barking (Buccino *et al* 2004). Thus, mirror neurons could provide the basis for an observation-execution pairing system, and permitting comprehension of the actions of others.

The existence of a mirror neuron system in humans could also provide support for gestural theories of speech perception; mirror neurons in motor or motor-planning regions, particularly in the left inferior frontal gyrus (IFG) and

ventral premotor cortex (PMv), sometimes considered homologous to F5 in monkeys (see Petrides *et al* (2005) for arguments that IFG should not be included), would theoretically respond both to executing and to perceiving articulatory gestures (Rizzolatti & Arbib 1998). Indeed, during the early experiment above, viewing con- and non-specifics making various oral gestures was associated with activity in IFG/PMv (Buccino *et al* 2004). This mirror system for action perception was initially used to provide a mechanism for the MTSP and was later developed to explore the general role of a mirror system in human speech perception and production (some adherents further claim that the human language system was built upon a system for grasping, which became a system for action understanding and execution through evolutionary mechanisms; Rizzolatti & Arbib 1998). *Protosign* is said to have developed from this system for communicative or pantomimic gestures; from there verbal *protolanguage* co-opted these brain systems, and eventually developed into modern human language ability (Arbib 2005). This close association between observation, understanding, and execution would theoretically promote pathways common to both perceiving and producing speech and help to explain the high degree of language competence found in humans (for alternative theories on the evolution of speech see MacNeilage 1998, Fitch 2000). Recent research into the mechanisms and neural correlates of speech perception has been influenced by the above findings and theories, and three main avenues of inquiry have been pursued.

Action word perception and action execution

Action-descriptive sentences and verbs have been used to look for effector-specific somatotopic activation of the motor cortex. If hand-related verbs activate hand motor regions, foot-related verbs activate foot regions, and lip-related verbs activate lip regions, it would create a link between the action observation/execution pairing of mirror neurons in apes and more abstract pairing of action observation/execution concepts in humans. For example, listening to action sentences, compared to abstract sentences, caused reduced motor-evoked potentials (MEPs) in participants' effector-specific left-hemisphere primary motor

regions (M1) when stimulated using single-pulse transcranial magnetic stimulation (TMS; Buccino *et al* 2005). TMS is a technique which uses a magnetic coil to selectively enhance (using single pulses time-locked to stimuli of interest) or inhibit (using repeated pulses prior to the experiment) relatively specific regions of the cerebral cortex. When stimulating a muscle via brain motor regions, TMS can modulate the activity of that muscle as measured by electromyographically-recorded MEPs. A lower MEP than control is considered evidence of inhibition of the stimulated region by the experimental stimulus or intervention; a higher MEP is considered evidence of excitation or facilitation (Levy 1987). In a behavioural portion of the same TMS study above, participants were asked to identify sentences as action or non-action, using either a hand-press or a foot-press (Buccino *et al* 2005). The participants displayed slower response times when responding with their hand during hand-action sentences, and faster response times when the response-effector and described-effector did not match. A similar TMS study found that effector-specific stimulation of left M1 instead *facilitated* response times during reading of action verbs, compared to distractor words and pseudowords (Pulvermuller *et al* 2005). Neither group found a significant effect when stimulating right M1, suggesting a left hemisphere advantage for processing action words or mental simulation of actions.

Reading action words, as compared to meaningless hash marks, was associated with increases in activity in left and right IFG, bilateral precentral gyrus (PrcG), and the supplementary motor area (SMA) during fMRI (Hauk *et al* 2004). The authors claimed that effector-specific words for face, arms, and legs were associated with a somatotopic pattern of activation for face words in bilateral IFG, arm words in bilateral middle frontal gyrus (MFG) and PrcG and postcentral gyrus (PocG), and leg words in dorsal PrcG/PocG. They further compared action-word activations with activity caused by deliberate movements of the tongue, finger, and foot. This comparison revealed (small-volume corrected) overlapping activity for arm words/finger movements in right MFG and left PrcG, and for leg words/foot movements in left PrcG/PocG and central superior frontal gyrus (SFG; Hauk *et al* 2004). A similar fMRI study by

Tettamanti and colleagues (2005) compared action sentences using mouth, hand, and leg verbs with matched abstract sentences within regions active during movement of the relevant effectors. They found increased activity for action vs. abstract sentences in left IFG *pars opercularis*; by masking out the regions active for the other two effectors, they found overlap between movement and speech perception for mouth sentences in left IFGop, IFG *pars triangularis*, and left inferior parietal lobule (IPL; among others), for hand sentences in left PrcG and other regions, and for leg sentences in left superior frontal sulcus (SFS) and left IPL. It is worth noting that the inferior-superior progression of activation for mouth, hand, and left applied only along the prefrontal strip including IFG and PM, not to the more posterior regions of activation, and ignores hand activity in left PMv (Tettamanti *et al* 2005). A third study of this type compared activity during observation of effector-specific actions and effector-related literal and metaphorical action sentences to a passive resting baseline (Aziz-Zadeh *et al* 2006). They found different locations of peak response for mouth, hand, and foot action observation within bilateral IFG/PMv, and also found that these peaks responded while reading effector-specific action sentences. Metaphorical sentences were not associated with significant effector-specific activity, though they showed more activity than literal action sentences in left IFG *pars orbitalis*.

These three studies seem to demonstrate that there is common neural machinery involved in executing actions, perceiving actions, and perceiving words associated with those actions. All three found similar frontal and motor activity for these tasks, with a seeming somatotopic organization of activity, with mouth actions localized in the IFG and inferior PMv, arm actions in the superior PMv, leg actions in dorsal PMv and SMA, and general localization in the left hemisphere (Hauk *et al* 2004, Tettamanti *et al* 2005, Aziz-Zadeh *et al* 2006). However, a close examination of these studies shows some inconsistencies and unresolved questions. The overlapping areas of activation between action execution and action word perception in Hauk *et al* (2004) are extremely small, and the whole-brain maps for each task are very different, with broad overlapping activations in bilateral PM/M1 and SMA for all three effectors; by contrast, action

words are associated with non-effector-specific activation in bilateral IFG, with minor activations in some motor areas. The study by Tettamanti *et al* (2005) did not correct for multiple comparisons, creating the possibility that some of the activity they found was spurious. In addition, while they compared activation for action sentences and abstract sentences, the lack of any action execution or perception task means effector-specific activity could be due to covert movement or action-planning provoked by perceiving the action sentences, rather than representing their involvement in perception or comprehension. And Aziz-Zadeh *et al* (2006) found overlap between observation and language, but their findings are similar to those of Hauk *et al* (2004), whereby action observation was associated with activity in PM/M1 and language perception was associated with more activity in inferior frontal regions. When comparing effector-specific word activation compared to rest, the group-level clusters for each effector were approaching or marginally significant, and were close to each other. Thus, they took peak voxels for each individual, for each effector-specific word condition, within functionally-defined PM regions (active during action observation). But this method artificially inflates significance levels by selecting the most active voxels to test a second time. When formally plotted out, it becomes obvious that there is very little overlap among the three studies described above for effector-specific action-word activity; there is only a rough-grained somatotopic organization. When the peak activations from each study were plotted on cytoarchitectonically-defined maps of PM and M1 cortices, they frequently fell outside of actual motor and premotor regions (Postle *et al* 2008).

Another study examined the same phenomena using effector-specific action execution, action observation, and passive reading of mouth/arm/leg words, concrete nouns, non-words, and hash marks (Postle *et al* 2008). They examined the activity for each condition within Brodmann areas 4 (M1) and 6 (PM). Action observation and execution revealed a somatotopic organization of effector-specific activity in PM and M1, with ventral mouth activity, dorsal leg activity, and middle arm activity. For action-execution region of interests (ROI), only the mouth words > hash marks contrast within the mouth ROI was significant; within

action-observation regions, only the foot ROI showed significant effects, with foot words being associated with higher activity than hand words, concrete words, non-words, and hash marks. Both ROIs showed some non-significant trends for other effectors. By combining all effector-specific execution, observation, and words, Postle *et al* looked broadly at action words vs. other lexical-visual stimuli. They found significantly higher activity for action words over hashes and non-words (though not concrete words) in PM action execution regions; action words were also associated with significantly higher activity than all three other lexical-visual stimuli in pre-SMA regions also active under action observation. pSMA was the only region where action words showed more activity than concrete, non-action words (Postle *et al* 2008).

In the above studies which used action sentences, little attention was paid to the possible confound of differences between action and object perception (Buccino *et al* 2005, Tettamanti *et al* 2005, Aziz-Zadeh *et al* 2006). The original work on mirror neurons found activity primarily when monkeys viewed actions being carried out on or with an object, and there are suggestions that mirror neurons and canonical neurons process object and action perception differently (Jeannerod *et al* 1995). To address this, Tremblay & Small (2010) used fMRI to look at activity within PM cortex across both visual observation of actions and objects, and while listening to and repeating action- and object-related sentences, and generating object-prompted sentences. They found significant activation in non-overlapping areas of the left PMv during action observation and action sentence perception; in contrast, left superior PMv showed overlap between object observation and object sentence perception. Left superior PMv was also more sensitive to object observation than to action observation. Notably, they found no significant activity during either type of sentence perception or production in IFG. Like Postle *et al* (2008), they found differences in action and non-action activity within the pSMA, but they found this region was more sensitive to object perception than action perception. Tremblay & Small (2010) claim their results demonstrate an important role for motor regions during speech comprehension, but do not support a strong link between action observation and sentence

comprehension. This link is crucial for current theories of speech perception based on a mirror neuron system. Instead, they argue their findings reflect an internal, mental simulation of actions; and whereas watching a video of an action or reading an action sentence constrains the perceiver to the action described, object perception allows a broader range of potential motor plans to become active. This would explain the increased activity for object perception and object sentence perception in motor planning areas like PMv and pSMA. They suggest similar patterns of activation ought to be looked for during abstract sentence comprehension, which would further clarify whether premotor areas are necessary for language comprehension in general, or if they are engaged only during perceptually-induced mental simulation of actions.

Phoneme-level speech perception and production

A second line of inquiry, inspired by the original motor theory's focus on phoneme discrimination, involves investigating the contribution of motor and frontal activity in traditional speech production regions to phoneme-level speech perception. It is important to remember that this avenue of investigation is considering a lower level of speech than the action understanding account described above. Phoneme-level research considers whether there is differential and specific neural activation in speech production regions during perception of different phonemes; at this level, the articulation required to produce the phoneme is considered the percept, or perceived action. But for action understanding at the word or sentence level, the percept is the semantically-described action or object, and it is this which is supposed to cause motor activity in effector-specific regions. As can be seen in the research described here, different studies use syllables, single words, sentences, and even short stories as stimuli, and researchers often do not consider potential differences due to the investigated level of speech while citing previous work or while discussing the implications of their own. Mirror mechanisms might be at work when perceiving one level of speech, such as while listening to phonemes or syllables, but not while listening to higher-level speech, or may have differing levels of activity at different speech

levels. Alternatively, the same mechanisms might occur across different speech levels, without being influenced by the object of perception.

PET studies by Zatorre *et al* (1992, 1996) found activity during perception of CVC syllables compared to white noise in left IFGtr, during phoneme discrimination / monitoring compared to passive syllable perception in left IFGop, and in right IFG during pitch discrimination compared to passive syllable perception. These early imaging studies caused investigators to suggest that left IFG was more active during low-level speech perception, and that right IFG was more active during low-level pitch perception (Zatorre *et al* 1992, 1996). This established a potential role for IFG in low-level speech perception, provided support for theories of hemispheric asymmetry, and also suggested some distinction within IFG between passive perception (IFGtr) and discrimination (IFGop) (Zatorre *et al* 2002).

Using CVCCVC words and pseudo-words that required either low tongue movement for the middle consonant pair (ff) or high tongue movement (rr), Fadiga *et al* (2002) used single-pulse TMS over the tongue region of M1. Listening to 'rr' words significantly increased MEPs in tongue muscles for both words and non-words, possibly because the stimuli emphasized the articulator motion associated with the 'ff' and 'rr' sounds. Following up on these findings, Pulvermüller and colleagues (2006) examined brain activation patterns for lip and tongue movement, silently articulating lip (/p/) and tongue (/t/) CV syllables, and perceiving the same syllables compared to perceiving spectrotemporally matched noise. They found bilateral activity in PM/M1 for lip and tongue motion, which broadened to include MFG and IFG (more in the left hemisphere) during silent articulation. Phoneme perception was associated with activity in bilateral superior temporal gyrus. ROI analyses from a set of spheres along M1 and along PM revealed somatotopic organization for lip and tongue movement (in M1), and articulating /p/ and /t/ (in PM), but phoneme perception only showed marginal differences in activity in PM (results were not corrected for multiple comparisons). Still, the results do suggest some degree of somatotopic organization of PM response to differently articulated phonemes.

A TMS study used lip-phonemes (/b/ and /p/) and tongue phonemes (/d/ and /t/) in a phoneme identification task (D'Ausilio *et al* 2009). During the task, double-pulse TMS was delivered to $M1_{lip}$ and $M1_{tongue}$ in two separate blocks. Stimulation of $M1_{lip}$ lowered reaction times for identifying lip phonemes, and increased reaction times for tongue phonemes; stimulation of $M1_{tongue}$ caused the converse effect, as well as increasing identification accuracy for tongue phonemes and lowering accuracy for lip phonemes. Thus, increased activity in one articulatory part of M1 (or PM, as excitation/inhibition might spread from one to the other) inhibits activity in others. This is crucial if phonemic representations in motor areas are to become active while perceiving specific phonemes and not others; this inhibition is a central feature of some theories of speech production as well (i.e. Bohland *et al* 2010).

Further evidence for the involvement of motor regions in speech perception comes from a study in which the virtual lesioning of $M1_{lip}$ via rTMS disrupted categorical perception of a /ba/-/da/ and a /pa/-/ta/ continuum, but did not disrupt categorical perception of a /ka/-/ga/ continuum (Möttönen & Watkins 2009). This suppression of $M1_{lip}$ did not shift the category boundaries, as one might have expected from the selective adaptation study described above (Eimas & Corbit 1973), in which repetitive listening to a phoneme from one category, suppressing that phoneme's neural representation, caused the category boundary to shift in the direction of the suppressed phoneme. However, the slope of the /ba/-/da/ category boundary became more shallow, meaning participants had increased difficulty in detecting stimuli near the boundary; furthermore, across-category discrimination was disrupted for the lip continua but not the velar continua. rTMS to the hand area did not disrupt categorical perception for lip or non-lip phonemic continua, supporting the finding that at the phoneme level, articulator regions are being activated by their phonemes, rather than phoneme perception being associated with general PM/M1 activation.

Within phoneme-level mirror neuron research, some researchers have suggested that rather than becoming active generally, motor regions are mostly active when a phonetic contrast is ambiguous, causing the listener to draw upon

internal articulatory knowledge to disambiguate the phoneme. Callan *et al* (2004) compared Japanese listeners and English listeners on a phoneme identification task using /r/ and /l/ CV syllables. This is an extremely difficult contrast for Japanese listeners, but fairly simple for English listeners. Using fMRI, they found increased activity over rest in a wide range of regions for both groups of listeners, including bilateral IFG, PM, auditory regions, and the cerebellum. However, Japanese listeners showed significantly higher activity than English speakers in IFG and PM; English speakers showed higher activity in auditory regions. This is consistent with the idea that easy phonemic contrasts are handled primarily by auditory-acoustic means, whereas difficult or novel ones are handled using articulatory and acoustic-articulatory regions (Sato *et al* 2009). However, other researchers have found no difference within frontal motor areas between perception of native and non-native phonemes, when one would expect speech production-related activity to help disambiguate the unfamiliar phonemes (Wilson *et al* 2006).

Another study by the same group required participants to produce /b/ and /d/ phonemes, discriminate /b/ and /d/ phonemes, and to passively listen to them (Callan *et al* 2010). They found broad, overlapping networks of activation with common activity in bilateral inferior and superior PMv (edging into IFGop), SMA, and STG. They used these findings as regions of interest for another experiment, in which participants discriminated between /b/ and /d/ or /a/ and /o/, or listened to white noise. They compared activity during correct consonant trials with incorrect consonant trials and with the easier vowel trials, and found activity in the same PMvi and PMvs regions as before. This, they conclude, suggests that while these regions are active during both speech production and perception, they play a particular role in predictive coding during phoneme discrimination, by matching auditory input with articulatory knowledge to constrain the phonetic search space. Because these regions show greater activity for the difficult consonant discrimination task than the easier vowel discrimination task, they also suggest that activity in these regions is dependent on task difficulty, rather than the amount of information available in the speech signal.

As described above, many studies seem to indicate that motor and frontal activity is highly task-dependent rather than being automatic (ex. Gold *et al* 2005), and most of the studies described above required some kind of discrimination or identification task. To look at task effects, Sato *et al* (2009) applied rTMS over PMvs during the performance of three different tasks: a phoneme identification task, where participants judged whether a CVC syllable began with a /p/ or /b/; a syllable identification task, in which the listeners distinguished between syllables differing only in their initial consonant; and a phoneme segmentation/discrimination task, in which listeners determined whether the initial consonant of two different syllables were the same or not. rTMS resulted in significantly longer reaction times only for the discrimination task, which had higher phonological and working memory demands than the other two tasks. These results were seen as congruent with the early studies by Zatorre *et al* (1992, 1996) described above, which used a similar phoneme segmentation/discrimination task. The results are, as noted by the authors, also compatible with the dorsal-ventral stream model described proposed by Hickok & Poeppel (2004), in which articulatory-auditory processing only becomes relevant during speech perception under increased working memory and articulatory rehearsal demands, and incompatible with the suggestion that speech production regions contribute to speech perception under normal listening conditions. Other researchers have made similar proposals with regards to speech production contributions to speech perception (Schwartz *et al* 2008).

Although Callan *et al* (2010) suggest it is task difficulty and not information availability in the speech signal which engages speech production regions, a consideration of the stimuli used in many of the above experiments suggests this may not be entirely accurate. As noted by Sato *et al* (2009), many of the studies described above embed their stimuli in white noise to increase the difficulty of the task used and thereby reduce ceiling effects. However, embedding a signal in noise increases task difficulty by decreasing the ability of the listener to perceive the signal, essentially lowering the available information (Postma & Kolk 1992). The stimuli in D'Ausilio *et al's* (2009) and Callan *et al's*

(2010) studies were both embedded in noise. This impacts the interpretation of results from those studies, as it changes passive perception conditions into speech-in-noise conditions, and makes stimuli in identification or discrimination tasks more ambiguous than they would be under normal listening conditions.

Multimodal speech perception

Finally, there have been studies of multimodal speech perception and production. These studies typically look at the potential top-down effects of articulatory knowledge on unimodal acoustic and visual speech from a forward-model or analysis-by-synthesis perspective, and also consider how different speech modalities are weighted during perception.

One prediction coming from a mirror neuron-motor theory is that viewing or listening to phonemes produced with a specific articulator will activate motor regions associated with producing that phoneme. In a multi-pulse TMS study, participants viewed silent visual productions of /ba/ and /ta/, auditory productions of /ba/, audiovisual /ba/, and synchronized auditory /ba/ and visual /ta/ (i.e. the McGurk-MacDonald effect producing a /da/ percept) while PM/$M1_{lip}$ was stimulated (Sundara *et al* 2001). Visual and audiovisual /ba/ were associated with significantly higher MEPs than the other stimuli, but not incongruent or auditory-only /ba/. This suggests that visual speech signals are associated with speech motor activity, but not auditory speech. This is consistent with some analysis-by-synthesis accounts, which propose that it is visual speech (preceding auditory speech by several to several hundred milliseconds) which cues articulatory regions with the particular gestures to compare to the incoming auditory signal (van Wassenhove *et al* 2005). This would be in keeping with the finding that perceiving visual speech assists the perception of noise-masked speech (MacLeod & Summerfield 1987).

Other studies, however, have found that stimulation of left PM/$M1_{lip}$ during auditory speech perception of continuous prose + pixellated visual white noise, auditory non-speech perception + visual noise, watching a speaking mouth with auditory white noise, or watching eye and brow movements during speech

with auditory white noise, results in higher MEPs for both auditory speech and visual speech-related lip movements (Watkins *et al* 2003). These results conflict with the above finding that only visual speech mediates responsiveness in motor cortex, though the authors suggest this might be due to their use of higher stimulation strengths and continuous speech stimuli. A follow-up study by the same investigators used the same TMS paradigm, while also recording PET to correlate differences in cerebral blood flow (cbf) with changes in excitability as measured using MEPs from a lip muscle (Watkins & Paus 2004). They found significantly greater MEPs only for listening to auditory speech, not for any of the visual or control conditions. They found standard cbf differences between auditory and visual conditions in primary auditory and visual cortices. Additionally, there was activity within left IFGtr for speech and left IFGtr/IFGor for lip perception. There was a significant positive correlation between left IFGop/IFGtr and MEP size for auditory speech, and significantly negative correlations between right PMv and both auditory speech and lip perception. Watkins & Paus (2004) proposed a model in which information is passed from auditory regions through the parietal operculum to the posterior IFG, which modulates the excitability of the motor cortex; this may reflect internal repetition of the speech without overt production. They suggested a similar pathway, from the occipital lobe through superior temporal and inferior parietal regions, to IFG and then PM/M1, is active for visual perception of speech gestures. They explain the more anterior IFGtr and IFGor activity during auditory and visual speech perception as being due to semantic processing and visual speech reading, respectively, and not as part of this sensory-to-motor speech perception network.

Skipper and colleagues (2005) used fMRI to measure patterns of brain activity while participants listened passively to short (18-24s) stories presented in audio-only, video-only, or audiovisual formats. Audiovisual and auditory stimuli activated broad, overlapping clusters of activity; however, where audiovisual stimuli were associated with several clusters in left and right IFG, PM/M1/S1 and SFG, auditory stimuli showed only a single cluster in left IFG and visual stimuli weren't associated with activity outside of occipitotemporal visual areas. The

authors suggested that production regions are not necessarily active during auditory-only speech perception, but become quite active when integrating gestural information conveyed by visual speech; they also suggested that the lack of speech motor activation while watching visual speech may have been due to a lack of any clear perceptual or linguistic goal, and that activity might have been seen if the participants had either been instructed to attend to the visual stimuli in a specific way, or had had a task of some sort during visual speech perception (Skipper *et al* 2005).

A second fMRI study by this group made use of the McGurk-MacDonald effect (Skipper *et al* 2007). In this study, they proposed a more explicit analysis-by-synthesis model in which motor goal regions send (via PMv/M1) a multisensory hypothesis about the potential auditory and somatosensory consequences of the observed speech to superior temporal and parietal regions involved in secondary sensory processing. They presented audio and visual /pa/, /ka/ and /ta/ to participants, as well as $A_{pa}V_{ka}$ (producing a /ta/ percept) and congruent audiovisual /pa/, /ka/, and /ta/. In separate tasks, participants were also asked to say /pa/, /ka/, and /ta/, and to judge which phoneme was being produced during a second set of the audiovisual stimuli. They found more overlapping activation of speech production regions for audiovisual and visual speech perception than for auditory perception, and a higher correlation in frontal regions with perceiving a fused McGurk-MacDonald percept than for perceiving only the auditory or visual phoneme. Again, this suggests that it is the visual part of the speech signal which activates motor regions, either as an efferent copy sent to other sensory regions, or when those regions send information which is mapped to phonemic representations in speech production regions (Skipper *et al* 2007).

Hasson and colleagues (2007) used the data from the above experiment to see which regions of the brain showed less activity for $A_{pa}V_{ka}$ when the stimulus was preceded by audiovisual /pa/, /ka/, or /ta/, assuming that neural activity would be lower when a component of the preceding stimulus matched a component in $A_{pa}V_{ka}$. They found that left IFGop was strongly suppressed when the preceding stimulus was either the same signal ($A_{pa}V_{ka}$) or the same percept (/ta/). They

concluded that, because IFGop was equally suppressed by repetition of the signal or a matching percept, but not by the unimodal stimuli, that IFGop is involved in abstract coding of audiovisual speech, and may be the specific region involved in generating articulatory-acoustic hypothesis to send to primary perceptual regions.

In another fMRI study, multimodal speech was compared to matched non-speech stimuli (Hertrich *et al* 2010). Participants listened to an ambiguous /p//t/ synthetic phoneme with either a static face or a video of a man articulating /p/ or /t/, or watched the videos silently. They also listened to non-speech tones, with the facial video replaced by the expansion of a large (/p/) or small (/t/) circle. In both cases there were silent videos, auditory stimuli with a static face, and synchronized audiovisual stimuli. Participants were asked to passively attend to the final pitch shift for each auditory stimulus, to direct their attention to the auditory component of the stimuli (this task was not controlled, and may have influenced their final results). Bilateral IFG and PM/M1 were active during all three modalities, more extensively for speech than non-speech stimuli. Right IFGtr was more sensitive to the very visible /p/ gesture than /t/. Activity here showed a subadditive effect, whereby audiovisual activity was lower than the combined activity during auditory and visual perception; in fact, auditory perception was associated with higher activity in right IFGtr than audiovisual perception, which the authors suggest represents an inhibitory effect of visual stimuli. Similar findings applied to right IFGop, and, in general, left IFG showed more activity during auditory than audiovisual speech perception. The authors suggest their results may reflect the acoustically ambiguous auditory stimulus and attention to final pitch changes, but also claim that their results cast doubt on the visually-based analysis-by-synthesis account given by Skipper and colleagues (Skipper *et al* 2007, Hertrich *et al* 2010).

As described above, it is possible that embedding stimuli in noise may be responsible for some of the motor activity described in a number of studies. A similar issue may confound some of the results described in all three avenues of mirror neuron-based investigation described above. These studies used continuous sampling fMRI, in which complete brain volumes are acquired one

directly after the other. The noise from the scanner creates very loud background noise, and this may have an effect similar to intentionally embedding stimuli in noise. Of the above studies, only Tremblay & Small (2010) and Pulvermuller *et al* (2006) made use of a technique to reduce the effect of scanner noise during auditory trials. Both studies used *sparse-sampling* fMRI, in which there is a silent gap between brain volume acquisitions; while not as accurate at capturing the peak of the blood oxygen level dependent (BOLD) signal, it allows auditory stimuli to be heard without loud background noise (Gracco *et al* 2005). All other fMRI studies discussed above, therefore, have potentially confounding noise masking the auditory stimuli presented during scanning, which may contribute to motor activity seen during seemingly passive perception or simple identification tasks.

Future avenues of inquiry

Motor and frontal involvement in speech may not only be confined to action understanding, lower-level speech perception of phonemes, or visual speech. A recent study examined the spatial-temporal correlation between the brain activity of a storyteller and her listeners, using a foreign-language storyteller as a control (Stephens *et al* 2010). They found that frontal/anterior activity in the listeners mirrored, but preceded, the same activity in the speaker. Parietal and occipital activity mirrored, but followed, the speaker's patterns of brain activity, and activity in primary auditory regions was synchronized between the speaker and listeners. When the listeners were played a similar spontaneous story told by a foreign-language speaker, there were no significant spatio-temporal correlations between speaker and listeners. This suggests that the activity seen while listening to and comprehending a spoken story in some cases anticipates, and in other cases follows, that of the speaker's, creating a tight neural coupling (Stephens *et al* 2010). A logical step for future studies would be to apply the same techniques to a dialogue, to see how the neural activity of one speaker couples to that of the other, depending on the timing and content of the discussion.

Objectives and Hypotheses:

A large number of studies have, in recent years, looked at the contribution of frontal and motor regions during speech perception. However, there are problems and inconsistencies, as described above. Debate still exists over whether motor activity is a necessary part of phoneme-level perception, or if it is modulated only by task requirements and the difficulty or novelty of the phonemes. At the word and sentence-level, the debate centres around the difference between action-speech and object-speech, whether abstract sentence processing also involves motor processing, and whether this motor activity is effector-specific or generalized to some part of PM/M. During multimodal speech perception, questions still exist regarding the relative contribution of auditory and visual modalities to motor activity during speech perception. Furthermore, much of the research has been confounded by the embedding of stimuli in noise, either intentionally or by using continuous-sampling fMRI methodologies.

Our intention with this current study is to pull back from the specific claims of mirror neuron-based theories of phoneme discrimination or action understanding, and to consider instead whether frontal and motor regions active during simple word-level speech production are active both during word-level passive perception, as well as during more complex perception of non-action, sentence-level speech. If word-level production regions are active during higher-order speech perception, these regions can then be examined on a modality- and task-specific basis to see if their activation patterns can be linked to the functional requirements of those conditions.

This is tested by using two studies, described below. Single-word speech perception activity is examined within those regions activated during a simple speech production task at the same level. Then, higher-level speech perception is examined within the same production regions for auditory, visual, and audiovisual sentences, to see if the activity is similar across speech levels. This latter analysis will also reveal potential differences in production regions between different

speech modalities. If either level of speech perception corresponds poorly with speech production, secondary whole-brain analyses can be conducted to explore the perceptual network at that level to see where it departs from that involved in speech production. This is particularly likely if speech production involves a relatively limited network of activity, with modality-specific activity outside of these regions.

We expect to find similar functional networks active during production of single words and simple passive word perception, and higher-level sentence perception, assuming it is possible to generalize about production regions' contributions across speech levels. For high-level speech, following Skipper and colleagues' (2005, 2007) analysis-by-synthesis approach, we expect that the frontal-motor components of production networks will be more associated with visual speech perception than with auditory speech perception.

Methods

Experiment 1: Production and perception of multi-modal lexical stimuli

The first experiment was intended to reveal the neural correlates of speech perception and production at the word-level, using simple concrete objects as the stimuli to avoid potentially confounding activity from action words, and to ensure participants used consistent names for the pictured objects. The perceptual modalities were varied to reveal differences and similarities across the neural networks responsible for auditory perception and repetition, object retrieval and naming, and reading silently and aloud.

Participants consisted of 12 undergraduate student volunteers (6 male, mean age 22.06). All participants were native/primary speakers of North American English, with normal hearing and normal/corrected-to-normal vision, no history of neurological disorder, and were right-handed as determined by scores on the Annett Handedness self-report inventory (Briggs & Nebes 1975). All participants gave informed consent in accordance with the review and ethics board of the Montreal Neurological Institute (MNI). The study was carried out

with the approval of the Magnetic Resonance Research Committee (MRRC) and the MNI Research Ethics Board.

Fifty (50) concrete nouns in written, picture, or auditory format were presented to participants in two blocks (one of 12 trials and one of 13 trials) for each condition (as well as for a passive rest condition), in random order, during each of two separate functional runs. Before each block, participants were instructed to either rest, to passively read / view / listen to the stimulus, or to read / name / repeat the stimulus aloud. All nouns were selected from a standardized set of simple line drawings of objects (Snodgrass & Vanderwart 1980). Visual stimuli were presented on an LCD projector viewed through a mirror attached to the head coil. Auditory stimuli were presented binaurally using MRI-compatible headphones. A Dell Precision laptop was used to present the stimuli, using Presentation software (Neurobehavioural System, Albany, CA, USA).

All stimuli were presented during silent intervals between volume acquisitions to enable clear hearing of the auditory stimuli and to reduce motor artefacts during speech production (Gracco *et al* 2005). Image acquisition was performed using a 3T Siemens Trio scanner at the MNI. For each of the two experimental runs, 191 $4mm^3$ volumes were acquired with a T2*-weighted multi-slice EPI descending interleaved sequence (TE = 30ms, TR = 2.16s, 2.84s TR delay, flip angle 90 º, matrix 64x64, fov 256x256mm). Interleaved acquisition reduces slice overlap saturation with the relatively short TR, and better represents the temporal distribution of activity; descending order avoids any chance of magnetically saturating ascending cerebral blood flow. Each volume consisted of 36 axial slices oriented parallel to the AC-PC line (no gap). High-resolution $1mm^3$ T1-weighted anatomical images were also acquired between runs #1 and #2. Participants had their head immobilized using a polystyrene-filled vacuum bag, to minimize motion artefacts caused by head movements during volume acquisition.

Experiment 2: Multimodal perception of higher-level speech

The second experiment was intended to reveal the various brain networks active during perception of affective and linguistic sentence prosody, for unimodal auditory and visual sentences, as well as multimodal audiovisual sentences. Prosodic processing was assessed not as simple, passive perception, but by using a forced-choice discrimination task within both types of prosody.

Participants consisted of 10 volunteers (5 female, mean age 26.0). All participants were native/primary speakers of North American English, with normal hearing and normal/corrected-to-normal vision, no history of neurological disorder, and were right-handed as determined by scores on the Annett Handedness self-report inventory (Briggs & Nebes, 1975). All participants gave informed consent in accordance with the review and ethics board of the MNI. The study was carried out with the approval of the MRRC and the MNI Research Ethics Board.

Stimuli were digitally recorded by a female native English speaker using clear enunciation and regular lip movement. The speaker was shown with her head and shoulders against a dark background during stimulus recording. For audio-only trials, a still image of the speaker with a neutral expression was used during the presentation of the auditory stimuli. For video-only trials, all auditory input was removed. Audiovisual trials presented both the video and audio streams. Recordings were conducted using a Hi8 Sony digital camera and edited using Adobe Creative Suite 3 audio-visual editing software.

The stimulus sentences were divided into 120 affective (60 angry, 60 happy) sentences and 120 linguistic (60 question, 60 statement) sentences, for a total of 240 stimuli. Affective sentences were unique; linguistic sentences were matched for question and statement sentences, as question sentences were created by digitally raising the final pitch of each statement sentence. All sentences were of noun phrase-verb-noun phrase structure and were semantically neutral, with prosodic category information conveyed through visual and intonational cues. All sentences were digitally edited to last between 2.5 and 2.8 seconds, and each sentence was presented only once.

Testing sessions lasted approximately 2 hours and consisted of two runs each for affective and linguistic sentences, counterbalanced for run order across participants. During each run, 80 stimulus sentences were presented in event-related, pseudorandom order. Each run consisted of 60 test trials bracketed by 10 rest trials, with no more than two stimuli of the same modality played in a row. Each trial began with a 3.0s volume acquisition, followed by a 0.2-1.0s jitter, the 2.5-2.8s stimulus, and then a 2.2-3.3s response period prompted by a fixation cross. Each trial therefore lasted 9.0s. See Figure 2. for a representation of a single trial. Participants were asked before each run to identify sentences as angry vs. happy for the affective runs, or question vs. statement for the linguistic runs, using a two-button response box under their right hand. Stimuli were presented using the same experimental setup as in Experiment #1. Sparse-sampling was also used, as in Experiment #1, to minimize motion artefacts and ensure speech perception was unhindered by scanner-related noise.

Functional and anatomical data were collected using a 1.5T Siemens Sonata scanner at the MNI. For each functional run, 80 $4mm^3$ volumes were acquired in descending sequential order with a T2*-weighted multi-slice EPI sequence (TE = 50ms, TR = 3.0s, delay in TR 6.0s, flip angle 90^o, matrix 64x64, fov 256x256mm). Descending sequential order was used in this experiment because a longer TR reduces the likelihood of slice overlap saturation, and it was preferable to avoid spin history artefacts. Each volume consisted of 35 axial slices oriented parallel to the AC-PC line (no gap). A high-resolution anatomical scan was conducted after the first two functional runs, with a T1-weighted volume acquired for the whole head (TE = 9.2ms, TR = 22ms, fov 256x256mm, $1mm^3$).

fMRI analysis

The primary goal in analyzing these two experiments was to examine brain activity associated with speech perception within the speech production network. Thus, a functional mask was created by taking a conjunction of the naming, repeating, and reading speech production tasks from Experiment #1 at p = 0.005 (uncorrected) to find only those regions active under all three tasks; this

revealed areas fundamental to speech production and not those specific to each task. This conjunction map was used as a functional mask of single-word production regions during group-level analysis of the perception tasks in Experiment #1 and Experiment #2, using mixed-effects meta-analysis (MEMA) as described below. Activation within the mask was considered significant using $p = 0.01$ corrected to $p = 0.05$ using FWE cluster-size correction (minimum cluster size 12 voxels). For Experiment #1, perception and production conditions were also directly contrasted to reveal differences in the spread and intensity of motor activity between the conditions ($p = 0.001$ corrected to $p = 0.005$ using FWE at 20 voxels). For Experiment #2, a series of secondary whole-brain analyses were run to determine task- and modality-specific effects in regions outside of the production mask from Experiment #1.

In order to compare findings across experiments, the same pre-processing procedure was followed for both, allowing only for experiment-specific differences (as noted below). The anatomical images for each participant were skull-stripped using a brain-extraction algorithm, and the volume acquired closest to the anatomical image (for each functional run) was aligned to the anatomical using a modified localized Pearson correlation function (Saad *et al* 2009). Each volume from the functional runs was then registered to the anatomically-aligned volume using Fourier interpolation. Volumes with movement artefacts were recorded for censoring during individual participant regressions. Each run was de-spiked and spatially smoothed to 6.0mm FWHM using a Gaussian kernel to decrease spatial noise, and the blurred volumes had their signals scaled to a mean signal intensity of 100. The anatomical images were warped to standard MNI space, using cubic interpolation, to match the MNI152 linearly-averaged brain; the functional runs were then warped to match the new anatomical space at $3mm^3$ voxel resolution (Cox *et al* 1996).

For Experiment #1, the haemodynamic response function (HRF) from each functional run was modelled as a block function for both 12-trial and 13-trial blocks, which were combined using a general linear test after regression. For Experiment #2 the HRF was deconvolved with a piecewise linear spline, known

as a *tent function*, with each of the 12 stimulus types (4 sentence types x 3 modalities) as separate regressors. The affective and linguistic categories were collapsed to create betas and t-values for audiovisual (AV), auditory-only (A), and visual-only (V) modalities using a general linear test after regression. The six motion parameters, instruction breaks, and other non-stimuli non-rest events were modeled as regressors of non-interest. The resulting design matrix was fitted to a generalized least-squares time series, using a restricted maximum likelihood (REML) autoregressive moving average (ARMA) to correct for serial correlation in noise over the time series.

Beta co-efficients and t-values were extracted from single subject analyses and used for group analysis. Group analyses for each experiment were conducted using mixed-effects meta-analyses (MEMA) to better control for within-subject variability by weighting results by beta estimate reliability, using the associated t-values (Chen *et al* 2010). Both whole-brain and functionally masked analyses were conducted. Whole-brain activation in Experiment #2 was considered significant using $p = 0.005$ corrected to $p = 0.05$ using FWE cluster-size correction (minimum cluster size 41 voxels) as calculated by 3dClustSim, an AFNI program which uses brain dimensions and smoothing estimates to conduct Monte Carlo probability simulations of random noise fields at given thresholds. Conjunction analyses (signified by 'Λ') used only those regions significantly active *above baseline* (not including "deactivations") under each condition individually at the uncorrected p-value; conjunction maps were then corrected to $p < 0.05$ using the FWE method described above. Results approaching significance are reported due to the conservative nature of the conjunction.

Results

Production mask

The conjunction analysis from Experiment #1 revealed activity within the following core speech production regions: bilateral medial superior frontal gyrus (SFGm aka pSMA/SMA), bilateral strips along the ventral pre- and post-central

gyri (M1v and S1v), bilateral clusters in the dorsal post-central gyrus (S1), bilateral putamen, bilateral superior cerebellum (Cbm), bilateral transverse temporal gyrus (TTG), and one cluster in left TTG just posterior to the bilateral cluster (Fig. 3). Notably, there was no significant activity within inferior frontal regions, posterior parietal regions, or secondary auditory regions. See Table 1 for coordinates and cluster sizes for all production-masked results.

Word-level speech perception within production regions

The passive listening task from Experiment #1 was examined using the above production mask, to find activity within production regions during passive perception. The other two passive conditions (picture naming and word reading) were not analysed here, as they were less comparable to the stimuli from Experiment #2. Passive listening was associated with bilateral activity along M1v and S1v, bilateral dorsal S1 (extending anteriorly into the central sulcus), bilateral TTG, and left posterior TTG (Fig. 4a). A whole-brain contrast analysis of Repeat - Listen, conducted to determine whether activity during speech perception was equal to that during overt production, revealed significantly higher activity for word production in the same bilateral strips of M1v/S1v, as well as bilateral superior Cbm, a superior dorsal portion of the brainstem, and left SMA; word perception showed higher activity than production in the cingulate gyrus (CG) anterior to the genu of the corpus callosum, and in right anterior middle frontal gyrus (MFG) (Fig. 4b).

Sentence-level speech perception within production regions

Each stimulus modality from Experiment #2 was analysed within the production mask from Experiment #1. Audiovisual sentence perception was associated with clusters of activity in left superior M1v/S1v, right superior Cbm, left ventral post-central gyrus, left posterior TTG, and right middle TTG (Fig. 5a). Auditory perception showed similar patterns of activity (Fig. 5b). Visual perception was associated with activity only in left superior M1v/S1v and right superior Cbm (Fig. 5c). No modality-specific differences were found within the

frontal or motor portions of the production mask. Left superior M1v/S1v and right superior Cbm were commonly active across all three modalities; the only modality-specific differences were in primary auditory regions, which were not active for visual speech perception.

Sentence-level speech and modality

Only a single frontal/motor cluster in M1/S1 was found within production regions during high-level perception, which did not differ by modality and was likely due to the button-pressing response (Wildgruber *et al* 2004). Contrast analyses (Fig. 12, 13, and 15) between the various modalities in Experiment #2 revealed no significant difference in these regions among the different modalities. Because of the lack of activity within low-level production regions for high-level perception, we conducted whole-brain analyses of Experiment #2, independently of Experiment #1, to reveal modality-specific differences outside of core low-level speech production regions. See Table #2 for cluster locations described below; individual figures are noted with their descriptions.

Amodal speech perception activity

A three-way conjunction of all three modalities vs. the resting baseline revealed higher activity during speech perception in middle/inferior occipital gyrus (M/IOG) extending anteriorly and inferiorly along the inferior temporal gyrus (ITG), and spreading medially in places across the lateral occipito-temporal sulcus into the fusiform gyrus (FG). Activation also ran anterior and superior from M/IOG into posterior middle and superior temporal gyrus (M/STGp), and posterior superior temporal sulcus (STSp), but did not extend superiorly as far as the angular gyrus or IPL. All three conditions further included separate clusters of activity in right middle STS, the left superior M1v/S1v, right superior Cbm, and bilateral inferior Cbm (Fig. 6). The results described below do not include these commonly-activated regions.

Auditory components of speech perception

The auditory components of speech perception, found via a conjunction analysis of audiovisual and auditory perception (AV Λ A), were associated with activity in the bilateral superior temporal plane, extending inferiorly to the STS and in some places to the middle temporal gyrus. Additional clusters of activity were found in left SFGm and right IFGtr (Fig. 7). Unimodal auditory speech also included activity along the bilateral striate and precuneus, and right IFGop (Fig. 8). Note that all auditory conditions show substantial activation spread across the Sylvian fissure from superior temporal regions into the parietal operculum. Examination of the images at arbitrarily high thresholds showed that this activity was due to spreading and was not independently significant.

Visual components of speech perception

The visual components (AV Λ V) of speech perception were only associated with activity (beyond that shown in all three modalities) within MTGp (Fig. 9). Unimodal visual speech showed an additional cluster of activity in right IFGop, spreading into right middle frontal gyrus (MFG; Fig. 10).

Multimodal and unimodal speech perception

Audiovisual speech perception included clusters along the bilateral superior temporal plane and in left SFGm; a cluster in right IFGtr approached significance (Fig. 11). Audiovisual perception showed more robust activation than auditory perception in a contrast analysis within the bilateral posterior-inferior occipitotemporal network, a portion of the right STG/Sp around the anterior ascending branch of the STSp, and in the central straight gyrus (SG) and medial orbital gyrus (MOrG; Fig. 12). A contrast analysis revealed significantly higher activity for audiovisual perception than visual in the bilateral superior temporal plane, anterior insula, left amygdala and left medial geniculate nucleus (MGN), right MTGa, central SG and MOrG, and in left IFGtr (the latter cluster only approaching significance; Fig. 13).

Unimodal stimuli (A Λ V) were associated with common activity in right IFGop (Fig. 14). Auditory perception showed more activity than audiovisual only

in the striate and pre-cuneus regions of the primary visual cortex (Fig. 12). The same region was also more active during auditory perception than during visual perception (Fig. 15). Additionally, auditory perception showed more activity than visual in the superior temporal plane, left IFGor, and left SFGm. Visual perception showed more activity than audiovisual in a region of MTGp often described as area MT/V5 (Fig. 13). It was associated with higher activity than auditory perception in the bilateral posterior-inferior occipitotemporal network, particularly on the right, again including area MT/V5 (Fig. 15).

Discussion

This work compares two studies. In one, participants either passively perceived single words or repeated single words. The difference between these two conditions lay in passive perception vs. overt speech production. In the other, participants listened to or viewed complete sentences, and were required to extract suprasegmental emotional or linguistic information from auditory, visual, or audiovisual speech streams. Here, the difference lay in the different neurocognitive and perceptual demands of the same task across different modalities. The two studies differed most significantly in speech level. Listeners in the first were required only to perceive and produce a single word, where listeners in the second had to perceive an entire sentence, and extract additional information from the available perceptual streams.

Speech production

Our investigation of speech production identified a network of activity which resembles those found in other studies of overt syllable, word, and narrative speech production (Wise *et al* 1999, Blank *et al* 2002, Bohland & Guenther 2006, Sörös *et al* 2006, Chang *et al* 2009). We found some overlapping activity between single-word speech production and single-word speech perception, but the various modalities in Experiment #2 only showed limited activity within the production mask, which did not closely resemble the activation

patterns from Experiment #1. Notably, there was no activity within the production mask anterior to the precentral sulcus (excepting pSMA/SMA). This suggests that activation of the left IFG, which is commonly treated as a speech production region in the literature, is not an essential part of the speech production network activated in Experiment #1. Studies have generally found production-related activity within left IFG during complex vocalizations and explicit speech planning (Sörös *et al* 2006, Chang *et al* 2009), so the lack of activity here may be reflective of the cued nature of the speech, which did not particularly require any internal selection of a verbal response for the listen-and-repeat condition in Experiment #1 (Tremblay & Gracco 2006). There was more widespread activity in the production mask within primary motor regions than there was for either passive perception in Experiment #1 or for the various modalities in Experiment #2.

Word-level speech perception within production regions

Passive perception in Experiment #1 showed activity within the production mask congruent with that found in studies of speech perception-production overlap and with mirror system explanations of speech perception. We found activity during passive listening within the central sulcus bilaterally (including the caudal part of the precentral gyrus and the rostral part of the postcentral gyrus) in a region commonly associated with sensorimotor control of the face and mouth (Fox *et al* 2001, Postle *et al* 2008). Our activation clusters for passive perception within production regions are not close to those reported in studies of action words (ex. Hauk *et al* 2004), but are in the same general area identified as being active during speech production, and to a lesser degree speech perception, in a study of production and perception of non-word syllables (Wilson *et al* 2004).

There are several potential explanations for activity in primary motor regions during speech perception. Activation of speech production regions during word perception could prime such regions to activate more quickly during reproduction of speech than when producing a fixed response or a non-matching

word (Fowler *et al* 2003). This would explain the short latencies found during speech shadowing studies (Porter & Lubker 1980). Again, the level of representation is an important factor in examining these theories. The priming explanation can be interpreted within the action perception framework, with the perceived action being phonetic articulatory gestures which activate the regions involved in producing these gestures. At the word level and above, the perceived action could still be these articulatory gestures, or it could be the actions being semantically described. As described for Tremblay & Small (2010), some mirror system theories suggest that perception-coupled motor activity derives from mental simulation or imaging of actions, rather than an obligatory coupling of effector-specific motor regions with semantically related words. Tomasino and colleagues (2007), using an idea related to this simulation account, claim that enhanced M1 activity in some previous studies of action words is due to the adoption of a strategy of mentally imagining described actions, and that in studies which did not find enhanced M1 activation for action words, participants were not adopting a mental action simulation strategy to meet the task demands. Some studies have interpreted this to mean participants must be explicitly instructed to engage in mental imagery, but Tomasino *et al* suggest participants can adopt an internal simulation strategy either implicitly or explicitly (Tomasino *et al* 2007, Tremblay & Small 2010). Given that participants may or may not adopt an action simulation strategy, this may also explain inconsistent findings of M1 activity during perception studies.

Using object words and not action words, Tremblay & Small (2010) found clusters of activity in left PMv for object sentences (both passive listening and repetition); and while this was anterior to our own cluster in left M1/S1, both their study and ours found motor activity centring around face and mouth regions, not those of more distal effectors. They explain the activity they found, which was significantly higher for object than for action sentences, by suggesting their participants had a wider range of potential motor plans or mental simulations available when comprehending an object than when comprehending a single predefined action (Tremblay & Small 2010).

This might explain Tremblay & Small's (2010) findings at the sentence level, as their frontal-motor regions of activation were primarily in motor planning regions such as left PMv and pSMA. But at the word-level and below there is another, perhaps simpler explanation for our own bilateral M1/S1 clusters during the listen-repeat phase. Phonological working memory (PWM) creates a phonological loop between motor regions (subvocal articulatory rehearsal) and the inferior parietal lobule (storage area) to maintain verbal or lexical information for short-term recall and manipulation (Baddeley 2003). The overlap between mirror system and PWM explanations for the IFG/PM activity sometimes observed during speech perception has not been well explored. For example, a recent study, building on the work done by Fadiga *et al* (2002), examined the effects of phonological and lexical properties of disyllabic Italian pseudowords containing a doubled labial or alveolar (tongue) consonant (as well as rare and frequent Italian words containing a doubled tongue consonant) in the middle of the word on the excitability of the tongue part of left M1, using single-pulse TMS at 0, 100, 200, and 300ms from the beginning of the doubled consonant (Roy *et al* 2008). Random trials (25%) required a word vs. non-word lexical decision response. At 0ms, there was no difference in MEPs for any of the words or pseudowords; at 100ms, pseudowords requiring tongue articulation had significantly higher MEPs than pseudowords not requiring tongue articulation, and the real words trended closer to the tongue-articulated pseudowords; at 200ms, rare words had significantly higher MEPs than frequent words, and a trend remained for the two classes of pseudowords; by 300ms, only the difference between rare and frequent words remained.

The authors explain their findings from both studies as showing an automatic motor resonance for the phono-articulatory content of words, followed by this system assigning meaning to words, and being more active during this process for rare words than frequent words due to the increased difficulty of deciding whether a rare word is a word or not (Roy *et al* 2008). They also cite literature showing left IFG/PM activity during various lexico-semantic decision tasks. Much of their discussion could equally suit an explanation based in PWM,

especially given the task demands. As soon as the stimulus was heard, participants would have had to hold the word in PWM, knowing they might have had to make a judgment on it. Covert articulation would activate parts of M1 involved in articulating the word, hence the difference between tongue-articulated and non-tongue-articulated pseudowords at 100ms. The earlier study shows that at 100ms tongue-articulated words of the types used in Roy *et al* (2008) are also associated with significantly higher MEPs than non-tongue-articulated words (Fadiga *et al* 2002). Behavioural studies have shown that reaction times for lexical decision tasks are shortest for high frequency words, followed by non-words and then low frequency words (Rubenstein *et al* 1970, Forster & Chambers 1973). In combination with the results from Roy *et al* (2008), this suggests that in early processing all tongue-articulated words are maintained in PWM, with rare words being maintained longer than pseudowords or high frequency words in order to attempt to match them to words in the lexicon.

Imaging and neuropsychological studies of subvocal or covert articulation processes in PWM have commonly localized these mechanisms in IFG and PMv (Huang *et al* 2002, Baddeley 2003); this shows an overlap with the regions implicated by mirror system theories. An fMRI study was used to localize regions involved in the Sternberg recognition task, which requires participants to maintain one or six letters in PWM, until prompted by a probe to state if the probe includes a letter they saw previously (Herwig *et al* 2003). They found task-related activity in bilateral middle frontal gyrus, left IFG/PMv, left superior PMv, bilateral SMA, and left IPS. TMS studies explicitly investigating PWM, unlike those examining mirror or motor contributions to speech perception, are lacking. One such study found that rTMS over either PMv/IFGop or IPL reduced accuracy and increased reaction times on stress difference, initial sound difference, and digit span same/different judgment tasks, but not a visual pattern span task of the same type (Romero *et al* 2006). The first two tasks are thought to involve mental rehearsal, and the third to use both the phonological store and the articulatory process posited by PWM. Event-related short-train TMS applied to PM and the intraparietal sulcus (IPS; approximately) decreased accuracy on the same task for

left PM compared to right PM, but not for the IPS or for reaction times in either region. Another study found rTMS over left IFG disrupted the delay phase of a delayed phonological matching task, in which participants had to maintain a word in working memory in order to match it to a phonologically similar pseudoword (Nixon *et al* 2006). TMS studies examining mirror systems of speech perception, which may be confounded with PWM, commonly act on M1, and it is possible that they are also facilitating or inhibiting PWM processes. At least one imaging study explicitly considering PWM has found activity within M1/S1 during PWM tasks. McGettigan *et al* (2011) found activity in left M1 very close to our location while participants maintained pseudowords during a delay between perception and reproduction, compared to the same task using tones, and found this activity increased in proportion to the number of syllables in the pseudoword. They found no activity in IFG during this task, and little in premotor areas. An aggregate look at studies that used an encoding -> rehearsal -> probe paradigm showed high activity in bilateral precentral gyrus and central sulcus, as well as IFG, for both encoding and rehearsal phases (Buchsbaum *et al* 2011). This further suggests that articulatory rehearsal may not be realized only within PMv or IFG alone.

While a PWM explanation of our findings of M1 overlap between word perception and repetition is intriguing, all of the research on PWM, and many similar mirror system studies, use tasks which could contain a PWM component (Herwig *et al* 2003, Romero *et al* 2006, Roy *et al* 2008, McGettigan *et al* 2011). Experiment #1, however, did not. It is possible that articulatory rehearsal is in some way automatic at lower speech levels, that participants mentally recite words or syllables whenever they hear them. But such a view would be difficult to support, since there is no particular reason for the brain to tie up neuronal resources in this way. At the segmental or word level, it is possible that heard speech activates motor regions in an articulatory-specific manner, but only in the absence of a particular task. This would account for the findings of some studies which find TMS modulates activity within specific parts of M1 during effector-emphasized word perception, and fMRI studies which find overlap between perception and production of syllables within M1 (ex. Fadiga *et al* 2002, Wilson

et al 2004). It would also account for studies where an explicit task was not associated with M1 activity, such as Zatorre *et al* (1992), where phonetic discrimination was associated with left IFG and pitch discrimination with right IFG. During increased task demands, primary motor activity may be suppressed in favour of other regions, particularly PM and IFG. This is congruent with studies which find increased activity within PM/IFG during difficult low-level speech tasks, and no effect of TMS inhibition of PM on simple low-level speech tasks (Sato *et al* 2009, Callan *et al* 2010). Furthermore, this explanation does not rule out task-related M1 activity, such as that found in some PWM studies, as primary motor activity would not be suppressed if it facilitated the task at hand (such as covertly maintaining a syllable in working memory; McGettigan *et al* 2011).

Sentence-level speech perception within production regions

In contrast to word-level perception, there was little overlap between activity within the production mask and activity associated with sentence-level perception. Primary auditory cortex was active in the production mask due to auditory feedback (McGuire *et al* 1996), and during auditory and audiovisual speech perception in Experiment #2 (Zatorre *et al* 1992). The hand area of M1/S1 was involved in the button-pressing response in Experiment #2 (see also Wildgruber *et al* 2004); the focus of activity in PM/M1/S1 in Experiment #1 was inferior and slightly anterior to this activity for both production and perception tasks, and was active as part of vocal production, possibly as covert articulatory rehearsal or resonant articulatory activity (as discussed above). Superior Cbm was also active for both production and perception in Experiment #2, but not for perception in Experiment #1, suggesting that, as for M1/S1, its activity was related to the button-pressing response (Ackermann *et al* 1998).

Both experiments showed activity for speech production and sentence-level perception in the pSMA/SMA, but the active region did not overlap significantly between the two experiments, being more anterior and inferior for sentence-level perception. The lack of activity in this region for passive word

perception in Experiment #1 is not a surprise given the lack of any overt response. The lack of significant overlap in pSMA/SMA between the two experiments is consistent with studies showing more anterior/pSMA activity in this region for internally generated responses (Deiber *et al* 1996, Tremblay *et al* 2006, Karch *et al* 2009) and more posterior/SMA activity during repetition (Crosson *et al* 2001, Tremblay & Gracco 2006, Tremblay & Gracco 2010).

There was meaningful overlap between word production and auditory sentence perception only within primary auditory regions; overlap within motor and cerebellar regions was coincidental and not reflective of similar neural processes during the two tasks. None of the 'traditional' mirror regions such as IFG, PMv, or IPL were active during both word production and sentence perception, and *contra* Skipper *et al* (2005, 2007), there were no modality-specific differences within the motor/frontal regions for perception. Other studies have also noted discrepancies between findings from studies using low-level speech and those using higher-level speech on perceptual tasks similar to ours (Postle *et al* 2008). Given the overlap between low-level production and perception, and the lack of overlap between low-level production and high-level perception, we must conclude that any generalization across speech levels, at least for perception, should be done cautiously and take into consideration the different task demands being placed on participants at different speech levels.

High-level speech perception, modality, and task

We found no significant overlap between word production and sentence perception within frontal or motor regions, nor did we find any differences between modalities outside of primary sensory regions. It is possible, however, that at higher speech levels, with different task demands, there is frontal or motor activity which does not overlap with the more primary motor activity found during word production. Thus, we looked specifically at whole-brain analyses of the perception tasks from Experiment #2.

Our results bear some resemblance to similar studies of uni- and multi-modal speech perception, which typically contrast passive or discrimination

listening to a resting baseline (Skipper *et al* 2005, Dick *et al* 2009, Hertrich *et al* 2010). Each of these studies, like our own, found similar patterns of activity for audiovisual speech perception in occipital, occipitotemporal, and superior temporal regions. These include activation within the putative ventral visual stream (bilateral M/IOG through ITG and FG), which plays a hypothesized role in recognition of objects, visual motion and faces (Halgren *et al* 1999, Kourtzi *et al* 2000); area MT/V5, which is involved in visual processing of motion (Kourtzi *et al* 2000) and eye gaze directed towards the viewer (Watanabe *et al* 2006); the superior temporal plane, extending from the anterior temporal pole to the bifurcation of the STSp, which is involved in such a broad variety of auditory and speech tasks that any comprehensive list would be impossible in this space (see Price 2010 for a review); and the pSMA/SMA, which is involved in response selection and action sequencing (Karch *et al* 2009, Tremblay *et al* 2009, 2010), and processing task-related temporal intervals (Coull *et al* 2004, Geiser *et al* 2008).

Beyond this broad resemblance, there are a range of key differences between previous findings and our own. Firstly, and most critically with respect to our expectations, neither audiovisual (as found by Skipper *et al* 2005) nor visual (as predicted by Skipper *et al* 2007) speech perception preferentially activated motor or inferior frontal areas traditionally associated with speech production, over auditory speech perception alone. This finding was consistent both for the production-masked results discussed above, as well as for whole-brain analyses. Our findings paint a more complex picture showing that activation is associated with modality in ways partly inconsistent with an efferent-copy motor or mirror system explanation.

We found activation of right IFGtr for audiovisual and auditory perception, but not for visual speech perception. This suggests that IFGtr is involved in the processing of the auditory information stream in speech; previous studies have indicated that right IFGtr plays a role in pitch perception (Zatorre *et al* 1992). Klein and colleagues (2001) also found activation in right IFGtr for English speakers, but not Mandarin speakers, during a sameness task using

Mandarin words with lexical tones; this suggests that the activity in our study reflects perception of pitch differences over semantic differences in the stimuli, as would be expected with our use of semantically neutral sentences, which differed only in prosodic content. This activity in right IFGtr is therefore also driven by the prosodic task demands of Experiment #2. Both IFGtr and right IFG more generally, have been found to be more active while discriminating between different prosodic intonations, compared both to rest as well as to listening to neutral or semantically emotional speech (George *et al* 1996, Plante *et al* 2002, Kotz *et al* 2003).

In our study, right IFGop was active during auditory and visual sentences. A lowered threshold revealed activity in right IFGop for audiovisual speech as well, but well below our modestly stringent level of significance. Visual perception of sentences has been associated with higher activity in right IFGop than audiovisual perception (Skipper *et al* 2005). Skipper *et al* suggest this might be due to passive speech-reading by their participants. While this may explain its activity during visual speech perception in our study, right IFGop was active for both visual and auditory speech perception. Previous studies have found right IFGop activity during emotional prosody judgements (compared to a word discrimination task), and a part of right anterior IFGop has been associated, along with pSMA, with explicit judgements of auditory speech rhythm (Buchanan *et al* 2000, Geiser *et al* 2008). Given this established association between right IFGop and prosodic or speech rhythm discrimination, its activity may be more accurately associated with our prosody discrimination task *per se* than with speech reading generally. In this case, our results suggest it is more active when less information is available to make a judgement (for auditory and visual speech alone), and less active when more information is available to make a judgement (during audiovisual speech). Studies of emotional sentence prosody recognition have shown that hit rates increase with the number of available modalities (Paulmann & Pell 2011). However, other studies have found increased activity related to lower stimulus saliency more medially and anteriorly, near the border of the MFG and IFGtr (Leitman *et al* 2010).

In both cases, the right-lateralization of frontal activity may be attributed to the nature of the prosodic discrimination task in Experiment #2. Sentence-level prosody is typically associated with bilateral or right-lateralized activity in imaging studies, such as those comparing perception of lexical (temporally local) and intonational (sentence-level) prosody in native speakers and non-speakers of tonal languages (Klein *et al* 2001, Gandour *et al* 2003) or comparing prosodic identification with phonetic monitoring (Wildgruber *et al* 2005). Such studies typically link perception of lexical prosody or judgements of linguistic prosody to the left hemisphere, and affective or intonational prosody to the right (Wildgruber 2006). As described previously, there are theories suggesting that the left and right hemisphere are specialized for temporal auditory information (left) and spectral auditory information, which are in line with the above findings (Zatorre *et al* 2002). Findings of lateralization from lesion studies have typically been more ambiguous, showing impairment for prosodic perception in both left and right brain-damaged patients, with particular impairment for left brain-damaged patients in identifying linguistic prosody (Pell & Baum 1997). These studies typically consider only auditory responses to speech prosody, but at least one right-hemisphere damaged patient has shown both auditory and visual prosodic impairments (Nicholson *et al* 2002). rTMS studies using more precise virtual lesions have not clarified the picture; inhibition of both left and right IFG during identification of emotional sentence prosody significantly reduces reaction times on a forced-choice task (Hoekert *et al* 2010). The likely explanation between different findings from TMS and lesion data on the one hand, and some fMRI studies on the other, is that left and right hemispheres engage different sub-processes during prosodic identification and discrimination, and that these processes interact to some degree with semantic processes (Baum & Dwivedi 2003, Pell 2006). This is compatible with the asymmetrical sampling in time theory described above (Gandour *et al* 2003, Poeppel 2003).

Given the nature of the task in Experiment #2, one might wonder whether the PWM mechanisms described for low-level speech perception might be expected to play a role during the auditory conditions of Experiment #2.

Certainly, participants would need to hold the sentences in working memory in order to complete the identification task at the end of each trial, so unlike the word perception condition from Experiment #1, there is a role for working memory in Experiment #2. However, unlike Experiment #1, the focus in Experiment #2 was on a suprasegmental component of speech, one largely independent of any sort of articulation, covert or overt. Because the sentences were selected to be semantically neutral, participants also would not need to attend strongly to the specific words used. Covert articulation would therefore be less relevant to the task than a working memory mechanism which could maintain suprasegmental information about the sentence prosody. This mechanism need not be instantiated within motor regions, unlike covert articulation. Right IFGtr could well represent an instantiation of this higher-level working memory, and this would correspond well with the above discussion of hemisphere-specific processing intervals, if high temporal-resolution motor information (articulation) is maintained in part of the left PM/M1 and low temporal-resolution information (pitch changes over a sentence) is maintained in right IFGtr (Gandour *et al* 2003, Poeppel 2003).

In the discussion of general whole-brain activity for audiovisual speech perception, pSMA/SMA activity was partly attributed to response selection on the discrimination task. However, it was most active for audiovisual and auditory perception, while activity in this region did not achieve significance for visual speech perception. This is likely due to the differing task demands of auditory versus visual prosody perception. pSMA/SMA, along with right anterior IFG, have been shown to be more active when participants attend to timing of events during stimulus presentation, such as comparing length of presentation of coloured stimulus pairs vs. comparing their colour (Coull *et al* 2004). Similarly, pSMA/SMA and right IFG have also been shown to be active during explicit judgement of speech rhythm (Geiser *et al* 2008). If pSMA plays a preferential role in tasks where timing judgments are involved, this might explain its greater activity for auditory speech perception. A judgement of verbal prosody relies on intonation across much of the utterance for both affective and linguistic prosody.

For example, a study of the priming effects of emotional utterances on judgement of facial expression shows that listeners must hear more than 300ms of the utterance for a priming effect to occur (Pell 2005). In contrast, facial movement is not typically required for an affective judgement, and studies using audiovisual prosody often use still images of faces (Ethofer *et al* 2006b). Thus, when making a prosodic judgement, the visual condition does not require much attention to temporal features of the signal, whereas the auditory conditions do, hence the greater association of these temporally-sensitive regions with auditory conditions. This would similarly explain why IFGtr was not active for the visual condition, as the working-memory requirements of visual prosody / facial expression are lower than those for auditory prosody, and rely more on spatial relationships than changes in pitch or sound spectra over time.

The difference among the three modalities was also associated with activation within more posterior regions involved in voice and facial processing. All three conditions showed activity in the ventral visual processing stream, but audiovisual and visual speech perception of moving faces were associated with significantly higher activity in those regions than auditory speech perception. And, as in previous studies, when participants viewed a still face during auditory speech perception, activity increased in the striate and precuneus (primary visual areas; Calvert *et al* 2003). This may be due to increased processing in primary visual areas in an attempt to extract visual information to send on to secondary visual areas, which are more active during conditions which actually provide information for them to process (Calvert *et al* 2003). Area MT/V5, which was active during the visual conditions but not auditory speech perception, is known to be active during perception of dynamic faces as well as non-biological motion (Halgren *et al* 1999, Calvert *et al* 2003, Miki *et al* 2004). In particular, right MT/V5 was more active for audiovisual than auditory perception, and more active for purely visual perception than for audiovisual perception, a finding concordant with studies of audiovisual matching tasks in speech (Saito *et al* 2005).

We also found modality-specific activity in posterior temporal sulcus and gyrus, particularly in the posterior ascending branch of the superior temporal

sulcus for visual conditions, and the trunk of the STS for auditory conditions. Our results are similar to those of studies looking for an integration centre for audiovisual sentences near the bifurcation of the pSTS (Kreifelts *et al* 2009, Beauchamp *et al* 2010). We found an area in bilateral middle STG/S (larger on the right) which responded for all three modalities, a finding which closely resembles those of some studies looking at multi-modal audiovisual speech (Stevenson *et al* 2009), though unlike some other studies, we did not find significantly higher activity for audiovisual sentences in this region (Szycik *et al* 2008, Kreifelts *et al* 2009).

Conclusions

Single-word cued speech production showed neural activity which significantly overlapped with that evoked by single-word perception, but did not show significant overlap with activity associated with higher-level speech perception. In light of the lack of any task demands during word perception, while our findings for low-level perception could be explained by phonological working memory and covert articulation (Baddeley 2003, McGettigan *et al* 2011), or mirror system-based speech theories (Fadiga *et al* 2002, Pulvermuller *et al* 2006), the best explanation is one based loosely on the MTSP. At lower speech levels, articulation is perceptually relevant, and activates articulator-specific motor regions. In the presence of tasks which have demands which cannot be met by primary motor regions, this activity is suppressed or inhibited in favour of other regions whose activity is relevant to the task. At higher speech levels, articulation becomes less relevant to the speech signal, and motor regions may only come online in the presence of specific task demands. Given that none of the three modalities from Experiment #2 were preferentially associated with activity in frontal or motor regions commonly associated with speech perception, our findings do not support a robust interpretation of Skipper *et al*'s theory that the visual component of speech sends an efferent copy of the signal to be matched to the auditory stream, at the sentence level. This may be true for low-level speech

(which was used in Skipper *et al* 2007), if the visual modality provides information particularly relevant for articulation. But even compared to Skipper *et al* 2005, which used short stories, the frontal and motor activity we found at the sentence level reflects strict task requirements, and cannot be easily divided up by modality.

The interaction of modality and task is complex. Both right inferior frontal regions (IFGtr and IFGop) were involved in the prosodic aspect of the task in Experiment #2, but IFGop was more active for low-information modalities, and thus likely associated with the discrimination component of the task. IFGtr, on the other hand, was more active for modalities with temporally-relevant and spectral components, and is probably more involved in classifying the parts of the prosodic signal which vary over time, such as pitch; it may also play a role in auditory working memory. Our inferior frontal results were right-lateralized, contrary to many studies of non-prosodic speech perception. The likeliest explanation is again the task demands of Experiment #2, which required an explicit discrimination between different types of prosody, and therefore engaged a more right-lateralized network than might a more lexical or semantic task, which would demand processing over shorter time intervals. The only premotor region which was active during sentence level perception, pSMA/SMA, has well-defined cognitive and motor roles which do not speak to mirror or MTSP explanations, but to temporal sequencing and response selection, the latter explaining the hand area PM/M1/S1 activity seen under all conditions.

Rather than the obligatory articulatory motor processing suggested by the MTSP (Liberman & Mattingly 1985) or some mirror-system theorists (Skipper *et al* 2007, Callan *et al* 2010), or the implicit motor imagery / action representation suggested by others to play a level-general and modality-specific role in speech perception (Hauk *et al* 2004, Tomasino *et al* 2007, Tremblay & Small 2010), our results show that different speech levels inherently engage different frontal and motor mechanisms, and that these mechanisms depend on the specific demands placed upon the participants by the speech level, task, and modality.

References:

Ackermann H, Wildgruber D, Daum I, Grodd W. Does the cerebellum contribute to cognitive aspects of speech production? A functional magnetic resonance imaging (fMRI) study in humans. *Neurosci Lett*. 1998;**247**(2-3):187-90.

Adolphs R, Tranel D. Intact recognition of emotional prosody following amygdala damage. *Neuropsychologia*. 1999;**37**(11):1285-92.

Adolphs R, Tranel D, Damasio H. Emotion recognition from faces and prosody following temporal lobectomy. *Neuropsychology*. 2001;**15**(3):396-404.

Arbib MA. From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behav Brain Sci*. 2005;**28**(2):105-24.

Aziz-Zadeh L, Wilson SM, Rizzolatti G, Iacoboni M. Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Curr Biol*. 2006;**16**(18):1818-23.

Baddeley A. Working memory: looking back and looking forward. *Nat Rev Neurosci*. 2003;**4**(10):829-39.

Baum SR, Dwivedi VD. Sensitivity to prosodic structure in left- and right-hemisphere-damaged individuals. *Brain Lang*. 2003;**87**(2):278-89.

Beauchamp MS, Nath AR, Pasalar S. fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *J Neurosci*. 2010;**30**(7):2414-7.

Blank SC, Scott SK, Murphy K, Warburton E, Wise RJ. Speech production: Wernicke, Broca and beyond. *Brain*. 2002;**125**(Pt 8):1829-38.

Boemio A, Fromm S, Braun A, Poeppel D. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci*. 2005;**8**(3):389-95.

Bohland JW, Guenther FH. An fMRI investigation of syllable sequence production. *Neuroimage*. 2006;**32**(2):821-41.

Bohland JW, Bullock D, Guenther FH. Neural representations and mechanisms for the performance of simple speech sequences. *J Cogn Neurosci*. 2010;**22**(7):1504-29.

Briggs GG & Nebes RD. Patterns of hand preference in a student population. *Cortex*. 1975;**11**(3):230-238.

Buccino G, Lui F, Canessa N, Patteri I, Lagravinese G, Benuzzi F, Porro CA, Rizzolatti G. Neural circuits involved in the recognition of actions performed by nonconspecifics: an FMRI study. *J Cogn Neurosci*. 2004;**16**(1):114-26.

Buccino G, Riggio L, Melli G, Binkofski F, Gallese V, Rizzolatti G. Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioral study. *Brain Res Cogn Brain Res*. 2005;**24**(3):355-63.

Buchanan TW, Lutz K, Mirzazade S, Specht K, Shah NJ, Zilles K, Jäncke L. Recognition of emotional prosody and verbal components of spoken language: an fMRI study. *Brain Res Cogn Brain Res*. 2000;**9**(3):227-38.

Buchsbaum BR, Baldo J, Okada K, Berman KF, Dronkers N, D'Esposito M, Hickok G. Conduction aphasia, sensory-motor integration, and phonological

short-term memory - An aggregate analysis of lesion and fMRI data. *Brain Lang*. 2011 Jan 20.

Callan DE, Jones JA, Callan AM, Akahane-Yamada R. Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *Neuroimage*. 2004;**22**(3):1182-94.

Callan D, Callan A, Gamez M, Sato MA, Kawato M. Premotor cortex mediates perceptual performance. *Neuroimage*. 2010;**51**(2):844-58.

Calvert GA, Campbell R. Reading speech from still and moving faces: the neural substrates of visible speech. *J Cogn Neurosci*. 2003;**15**(1):57-70.

Chang SE, Kenney MK, Loucks TM, Poletto CJ, Ludlow CL. Common neural substrates support speech and non-speech vocal tract gestures. *Neuroimage*. 2009;**47**(1):314-25.

Chen G, Saad ZS, Cox RW. Modeling multilevel variance components and outliers in group analysis. 16th Annual Meeting of the Organization for Human Brian Mapping. Barcelona, Spain, 2010.

Cooper FS, Liberman AM, Borst JM. The interconversion of audible and visible patterns as a basis for research in the perception of speech. *Proc Natl Acad Sci*. 1951;**37**(5):318-325.

Cooper FS, Delattre PC, Liberman AM, Borst JM, Gerstman LJ. Some experiments on the perception of synthetic speech sounds. *J Acoust Soc Am*. 1952;**24**: 597-606.

Coull JT, Vidal F, Nazarian B, Macar F. Functional anatomy of the attentional modulation of time estimation. *Science*. 2004;**303**(5663):1506-8.

Cox RW. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*. 1996;**29**:162-173.

D'Ausilio A, Pulvermüller F, Salmas P, Bufalari I, Begliomini C, Fadiga L. The motor somatotopy of speech perception. *Curr Biol*. 2009;**19**(5):381-5.

Deiber MP, Ibañez V, Sadato N, Hallett M. Cerebral structures participating in motor preparation in humans: a positron emission tomography study. *J Neurophysiol*. 1996;**75**(1):233-47.

Dick AS, Goldin-Meadow S, Hasson U, Skipper JI, Small SL. Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Hum Brain Mapp*. 2009;**30**(11):3509-26.

Diehl RL, Lotto AJ, Holt LL. Speech perception. *Annual Review of Psychology*. 2004;**55**:149-79.

di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G. Understanding motor events: a neurophysiological study. *Exp Brain Res*. 1992;**91**(1):176-80.

Eimas PD, Corbit JD. Selective Adaptation of Linguistic Feature Detectors. *Cog. Psych.* 1973;**4**:99-109.

Ethofer T, Anders S, Erb M, Droll C, Royen L, Saur R, Reiterer S, Grodd W, Wildgruber D. Impact of voice on emotional judgment of faces: an event-related fMRI study. *Hum Brain Mapp*. 2006a;**27**(9):707-14.

Ethofer T, Pourtois G, Wildgruber D. Investigating audiovisual integration of emotional signals in the human brain. *Prog Brain Res*. 2006b;**156**:345-61.

Fadiga L, Craighero L, Buccino G, Rizzolatti G. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur J Neurosci*. 2002;**15**(2):399-402.

Fitch WT. The evolution of speech: a comparative review. *Trends Cogn Sci*. 2000;**4**(7):258-267.

Foulke E. Listening comprehension as a function of word rate. *J Commun*. 1968;**18**(3):198-206.

Fowler CA. Segmentation of coarticulated speech in perception. *Percept Psychophys*. 1984;**36**(4):359-68.

Fowler CA. An event approach to the study of speech perception from a direct—realist perspective. *J Phonetics*. 1986;**14**:3-28.

Fowler CA, Dekle DJ. Listening with eye and hand: cross-modal contributions to speech perception. *J Exp Psychol Hum Percept Perform*. 1991;**17**(3):816-28.

Fowler CA, Brown JM, Sabadini L, Weihing J. Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *J Mem Lang*. 2003;**49**(3):396-413.

Fox PT, Huang A, Parsons LM, Xiong JH, Zamarippa F, Rainey L, Lancaster JL. Location-probability profiles for the mouth region of human primary motor-sensory cortex: model and validation. *Neuroimage.* 2001;**13**(1):196-209.

Galantucci B, Fowler CA, Turvey MT. The motor theory of speech perception reviewed. *Psychon Bull Rev*. 2006;**13**(3):361-77.

Gandour J, Dzemidzic M, Wong D, Lowe M, Tong Y, Hsieh L, Satthamnuwong N, Lurito J. Temporal integration of speech prosody is shaped by language experience: an fMRI study. *Brain Lang.* 2003;**84**(3):318-36.

Geiser E, Zaehle T, Jancke L, Meyer M. The neural correlate of speech rhythm as evidenced by metrical speech processing. *J Cogn Neurosci*. 2008;**20**(3):541-52.

George MS, Parekh PI, Rosinsky N, Ketter TA, Kimbrell TA, Heilman KM, Herscovitch P, Post RM. Understanding emotional prosody activates right hemisphere regions. *Arch Neurol*. 1996;**53**(7):665-70.

Gick B, Derrick D. Aero-tactile integration in speech perception. *Nature.* 2009;**462**(7272):502-4.

Giraud AL, Kleinschmidt A, Poeppel D, Lund TE, Frackowiak RS, Laufs H. Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron*. 2007;**56**(6):1127-34.

Gold BT, Balota DA, Kirchhoff BA, Buckner RL. Common and dissociable activation patterns associated with controlled semantic and phonological processing: evidence from FMRI adaptation. *Cereb Cortex*. 2005;**15**(9):1438-50.

Gracco VL, Löfqvist A. Speech motor coordination and control: evidence from lip, jaw, and laryngeal movements. *J Neurosci*. 1994;**14**(11 Pt 1):6585-97.

Gracco VL, Tremblay P, Pike B. Imaging speech production using fMRI. *Neuroimage*. 2005;**26**(1):294-301.

Griffiths TD, Uppenkamp S, Johnsrude I, Josephs O, Patterson RD. 3. Encoding of the temporal regularity of sound in the human brainstem. *Nat Neurosci*. 2001;**4**(6):633-7.

Guenther FH. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychol Rev*. 1995;**102**(3):594-621.

Halgren E, Dale AM, Sereno MI, Tootell RB, Marinkovic K, Rosen BR. Location of human face-selective cortex with respect to retinotopic areas. *Hum Brain Mapp*. 1999;**7**(1):29-37.

Hauk O, Johnsrude I, Pulvermüller F. Somatotopic representation of action words in human motor and premotor cortex. *Neuron*. 2004;**41**(2):301-7.

Hertrich I, Dietrich S, Ackermann H. Cross-modal interactions during perception of audiovisual speech and nonspeech signals: an fMRI study. *J Cogn Neurosci*. 2011;**23**(1):221-37.

Herwig U, Abler B, Schönfeldt-Lecuona C, Wunderlich A, Grothe J, Spitzer M, Walter H. Verbal storage in a premotor-parietal network: evidence from fMRI-guided magnetic stimulation. *Neuroimage*. 2003;**20**(2):1032-41.

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

Hoekert M, Vingerhoets G, Aleman A. Results of a pilot study on the involvement of bilateral inferior frontal gyri in emotional prosody perception: an rTMS study. *BMC Neurosci*. 2010;**11**:93.

Hornak J, Bramham J, Rolls ET, Morris RG, O'Doherty J, Bullock PR, Polkey CE. Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices. *Brain*. 2003;**126**(Pt 7):1691-712.

Hyde KL, Peretz I, Zatorre RJ. Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia*. 2008;**46**(2):632-9.

Imaizumi S, Mori K, Kiritani S, Kawashima R, Sugiura M, Fukuda H, Itoh K, Kato T, Nakamura A, Hatano K, Kojima S, Nakamura K. Vocal identification of speaker and emotion activates different brain regions. *Neuroreport*. 1997;**8**(12):2809-12.

Jeannerod M, Arbib MA, Rizzolatti G, Sakata H. Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends Neurosci*. 1995;**18**(7):314-20.

Joos M. Acoustic Phonetics. *Language*. 1948;**24**(2):1-136.

Karch S, Mulert C, Thalmeier T, Lutz J, Leicht G, Meindl T, Möller HJ, Jäger L, Pogarell O. The free choice whether or not to respond after stimulus presentation. *Hum Brain Mapp*. 2009;**30**(9):2971-85.

Kelso JA, Tuller B, Vatikiotis-Bateson E, Fowler CA. Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. *J Exp Psychol Hum Percept Perform*. 1984;**10**(6):812-32.

Klein D, Zatorre RJ, Milner B, Zhao V. A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *Neuroimage*. 2001;**13**(4):646-53.

Köhler E, Keysers C, Umiltà MA, Fogassi L, Gallese V, Rizzolatti G. Hearing sounds, understanding actions: action representation in mirror neurons. *Science*. 2002;**297**(5582):846-8.

Kotz SA, Meyer M, Alter K, Besson M, von Cramon DY, Friederici AD. On the lateralization of emotional prosody: an event-related functional MR investigation. *Brain Lang*. 2003;**86**(3):366-76.

Kourtzi Z, Kanwisher N. Activation in human MT/MST by static images with implied motion. *J Cogn Neurosci*. 2000;**12**(1):48-55.

Kreifelts B, Ethofer T, Shiozawa T, Grodd W, Wildgruber D. Cerebral representation of non-verbal emotional perception: fMRI reveals audiovisual integration area between voice- and face-sensitive regions in the superior temporal sulcus. *Neuropsychologia*. 2009;**47**(14):3059-66.

Kuhl PK, Miller JD. Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science*. 1975;**190**(4209):69-72.

Kuhl PK, Padden DM. Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *J Acoust Soc Am*. 1983;**73**(3):1003-10.

Leitman DI, Wolf DH, Ragland JD, Laukka P, Loughead J, Valdez JN, Javitt DC, Turetsky BI, Gur RC. "It's Not What You Say, But How You Say it": A Reciprocal Temporo-frontal Network for Affective Prosody. *Front Hum Neurosci*. 2010;**4**:19.

Levy WJ. Transcranial stimulation of the motor cortex to produce motor-evoked potentials. *Med Instrum*. 1987;**21**(5):248-54.

Liberman AM, Delattre P, Cooper FS. The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *Am J Psychology*. 1952;**65**:497-516.

Liberman AM, Delattre P, Cooper FS, Gerstman L. The role of consonant–vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs: General & Applied.* 1954;**68**:1-13.

Liberman AM. Some results of research on speech perception. *Journal of the Acoustical Society of America.* 1957;**29**:117-123.

Liberman AM, Cooper FS, Harris KS, MacNeilage PF. A motor theory of speech perception. Proceedings of the Speech Communication Seminar, Stockholm. 1962:1-12.

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of speech code. *Psychological Review*. 1967;**74**:431-461.

Liberman AM, Mattingly IG. The motor theory of speech perception revised. *Cognition*. 1985;**21**(1):1-36.

Liberman AM, Whalen DH. On the relation of speech to language. *Trends Cogn Sci.* 2000;**4**(5):187-196.

Lindblom B, Lubker J, Gay T. Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *J. Acoust. Soc. Am*. 1977;**62**(S1):S15-S15.

Lippmann RP. Speech recognition by machines and humans. *Speech Comm*. 1997;**22**(1):1-15.

MacLeod A, Summerfield Q. Quantifying the contribution of vision to speech perception in noise. *Br J Audiol*. 1987;**21**(2):131-41.

MacNeilage PF. The frame/content theory of evolution of speech production. *Behav Brain Sci*. 1998;**21**(4):499-511; discussion 511-46.

McGettigan C, Warren JE, Eisner F, Marshall CR, Shanmugalingam P, Scott SK. Neural correlates of sublexical processing in phonological working memory. *J Cogn Neurosci*. 2011;**23**(4):961-77.

McGuire PK, Silbersweig DA, Frith CD. Functional neuroanatomy of verbal self-monitoring. *Brain*. 1996;**119**(Pt 3):907-17.

McGurk H, MacDonald J. Hearing lips and seeing voices. *Nature.* 1976;**264**:746–47.

McNeill D, Lindig K. The perceptual reality of phonemes, syllables, words, and sentences. *J Verb Learn Verb Behav*. 1973;**12**(4):419-430.

Miki K, Watanabe S, Kakigi R, Puce A. Magnetoencephalographic study of occipitotemporal activity elicited by viewing mouth movements. *Clin Neurophysiol*. 2004;**115**(7):1559-74.

Möttönen R, Watkins KE. Motor representations of articulators contribute to categorical perception of speech sounds. *J Neurosci*. 2009;**29**(31):9819-25.

Nicholson KG, Baum S, Cuddy LL, Munhall KG. A case of impaired auditory and visual speech prosody perception after right hemisphere damage. *Neurocase*. 2002;**8**(4):314-22.

Nixon P, Lazarova J, Hodinott-Hill I, Gough P, Passingham R. The inferior frontal gyrus and phonological processing: an investigation using rTMS. *J Cogn Neurosci*. 2004;**16**(2):289-300.

Oden GC, Massaro DW. Integration of featural information in speech perception. *Psych. Rev*. 1978;**85**(3):172-191.

Pallett D. Performance assessment of automatic speech recognizers. *J Res NIST*. 1985;**90**(5):1-17.

Paulmann S, Seifert S, Kotz SA. Orbito-frontal lesions cause impairment during late but not early emotional prosodic processing. *Soc Neurosci*. 2010;**5**(1):59-75.

Paulmann S, Pell MD. Is there an advantage for recognizing multi-modal emotional stimuli? *Motivation and Emotion*. 2011;**35**(2):192-201.

Pell MD, Baum SR. Unilateral brain damage, prosodic comprehension deficits, and the acoustic cues to prosody. *Brain Lang*. 1997;**57**(2):195-214.

Pell MD. Prosody-Face interactions in emotional processing as revealed by the facial affect decision task. *J Nonverbal Behav*. 2005;**29**(4):193-215.

Pell MD. Cerebral mechanisms for understanding emotional prosody in speech. *Brain Lang*. 2006;**96**(2):221-34.

Petrides M, Cadoret G, Mackey S. Orofacial somatomotor responses in the macaque monkey homologue of Broca's area. *Nature*. 2005;**435**(7046):1235-8.

Plante E, Creusere M, Sabin C. Dissociating sentential prosody from sentence processing: activation interacts with task demands. *Neuroimage*. 2002;**17**(1):401-10.

Poeppel D. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*. 2003;**41**(1):245-55.

Porter RJ Jr, Lubker JF. Rapid reproduction of vowel-vowel sequences: evidence for a fast and direct acoustic-motoric linkage in speech. *J Speech Hear Res*. 1980;**23**(3):593-602.

Postle N, McMahon KL, Ashton R, Meredith M, de Zubicaray GI. Action word meaning representations in cytoarchitectonically defined primary and premotor cortices. *Neuroimage*. 2008;**43**(3):634-44.

Postma A, Kolk H. The effects of noise masking and required accuracy on speech errors, disfluencies, and self-repairs. *J Speech Hear Res.* 1992;**35**(3):537-44.

Potter RK, Kopp GA, Green HC. Visible Speech. D. Van Nostrand Co., New York NY, 1947.

Price CJ. The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann N Y Acad Sci*. 2010;**1191**:62-88.

Pulvermüller F, Hauk O, Nikulin VV, Ilmoniemi RJ. Functional links between motor and language systems. *Eur J Neurosci*. 2005;**21**(3):793-7.

Pulvermüller F, Huss M, Kherif F, Moscoso del Prado Martin F, Hauk O, Shtyrov Y. Motor cortex maps articulatory features of speech sounds. *Proc Natl Acad Sci USA*. 2006;**103**(20):7865-70.

Rizzolatti G, Fadiga L, Gallese V, Fogassi L. Premotor cortex and the recognition of motor actions. *Brain Res Cogn Brain Res*. 1996;**3**(2):131-41.

Rizzolatti G, Arbib MA. Language within our grasp. *Trends Neurosci.* 1998;**21**:188-194.

Romero L, Walsh V, Papagno C. The neural correlates of phonological short-term memory: a repetitive transcranial magnetic stimulation study. *J Cogn Neurosci*. 2006;**18**(7):1147-55.

Roy AC, Craighero L, Fabbri-Destro M, Fadiga L. Phonological and lexical motor facilitation during speech listening: a transcranial magnetic stimulation study. *J Physiol Paris*. 2008;**102**(1-3):101-5.

Saad ZS, Glen DR, Chen G, Beauchamp MS, Desai R, Cox RW. A new method for improving functional-to-structural alignment using local Pearson correlation. *NeuroImage* 2009;**44**:839-848.

Sabri M, Binder JR, Desai R, Medler DA, Leitl MD, Liebenthal E. Attentional and linguistic interactions in speech perception. *Neuroimage*. 2008;**39**(3):1444-56.

Saito DN, Yoshimura K, Kochiyama T, Okada T, Honda M, Sadato N. Cross-modal binding and activated attentional networks during audio-visual speech integration: a functional MRI study. *Cereb Cortex*. 2005;**15**(11):1750-60.

Sato M, Tremblay P, Gracco VL. A mediating role of the premotor cortex in phoneme segmentation. *Brain Lang*. 2009;**111**(1):1-7.

Schwartz JL, Sato M, Fadiga L. The common language of speech perception and action: a neurocognitive perspective. *Revue française de linguistique appliquée*. 2008;**2**:9-22.

Skipper JI, Nusbaum HC, Small SL. Listening to talking faces: motor cortical activation during speech perception. *Neuroimage*. 2005;**25**(1):76-89.

Skipper JI, van Wassenhove V, Nusbaum HC, Small SL. Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb Cortex*. 2007;**17**(10):2387-99.

Snodgrass JG, Vanderwart M. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J Exp Psychol Hum Learn*. 1980;**6**(2):174-215.

Sörös P, Sokoloff LG, Bose A, McIntosh AR, Graham SJ, Stuss DT. Clustered functional MRI of overt speech production. *Neuroimage*. 2006;**32**(1):376-87.

Stephens GJ, Silbert LJ, Hasson U.Speaker-listener neural coupling underlies successful communication. *Proc Natl Acad Sci USA*. 2010;**107**(32):14425-30.

Stevens KN, Halle M. Remarks on analysis by synthesis and distinctive features. In: Wathen-Dunn W, editor. Models for the perception of speech and visual form. Cambridge, MA: MIT Press.

Stevenson RA, James TW. Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage*. 2009;**44**(3):1210-23.

Sundara M, Namasivayam AK, Chen R. Observation-execution matching system for speech: a magnetic stimulation study. *Neuroreport*. 2001;**12**(7):1341-4.

Szycik GR, Tausche P, Münte TF. A novel approach to study audiovisual integration in speech perception: localizer fMRI and sparse sampling. *Brain Res*. 2008;**1220**:142-9.

Taler V, Baum S, Saumier D, Chertkow H. Comprehension of grammatical and emotional prosody is impaired in Alzheimer's disease. *Neuropsychology* 2008;**22**(2):188-195.

Tatham M, Morton K. Speech Perception and Production. Palgrave MacMillan, New York NY, USA, 2006.

Tettamanti M, Buccino G, Saccuman MC, Gallese V, Danna M, Scifo P, Fazio F, Rizzolatti G, Cappa SF, Perani D. Listening to action-related sentences activates fronto-parietal motor circuits. *J Cogn Neurosci*. 2005;**17**(2):273-81.

Tomasino B, Werner CJ, Weiss PH, Fink GR. Stimulus properties matter more than perspective: an fMRI study of mental imagery and silent reading of action phrases. *Neuroimage*. 2007;**36**(S2):T128-41.

Tremblay P, Gracco VL. Contribution of the frontal lobe to externally and internally specified verbal responses: fMRI evidence. *Neuroimage*. 2006;**33**(3):947-57.

Tremblay P, Gracco VL. Contribution of the pre-SMA to the production of words and non-speech oral motor gestures, as revealed by repetitive transcranial magnetic stimulation (rTMS). *Brain Res*. 2009;**1268**:112-24.

Tremblay P, Gracco VL. On the selection of words and oral motor responses: evidence of a response-independent fronto-parietal network. *Cortex*. 2010;**46**(1):15-28.

Tremblay P, Small SL. From Language Comprehension to Action Understanding and Back Again. *Cereb Cortex*. 2010 Oct 12.

Tyler LK, Marslen-Wilson W. Fronto-temporal brain systems supporting spoken language comprehension. *Philos Trans R Soc Lond B Biol Sci*. 2008;**363**(1493):1037-54.

van Wassenhove V, Grant KW, Poeppel D. Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci USA*. 2005;**102**(4):1181-6.

von Kriegstein K, Patterson RD, Griffiths TD. Task-dependent modulation of medial geniculate body is behaviorally relevant for speech recognition. *Curr Biol*. 2008;**18**(23):1855-9.

Warrier C, Wong P, Penhune V, Zatorre R, Parrish T, Abrams D, Kraus N. Relating structure to function: Heschl's gyrus and acoustic processing. *J Neurosci*. 2009;**29**(1):61-9.

Watanabe S, Kakigi R, Miki K, Puce A. Human MT/V5 activity on viewing eye gaze changes in others: A magnetoencephalographic study. *Brain Res*. 2006;**1092**(1):152-60.

Watkins KE, Strafella AP, Paus T. Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*. 2003;**41**(8):989-94.

Watkins K, Paus T. Modulation of motor excitability during speech perception: the role of Broca's area. *J Cogn Neurosci*. 2004;**16**(6):978-87.

Wildgruber D, Hertrich I, Riecker A, Erb M, Anders S, Grodd W, Ackermann H. Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. Cereb Cortex. 2004;14(12):1384-9.

Wildgruber D, Riecker A, Hertrich I, Erb M, Grodd W, Ethofer T, Ackermann H. Identification of emotional intonation evaluated by fMRI. *Neuroimage*.

2005;**24**(4):1233-41.

Wildgruber D, Ackermann H, Kreifelts B, Ethofer T. Cerebral processing of linguistic and emotional prosody: fMRI studies. *Prog Brain Res*. 2006;**156**:249-68.

Wilson SM, Saygin AP, Sereno MI, Iacoboni M. Listening to speech activates motor areas involved in speech production. *Nat Neurosci*. 2004;**7**(7):701-2.

Wilson SM, Iacoboni M. Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *Neuroimage*. 2006;**33**(1):316-25.

Wise RJ, Greene J, Büchel C, Scott SK. Brain regions involved in articulation. *Lancet*. 1999;**353**(9158):1057-61.

Zatorre RJ, Evans AC, Meyer E, Gjedde A. Lateralization of phonetic and pitch discrimination in speech processing. *Science*. 1992;**256**(5058):846-9.

Zatorre RJ, Meyer E, Gjedde A, Evans AC. PET studies of phonetic processing of speech: review, replication, and reanalysis. *Cereb Cortex*. 1996;**6**(1):21-30.

Zatorre RJ, Belin P, Penhune VB. Structure and function of auditory cortex: music and speech. *Trends Cogn Sci*. 2002;**6**(1):37-46.

Appendix A: Figures

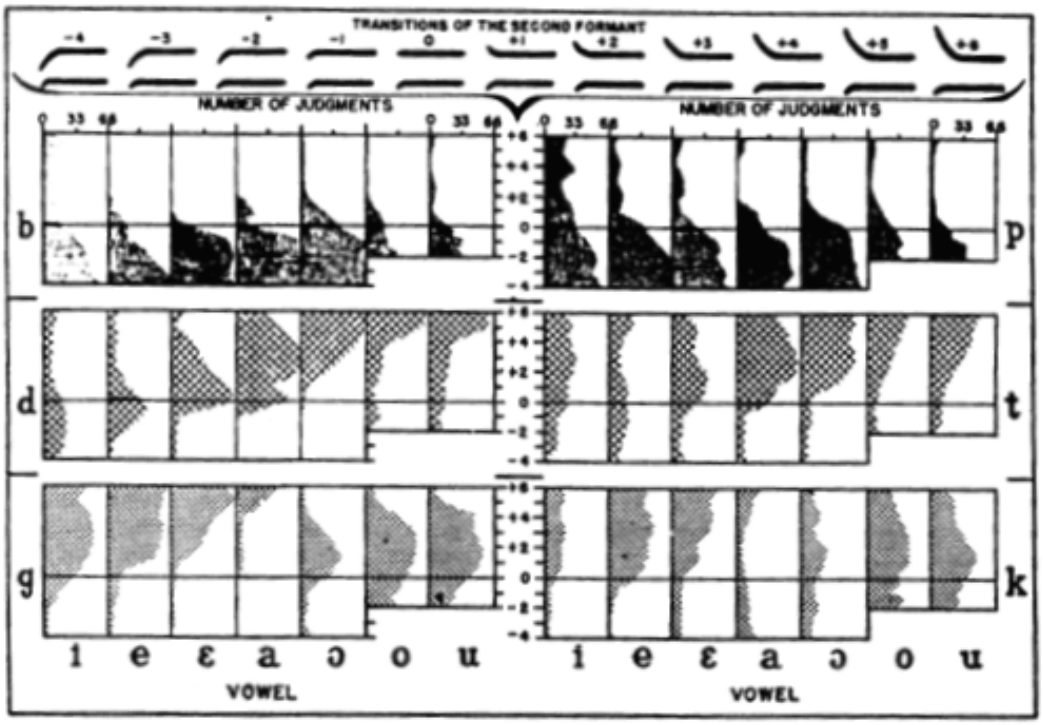


Figure 1. Identification of voiced and unvoiced stop consonants across a range of F1 transitions, from large rising (-4) to large falling (+6). Boxes represent plots of number of subject identifications of each consonant when paired with succeeding vowel (bottom) for various F1 transitions (vertical scales). Horizontal line represents flat (0) F1 transition. Figure is from Figure 3 of Liberman *et al* (1954).

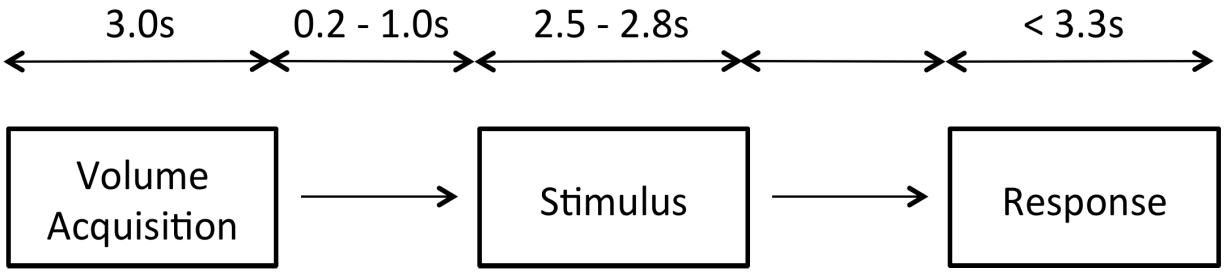


Figure 2: Stimuli and imaging sequence from experiment #2 for a single trial.

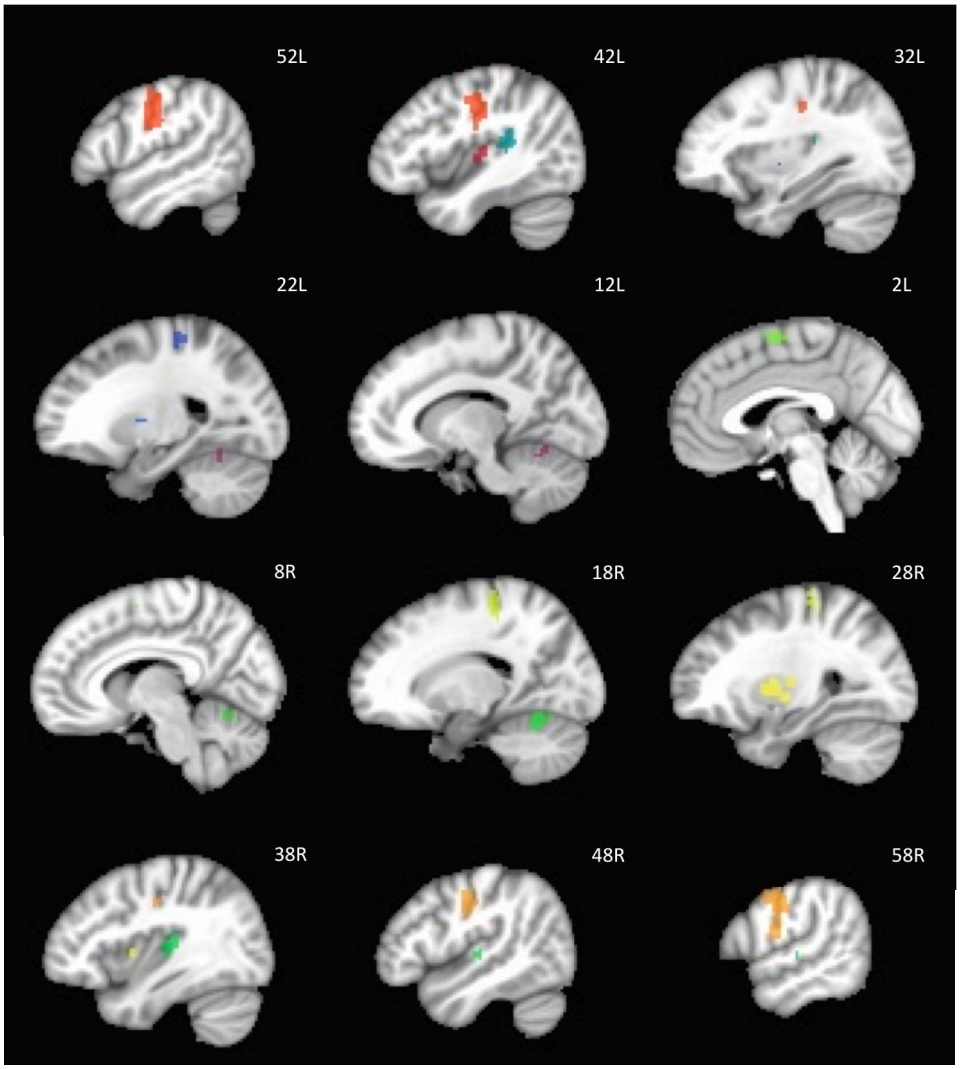


Figure 3: Production mask created by a three-way conjunction of the production tasks from experiment #1. Images are set at $p < 0.005$ uncorrected. Proceeds left to right from L -> R.

*All functional images following are shown in standard MNI space and overlaid onto the MNI 152 non-linearly averaged brain. Right and left hemispheres are on their respective sides for all subsequent images.

** In all subsequent images, positive t-values go from dark orange to light yellow with increasing positivity; negative t-values go from dark blue to light blue with increasing negativity.

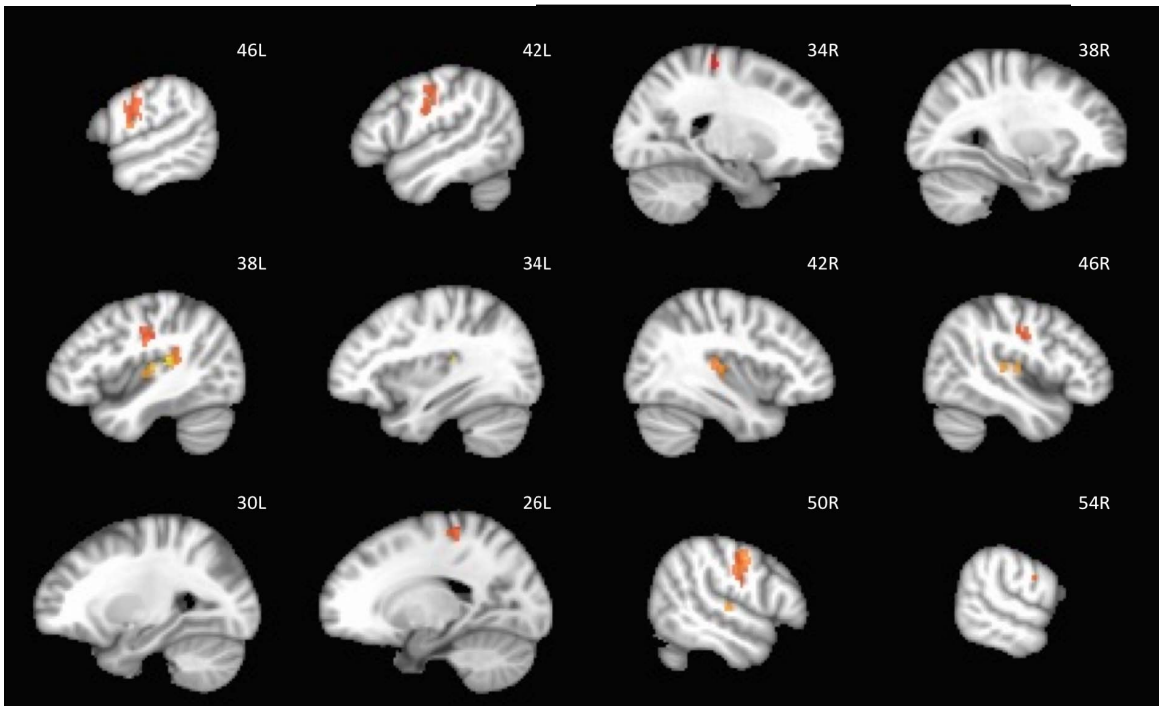


Figure 4a: Production-masked analysis of passive listening condition vs. resting baseline. Images are set at $p < 0.01$ uncorrected, corrected to $p < 0.05$ using a minimum cluster size of 12 voxels.

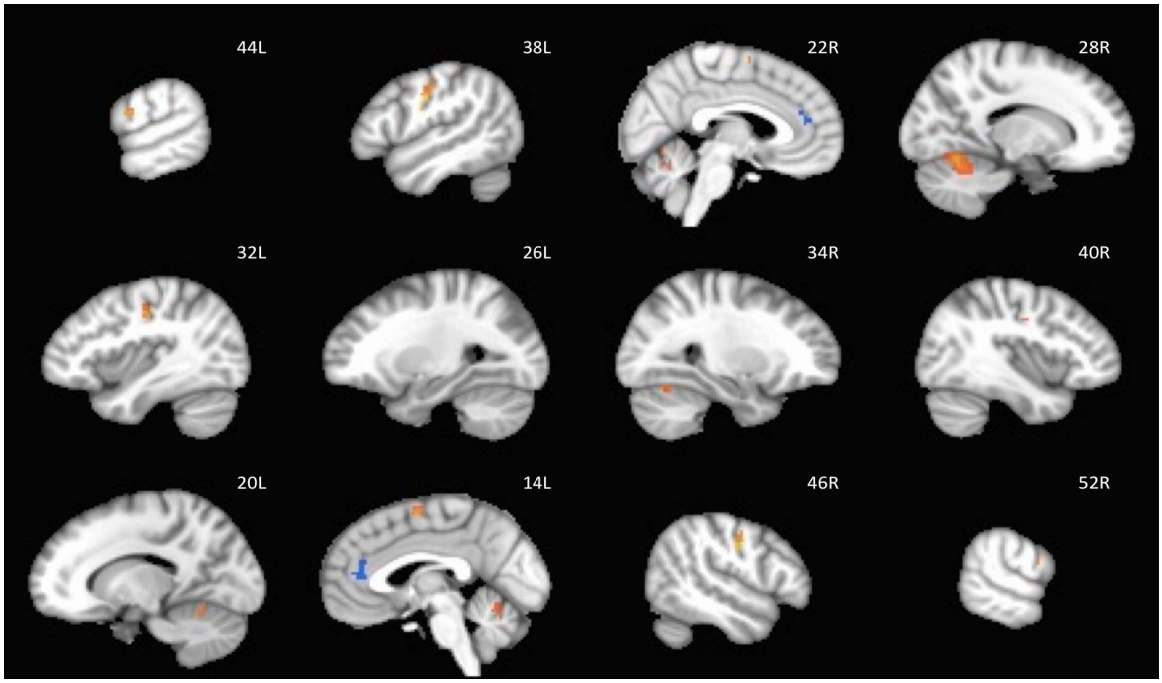


Figure 4b: Contrast analysis of repetition - listening conditions. Images are set at $p < 0.001$ uncorrected, corrected to $p < 0.05$ using a minimum cluster size of 20 voxels.

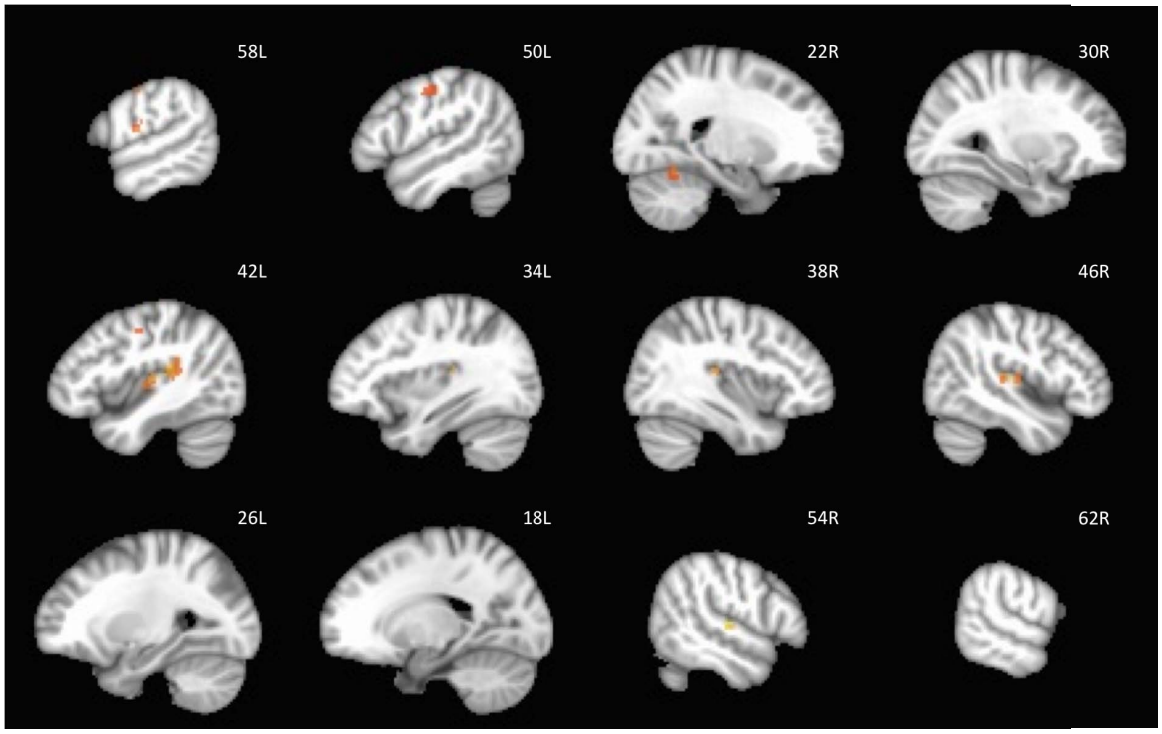


Figure 5a: Production-masked analysis of audiovisual sentence perception vs. resting baseline. Images are set at $p < 0.01$ uncorrected, corrected to $p < 0.05$ using a minimum cluster size of 12 voxels.

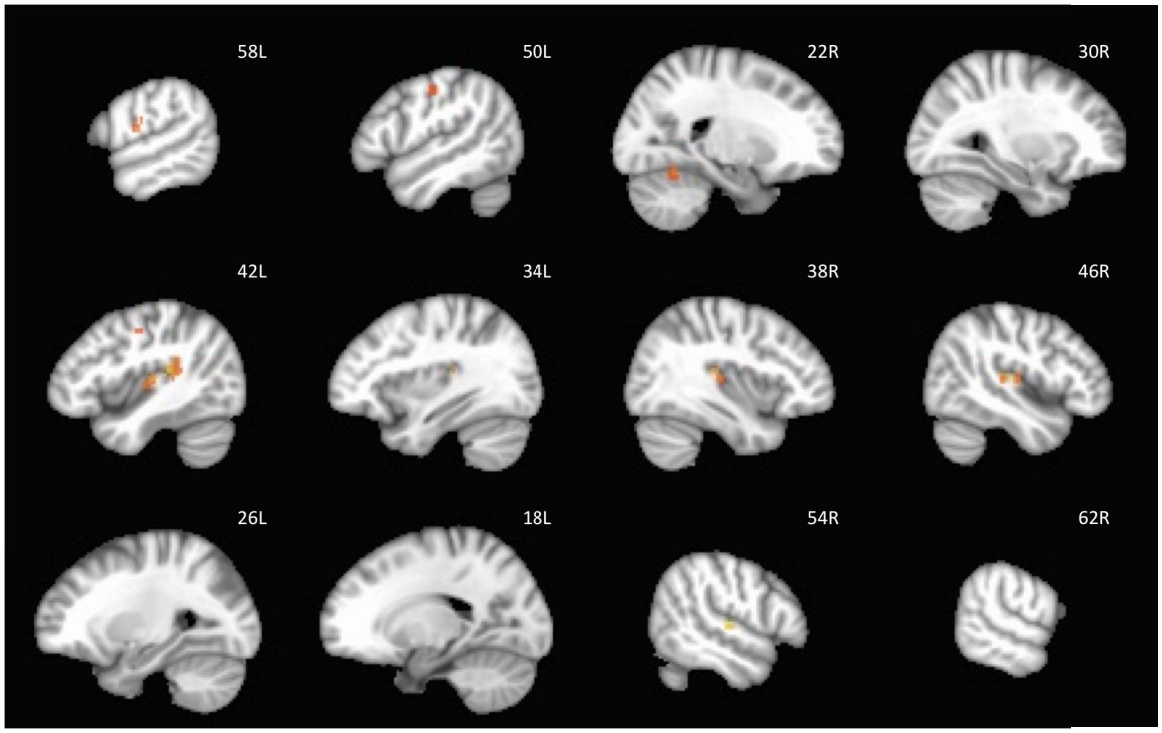


Figure 5b: Production-masked analysis of auditory sentence perception vs. resting baseline. Images are set at $p<0.01$ uncorrected, corrected to $p< 0.05$ using a minimum cluster size of 12 voxels.

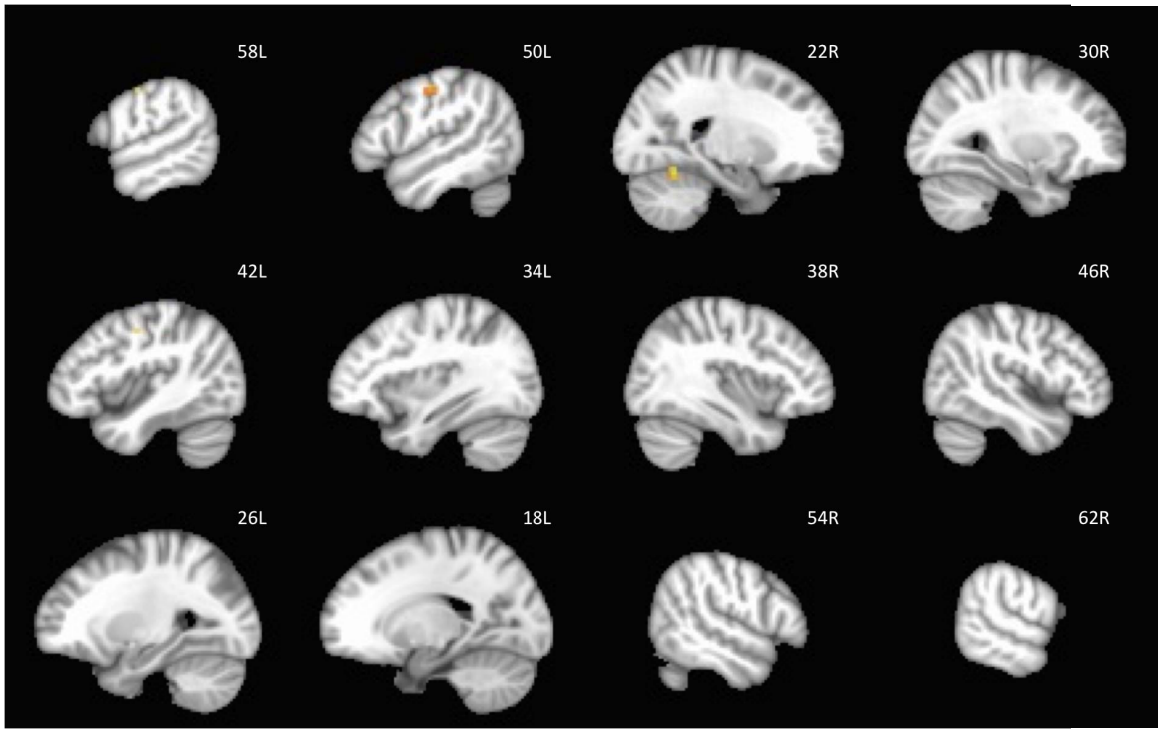


Figure 5c: Production-masked analysis of visual sentence perception vs. resting baseline. Images are set at $p < 0.01$ uncorrected, corrected to $p < 0.05$ using a minimum cluster size of 12 voxels.

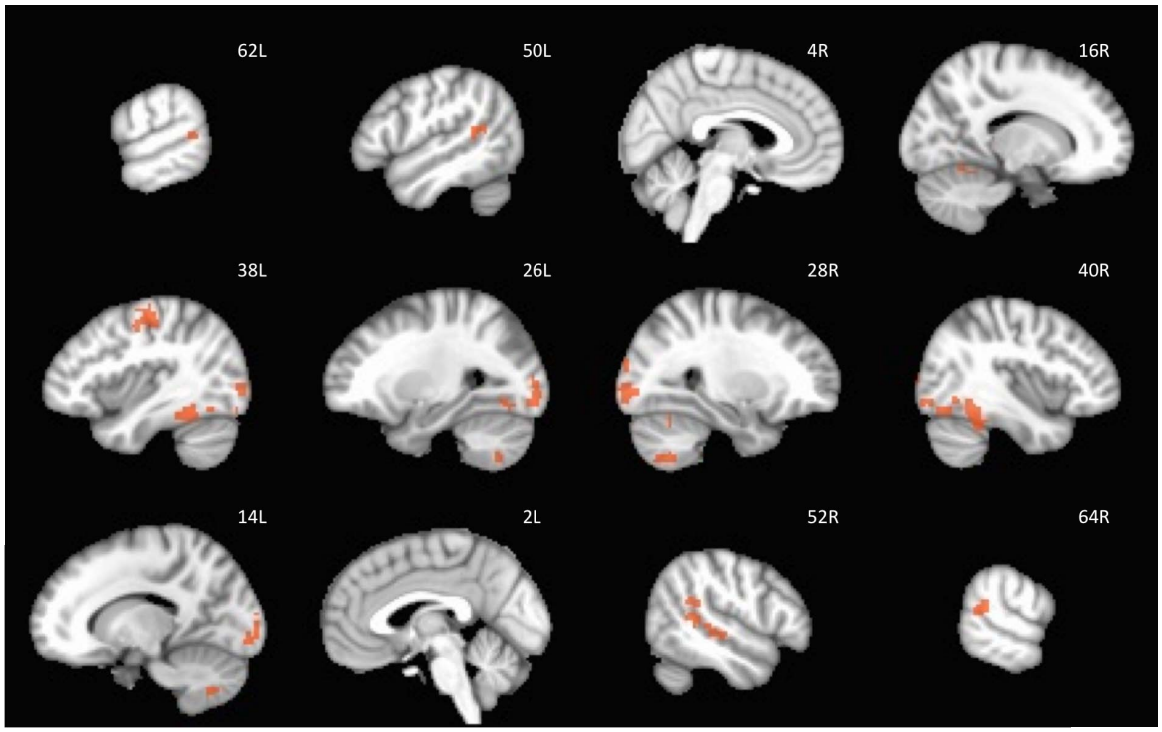


Figure 6: Three-way conjunction of audiovisual, auditory, and visual speech perception vs. resting baseline. Images are set at $p < 0.01$ uncorrected, corrected to $p < 0.05$ using a minimum cluster size of 12 voxels.

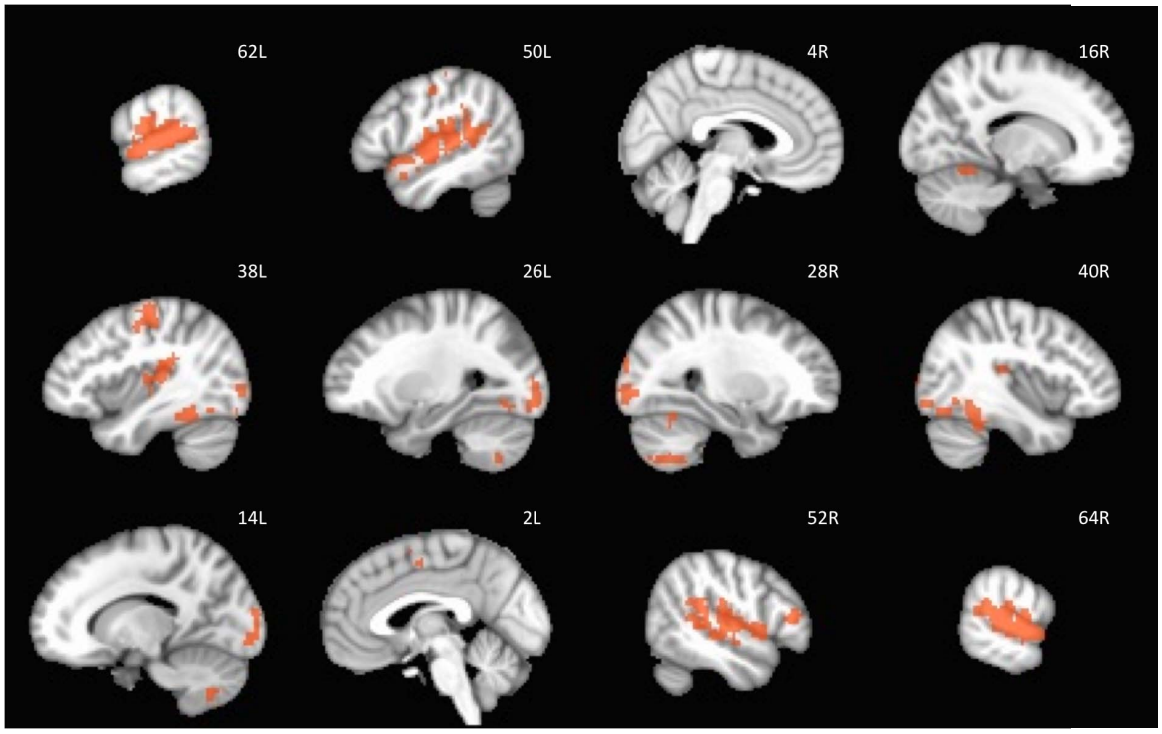


Figure 7: Conjunction of audiovisual and auditory sentence perception vs. resting baseline. Images are set at $p < 0.005$ uncorrected, corrected to $p < 0.05$ using a minimum cluster size of 42 voxels.

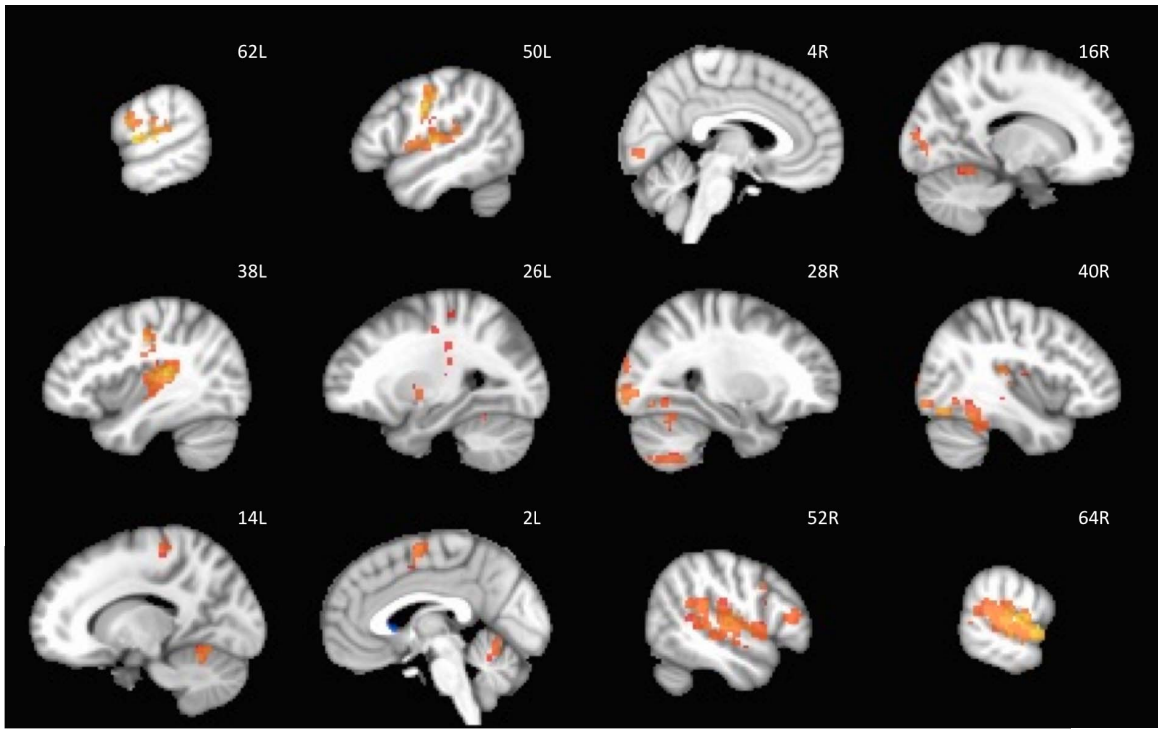


Figure 8: Whole-brain analysis of auditory sentence perception vs. resting baseline. Images are set at $p < 0.005$ uncorrected, corrected to $p < 0.05$ using a minimum cluster size of 42 voxels.

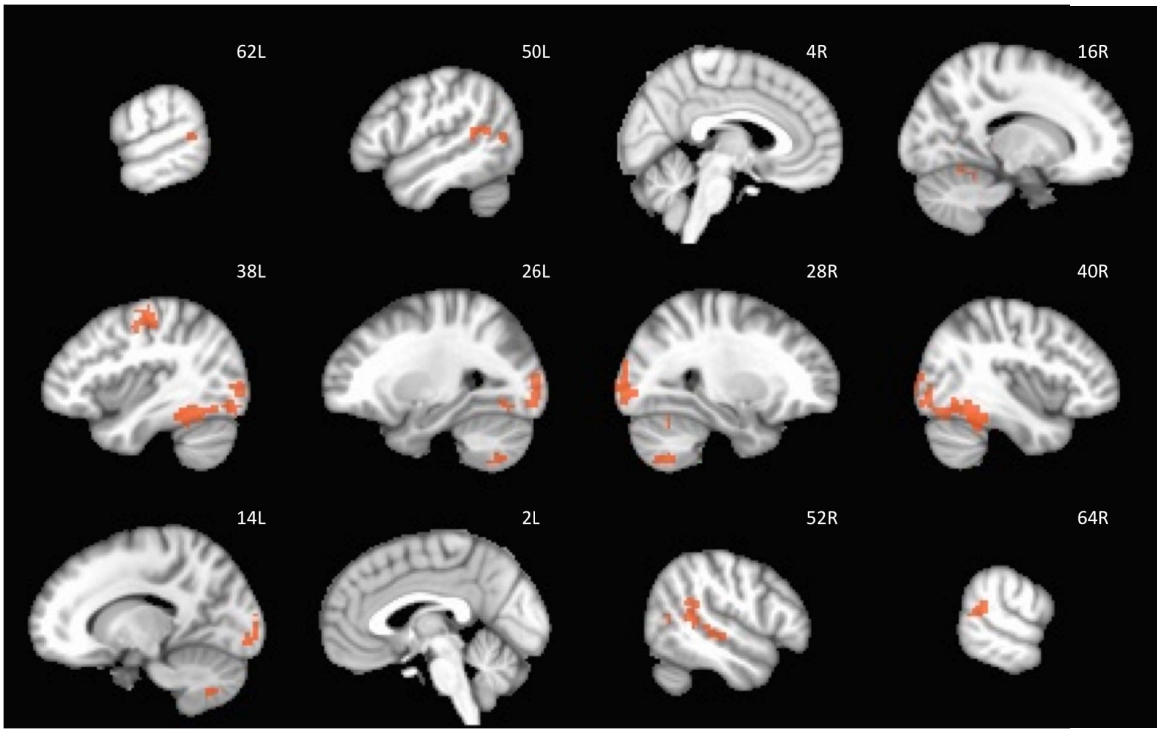


Figure 9: Conjunction of audiovisual and visual sentence perception vs. resting baseline. Images are set at $p < 0.005$ uncorrected, corrected to $p < 0.05$ using a minimum cluster size of 42 voxels.

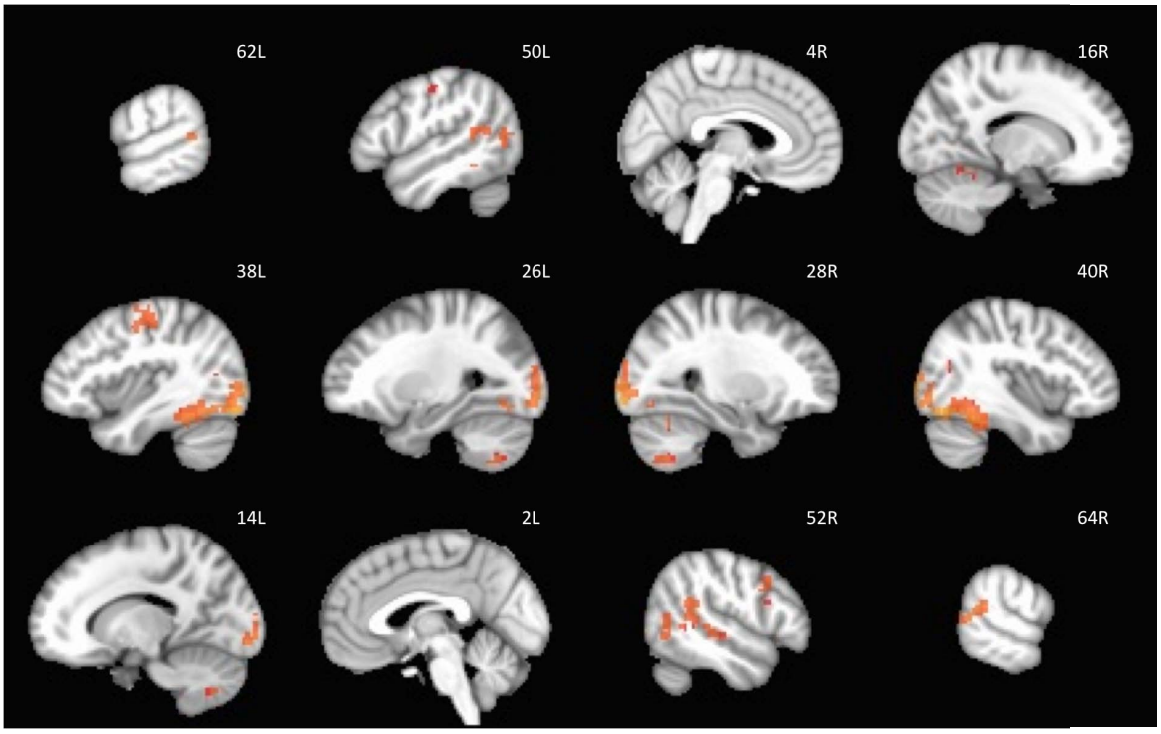


Figure 10: Whole-brain analysis of visual sentence perception vs. resting baseline. Images are set at $p < 0.005$ uncorrected, corrected to $p < 0.05$ using a minimum cluster size of 42 voxels.

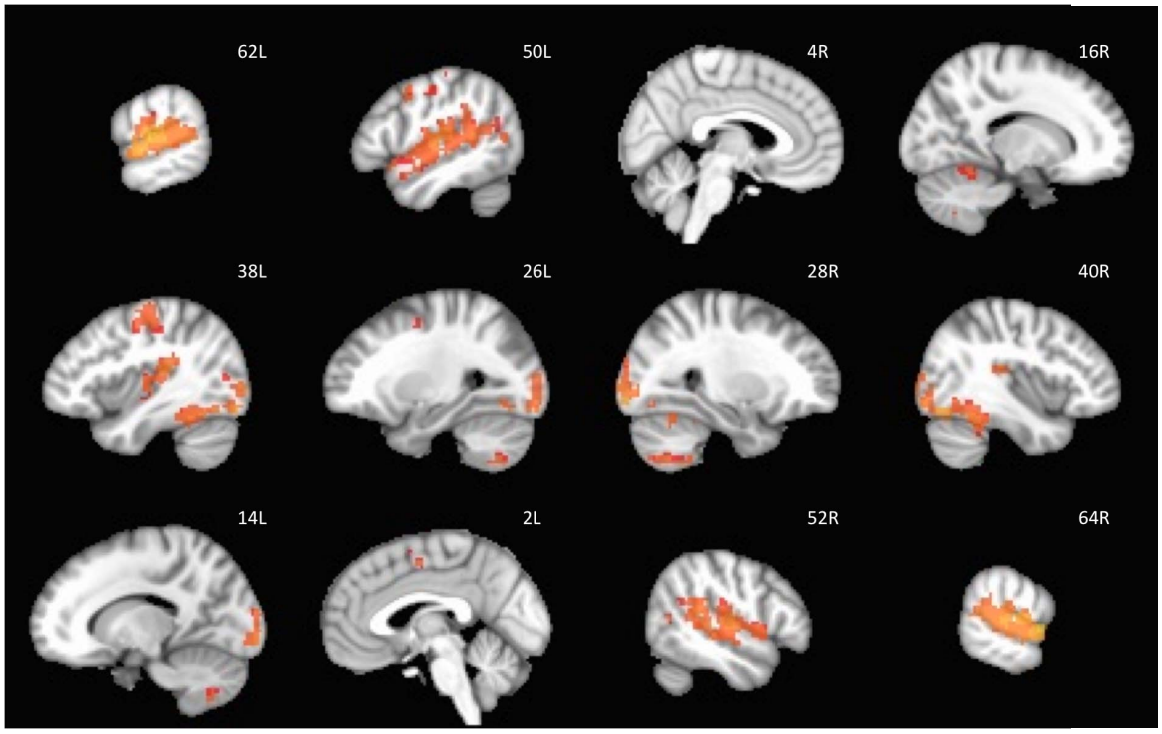


Figure 11: Whole-brain analysis of audiovisual sentence perception vs. resting baseline. Images are set at $p < 0.005$ uncorrected, corrected to $p < 0.05$ using a minimum cluster size of 42 voxels.

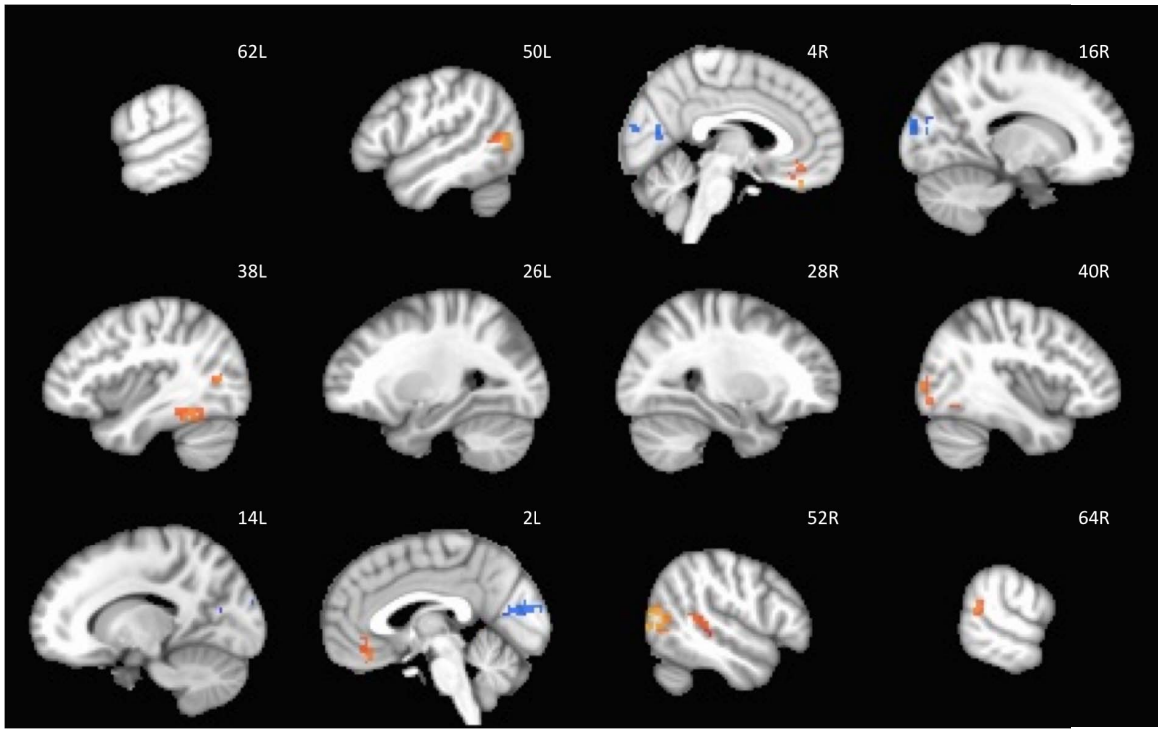


Figure 12: Contrast analysis of audiovisual vs. auditory sentence perception. Images are set at $p < 0.005$ uncorrected, corrected to $p < 0.05$ using a minimum cluster size of 42 voxels. Orange regions represent AV > A; blue represent A > AV.

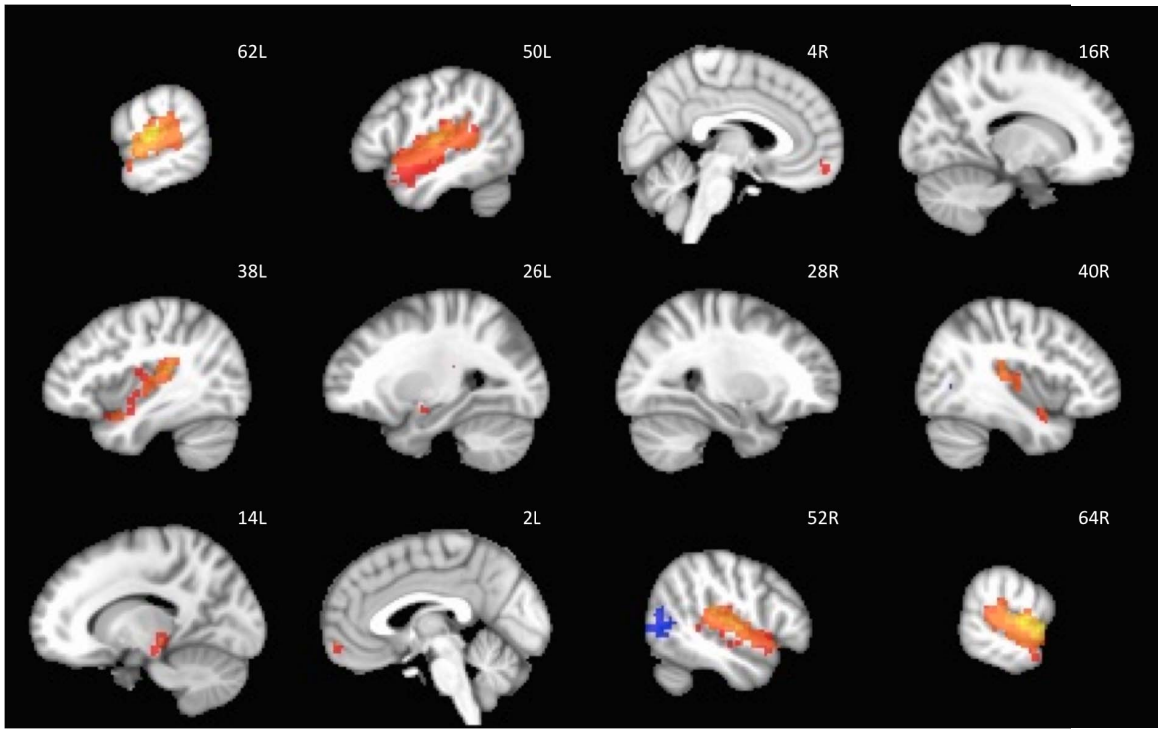


Figure 13: Contrast analysis of audiovisual vs. visual sentence perception. Images are set at $p < 0.005$ uncorrected, corrected to $p < 0.05$ using a minimum cluster size of 42 voxels. Orange regions represent AV > V; blue represent V > AV.

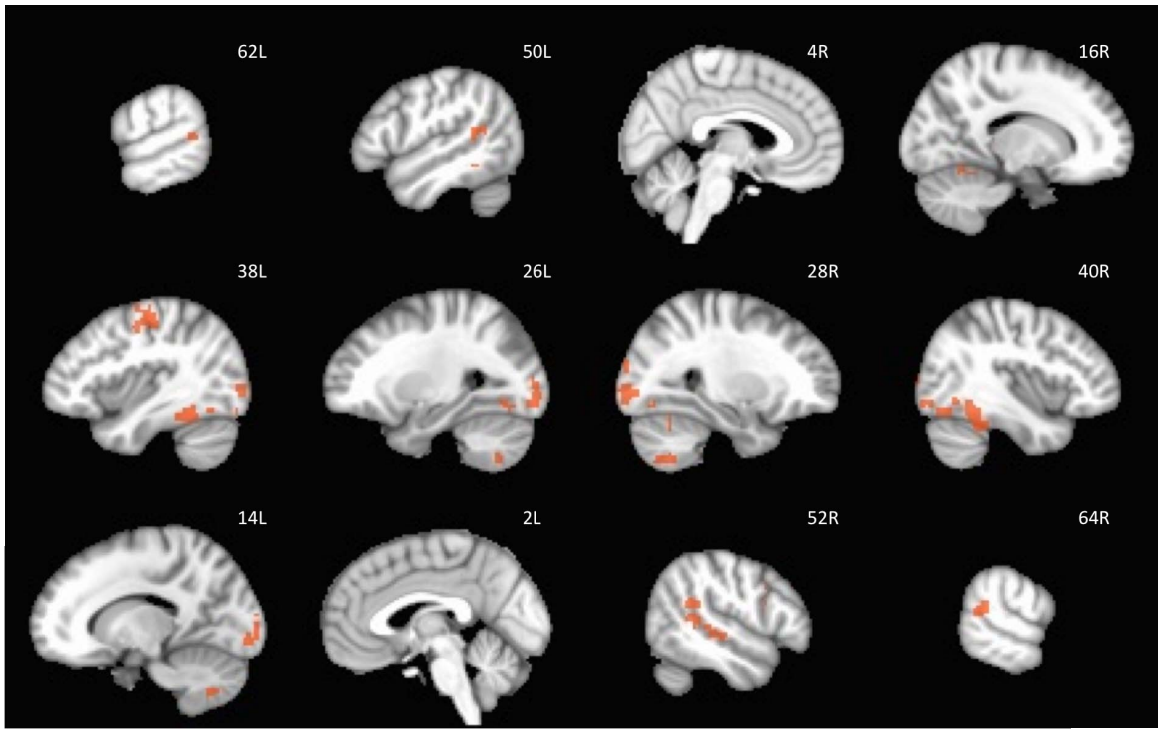


Figure 14: Conjunction of auditory and visual sentence perception vs. resting baseline. Images are set at $p < 0.005$ uncorrected, corrected to $p < 0.05$ using a minimum cluster size of 42 voxels.

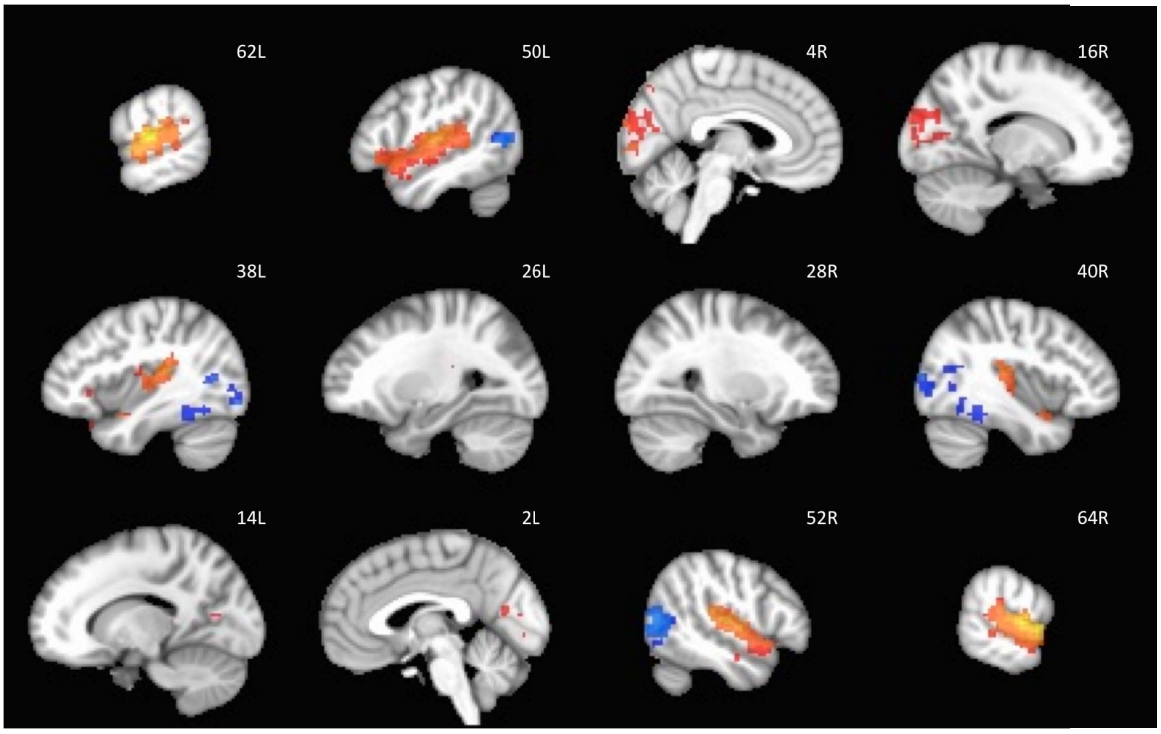


Figure 15: Contrast analysis of auditory vs. visual sentence. Images are set at $p < 0.005$ uncorrected, corrected to $p < 0.05$ using a minimum cluster size of 42 voxels. Orange regions represent A > V; blue represent V > A.

Appendix B: Tables

Table 1: Activation clusters from production mask and production-masked analyses.

	#Voxels*	Centre of Mass			Region	
	------	X	Y	Z		
Production Mask	232	-50.3	10.5	32.7	Ventral pre/post central gyrus	L
	172	53.9	6.6	32.2	Ventral pre/post central gyrus	R
	75	27.6	3.1	4.5	Putamen	R
	70	20.2	28.4	61.3	Dorsal posterior pre/post central gyrus	R
	69	0.3	2.8	62.7	Superior frontal gyrus	R/L
	60	16.2	59.4	-18	Superior cerebellum	R
	56	43	21.7	9.1	Transverse temporal gyrus	R
	50	-39.4	34.1	15.1	Transverse temporal gyrus	L
	34	-25.9	5.6	0.4	Putamen	L
	31	-18.2	29.2	61.7	Dorsal posterior pre/post central gyrus	L
	22	-16.5	58.6	-17.2	Superior cerebellum	L
	20	-43	18	6.5	Transverse temporal gyrus	L
Low-level Perception	150	-52	10.1	31.6	Ventral central sulcus	L
	103	53.6	6.4	36.9	Ventral central sulcus	R
	54	44.1	21.1	8.4	Transverse temporal gyrus, insula	R
	50	-39.2	33.3	14.6	Transverse temporal gyrus	L
	23	18.9	27	62.8	Dorsal pre/post central gyrus	R
	20	-43.2	18.1	6.6	Transverse temporal gyrus	L
	17	-18.4	30.3	61.5	Dorsal pre/post central gyrus	L
AV	39	-39.6	32.9	15.5	Transverse temporal gyrus	L
	30	-50	11.4	45	Inferior dorsal precentral gyrus	L
	22	19.8	57.6	-18	Superior cerebellum	R
	20	50	17.2	5	Transverse temporal gyrus	R
	20	-43.4	18	6.5	Transverse temporal gyrus	L
	14	42.4	25.5	12	Transverse temporal gyrus	R
	13	-58.2	9.2	17.5	Inferior ventral post-central gyrus	L
Auditory	39	46.5	20.5	7.7	Transverse temporal gyrus	R
	38	-39.5	32.8	15.5	Transverse temporal gyrus	L
	23	-48.6	12.5	45.4	Inferior dorsal precentral gyrus	L
	21	19.9	57.8	-17.9	Superior cerebellum	R
	20	-43.2	18	6.5	Transverse temporal gyrus	L
	16	-58.4	8.9	17.2	Inferior ventral post-central gyrus	L
Visual	23	-49.9	11.3	45.3	Inferior dorsal precentral gyrus	L
	21	19.8	57.7	-17.9	Superior cerebellum	R

*All clusters are significant at $p < 0.01$, corrected to $p < 0.05$ using a minimum cluster size of 12 voxels.

Table 2: Whole-brain analyses of audiovisual, auditory, and visual sentence perception

	#Voxels*	Centre of Mass**			Region	H
	------	X	Y	Z		
3-way	177	33.8	55.3	-18.1	Inferior occipital gyrus, fusiform gyrus, inferior temporal gyrus, superior cerebellum	R
conj	146	-26.2	92.5	-2.5	Inferior and middle occipital gyrus	L
	121	56.9	44.1	15.7	Posterior superior temporal gyrus, posterior superior temporal sulcus (trunk, bifurcation and anterior ascending), planum temporale, middle temporal gyrus	R
	104	-34.9	56.4	-16.4	Inferior temporal gyrus, fusiform gyrus	L
	94	30.1	91.1	1.4	Inferior and middle occipital gyrus	R
	55	-40.3	17.9	53.4	Pre/post central gyri	L
	47	-55.5	50	12.4	Posterior superior temporal gyrus, posterior superior temporal sulcus (trunk, bifurcation and posterior ascending), middle temporal gyrus	L
	43	53.7	22.6	-1.7	Superior bank of the superior temporal sulcus (trunk)	R
	41	25.6	63.2	-51.7	Inferior cerebellum	R
AV∧A	1003	-53	23.5	8.8	Superior temporal plane including STp clusters above	L
	857	56.4	21.5	7.7	Superior temporal plane including STp clusters above	R
	200	32.9	54.9	-18.2	Inferior occipital gyrus, fusiform gyrus, inferior temporal gyrus, superior cerebellum	R
	176	-25.2	92.9	-1.4	Inferior and middle occipital gyrus	L
	109	-34.8	56.4	-16.3	Inferior temporal gyrus, fusiform gyrus	L
	108	-42.2	17.6	51.9	Pre/post central gyri	L
	103	29.9	91.1	2.1	Inferior and middle occipital gyrus	R
	62	25.2	63	-51.2	Inferior cerebellum	R
	39	-20.2	64.1	-49.5	Inferior cerebellum	L
	39	-6.8	-2.1	54.4	Superior frontal gyrus	L
	33	52.6	-32.1	9.4	Inferior frontal gyrus pars triangularis	R
AV∧V	379	33.6	68.4	-10.4	Inferior and middle occipital gyrus, fusiform gyrus, inferior temporal gyrus, superior cerebellum	R

	253	-32.2	86.7	-1.1	Middle and inferior occipital gyri, middle and inferior temporal gyri	L
	145	-35.9	55.4	-16.7	Inferior temporal gyrus, lateral occipitotemporal gyrus, fusiform gyrus	L
	144	56.5	44.3	15	Posterior superior temporal gyrus, posterior superior temporal sulcus (trunk, bifurcation and both ascending), planum temporale, middle temporal gyrus	R
	60	-39.9	17.8	53.4	Pre/post central gyri	L
	55	-55.4	51.1	12.8	Posterior superior temporal gyrus, posterior superior temporal sulcus (bifurcation and posterior ascending), middle temporal gyrus	L
	45	-21.6	63.8	-49.7	Inferior cerebellum	L
	44	53.7	22.5	-1.8	Superior bank of the superior temporal sulcus (trunk)	R
	41	25.6	63.2	-51.7	Inferior cerebellum	R
	37	53.7	60.6	11.2	Middle temporal gyrus, posterior superior temporal sulcus (posterior ascending branch)	R
A∧V	193	33.4	56	-17.9	Inferior occipital gyrus, fusiform gyrus, inferior temporal gyrus, superior cerebellum	R
	147	-26.1	92.5	-2.5	Middle and inferior occipital gyri	L
	125	56.7	44.2	15.8	Posterior superior temporal gyrus, posterior superior temporal sulcus (trunk, bifurcation and both ascending), planum temporale, middle temporal gyrus	R
	109	-35.5	55.9	-16.2	Inferior temporal gyrus, fusiform gyrus	L
	95	30.2	91	1.3	Middle and inferior occipital gyrus	R
	57	-40.3	17.5	53.5	Pre/post central gyri	L
	47	-55.5	50	12.4	Posterior superior temporal gyrus, posterior superior temporal sulcus (bifurcation and posterior ascending), middle temporal gyrus	L
	45	25.3	63.6	-51.3	Inferior cerebellum	R
	43	53.7	22.6	-1.7	Superior bank of the superior temporal sulcus (trunk)	R
	36	-20.8	63.9	49.3	Inferior cerebellum	L
	35	7.8	3.7	55.4	Inferior frontal gyrus pars opercularis dorsal	R
AV	1150	-54.2	23.9	8.8	Superior temporal plane, posterior inferior superior temporal sulcus	L
	1031	57.5	22.8	7.6	Superior temporal plane, back across the inferior superior temporal sulcus into the middle temporal gyrus	R
	477	-32.3	76.7	-7.7	Fusiform gyrus, inferior temporal gyrus, middle occipital gyrus, inferior occipital gyrus	L

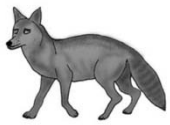
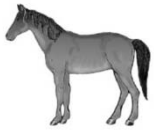
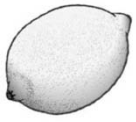
	457	33.9	71.6	-10	Fusiform gyrus, inferior temporal gyrus, middle occipital gyrus, inferior occipital gyrus, cerebellum	R
	206	-43.3	14.2	50.9	Pre/post central gyrus	L
	90	24.4	62.5	-51.2	Inferior cerebellum	R
	56	-22.1	64.2	-50.7	Inferior cerebellum	L
	46	-6.6	-2	53.9	Medial superior frontal gyrus	L
	39	53.2	-31.9	9.6	Inferior frontal gyrus, pars triangularis	R
Audio	1134	-54	22.3	8.8	Superior temporal plane, posterior inferior superior temporal sulcus	L
	917	57.6	20.3	7.4	Superior temporal plane, back across the inferior superior temporal sulcus into the middle temporal gyrus	R
	245	33.2	57.9	-17.6	Fusiform gyrus, inferior temporal gyrus, cerebellum	R
	218	20.5	89.3	2.5	Inferior occipital gyrus, middle occipital gyrus, striate area	R/L
	198	-25.3	92.6	-2.5	Middle occipital gyrus, inferior occipital gyrus, inferior temporal gyrus	L
	131	-33.8	58.1	-16.2	Inferior occipital gyrus, fusiform gyrus, inferior temporal gyrus	L
	125	-42.1	16.7	52.6	Pre/post central gyrus	L
	70	25.1	63.7	-51.5	Inferior cerebellum	R
	53	-7	-2.3	54.8	Medial superior frontal gyrus	L
	48	49.4	-12	28.1	Inferior frontal gyrus, pars opercularis	R
	46	52.1	-32.2	9.6	Inferior frontal gyrus, pars triangularis	R
	45	-20.1	64.2	-50.2	Inferior cerebellum	L
Video	777	41.3	64.7	-2.7	Middle occipital gyrus down and forward to fusiform gyrus, inferior temporal gyrus; up and forward to posterior superior & middle temporal gyrus / posterior superior temporal sulcus	R
	483	-34.2	76.4	-6.7	Middle and inferior occipital gyrus down to fusiform gyrus, inferior temporal gyrus; up and forward to middle temporal gyrus	L
	85	-42.9	16.4	52.1	Pre/post central gyrus	L
	63	-55	52.1	12.2	Posterior superior temporal sulcus	L
	56	48.5	-11.2	31	Inferior frontal gyrus, pars opercularis	R
	47	-21.8	63.8	-50.1	Inferior cerebellum	L

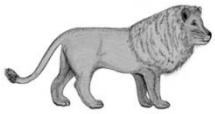
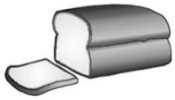
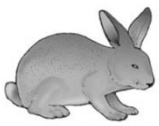
	46	25.4	63.9	-51.8		Inferior cerebellum	R
	44	53.7	22.7	-1.8		Superior bank of the superior temporal sulcus	R
AV-A	158	6	84.8	17.3	A+	Striate area, precuneus	R/L
	142	46.6	72.6	7	AV+	Middle occipital gyrus, inferior temporal gyrus, middle temporal sulcus, posterior middle temporal gyrus	R
	105	-47.5	69.9	3.2	AV+	Middle temporal and inferior temporal gyrus, middle temporal sulcus	L
	87	57.1	40.8	11.9	AV+	Posterior superior temporal gyrus / sulcus anterior ascending branch, trunk	R
	56	-38.2	47.4	-18.4	AV+	Inferior temporal and fusiform gyri	L
	48	0.9	-36.9	-17.4	AV+	Straight gyrus and medial orbital gyrus	R/L
	47	46.2	63.8	-14.4	AV+	Inferior temporal gyrus	R
AV-V	1339	-53.3	18.6	5.2	AV+	Superior temporal plane, temporal pole, laterally to amygdala, anterior insula, anterior middle temporal gyrus posterior superior temporal sulcus trunk	L
	996	56.5	14	3.7	AV+	Superior temporal plane, temporal pole, anterior insula, anterior middle temporal gyrus	R
	100	48	65.6	5.4	V+	Posterior middle temporal gyrus, posterior superior temporal sulcus posterior ascending branch	R
	48	-4.3	-55.1	-11.7	AV+	Straight gyrus and medial orbital gyrus	L/R
	42	-12.5	28.3	-6.4	AV+	Medial geniculate	L
NS	32	-45.9	-26.1	-2.7	AV+	Inferior frontal gyrus, pars triangularis	L
A-V	1148	-53.6	17.4	5.7	A+	Superior temporal plane to inferior frontal gyrus pars triangularis	L
	837	56.3	13.6	4.3	A+	Superior temporal plane	R
	410	46.7	65.7	1.5	V+	Inferior occipital up to middle temporal and down to inferior temporal	R
	350	7.6	85.2	15.4	A+	Striate cortex to precuneus	R
	145	-45.8	71.7	5.3	V+	Middle occipital, middle temporal, inferior temporal gyrus	L
	50	-38.3	51.1	-16.7	V+	Inferior temporal gyrus	L
NS	33	-2.3	-12.5	59	A+	Superior frontal gyrus	L

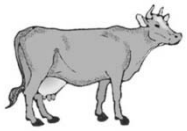
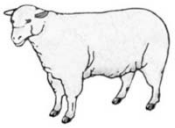
* All clusters are significant at $p < 0.005$, corrected to $p < 0.05$ using a minimum cluster size of 42 voxels. Rows with 'NS' fall below this threshold but are included because of the conservative nature of conjunction analysis.

** All coordinates are given in MNI standard space.

Appendix C: Experiment #1 Stimulus Words

Word	Image	Word	Image
balloon		flag	
barrel		foot	
basket		fox	
bear		hat	
bed		horse	
bell		lamp	
belt		leaf	
boot		leg	
bottle		lemon	

box		lion	
bread		moon	
button		mountain	
cake		needle	
candle		nose	
cat		pen	
chain		pencil	
chair		pepper	
chicken		pipe	
clock		rabbit	
corn		ring	

cow		sheep	
crown		skirt	
desk		tree	
dog		wheel	
fence		window	

Appendix D: Experiment #2 Stimulus Sentences

Affective Stimuli

Happy

I am next in line at the bank
The baker is making bread
Paul took a shower before breakfast
Josh is working on the computer
The star witness was a dentist
Christopher is at work until eight
Meat from a sheep is called mutton
The water is boiling now
The window upstairs is open
Jennifer is eating outside
The blue-jay is sitting on his perch
Natalie's dress is dark purple
All the horses are in the stable
Jane walked down to the corner store
The store is open on Sunday
I just saw a bird over there
The mechanic changes the car's tyres
Kate prefers chocolate to vanilla
Hit the tennis ball against the wall
They painted all the walls yellow
Lisa's birthday is in august
We are crossing over a bridge
I made sure that it was clear
Toby takes cream in his coffee
The dishes are on the counter
The dog slept in the chair all day
She put away all the dishes
Katherine is buying kitty litter
The chairs were put back in their place
My mom is doing crossword puzzles

Zoe turned on the radio
All the worms come out when it rains
The cupboards in the kitchen are red
Francis is at the restaurant
Vanessa is closing up the store
Her pants were hemmed by a tailor
My dad left the book on the table
I watched a programme on TV
Sarah bought all the vegetables
Justine turned on the hot water
The newspaper was in a bag
The blankets on the bed are blue
She photocopied the whole book
Heidi is reading a cookbook
Mark knows how to play the guitar
Melanie closed the gate behind her
We're going to see a movie
Amy is listening to music
A giraffe has a very long neck
I just got a phone call from him now
There's another cat outside
The photographer took the picture
I cooked the spaghetti in the pot
Sally is ironing her skirt
The pan is made of stainless steel
The soup is cooking on the stove
The results were shown on a bar graph
I ordered a gin and tonic
He sat down at the kitchen table
Shari is sitting beneath a tree

Angry

We waved gooodbye as Fred drove away
I brush my teeth at night before bed
Laurel set the table for breakfast
We had hamburgers for dinner
He filled up my tank with gasoline
Kelly's winter coat is quite long
We all took a nap after lunch
Helen made sandwiches for lunch
On our way, we stopped at the market
Ben goes to the gym in the mornings
She wrapped the scarf around her neck
The book is on top of the pile
Dan finished everything on his plate
The orchestra began to play
Shirley asked another question
The doctor walked into the room
Her parents have brunch on Sundays
Daisy poured another cup of tea
It takes an hour to get there
After an hour our time was up
Ron's clothes are neatly packed away
We bought five buns at the bakery
Amelia gave the dog a bath
Lee has never eaten squash before
We knocked on the door and rang the bell
These houses were built with red brick
This book has eleven chapters
Kids can start school at age five
Samantha's fence is painted green
Rachel put in a load of laundry

There are eight sheets of paper left
Valerie went to the library
Jason went to the museum
David watched a hockey game last night
The cake was topped with fruits and nuts
Leslie's new hair cut is quite short
Meredith will be home after six
The cat curled up and went to sleep
Renee is wearing a watch today
We placed our order with the waiter
Diane put all her things away
That girl's hair is very curly
Frankie washed her hands before dinner
Rose read the entire document
This loaf of bread cost two dollars
The small log cabin is empty
Sam is learning to speak German
Rosie peeled the apple with her knife
The big truck stopped at the red light
The bus is due in ten minutes
Courteney rode her bike to the park
All the frogs jumped into the pond
Stephen explored the road
The brown gloves belong to Shelley
Hank can name all the stars in the sky
Mozart wrote many sonatas
Tracy got her hair cut today
We might go to a movie tonight
Heather's shoe laces are untied
Erin wore her sunglasses to work

Linguistic

Question

All the books are alphabetized?
Tom is tightening the loose bolt?
She's feeding the birds day-old bread?
The dealer shuffled the deck of cards?
The hot-dogs are out on the counter?
Beth took a taxi to the dance?
Mark is sitting on Bobby's desk?

Jonas is playing with a dog?
Robin hung the clothes on the line?
Simon is peeling the orange?
The doctor gave John a prescription?
The budgie flew out the window?
During the flight two movies were shown?
Our planet is third from the sun?

The king was seated on his throne?
The toast is burning in the toaster?
They're going to the bank today?
Elisha is rollerskating?
Bridget took the bus home from work?
Tim changed the light bulb yesterday?
The hammer is in the basement?
Josh is riding his bicycle?
There's a red truck in the sandbox?
There are seven people on the bus?
The children are leaving for camp?
We're going to light the candles?
Sean is going to the circus?
He only eats organic food?
They're going to get coffee later?
Caroline turned on all the lights?
He's been in Paris for three months?
We're going to study math now?
Danny is eating a pastry?
Joe is allergic to peanuts?
Jason is eating a persimmon?
They're ordering a cheese pizza?
Max drew that picture on the poster?
Chrissy is completing the puzzle?
You're wearing your new running shoes?
The corner store is closed on Sundays?
The notes are up on the blackboard?
The cupboard in the kitchen is blue?
Marcus is colouring in his book?
Yeast is important for making bread?
Elizabeth is speaking French?
The dog is staying at the kennel?
Sasha is making lemonade?
He lit the fire with a match?
William is reading a novel?
Charles was left by himself last night?
The bears are hibernating now?
George is working on a group project?
Lucille is wearing a new ring?
Jade is sitting on a big rock?
Ted is driving a twelve-wheel truck?
Aiden is chasing his sister?
Joan is eating her spaghetti?
Jennifer is eating sushi?
Andrew is cooking dinner tonight?
Michael is staying home today?

Statement

The hammer is in the basement
Yeast is important for making bread
The children are leaving for camp
He only eats organic food
They're going to get coffee later
Sasha is making lemonade
He lit the fire with a match
He's been in Paris for three months
Charles was left by himself last night
The bears are hibernating now
Danny is eating a pastry
They're ordering a cheese pizza
George is working on a group project
Lucille is wearing a new ring
Jade is sitting on a big rock
Ted is driving a twelve-wheel truck
You're wearing your new running shoes
Aiden is chasing his sister
The corner store is closed on Sundays
Michael is staying home today
The notes are up on the blackboard
The cupboard in the kitchen is blue
Tom is tightening the loose bolt
Marcus is colouring in his book
She's feeding the birds day-old bread
The hot-dogs are out on the counter
Mark is sitting on Bobby's desk
Jonas is playing with a dog
Elizabeth is speaking French
The dog is staying at the kennel
William is reading a novel
We're going to study math now

Simon is peeling the orange
Joan is eating her spaghetti
Our planet is third from the sun
Jennifer is eating sushi
The toast is burning in the toaster
They're going to the bank today
Andrew is cooking dinner tonight
Elisha is rollerskating
All the books are alphabetized
Josh is riding his bicycle
There's a red truck in the sandbox
There are seven people on the bus
We're going to light the candles
The dealer shuffled the deck of cards
Sean is going to the circus
Beth took a taxi to the dance
Caroline turned on all the lights
Robin hung the clothes on the line
Joe is allergic to peanuts
Jason is eating a persimmon
Max drew that picture on the poster
The doctor gave John a prescription
Chrissy is completing the puzzle
The budgie flew out the window
During the flight, two movies were shown
The king was seated on his throne
Bridget took the bus home from work
Tim changed the light bulb yesterday