



## Review

## The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception



Jeremy I. Skipper<sup>a,\*</sup>, Joseph T. Devlin<sup>a</sup>, Daniel R. Lametti<sup>a,b</sup>

<sup>a</sup> Experimental Psychology, University College London, United Kingdom

<sup>b</sup> Department of Experimental Psychology, University of Oxford, United Kingdom

## ARTICLE INFO

## Article history:

Received 17 July 2016

Accepted 24 October 2016

## Keywords:

Articulation

Complex network

Phoneme

Neuroimaging meta-analysis

Speech production

## ABSTRACT

Does “the motor system” play “a role” in speech perception? If so, where, how, and when? We conducted a systematic review that addresses these questions using both qualitative and quantitative methods. The qualitative review of behavioural, computational modelling, non-human animal, brain damage/disorder, electrical stimulation/recording, and neuroimaging research suggests that distributed brain regions involved in producing speech play specific, dynamic, and contextually determined roles in speech perception. The quantitative review employed region and network based neuroimaging meta-analyses and a novel text mining method to describe relative contributions of nodes in distributed brain networks. Supporting the qualitative review, results show a specific functional correspondence between regions involved in non-linguistic movement of the articulators, covertly and overtly producing speech, and the perception of both nonword and word sounds. This distributed set of cortical and subcortical speech production regions are ubiquitously active and form multiple networks whose topologies dynamically change with listening context. Results are inconsistent with motor and acoustic only models of speech perception and classical and contemporary dual-stream models of the organization of language and the brain. Instead, results are more consistent with complex network models in which multiple speech production related networks and subnetworks dynamically self-organize to constrain interpretation of indeterminate acoustic patterns as listening context requires.

© 2016 Published by Elsevier Inc.

## Contents

1. Introduction	78
1.1. Rethinking the question	79
1.1.1. Complexity of speech production	79
1.1.2. Dynamics of speech perception	80
1.2. Towards more nuanced questions	80
2. Qualitative review	81
2.1. Speech production	81
2.2. Behaviour	82
2.3. Computational modelling	83
2.4. Nonhumans	83
2.5. Damage and disorders	84
2.6. Stimulation and recording	85
2.6.1. Specificity	85
2.6.2. Dynamics	86
2.7. Neuroimaging	86
2.7.1. Specificity	86
2.7.2. Mechanisms	87
2.7.3. Dynamics	88

\* Corresponding author at: University College London, Experimental Psychology, 26 Bedford Way, London WC1H 0AP, United Kingdom.

E-mail address: [jskipper@lab-lab.org](mailto:jskipper@lab-lab.org) (J.I. Skipper).

3.	Quantitative review	88
3.1.	Overview	88
3.1.1.	Hypothesis 1: Regional overlap	88
3.1.2.	Hypothesis 2: Network organization	89
3.1.3.	Hypothesis 3: Dynamic network topology	89
3.2.	Materials and methods	89
3.2.1.	Common criteria	89
3.2.2.	Hypothesis 1: Regional overlap	89
3.2.3.	Hypothesis 2: Network organization	90
3.2.4.	Hypothesis 3: Dynamic network topology	91
3.3.	Results	93
3.3.1.	Hypothesis 1: Regional overlap	93
3.3.2.	Hypothesis 2: Network organization	95
3.3.3.	Hypothesis 3: Dynamic network topology	95
4.	Discussion	95
4.1.	Summary	95
4.1.1.	Qualitative review	95
4.1.2.	Quantitative review	95
4.2.	Analysis-by-synthesis	96
4.2.1.	Overview	96
4.2.2.	Why speech production systems?	97
4.2.3.	Is it necessary?	97
4.3.	Neurobiological models	98
5.	Conclusions	98
	Acknowledgments	98
	Appendix A	99
A.1.	Common search criteria	99
A.2.	Natural (“passive”) listening to nonwords and words	99
A.3.	Speech production and silent articulation	99
A.4.	Network analyses	99
	References	99

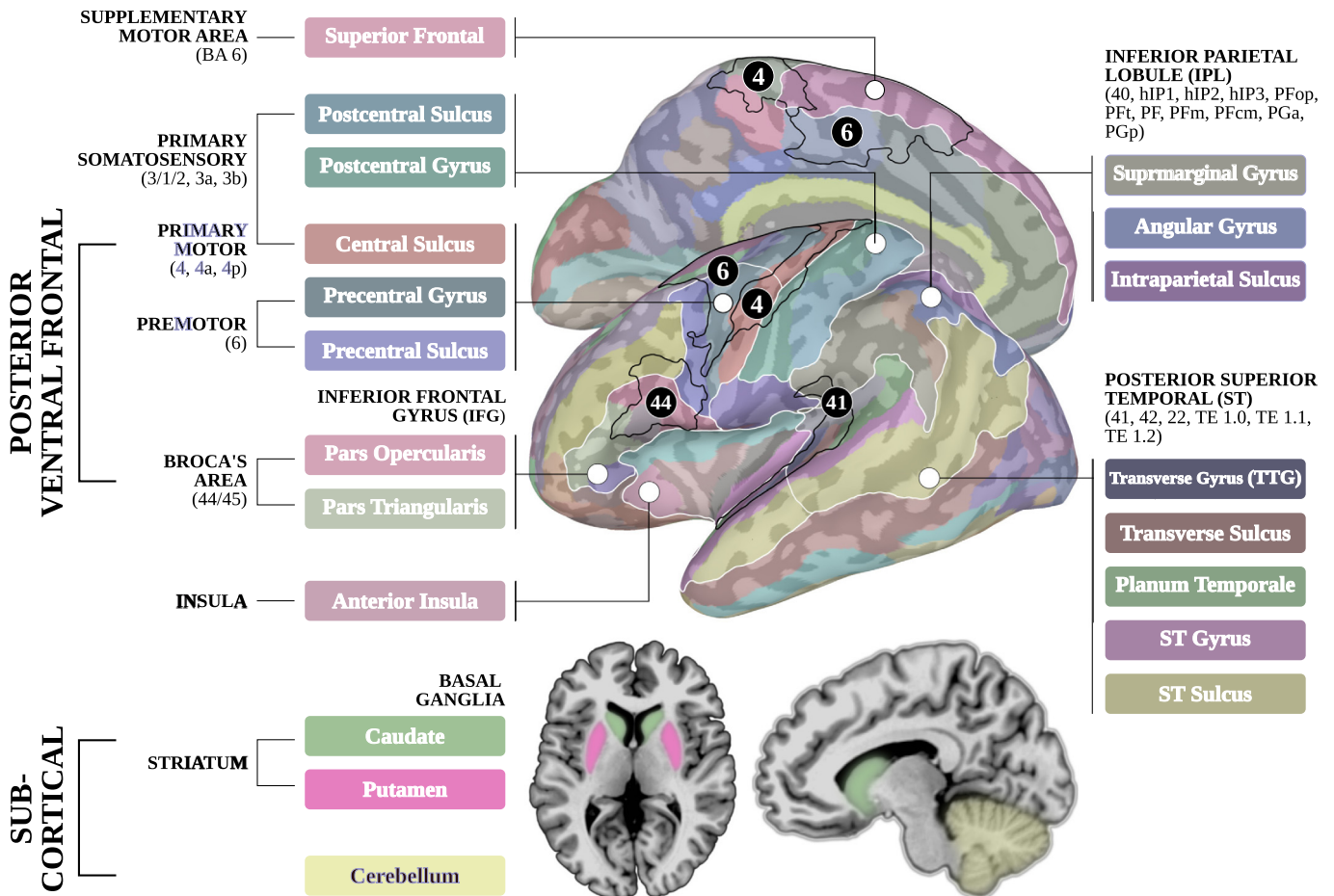
## 1. Introduction

Why does it matter if “the motor system” or, specifically here, brain regions supporting speech production, play a role in speech perception? It matters because, after decades of research, we still do not know how we perceive speech sounds even though this behaviour is fundamental to our ability to use language. One hindrance to this understanding has been an inability to specify how we hear sounds as particular (and putative) speech categories like phonemes or syllables. Indeed, no acoustic features have been found that can uniquely and consistently be used to characterize those units (Appelbaum, 1996; Appelbaum, 1999; Goldinger & Azuma, 2003; Port, 2010/1). Speech production systems matter in this context because it has long been proposed by theoretical models of speech perception that this problem of acoustic indeterminacy – or “lack of invariance” – can be addressed by making reference to the motor system. In particular, the motor theory of speech perception proposed that “sounds are not the true objects of perception ... rather, they only supply the information for immediate perception of the gestures” (Liberman & Mattingly, 1985). These gestures are “represented in the brain as invariant motor commands that call for movements of the articulators” and involve a “perception-production link [that] is a necessary condition for recognizing speech as speech” (Liberman & Mattingly, 1985). In contrast, the “analysis-by-synthesis” (AxS) model proposed that the motor system assists perception by providing production-based constraints on the interpretation of acoustic patterns as needed (Bever & Poeppel, 2010; Poeppel & Monahan, 2011; Skipper, Nusbaum, & Small, 2006; Stevens & Halle, 1967).

If either of these models were accurate, there are implications not only for our theories of speech perception, but also our

understanding of the organization of language in the brain. That is, though neither model is neurobiologically well specified, both can be used to make inferences about the brain basis of language. The motor theory of speech perception suggests that the motor system needs to play a role in the neurobiology of speech perception. In contrast, the AxS model suggests that speech perception is more distributed in the brain with the motor system contributing dynamically in an active, constructive, or predictive manner. Neither model is consistent with “textbook” or “classical” models of the organization of language in the brain because speech production and perception are presented as separable neurobiological processes in those models, with production occurring as a result of processing in Broca’s area and comprehension as a result of processing in Wernicke’s area (a model still taught to medical students; Geschwind, 1970).

Both models also seem to be inconsistent with the most cited contemporary model of the organization of language and the brain, the “dual-stream” model of speech and language (Hickok & Poeppel, 2007; Poeppel & Hickok, 2004). Though the “dorsal stream” in this model is proposed to support “sound to action”, it does “not appear to be a critical component of the speech perception process” (Poeppel & Hickok, 2004). As support, Hickok, Costanzo, Capasso, and Miceli (2011) analysed patients with damage to Broca’s area and concluded that the “motor speech system is not necessary for speech perception” (p. 214). Other scientists seem to support this view. For example, in their review, Scott, McGettigan, and Eisner (2009) state that “the motor cortex is not essential for perceiving spoken language” (p. 301) while Lotto, Hickok, and Holt (2009) posit that “there is no need to think that this interaction [with the production system] would be required of normal speech perception” (p. 3). In other words, we are left with a model of speech perception in the brain in which the motor



**Fig. 1.** Brain regions discussed in this review. Lateral and medial surface renderings of the left hemisphere (top) and individual axial and sagittal slices (bottom) showing the location of most of the distributed set of regions involved in both speech production and perception. The white outlines encompass gross anatomical zones (e.g., posterior ventral frontal regions). The actual anatomical regions comprising those zones are defined by colour. Within any given anatomical region, darker shades of a colour correspond to gyri and lighter shades to sulci. The functional synonyms often given to those regions or groupings of regions are provided. Approximate cytoarchitecture of these regions is defined parenthetically. White numbers refer to actual probabilistic cytoarchitectonic regions (given in Brodmann's areas) and the black outlines indicate the boundaries of those regions.

system is not “critical”, “necessary”, “essential”, or “required” for perceiving speech. This position has generated some debate (Hickok, Holt, & Lotto, 2009; Lotto et al., 2009; Wilson, 2009) and has some opponents (Fischer & Zwaan, 2008; Pulvermüller & Fadiga, 2010). There are, however, no comprehensive reviews of the role or roles of the motor system in speech perception that might provide a countervailing position. This review fills this knowledge gap by demonstrating that, inconsistent with these negative positions and more consistent with AxS like models, neurobiologically distributed speech production systems play ubiquitous, specific, and context dependent roles in normal speech perception (Bever & Poeppel, 2010; Pickering & Garrod, 2013; Poeppel & Monahan, 2011; Pulvermüller & Fadiga, 2010; Schwartz, Basirat, Ménard, & Sato, 2012/9; Skipper, 2015; Skipper et al., 2006).

### 1.1. Rethinking the question

Before conducting the review, however, it is worth carefully deconstructing the question that has been discussed in the literature: What role does the motor system play in speech perception, if any? This and similarly phrased questions have been the subject of philosophical discussion and empirical investigation since at least the 17th century. For example, the French philosopher

Géraud de Cordemoy wrote in (1668), “...whensoever any sound agitates the Brain, there flow immediately spirits towards the Muscles of the Larynx, which duly dispose them to form a sound altogether like that, which was just now striking the Brain.” We suggest, however, that the question and indeed the entire debate is misleading due to the complexity of the neurobiology of speech production and the dynamic nature of speech perception.

#### 1.1.1. Complexity of speech production

When scientists refer to the “motor speech system,” what do they mean? This term is rarely defined in the speech perception literature and neither explicit nor implicit definitions correspond to current theory and research on motor and speech production systems. Very often the motor system is discussed *only* in reference to Broca's area, presumably because the classic neurological model of language associates speech production deficits with damage to this region (e.g., Geschwind, 1970). When Broca's area is defined, it is usually said to comprise the pars triangularis, roughly Brodmann's cytoarchitectonic area (BA) 45 and pars opercularis, roughly BA 44 (though see Amunts et al., 1999) in the posterior inferior frontal gyrus (IFG). This focus on Broca's area is odd given that research shows that “Broca's aphasia” is often not caused by damage to those regions, and, when there is damage there, it also extends well beyond them, e.g., to include the insula, white matter tracts, and

even Wernicke's area (Dronkers, Plaisant, Iba-Zizen, & Cabanis, 2007; Fridriksson, Fillmore, Guo, & Rorden, 2015). Indeed, Broca's area seems to be fairly silent during the actual production of speech and plays more of a coordinative role prior to articulation (Flinker et al., 2015). When regions other than Broca's area are included in reference to the motor system they are usually limited to ventral premotor and primary motor cortices in the precentral gyrus and sulcus and the anterior half of the central sulcus (BAs 6 and 4). Henceforth, we will generally refer to this entire set of regions as "posterior ventral frontal" (Fig. 1).

The focus on posterior ventral frontal cortex is problematic because it is inconsistent with research on motor systems. That is, there is no monolithic motor system limited to such regions. Indeed, it would be difficult to find a suitable definition of the phrase "motor system" because the motor system is not *one* system (Rowe & Siebner, 2012). Rather, there are motor systems, each supporting specific ethological behaviours such as "the facial motor system" (Cattaneo & Pavesi, 2014; Graziano, 2016). Regions participating in any one of these motor systems inevitably involve a distributed set of coordinated regions that have been collectively called the "extended motor system" (Rowe & Siebner, 2012). Extended motor systems are "sensorimotor" systems, indicating that they include non-classically defined motor cortex and the view that production and perception of movement are inseparable (Lametti & Watkins, 2016; Ostry & Gribble, 2016). Indeed, early "sensory" regions like auditory cortex have multisensory response properties and connections from motor and somatosensory regions and, conversely, "motor" regions have multisensory perceptual properties (Ghazanfar, 2009; Graziano, Hu, & Gross, 1997; Nelson & Mooney, 2016; Nelson et al., 2013; Rizzolatti & Craighero, 2004; Schneider, Nelson, & Mooney, 2014).

These general principles apply equally to the specific case of speech production. Speech production is itself not a monolithic process but, rather, a complex set of processes that unfold in parallel (see Melinger, Branigan, & Pickering, 2014 for a review). It has even been argued that "speech production is very likely the most complicated motor act performed by any species" (Perkell et al., 2001, p. 1). For example, Price, Crinion, and Macsweeney (2011) characterize the subprocesses that comprise speech production as including:

... conceptualization of the intended message, word retrieval, selection of the appropriate morphological forms, sequencing of phonemes, syllables, and words, phonetic encoding of the articulatory plans, initiation, and coordination of sequences of movements in the tongue, lips, and laryngeal muscles that vibrate the vocal tract, and the control of respiration for vowel phonation and prosody. In addition... auditory, and somatosensory processing of the spoken output is fed back to the motor system for online correction of laryngeal and articulatory movements... (p. 1)

The many subprocesses of speech production involve a distributed set of sensorimotor brain regions that go well beyond posterior ventral frontal regions. To give one example, in their review, Indefrey and Levelt (2004) enumerate a number of subprocesses involved in speech production and associated regions, claiming that one such subprocess, "phonetic encoding and articulation", is associated with *seventeen* different brain regions.

To summarize, acknowledging the distributed sensorimotor nature of motor systems generally and speech production systems specifically leaves open the possibility that "the motor system" plays an important role or roles in speech perception despite conclusions otherwise. This is because speech production systems extend well beyond posterior ventral frontal regions and any of these other regions could play a role in perception.

Afterall, they are *sensorimotor* processes, perhaps different from perception only as a function of the timing of when components of the system are engaged. If we take seriously the complexity of speech production, any one production subprocess will involve its own distributed set of brain regions that might each play a role in speech perception and do so at different times, in different contexts.

### 1.1.2. Dynamics of speech perception

Like "the motor system", what is meant by "speech perception" is typically ill defined. It is often discussed in the neurobiological literature as if it is a static operation, the result of which are minimal categorical units of speech analysis, phonemes or syllables, from which we can then build words and put those words into sentences. This assumption is reflected in the way speech perception and the brain is studied using primarily isolated speech sounds like "da" and "ba", the latter only likely to be naturally heard in a meadow (see Skipper, 2015). However, there is much debate about what even constitutes "the" unit of analysis in speech (Goldinger & Azuma, 2003). Port (2010/1) concluded that units like the phoneme are only a descriptive by-product of literacy and have no psychological reality (Bybee & McClelland, 2005; see also Lotto & Holt, 2000). Rather, speech is likely an interactive processes between putative levels of linguistic analysis and non-linguistic information that might not be separable into discrete units (see Onnis & Spivey, 2012). What we call units of analysis may actually be "self-organizing dynamic states" (Goldinger & Azuma, 2003). This might occur through an adaptive window of perceptual analysis that expands and contracts to the demands of the listening situation (Nusbaum & Henly, 1992). That is, units are ad-hoc and usage based and can be of various sizes as context dictates, including even larger "chunks" and "prefabs" (Beckner et al., 2009).

Thus, speech is not simply a process by which minimal categorical units are specified because there seems to be a lack of a stable units to specify. This is related to the aforementioned "lack of invariance" problem of speech perception. Variance includes many-to-one mappings in which different acoustic patterns give rise to the perception of the same sound and one-to-many mappings where one acoustic pattern gives rise to different percepts. As mentioned, no acoustic features have been found that invariably specify any given acoustic pattern as a particular hypothetical speech category like a phoneme (Appelbaum, 1996, 1999; Goldinger & Azuma, 2003). This is perhaps not particularly surprising if listeners have a usage based and adaptive window that changes with context and there are, therefore, no stable units to invariably specify.

To summarize, recognizing the dynamic and contextually determined nature of speech also leaves open the possibility that "the motor system" plays an important role in speech perception despite conclusions otherwise. This is because different brain regions might dynamically contribute to speech perception as a function of the context available to listeners. Integrating this discussion with the previous section, the distributed set of sensorimotor brain regions associated with subprocesses of speech production might contribute at different times as context dictates (e.g., when discourse context is not yet available to constrain variance at the start of a conversation). Unfortunately, most neuroimaging speech and language stimuli do not include any or much context to permit evaluation of this claim (Skipper, 2014; see Skipper, 2015 for discussion; Skipper et al., 2006).

### 1.2. Towards more nuanced questions

Acknowledging the complexities of speech production and supporting neurobiological processes and the dynamic and contextually determined nature of speech perception requires us



to ask more nuanced questions than those currently addressed. A better question than “What role does the motor system play in speech perception?” might be, “What roles do sensorimotor networks supporting speech production play in speech perception and when?”. This involves more specific “where”, “how”, and “when” questions. Though none of these categories are independent, “where” refers to questions about implementation. That is, which speech production associated brain regions and networks are involved in speech perception? “How” pertains to questions about the nature of the representations, computations, algorithms, and/or mechanisms implemented by these regions and networks. Are computations specifically associated with speech production or are they unrelated and *de novo*? Is a predictive mechanism employed across some networks as suggested by the AxS model? Finally, “when” refers to questions about the dynamics of the regions and networks involved. Are they always engaged? Are some networks engaged more by some stimuli, tasks, or contexts?

Here we address these “where”, “how”, and “when” questions. We do so by first conducting a qualitative review of the literature that spans a large number of converging methods. To preview, a diverse body of research confirms that regions and networks involved in producing speech are also involved in perceiving speech (“where”). Furthermore, at least some of the representations, computations, and mechanisms associated with these regions and networks are specific to their role in speech production (“how”) and their involvement varies in a dynamic contextually determined way (“when”). Though the qualitative review is extensive, it was not possible to cover all methods equally. Thus, we focused most on adult human data. Most notably missing is a *separate* developmental section despite a burgeoning body of research suggesting the importance of sensorimotor processes in the development of speech perception (Bruderer, Danielson, Kandhadai, & Werker, 2015; Dehaene-Lambertz, Hertz-Pannier, Dubois, & Dehaene, 2008; Guellai, Streri, & Yeung, 2014; Kuhl, 2010; Kuhl, Ramirez, Bosseler, Lin, & Imada, 2014; Perani et al., 2011; Werker & Tees, 1999).

We then outline a number of possible sources of bias in the qualitative review that warrants a less biased and more quantitative review. We then conduct a series of neuroimaging meta-analyses to explicitly test “where”, “how”, and “when” hypotheses suggested by the qualitative review. This work confirmed conclusions from that review, demonstrating a *ubiquitous*, specific, and dynamic involvement of regions for silently moving the articulators and overt speech production during speech perception. We then extend existing knowledge using network meta-analyses and a new text-mining approach that allows characterization of the representations, computations, and dynamics of nodes in distributed networks. To preview, this approach demonstrates that there are multiple possible speech production networks (along multiple fibre tracts) that participate in speech perception (“where”) and that the topologies of these networks change dynamically with listening contexts (“how” and “when”).

Following qualitative and quantitative reviews, we summarize results and discuss them in relation to the AxS model for which they are most consistent. This discussion also addresses questions of why speech production networks might be so ubiquitously involved in speech perception and whether they are necessary. We then propose rejecting textbook/classical and dual-stream models of the organization of language and the brain for models that implement AxS *like* mechanisms in a complex network framework. We conclude by suggesting that, at minimum, it is time to put the old “if” question to rest and focus on addressing the more nuanced “where”, “how”, and “when” questions suggested herein.

## 2. Qualitative review

Taking seriously the complexity of the neurobiology of speech production leads us to first provide a review of the brain regions, subprocesses, and networks supporting speech production as any of these might play a role in speech perception. We then review behavioural experiments, computational modelling, nonhuman brain research, human brain damage/disorder, brain stimulation/recording, and some neuroimaging studies demonstrating “where”, “how” and “when” roles for speech production systems in speech perception.

### 2.1. Speech production

In speech perception studies, the speech motor system is problematically characterized as being centred around posterior ventral frontal regions like Broca’s area (see 1.1.1). Indeed, lesion and electrical stimulation studies demonstrate that speech production is reliant on a much larger and distributed set of brain regions. Modern voxel-based lesion-symptom mapping (VLSM) shows involvement of the posterior ventral frontal regions (the inferior frontal gyrus, including both the pars opercularis and triangularis and ventral precentral gyrus), anterior insula, supplementary motor area (SMA), postcentral gyrus, inferior parietal lobule (IPL, mostly the supramarginal gyrus), and posterior superior temporal cortex and underlying white matter tracts like the superior longitudinal fasciculus (Baldo, Wilkins, Ogar, Willock, & Dronkers, 2011; Bates et al., 2003; Borovsky, Saygin, Bates, & Dronkers, 2007; Mirman et al., 2015; Wilson et al., 2010). The best predictors of recovery of speech production skills after stroke are time post-stroke (the most significant predictor), lesion volume, and where in a distributed set of 36 regions damage occurs (Hope, Seghier, Leff, & Price, 2013).

Movement of speech production related muscles can be elicited by electrical stimulation of all of the regions mentioned in VLSM studies. Face movements, for instance, can be elicited by stimulation of the ventral pre- and primary motor cortex, anterior insula, SMA, cingulate cortex, IPL, and the facial motor nucleus (see Cattaneo & Pavesi, 2014 for a review). Stimulation studies also show that articulators like the lips, jaw, vocalic/laryngeal, and tongue muscles roughly form a dorsal to ventral somatotopic map in the central sulcus (i.e., in ventral primary motor and somatosensory cortex; Grabski et al., 2012; Penfield & Boldrey, 1937). Though the Penfield and Boldrey report is typically cited in reference to this somatotopy, their results also suggest that movement and sensation of the tongue, mouth, throat, and face also occur with electrical stimulation of pars opercularis, IPL, and posterior superior temporal regions. Lesion and stimulation studies indicate that different regions perform different roles in production. For example, both lesions to and stimulation of the SMA result in the involuntary production of consonant vowel syllable sequences like “dadadada” (see discussion and commentary in MacNeilage, 1998 for references). Lesion data also shows that the SMA is involved in other types of sequencing during production, like sentence embedding (Wilson et al., 2010).

Neuroimaging studies similarly demonstrate that speech production is reliant on a large distributed “core” (i.e., consistently activated) set of cortical and subcortical brain regions. In both label-based (Indefrey, 2011; Indefrey & Levelt, 2000, 2004; see also Munding, Dubarry, & Alario, 2016) and activation likelihood estimation (ALE) neuroimaging meta-analyses (Brown, Ingham, Ingham, Laird, & Fox, 2005; Eickhoff, Heim, Zilles, & Amunts, 2009; Turkeltaub, Eden, Jones, & Zeffiro, 2002), the most consistently activated regions are the posterior IFG (mostly pars opercularis), ventral pre- and primary motor cortex, SMA, anterior insula,

posterior superior and middle temporal cortex, thalamus, basal ganglia, and cerebellum. Producing speech results in changes in IPL and posterior superior temporal regions as demonstrated by manipulation of auditory feedback (Houde, Nagarajan, Sekihara, & Merzenich, 2002) and perturbations of the articulators (Golfnopoulos et al., 2011). This activity is likely independent of auditory or bone conduction feedback as demonstrated in studies of covert speech production (Shergill et al., 2002).

Like lesion and brain stimulation studies, neuroimaging studies indicate that many widely distributed regions differentially support the psycholinguistic subprocesses that comprise speech production. For example, in their label-based meta-analysis and review of the literature, Indefrey and Levelt (2004) suggest that “conceptual preparation” and “lexical selection” load on the middle temporal gyrus, “phonological code retrieval” on the anterior insula, SMA, and posterior superior and middle temporal cortices, “syllabification” on posterior IFG, and “phonetic encoding and articulation” on 17 regions including most of the above “core” regions. Clearly, more work on the latter is required to understand the role of these regions in specific subprocesses associated with speech production. Perhaps a more fruitful parcellation that corresponds to research on the network organization of the brain (e.g., Anderson, 2015; Bullmore & Sporns, 2009; Knight, 2007; van den Heuvel & Sporns, 2013) would be to map these subprocesses onto specific networks or subnetworks rather than individual regions (Collard et al., 2016; van de Ven, Esposito, & Christoffels, 2009). Subnetworks would form through known anatomical connections between many of the “core” regions, perhaps aided by enhanced connectivity between laryngeal motor cortex and the SMA, temporal, and parietal regions in humans compared to non-human primates (Dick, Bernal, & Tremblay, 2014; Jürgens, 2002, 2009/1; Kumar, Croxson, & Simonyan, 2016). Indeed, many “core” regions are flexible “hubs”, that is, regions of high connectivity that can adaptively interact with other networks (e.g., Fuertinger, Horwitz, & Simonyan, 2015). Any of these dynamic subnetworks might then also support aspects of speech perception.

## 2.2. Behaviour

A great many behavioural experiments support the claim that some or all of these speech production related regions or networks play a role in speech perception. Before turning to these, however, it is instructive to dissect the logic underlying behavioural experiments that constitute a common argument against this proposal. The argument is that, if nonhuman animals or speechless developing humans can perceive speech, then vocal or speech production regions are not necessary for speech perception. Indeed, non-human rodents (e.g., chinchillas, Kuhl & Miller, 1975), birds (e.g., Japanese quail, Kluender, Diehl, & Killeen, 1987) and human infants (e.g., Jusczyk, 1981) can perceive speech sounds categorically. That is, they classify sets of similar but acoustically distinct speech sounds as a single discrete speech category, despite not being able to produce speech. Though these categorical speech perception studies are often revered because they suggest the reality of speech units like phonemes, they have been criticized. Problems include that the tasks assume the units under study and that within category differences are actually readily discernible and meaningful (Gerrits & Schouten, 2004; Schouten, Gerrits, & van Hessen, 2003; Toscano, McMurray, Dennhardt, & Luck, 2010).

In addition to more general problems with categorical speech perception tasks, the empirical evidence is not consistent with the argument that the perception of speech by non-speaking animals demonstrates that vocal systems are not necessary for speech perception. Neurobiologically, the argument is unsound because,

in the early work frequently used to support this argument (e.g. Jusczyk, 1981), the brain was not directly observed. Yet it has been suggested that premotor cortex is involved in processing sounds that we cannot produce in ways that make use of the underlying computational mechanisms that would also be involved in movement (Dehaene & Cohen, 2007; Schubotz, 2007). Indeed, posterior ventral frontal regions *are* activated in nonhuman primates by sounds they cannot produce (Gil-da-Costa et al., 2006; Joly et al., 2012), including human speech (Romanski & Goldman-Rakic, 2002). With regard to nonspeaking neonates and infants, Bruderer et al. (2015) showed that perturbing the articulators of 6-month old infants disrupts their ability to perceive speech sounds. In support of this behavioural finding, frontal cortex, including the pars opercularis, seems to be involved in prelinguistic infants’ speech perception abilities (Gervain & Werker, 2008; Homae, Watanabe, Nakano, & Taga, 2011; Shultz, Vouloumanos, Bennett, & Pelphrey, 2014). Furthermore, infants have stronger functional connectivity between frontal and temporal regions when listening to sentences (Homae, Watanabe, & Taga, 2014; Homae et al., 2011), perhaps corresponding to a surprisingly mature structural connectivity in infants between frontal, including premotor, and posterior temporal regions (Leroy et al., 2011; Perani et al., 2011).

A computationally informed or nascent role of vocal or speech production systems in non- or pre-vocal animals is also more consistent with other behavioural data that demonstrates strong reciprocal links between speech production and perception throughout the lifespan. To give 10 striking examples:

- I. Articulations produced by infants affect both their auditory and audiovisual speech perception (Bruderer et al., 2015; Yeung & Werker, 2013).
- II. In both children and adults perceptual learning of speech is transferred to the production domain and vice versa (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Kittredge & Dell, 2016; Lametti, Krol, Shiller, & Ostry, 2014; Shiller & Rochon, 2014).
- III. Imitation or shadowing speech tasks are much faster than would be predicted, e.g., when compared to simply repeating one sound or responding with button presses (Fowler, Brown, Sabadini, & Weihing, 2003; Nye & Fowler, 2003/1; Scarbel, Beautemps, Schwartz, & Sato, 2014).
- IV. During production, participants are faster to respond when they also hear or see a sound that is produced by the same articulators as the sound they are asked to produce (Galantucci, Fowler, & Goldstein, 2009; Roon & Gafos, 2015).
- V. The “verbal transformation effect” (VTE) has been argued to have a speech production related etiology (Basirat, Schwartz, & Sato, 2012). In particular, the VTE is the experience of illusory sounds in speech when rapidly repeated over time, e.g., “tress” repeated for three minutes might be heard as “dress”, “floris”, “florist”, “Jewish”, “Joyce”, “stress” and “purse” (see Warren, 1968). When self generated, the effect is increased with increasing articulatory involvement. When the VTE is experienced during listening, illusions are most likely to correspond to stable articulatory sequences.
- VI. Real-time manipulations of auditory or somatosensory feedback change production and corresponding perceptual representations (Houde & Jordan, 1998; Lametti, Rochet-Capellan, Neufeld, Shiller, & Ostry, 2014; Lane & Tranel, 1971; Nasir & Ostry, 2009; Shiller, Sato, Gracco, & Baum, 2009).
- VII. Producing speech can modify speech perception online in specific ways and even cause a McGurk-MacDonald effect (McGurk & MacDonald, 1976). For example, participants

hear an illusory speech percept when the sound they are actually producing is replaced with a speech sound other than the one they observe themselves producing (Mochida et al., 2013; Sams, Möttönen, & Sihvonen, 2005).

- VIII. Speech discrimination can be biased towards hearing certain sounds by stretching the face in a way that is similar to producing those sounds, or by applying inaudible air puffs to the skin to mimic production-associated aspiration of those sounds (Gick & Derrick, 2009; Ito, Tiede, & Ostry, 2009).
- IX. Distractor sounds influence the way the tongue makes contact with the roof of the mouth in ways specific to that sound when producing different sounds (Yuen, Davis, Brysbaert, & Rastle, 2010) (for discussion, see McGettigan, Agnew, & Scott, 2010; Rastle, Davis, & Brysbaert, 2010).
- X. Predictions associated with hearing, e.g., “The pond was full of croaking...” influences tongue position, affecting the subsequent forced production of a word like “toad” compared to “frogs” (Drake & Corley, 2015).

Considered in isolation, any one of these behavioural phenomena would not be enough to decisively link speech production to perception. Indeed, there have been plenty of arguments (e.g., see the four critiques referenced in McGowan & Faber, 1996) and a small number of studies that claim from behavioural evidence that “the motor system” does not play a role in speech perception. For example, Matchin, Groulx, and Hickok (2014), in a failure to replicate Sams et al. (2005) argue from null results that “the motor system does not play a role in AV [audiovisual] speech integration” (pg. 619). Furthermore, a look at the effect sizes in the cited studies (when reported) indicates that results often do not reach a medium effect (though this is true of psychology generally where the average effect size is typically small, e.g., Richard, Bond, & Stokes-Zoota, 2003). Nonetheless, when taken together, these findings present a converging body of evidence that clearly indicates a role or roles for speech production systems in speech perception. The infrequent report of lack of effect and smaller effect sizes might rather suggest that these roles are not all or nothing but, rather, dynamic and contextually determined.

### 2.3. Computational modelling

Computational models of speech production are often sensorimotor in nature, with auditory cortex playing a crucial role in learning to produce speech and speaking (e.g., Guenther, Ghosh, & Tourville, 2006; Hickok, Houde, & Rong, 2011). The converse, that models of speech perception assume the involvement of speech production systems, is less common. However, an examination of the computational modelling work that has been done suggests that speech representations, computations, and mechanisms that are sensorimotor in nature better account for data than auditory-only counterparts.

“Silent” computer speech recognition systems suggest the possibility that speech perception can be achieved without audible acoustic information, using only speech production related information. For example, a recurrent neural network trained to read lips from video can achieve 80% word classification accuracy (Wand, Koutník, & Schmidhuber, 2016). High resolution movement dynamics recorded from the tongue, lip, and jaw are sufficient for word and sentence recognition accuracies greater than 93% with support vector machine classifiers (Wang, Samal, Green, & Rudzicz, 2012; Wang et al., 2012). Indeed, incorporating articulatory feature information yields more accurate computer speech and speaker recognition than strategies that take only acoustic information into account (Badino, Canevari, Fadiga, & Metta, 2012; Canevari, Badino, D’Ausilio, Fadiga, & Metta, 2013; King

et al., 2007; Li et al., 2016; Mitra, Sivaraman, Nam, Espy-Wilson, & Saltzman, 2014; Tan, Liu, Jiang, & Zheng, 2015).

Likewise, explicit computational models of speech perception improve when production related information is included in models. Properties of discreteness and compositionality in speech perception are said to derive from the coupling of production and perception in models using artificial agents (Oudeyer, 2005). Similarly, computational models of language acquisition develop sensorimotor representations that are subsequently used during speech perception (Howard & Messum, 2014; Kröger & Cao, 2015; Messum & Howard, 2015; Westermann & Reck Miranda, 2004). Principles codified in the AxS model have been implemented in speech recognition systems, resulting in recognition improvement (Bawab, Raj, & Stern, 2008; Blackburn, 1996). Similarly, it was found that a sensorimotor speech model in the COSMO Bayesian modelling framework outperforms both auditory and motor only models with motor only models performing better than auditory models under adverse conditions (Moulin-Frier, Diard, Schwartz, & Bessière, 2015; Moulin-Frier, Laurent, Bessière, Schwartz, & Diard, 2012). Finally, song production also plays an important role in a computational model of bird song recognition (Yildiz & Kiebel, 2011).

### 2.4. Nonhumans

Humans are obviously not the only animals that both produce and perceive sounds categorically or otherwise. Neurobiological research indicates a role of distributed vocal production systems in perception in rodents, birds, and nonhuman primates. In mice, motor cortex influences auditory cortical processing (Anderson & Linden, 2016; Nelson et al., 2013). This likely occurs through projections and corollary discharge from motor to auditory cortex (Nelson & Mooney, 2016; Nelson et al., 2013; Schneider et al., 2014). This feedback related activity is often said to be suppressive. Though this would appear to reduce auditory sensitivity (seemingly at the expense of survival value), it might actually increase sensitivity and signal-to-noise ratio and/or be part of a predictive mechanism to distinguish internally from externally generated sound (Schneider & Mooney, 2015; Schneider et al., 2014).

Songbirds have complex vocal systems comparable to human speech (Bolhuis, Okanoya, & Scharff, 2010; Petkov & Jarvis, 2012). “A motor theory of song perception” has been proposed in birds because neurons that innervate the syrinx show selective responses to song listening (Williams & Nottebohm, 1985). Indeed, damage to the auditory or vocal peripheries of birds “result in altered auditory tuning” (Pytte & Suthers, 1999; Roy & Mooney, 2007; Solis & Doupe, 2000) and damage to sensorimotor song regions can result in song discrimination problems (Brainard & Doupe, 2013; Prather, Nowicki, Anderson, Peters, & Mooney, 2009). Neurons in bird sensorimotor regions are active during both the production and perception of song (J. F. Prather, Peters, Nowicki, & Mooney, 2008). Furthermore, they show evidence for the categorical perception of song notes (Mooney, 2014; Prather et al., 2009).

Non-human primate vocal communication is typically viewed as limited and simple and, perhaps for this reason, is neurobiologically understudied (though this appears to be changing, Ghazanfar & Eliades, 2014). In the macaque, regions 44 and 45 (with area 44 perhaps corresponding to the pars opercularis in humans) contain single neurons that respond during both primate call production and listening (Hage & Nieder, 2015). Though there are auditory mirror neurons, no study has found auditory mirror neurons for macaque calls in area F5 (Kohler et al., 2002), likely corresponding to human ventral premotor cortex (BA 6) (Ferri et al., 2015; Nelissen, Luppino, Vanduffel, Rizzolatti, & Orban, 2005). Mirror neurons in F5, however, have been found that discharge during



the production and observation of communicative mouth movements (Ferrari, Gallese, Rizzolatti, & Fogassi, 2003). In the marmoset, a more vocal primate, neurons in ventral premotor and auditory cortex are involved in both listening to and producing vocalizations (Miller, Dimauro, Pistorio, Hendry, & Wang, 2010; Miller, Thomas, Nummela, & de la Mothe, 2015). Self-initiated vocalizations result in auditory cortical suppression and excitation, with suppression being a more general response, corresponding to an increase in sensitivity to vocal feedback (Eliades & Wang, 2002, 2008, 2013).

### 2.5. Damage and disorders

Studies involving brain damage and disorder are often afforded more weight in evaluating evidence pertaining to “the role of the motor system in speech perception” because results are from humans and considered “causal”. The logic is that, if “motor cortex” is destroyed and a patient can still perceive speech, “motor cortex” must not play a necessary role. This logic is weak on any number of grounds (see the earlier discussion about what constitutes “motor cortex” and the complexity of speech production systems; other issues with this logic include the large size of lesions, individual variability, comorbid deficits, plasticity and rapid reorganization, etc.). Nonetheless, given this weight, we spend more time (re-) considering this literature. Results suggest that patients with impaired speech production abilities because of stroke, lesion, and movement disorders usually have some speech perception impairment. These are the result of a distributed set of brain regions. Furthermore, some of this work suggests specificity (as with specific place of articulation deficits) and dynamic involvement of these distributed region (as suggested by variability in results).

In stroke, deficits in processing of acoustic cues, speech segments, and words are not limited to fluent (e.g., Wernicke’s) aphasics. Nonfluent patients (like Broca’s and conduction aphasias) showed differential impairment of spectral and temporal acoustic cues in a word-discrimination test (Leeper, Shewan, & Booth, 1986). Miceli, Gainotti, Caltagirone, and Masullo (1980) found a strong relationship between the ability to produce speech and discriminate syllables in 69 fluent and nonfluent aphasics. Specifically, contrasts between groups with and without a phonemic output disorder showed that patients with a disorder were worse at discriminating phonemes, particularly but not limited to those distinguished by place of articulation (see also, Martin & Saffran, 2002). Basso, Casati, and Vignolo (1977) found that 95% of nonfluent (Broca’s and global) patients (N = 21) had a phoneme identification defect in detecting the boundary zone between voiced and voiceless consonants on a voice onset time (VOT) continuum. Furthermore, this disorder was present in 100% of patients with, and only 25% of patients without, disorders of phonemic output. Similarly, Shewan (1980) found that 67% of their nonfluent Broca’s aphasics (N = 9) had some receptive phonological impairment. This group showed a correlation of .88 between their production impairment scores and phonological perception. Finally, comprehension can be impaired in Broca’s aphasics if word processing is made more difficult through acoustic distortion (Moineau, Dronkers, & Bates, 2005).

These studies provide little, if any, information about what brain damage leads to these strongly positive relationships between speech production and perception in aphasias. It is often assumed that nonfluent damage is associated with frontal and fluent aphasias with posterior lesions. However, this can be challenged for any number of reasons (e.g., Mohr et al., 1978 but see also, Kreisler et al., 2000). Indeed, speech perception deficits seem to result from nearly any type of aphasia, including those with damage to frontal, posterior, or both sets of regions

(Blumstein, 1994). To give some examples, aphasia associated with both frontal and posterior damage results in impaired syllable identification relative to controls (Kimura & Watson, 1989). Gainotti, Miceli, Silveri, and Villa (1982) found that patients with left frontal lobe damage (N = 6) made significantly more errors on phoneme discrimination compared to patients with temporal (N = 9) but not parietal lobe (N = 7) lesions.

This distribution of speech perception results coheres well with the proposal to take seriously the claim that speech production systems are distributed and, therefore, their involvement in perception will also include a distributed set of regions. Indeed, many of the reviewed “core” production regions have specifically been shown to be involved in both speech production and perception in stroke and lesion studies. In a thoughtful review, Schirmer (2004) showed that lesions to the left frontal cortex (variously resulting in non-fluent, Broca’s, transcortical motor, and conduction aphasia) and the cerebellum result in both speech production and perception problems in segmenting speech based on temporal differences (e.g., “ladder” and “latter” differ only in small temporal differences in closure and voice onset times; see Ackermann, Mathiak, & Riecker, 2007 for a cerebellum specific review).

Frontal cortex is quite vague and reflects the extent of damage typically caused by stroke. A more focal cortical excision study in epilepsy patients suggests that frontal cortex does include regions specifically involved in moving the articulators. In particular, Taylor (1979) showed that removal of cortex from around the face region (in individuals without aphasia) led to deficits in phoneme identification (as cited in Kimura & Watson, 1989; Taylor, 1979; see also, note 20 in Zatorre, Evans, Meyer, & Gjedde, 1992). Finally, there are case studies providing evidence that insular damage results in both speech production problems (like transient mutism, oral apraxia, or conduction aphasia) and auditory agnosia for both nonspeech and speech sounds (for a review of the role of the insula in auditory processing, see Bamiou, Musiek, & Luxon, 2003/5; Fifer, 1993; Habib et al., 1995/3; Hyman & Tranel, 1989).

Movement disorders caused by other means further indicate the distributed nature of the roles of speech production systems in perception. Cerebral palsy is a movement disorder that can result in anarthria and dysarthria and impacts sensorimotor networks (Lee et al., 2011; Scheck, Boyd, & Rose, 2012). Both children and adults with cerebral palsy have been shown to perform worse on phoneme discrimination and this is often related to articulatory abilities (Bishop, Brown, & Robson, 1990; Peeters, Verhoeven, de Moor, & van Balkom, 2009; Peeters, Verhoeven, van Balkom, & de Moor, 2008; Smith, 2001). Parkinson’s disease is a degenerative movement disorder that results in reductions in premotor, SMA, and parietal cortex metabolism, linked to the basal ganglia (Eidelberg, 2009). These regions comprise not only motor systems for moving distal limbs but also those involved in swallowing and speech production (Robbins, Logemann, & Kirshner, 1986). Speech discrimination and phonological perception is impaired in Parkinson’s disease (Cummings, Darkins, Mendez, Hill, & Benson, 1988; Elorriaga-Santiago, Silva-Pereyra, Rodríguez-Camacho, & Carrasco-Vargas, 2013; Vitale et al., 2016; Zanini et al., 2003). Schirmer (2004) also reviewed studies showing that damage associated with the basal ganglia from both Parkinson’s and Huntington’s disease results in problems segmenting speech during both production and perception. Speech production involvement in perception even extends to the periphery, e.g., as children with cleft palates have problems detecting specific place of articulation features in speech (Whitehill, Francis, & Ching, 2003).

Studies of brain damage and disorders show that there is a strong relationship between speech production, a distributed set of production regions, and perception. Some show, however, that damage does not uniformly result in speech perception deficits. In many patients, the ability to perceive speech is said to be *rela-*



tively preserved (Damasio, 1992; Goodglass & Kaplan, 1983; Hillis, 2007; Naeser, Palumbo, Helm-Estabrooks, Stiassny-Eder, & Albert, 1989; Stassen et al., 2015; Weller, 1993). For example, Blumstein, Cooper, Zurif, and Caramazza (1977) is interpreted (e.g., Hickok, Costanzo, et al., 2011) as not showing a relationship between production and perception (despite that arguably 2/5ths of the patients did). They found that, in a group of nonfluent patients, only 27% had discrimination and 37% labelling impairments on a VOT continuum compared to 0% and 75% in Wernicke's aphasia. They also found no clear relationship between production and perception of VOT. The sample size for this analysis, however, was only eight patients with a heterogeneous mix of five different types of aphasias. To give an example involving more precise localization of function, Hickok et al. (2011; see also Rogalsky, Love, Driscoll, Anderson, & Hickok, 2011) examined speech discrimination in patients that had frontal lobe lesions that included some aspects of Broca's area. These patients performed significantly worse on speech discrimination than a control group with temporal-occipital damage. Furthermore, in one of the two speech tasks, more severe speech fluency problems were associated with worse discrimination. Despite these results, given the relatively decent performance of the Broca's damaged group on both speech tasks, and the lack of relationship between discrimination and fluency on one of the speech tasks, the authors conclude from null results that "the motor speech system is not necessary for speech perception".

Thus, like the reviewed behavioural data, brain damage and disorder results suggest a strong contribution of speech production regions to perception but with some variability. Even this variability, however, is relative and still far removed from showing unimpaired performance (as we indicate in the prior paragraph). Furthermore, such variability is not surprising when considering the large amount of variability in recovery from aphasia generally (Lazar & Antonello, 2008) and that speech production involves not just posterior frontal regions like Broca's area (which, as reviewed, does not play as prominent a role in speech production as previously believed). Rather, speech production is supported by a distributed set of brain regions, any of which might be more or less preserved and, therefore, participate in speech perception as normal or might have been involved in recovery after damage. Indeed, aphasia is associated with recovery and brain reorganization that occurs at different timescales post stroke (Saur et al., 2006). For example, in one study, patients with even extensive damage to Broca's area recovered speech production within three days and this was associated with activity shifts to putatively homologous regions in the right hemisphere (Thulborn, Carpenter, & Just, 1999).

## 2.6. Stimulation and recording

Direct cortical stimulation (DCS) and recording (i.e., electrocorticography or ECoG) usually occurs when patients, under local anaesthesia, are stimulated to locate epileptogenic zones appropriate for surgical removal. In contrast, transcranial magnetic stimulation (TMS) and direct current stimulation (tDCS) are noninvasive and involve excitation or inhibition of brain regions through the skull. This is usually in concert with measurements of the resulting behavioural effects and/or, in the case of TMS, electromyographic (EMG) responses in muscles if the stimulation is over primary motor cortex. Like lesion studies, TMS studies are often considered to have special status over other methods presumably because they involve unimpaired human participants and focal stimulation that can be used to infer causality (Paus, 2005).

As discussed in Section 1.1.1, DCS across a distributed set of regions results in both movement *and* sensation of the articulators (Penfield & Boldrey, 1937). Ojemann (1991, 1983) and Ojemann

and Mateer (1979) noted a "striking association" between motor sites and identifying speech sounds. Specifically, they found that stimulating sites involved in motor sequencing and mimicry frequently led to deficits in phoneme identification. Furthermore, DCS and simultaneous surface recordings suggest that such effects could come about through bidirectional connectivity between posterior ventral frontal, parietal, and posterior superior temporal regions (Matsumoto et al., 2004). Such connectivity may explain why DCS to posterior ventral frontal regions can result in auditory hallucinations (Lesser, Lueders, Dinner, Hahn, & Cohen, 1984). Though frontal lobe stimulation infrequently yields these (see Boatman, 2004 for discussion of Lesser et al., 1984), converging neuroimaging evidence supports the claim that interactions between a distributed set of speech production and auditory regions underlies the experience of auditory hallucinations (Allen, Larøi, McGuire, & Aleman, 2008; Diederer et al., 2012; Dierks et al., 1999; Jardri, Pouchet, Pins, & Thomas, 2011; Linden et al., 2011; Shergill, Brammer, Williams, Murray, & McGuire, 2000).

Subsequent to foundational DCS work, ECoG (Cheung, Hamiton, Johnson, & Chang, 2016; Cogan et al., 2014; Edwards et al., 2010; Rhone et al., 2016), TMS (For a review, see Devlin & Watkins, 2007), and tDCS studies (Daniel Robert Lametti, Oostwoud Wijdenes, Bonaiuto, Bestmann, & Rothwell, 2016; Sehm et al., 2013) have all shown that brain regions involved in moving the articulators and producing speech are involved in speech perception. Most of these studies focus on posterior ventral frontal cortex but some suggest a more distributed set of regions. For example, using tDCS Lametti et al. (2016) showed that the cerebellum plays a timing role during speech perception. These newer methods also significantly advance arguments pertaining to the specificity of the role of production systems and show that involvement is dynamic, varying with listening context.

### 2.6.1. Specificity

ECoG nicely extends DCS results to address specificity of representations. Cogan et al. (2014) used ECoG to reveal significant sensorimotor responses to speech in ventral frontal, somatosensory, parietal, and posterior superior temporal regions. They used classifiers to demonstrate that these sensorimotor responses during listening carried information about both the sensory and motor content of speech responses and that they represent a transformation from sensory input into motor output.

Cheung et al. (2016) also found that posterior ventral frontal regions overlapped for speech production and perception. They did not, however, demonstrate a somatotopic organization of cortical responses or an organization consistent with place of articulation processing in these regions during perception. Rather, responses were more similar to the acoustic representation in the auditory cortex, though weaker. Thus the authors maintain that "motor cortex" plays an auditory role in speech perception without transformation to an articulatory representation. Though the authors state that participants engaged in "passive" listening, they seem to have been "asked to identify the syllable they heard by selecting from a multiple-choice question". Schomers and Pulvermüller (2016) suggest that reliance on a potentially obscuring motor response to measure perception and high gamma might have led to these results. If so, something closer to reality, as demonstrated by Cogan et al. (2014), is that representations in posterior ventral frontal regions are likely more complex than simply auditory or articulatory only (or visual only for that matter, see the ECoG study of Rhone et al., 2016). Indeed, there is a great deal of evidence across species that the representations in these regions are not so homogeneous (Long et al., 2016; Peh, Roberts, & Mooney, 2015).

TMS studies extend ECoG results pertaining to specificity of representations in a manner more consistent with sensorimotor representations than the auditory only representations suggested by Cheung et al. (2016). First, speech perception activates motor cortex in a somatotopic manner. That is, different regions of motor cortex used to produce different sounds (lip or tongue regions) are differentially involved in perception of speech sounds that more or less involve those effectors, particularly for speech that is difficult to perceive (D'Ausilio, Bufalari, Salmas, & Fadiga, 2012; D'Ausilio et al., 2009; Mottonen & Watkins, 2009; Smalle, Rogers, & Möttönen, 2015). Second, it has been found that the degree of motor recruitment depends on the perceived distance between the voice of the listener and voices of the talkers being listened to – supporting a claim that speech perception relies on the contribution of listeners' own motor knowledge (Bartoli et al., 2015).

### 2.6.2. Dynamics

TMS studies also show that the degree of influence of speech production systems on perception varies with listening context, e.g., as in a quiet versus noisy pub or listening to someone speaking while they are eating chips versus when they are not. Specifically, TMS studies find a decrease in speech discrimination primarily in noisy or difficult speech tasks (Boatman, 2004; Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Romero, Walsh, & Papagno, 2006). Somatotopic facilitation of motor evoked potentials was associated with listening to distorted speech (produced by tongue depression) but not naturally produced speech (Nuttall, Kennedy-Higgins, Hogan, Devlin, & Adank, 2016). Despite this generality, null effects for clear speech does not mean that there is no relationship between speech production and perception when listening is easy. Indeed, TMS influences have been shown for speech not in noise (Bartoli et al., 2015; Sato, Tremblay, & Gracco, 2009; Schomers, Kirilina, Weigand, Bajbouj, & Pulvermüller, 2015) and somatotopic effects have been found for clear speech without an accompanying task (Möttönen, Dutton, & Watkins, 2013). Furthermore, hearing and seeing clear speech facilitates the excitability of articulatory motor cortex (Watkins, Strafella, & Paus, 2003).

There have been criticisms of studies demonstrating a change in speech perception following TMS of motor regions. One is that the perceptual change could, more simply, be a result of TMS-induced change in response strategies or biases employed during psychophysical testing (Venezia, Saberi, Chubb, & Hickok, 2012). However, Smalle et al. (2015) demonstrated that disruptive TMS applied to articulatory motor cortex decreased the discriminability of speech sounds without significantly altering response bias. A second criticism centres on the nature of the stimuli used in these studies – typically, discrete phonemes, something one is not likely to encounter in the real-world (Hickok, 2014). Thus, small TMS effects, even if real, likely have no real-world effect on understanding. However, several studies have now demonstrated that TMS to articulatory motor cortex and other speech production regions alter word comprehension in a speech specific manner (Murakami, Kell, Restle, Ugawa, & Ziemann, 2015; Schomers et al., 2015). A third criticism is that facilitatory TMS may excite motor regions, priming a production-perception link that, under normal conditions, would not be involved in perception. However, though this idea might apply to facilitation studies, it fails to predict the impairments in perception that are consistently observed following disruptive TMS to motor regions (Mottonen & Watkins, 2009; Möttönen, van de Ven, & Watkins, 2014; Smalle et al., 2015).

### 2.7. Neuroimaging

Among others, neuroimaging methods include positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and source localized electro- (EEG) and magnetoencephalography

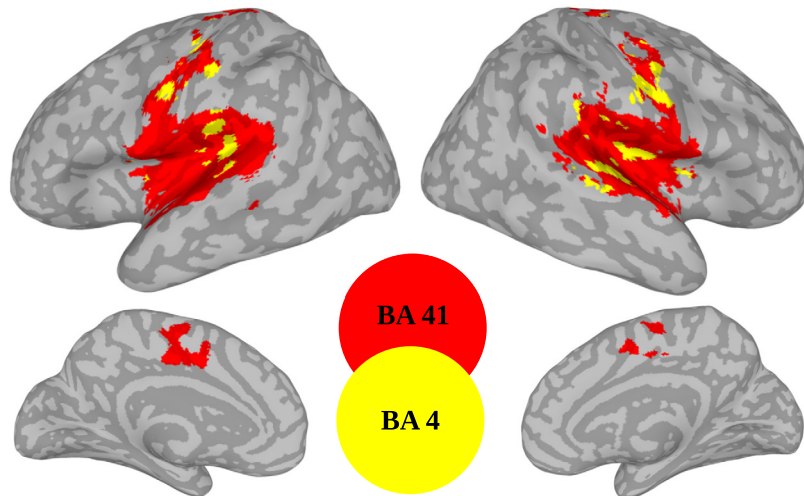
(MEG). Neuroimaging has variously demonstrated that speech production regions play a major role in speech perception using both “passive” and “active” listening tasks, with and without speech production localizers. Our review of these studies is truncated because quantitative meta-analyses of neuroimaging data are conducted in the second half of this manuscript. We focus here on studies that specifically address the specificity, mechanisms, and dynamics of the role of speech production systems in perception (i.e., the “how” and “when” questions).

#### 2.7.1. Specificity

Reviewed computational modelling, nonhuman brain, brain damage/disorder, and, particularly, stimulation/recording studies each suggest that *sensorimotor* representations and computations involved in producing speech are directly involved in speech perception (i.e., these regions are likely not performing unrelated and de novo computations). The majority of neuroimaging studies, starting with the earliest PET studies (Démonet et al., 1992; Zatorre et al., 1992), support this position. Specificity can be seen in “passive” speech listening studies, our primary focus as they do not require confounding motor responses that complicate interpretation. In particular, listening to and/or viewing audio (Aziz-Zadeh, Sheng, & Gheyntanchi, 2010), visual (Paulesu et al., 2003; Skipper, Goldin-Meadow, Nusbaum, & Small, 2007), and audiovisual (Skipper, Nusbaum, & Small, 2005; Skipper, van Wassenhove, Nusbaum, & Small, 2007; Tremblay & Small, 2011) speech engages primarily posterior superior temporal regions and, variously, all of the regions identified as constituting core speech production regions (though there is certainly individual variability, see e.g., Szenkovits, Peelle, Norris, & Davis, 2012). Some of these “passive” listening studies include a separate speech production task that can serve as a localizer for actual speech production regions, accounting for potential individual differences in patterns of activity when producing speech. Furthermore, these studies provide more specificity to the claim that speech perception across modalities activates regions also active during production (Grabski et al., 2013; Skipper et al., 2007; Tremblay & Small, 2011; Wilson & Iacoboni, 2006; Wilson, Saygin, Sereno, & Iacoboni, 2004).

Bolstering claims of specificity, “passively” listening to speech sounds activates pre- and primary motor regions in a somatotopic manner. That is, sounds that involve more movement of the lips when produced activate regions for actually moving the lips more than regions associated with other articulators (Pulvermüller et al., 2006). Arsenault and Buchsbaum (2015), however, failed to replicate these results using either univariate or multivariate pattern analysis. This might be in part because somatotopy is not as simplistic as typically portrayed in neuroimaging and TMS studies of speech (Meier, Aflalo, Kastner, & Graziano, 2008). Based on a review of 14 fMRI studies, Schomers and Pulvermüller (2016) suggest the failure to replicate is in part because Arsenault and Buchsbaum (2015) used a button-pressing task that produced “motor noise” or ceiling effects from continual motor preparation. They point out that Arsenault and Buchsbaum (2015) did find decoding of place of articulation in somatosensory cortex and argue that somatosensory activity would not be subject to the same ceiling effects and would require a speech motor explanation.

Perhaps supporting this argument, other “passive” studies using multivariate decoding do find evidence of specificity in posterior ventral frontal regions. Correia, Jansma, and Bonte (2015) trained classifiers to discriminate activity from listening to syllables that differed on place of articulation, manner, or voicing and then tested whether they could predict activity from a different set of syllables. Place, manner, and voicing were each found to be represented in superior temporal and somatosensory cortices. However, only place and manner were specifically found to be represented in a broader set of regions involved in producing speech that included



**Fig. 2.** “Intrinsic” sensorimotor connectivity of primary auditory cortex. Resting-state connectivity of left primary auditory cortex (BA 41) in a sample of 1000 participants (red; seed MNI coordinate is  $x = -44, y = -28, z = 8; r > .15$ ; clusters  $> 160 \text{ mm}^3$ ). Overlap of white matter structural connectivity originating in primary motor cortex (BA 4) with functional connectivity (yellow; data from ALE meta-analysis of 69 papers;  $N = 4565$ ;  $p < .05$  FDR corrected; clusters  $> 160 \text{ mm}^3$ ). See Fig. 1 for location of regions.

posterior ventral frontal cortex, the basal ganglia, and cerebellum (though other studies suggest a more “general” or abstract representation in these regions, Agnew, McGettigan, & Scott, 2011; Evans & Davis, 2015).

For these reasons, Schomers and Pulvermüller (2016) argue for a specific role of speech production regions in perception. Indeed, of the 10 fMRI task based studies in their review that analysed specificity for phonetic features (including Correia et al., 2015), 80% showed specific selectivity for mostly place of articulation in posterior IFG and pre- and primary motor cortex. This distribution is consistent with speech production studies in which sensitivities to features vary throughout the brain. For example, ECoG shows that place of articulation involves broad distributed networks throughout production, including posterior ventral frontal regions, whereas manner and voicing are dominated by auditory cortical responses after speech initiation (Lotte et al., 2015).

Finally, not all studies demonstrating selectivity involve listening to speech. “Resting state” studies, in which participants are not presented with stimuli and have no explicit task, suggest an “intrinsic” network organization of the human auditory cortex that includes all of the “core” speech production regions (Demertzi et al., 2014; Maudoux et al., 2012; Yakunina, Tae, Lee, Kim, & Nam, 2013). Fig. 2 is an example created for this review using publicly available meta-analytic data from 1000 participants. It shows that the resting state connectivity of a seed voxel in left primary auditory cortex overlaps with primary motor and somatosensory cortices, SMA, insula, and the IPL (Fig. 2 red; retrieved from [http://neurosynth.org/locations/-44\\_-28\\_8/](http://neurosynth.org/locations/-44_-28_8/); data from Yeo et al., 2011). Furthermore, the white matter connectivity from primary motor cortex (BA 4) across 69 studies includes significant connections to primary auditory cortex bilaterally (Fig. 2 yellow; data from ALE meta-analysis done with all available white matter voxel based morphometry studies using “normal” participants and having BA 4 connections; see Section 3.2 for methods description). Thus, the functional and structural connectivity of the auditory cortex is intrinsically sensorimotor (see Dick et al., 2014 for more information about the language connectome).

Using simultaneously acquired EEG and fMRI, Giraud et al. (2007) demonstrate the specificity of intrinsic network connections like these. They found that spontaneous EEG gamma oscillations, associated more with phonemic time scales and tongue

movement, correlated with resting BOLD fluctuations in auditory and ventral premotor *tongue* regions. In contrast, theta oscillations, associated more with syllabic time scales and movement of the jaw or mouth, was associated with auditory and ventral premotor *mouth* regions.

### 2.7.2. Mechanisms

Neuroimaging studies are somewhat unique compared to other approaches reviewed in that they can be used to look at the gross functional and/or structural connectivity of the whole brain. In doing so, questions about mechanisms can be addressed by making inferences based on which regions are connected and how they are connected. For example, for the AxS model, one might hypothesize a bidirectional or at least feedback relationship between posterior ventral frontal production regions and auditory cortices. This network should show some production related specificity (as seen for individual regions in the prior section). Finally, given that the AxS model is predictive, a temporal relationship would also be expected, i.e., posterior ventral activity should precede auditory cortices. Indeed, all of these criteria for an AxS like mechanism have been shown using network analyses.

Ventral premotor regions active during both speech production and “passive” speech listening have been shown to be reciprocally connected to superior temporal cortex (Wilson & Iacoboni, 2006). Furthermore, these and other speech production related regions (including somatosensory cortex and the cerebellum) form a stronger functional network during “passive” listening to pronounceable compared to unpronounceable stimuli (Londei et al., 2007, 2010). That is, they show some measure of specificity.

Several network studies have examined direction of connectivity. One shows bottom-up or feedforward connectivity from posterior superior temporal to premotor cortices during a categorization task (Chevillet, Jiang, Rauschecker, & Riesenhuber, 2013). Most studies, however, show top-down or feedback functional connectivity. For example, multimodal neuroimaging data (MRI/EEG/MEG) shows that pre- and primary motor cortex activity (in the ventral portion by our definition) is functionally connected with, precedes, and influences speech activation in posterior superior temporal regions (Gow & Segawa, 2009). Other studies show synchrony or oscillatory behaviour that can be characterized as both feedforward and feedback functional connectivity. For example,



fMRI studies show bidirectional connections between premotor and superior temporal cortices for speech over nonspeech sounds (Osnes, Hugdahl, & Specht, 2011). Using methods with better temporal resolution (MEG), reciprocal synchrony between auditory regions (like those at the temporoparietal junction) and ventral and dorsal motor regions have been shown to underlie speech sound categorization (Alho et al., 2014). Similarly, using joint fMRI and EEG, it has been shown that there is early reciprocal activity between posterior superior temporal, parietal, post and precentral gyrus for a syllable over chirp task that is not consistent with any straightforward feedforward hierarchical organization (Liebenthal, Sabri, Beardsley, Mangalathu-Arumana, & Desai, 2013).

### 2.7.3. Dynamics

Reviewed stimulation/recording studies suggested that posterior ventral frontal regions are dynamically engaged during speech perception, e.g., their contribution to speech perception increases as a function of discrimination difficulty. Task based neuroimaging experiments can be used to test whether this dynamic engagement occurs across speech production systems. Indeed, task based studies show similar results as those for “passive” listening, i.e., posterior superior temporal regions are co-active with speech production regions (Vigneau et al., 2006). Furthermore, this activity is not necessarily confounded with motor engagement associated with task demands. For example, speech production regions (most prominently, posterior ventral frontal regions) are active for phoneme categorization tasks (Alho et al., 2012). Moreover, they show phoneme category-selectivity even when that information is task-irrelevant and this predicts behavioural performance (Chevillet et al., 2013).

Like reviewed TMS studies, task based studies show that the more difficult speech is to understand, the more speech production regions are engaged. In particular, greater activity and/or connectivity has been shown for studies (some with speech production localizers) contrasting non-native and native speech (Callan, Callan, & Jones, 2014; Callan, Jones, Callan, & Akahane-Yamada, 2004; Wilson & Iacoboni, 2006), distorted and clear speech (Du, Buchsbaum, Grady, & Alain, 2014; Osnes et al., 2011), more and less temporally compressed speech (Adank & Devlin, 2010; Poldrack et al., 2001), speech in noise with lower and higher noise levels (Binder, Liebenthal, Possing, Medler, & Ward, 2004; Scott, Rosen, Wickham, & Wise, 2004; Zekveld, Heslenfeld, Festen, & Schoonhoven, 2006), and syllable complexity, operationalized by the presence or absence of a consonant cluster (Tremblay & Small, 2011; see Adank, 2012 for a related meta-analysis).

## 3. Quantitative review

### 3.1. Overview

Across converging methods, the qualitative review suggests a specific and dynamic role of distributed speech production systems in speech perception. This review, however, may have been biased in a number of ways. First, many of the reviewed studies used methods and tasks that have some unnatural motor component to them (like engaging motor systems electrically, button presses associated with tasks, or tasks that elicit subvocal rehearsal). This might result in the appearance that speech production regions are engaged when they are normally not. This, in combination with a possible article selection bias and “file-drawer effect” (Rosenthal, 1979), leaves open the possibility that speech production systems play less of a role in speech perception than was suggested by the qualitative review. Conversely, there is another source of bias in that many of the reviewed studies tend to report “motor system” engagement during speech perception as restricted to posterior

ventral frontal regions and not the distributed networks that actually support speech production. This leaves open the possibility that speech production systems play *more* of a role in speech perception than was suggested by the qualitative review.

These sources of bias warrant an unbiased quantitative review in the form of neuroimaging meta-analyses that address these concerns. Thus, we conducted a series of meta-analyses to test three overarching though interrelated hypotheses pertaining to the “what”, “how” and “where” of the role of speech production systems in perception defined in the Introduction and suggested by the preceding qualitative review. Some of these meta-analyses included neuroimaging studies whose tasks did not include an overt motor component, addressing concerns about unnatural engagement of motor systems. All meta-analyses included all available studies, ruling out an article selection bias. The vast majority of included studies, if any, were not designed to investigate the role of production systems in speech perception, helping rule out a file-drawer effect. Meta-analyses also allowed us to examine all speech production related regions and networks and report the full extent of their engagement during speech perception.

#### 3.1.1. Hypothesis 1: Regional overlap

The first of the three sets of meta-analyses was used to test the overarching hypothesis that a distributed set of brain regions involved in speech production, extending well beyond posterior ventral frontal regions, overlap with *natural* speech perception (“where”). We define “natural” as those studies in which participants “passively” perceived speech as they might in the real world. Thus, we initially excluded studies that included metalinguistic tasks that participants would not actually do outside of the laboratory (like the alternative forced choice paradigms used in most experiments). In addition to having more real-world validity, natural speech perception avoids the confound that task based studies necessarily engage motor systems (e.g., for button pressing). That said, in testing this and all other hypotheses, we did *not* simply analyze speech perception versus rest as the resulting brain activity could have any number of alternative explanations unrelated to speech. Rather, we used only comparisons with “high-level” controls (defined below) that should have the effect on whole of “subtracting out” processes not specifically associated with speech perception.

This first hypothesis also addresses a number of more specific subhypotheses. The qualitative review suggested that the role of speech production regions in perception is specific to the computations associated with production (i.e., activity patterns are not simply the same regions performing different tasks as demonstrated by, e.g., somatotopy analyses). We test this subhypothesis by analyzing the overlap of natural speech with both speech production and soundlessly moving the articulators. Because the latter does not involve auditory feedback, it more strongly suggests that any overlap is driven by computations specific to moving the articulators particularly if there is a similar overlap for overt speech production (“how”).

The qualitative review also suggested that the role of speech production regions is dynamic, e.g. the engagement of production regions increases in relation to the difficulty of speech to be perceived. We test this subhypothesis by analyzing both natural speech (words) and the more difficult to process nonwords. Overlap with natural speech would indicate that speech production systems are ubiquitously engaged (rather than typically not engaged). Greater overlap with nonwords (e.g., pseudowords), because they are harder to process than words, would indicate that speech production systems are more or less engaged as a function of listening context (“when”; see Skipper, 2014 for support).



### 3.1.2. Hypothesis 2: Network organization

The second overarching hypothesis we tested was that speech perception regions that overlap with regions associated with soundlessly moving the articulators and speech production have a network organization (“where”). Regional overlap associated with hypothesis 1 does not necessitate that those regions form a network. For example, temporal lobe regions might form a network independent of posterior ventral frontal regions involved in articulation during perception but form a single connected network during production (as might be predicted by Hickok & Poeppel, 2007). If all regions form connected networks during perception, however, a stronger case can be made that they are working together on a common production related (or included) computational goal (“how”). In contrast to the portrait given by the “classical” and “dorsal stream” models, complex networks generally – and speech production systems more specifically – are far less fixed and more distributed. Thus, we also tested the subhypothesis that regions of overlap have the capacity to dynamically organize into many possible networks along multiple possible white matter fibre bundles or tracts (“when”).

### 3.1.3. Hypothesis 3: Dynamic network topology

The final overarching hypothesis we tested was that the topologies of these speech production and perception networks dynamically organize around the context encountered by participants (“when”). That is, we test whether the contribution of regions (or nodes) and connections associated with speech production and perception networks changes as a function of the stimuli and tasks encountered by participants in studies intended to address questions about different cognitive subprocesses associated with language. To test this and Hypothesis 2, we included in our meta-analyses *all available* speech perception, language comprehension and production studies, including those with less natural (i.e., “active” meta-linguistic) tasks. This provides the necessary variability in stimuli and tasks (compared to just listening in Hypothesis 1) to analyze their impact on network topology. In addition, this variability allows us to start characterizing the computational role of each node in networks as a function of stimuli and tasks. Specifically, using a novel text-mining procedure we described network topology as a function of “unit” size (like whether participants were listening to phonemes or sentences), associated levels of processing (like whether what they were doing with those units could be characterized as speech or semantic processing) or more production related behaviours (like moving the tongue or sequencing).

## 3.2. Materials and methods

We tested these three overarching hypotheses with three sets of neuroimaging meta-analyses using the BrainMap database (<http://brainmap.org/>). Specifically, we queried that database for experiments meeting a set of common metadata criteria and sets of criteria specific to each hypothesis (described in the following sections colloquially; exact searches in BrainMap’s taxonomy are given in the Appendix A). These queries returned X/Y/Z stereotaxic coordinate space “locations”, that is, centres of mass or peaks of functional brain activity reported in neuroimaging papers (Fox & Lancaster, 2002; Fox et al., 2005; Laird, Lancaster, & Fox, 2005). Locations that were originally published in the Talairach coordinate space were converted to Montreal Neurological Institute (MNI) space (Laird et al., 2010; Lancaster et al., 2007). Then Activation Likelihood Estimation (ALE) meta-analyses were done by modelling each MNI location as a three-dimensional probability distribution and quantitatively assessing their convergence across experiments. Significance was assessed by permutation analysis of above-chance clustering between experiments (Eickhoff, Bzdok,

Laird, Kurth, & Fox, 2012; Eickhoff et al., 2009, 2011; Turkeltaub et al., 2012). Contrasts and conjunctions (i.e., overlaps) between ALE meta-analyses were all done using ten thousand permutations to derive p-values (Eickhoff et al., 2011). All resulting ALE maps were false discovery rate (FDR) corrected for multiple comparisons to  $p < 0.05$  and further protected by using a minimum cluster size of  $160 \text{ mm}^3$  (20 voxels). All results are displayed on Freesurfer (Fischl, 2012) based surface representations of the MNI aligned Colin27 brain (Holmes et al., 1998) using SUMA (Saad, Reynolds, Argall, Japee, & Cox, 2004). The Colin27 brain was automatically parcellated into anatomical regions of interest (Destrieux, Fischl, Dale, & Halgren, 2010) to serve as a guide to the location of results on surface representations (Fig. 1).

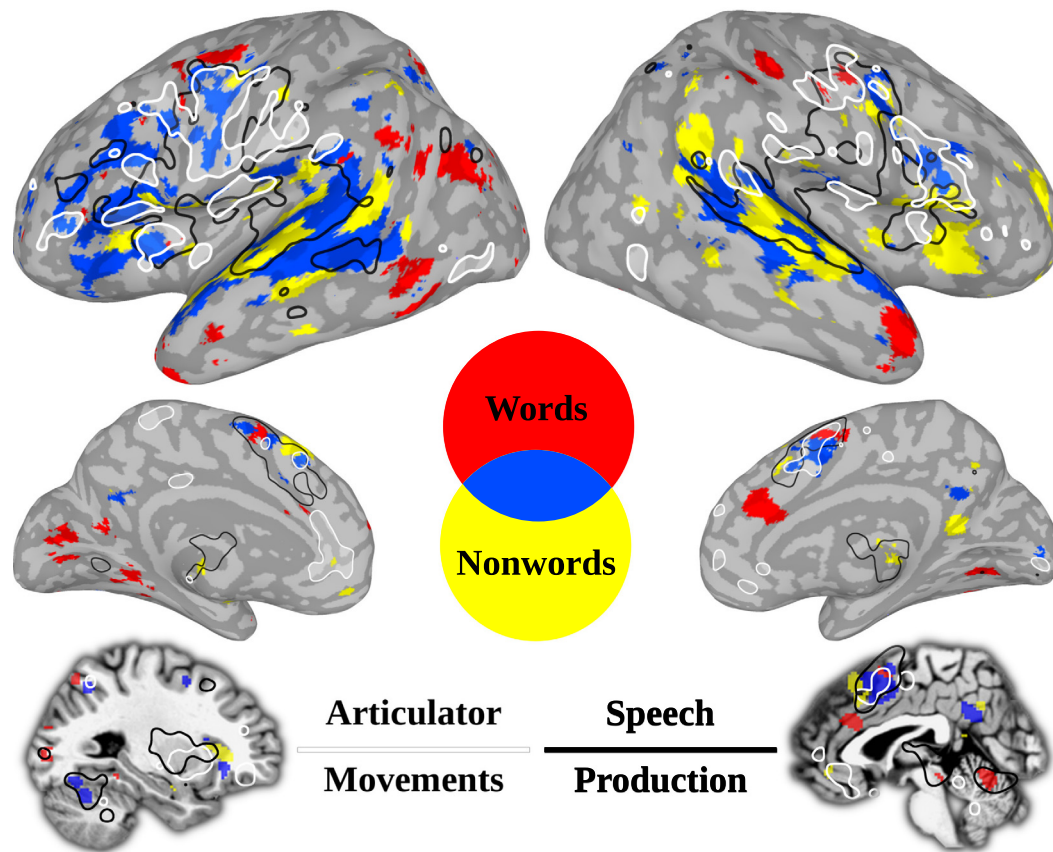
### 3.2.1. Common criteria

The BrainMap database was searched (in April, 2014) for a set of common criteria associated with all meta-analyses unless otherwise noted. In particular, experiments contributing to analyses included only “normal” participants who were right handed and older than seventeen. Locations from neuroimaging comparisons were required to be between experimental conditions and high-level controls (e.g., matched stimuli or tasks) and not rest or fixation (see rationale above). Finally, resulting brain activation was required to be positive and deactivations were excluded from analyses.

### 3.2.2. Hypothesis 1: Regional overlap

Overlap meta-analyses combined the common search criteria with four independent searches to test Hypothesis 1. First, a search to assess activation associated with natural or “passive” listening to words required there be no overt responses and that stimuli were auditorily presented words. Second, the natural listening to nonwords search required there be no overt responses and that stimuli were auditorily presented environmental sounds, music, pseudowords, reversed speech, or syllables. We use these multiple categories so as to have a comparable number of studies to “passive” word listening. Note that all categories were predicted to be harder to process than words because of familiarity (see Skipper, 2014) and that music did not include words. Third, a speech production search required all contributing experiments involve overt speaking. Fourth, to test for specificity, an articulatory movements without sound search required that the paradigm used involved overt chewing, breath-holding, drinking, eating, oral/facial, smiling, swallowing, or tasting movements and that any stimuli used were not auditory. The statistical contrast and conjunction of natural listening to words (search one) and nonwords (search two) were performed. The overlap of results with activity from speech production (search three) and articulatory movements without sound (search four) was then calculated.

We also qualitatively validated that regions of overlap from this analysis are involved in both speech production and perception by using the automated neuroimaging meta-analyses available from the neurosynth.org database (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011). At the time of writing, this database consisted of the association between 3107 high frequently terms and 413,429 activations reported in 11,406 studies. Neurosynth produces statistical z- and p-value brain maps in which each voxel corresponds to the likelihood of a given term being used in a study if activation is observed at that voxel. We found up to the first 10 most significant z-scores and associated terms related to production from the voxels in the Neurosynth maps corresponding to the centres of mass of activity from each overlapping cluster. We did the same for perception terms. The authors subjectively judged which terms were perceptual and which production related terms. In most cases this was unambiguous (e.g., “auditory” is a perceptual term whereas “production” is a production term). In many



**Fig. 3.** Neuroimaging meta-analyses of non-speech related movement of the articulators, speech production, and natural or “passive” listening. Red and yellow colours correspond to brain activity associated with listening to words versus nonwords relative to high-level control conditions, respectively. The conjunction/overlap of this activity is in blue. The outlined regions correspond to activity patterns associated with speech production (black outlines) and moving the articulators without producing speech (white outlines) relative to high-level control conditions. Sagittal slices show the same results but include subcortical regions like the cerebellum. The significance level for all analyses was .05, false discovery rate corrected for multiple comparisons with a cluster size of 160 mm<sup>3</sup> (20 voxels). See Fig. 1 for the location of activity patterns and Table 2 for the location and description of overlapping clusters.

cases perceptual and production terms are not language related (e.g., “pain” is a perceptual and “tapping” a production term) because this is a data-driven approach across any study. This descriptive analysis is intended to illustrate that regions showing production and perception overlap in the ALE analysis are also described as being both production and perception related using an independent method involving a large variety of studies. This qualitative analysis might also provide a suggestive window into the computations performed in each region.

### 3.2.3. Hypothesis 2: Network organization

To test the second hypothesis, we conducted ALE co-activation meta-analyses to find which networks are associated with overlapping production and perception regions from Hypothesis 1 results and the variability of the connectivity of those networks. These meta-analyses assume that regions that reliably co-activate with a seed region have formed a network with that region (see Eickhoff et al., 2011). Specifically, we started with the overlap of activity associated with speech production and naturally listening to sounds (the intersection of all colours and the black outline in Fig. 3) and divided these into a set of probabilistically defined cytoarchitectonic regions when available or anatomically defined regions. We chose “core” speech production overlap regions (see Section 2.1) but excluded “auditory regions” (like posterior superior temporal cortex) because, if speech production regions contribute to speech perception they should form a network with

these regions. Thus, we divided the original overlapping activity pattern into seven cytoarchitecturally defined regions, BA 45, BA 44, Lateral BA 6, Medial BA 6, BA 4 (including areas 4a and 4p), BAs 3/1/2 (including areas 3a and 3b) and the IPL (including areas HIP1, HIP2, HIP3, PFPop, PFT, PF, PFM, PFCM, PGA, and PGp; Eickhoff et al., 2005) and two anatomically defined regions, the insula (Makris et al., 2006) and cerebellum (Diedrichsen, Balsters, Flavell, Cussans, & Ramnani, 2009; see Fig. 1). These nine regions of interest (ROIs) accounted for 55% of the total original activity in the regional overlap analysis. Including the superior temporal gyrus, thalamus, and visual cortex would account for 93% of the activity.

Next, we returned to the BrainMap database and extended our search criteria in order to include a broader range of language stimuli and tasks (compared to natural or “passive” listening) for three additional sets of network meta-analyses. Specifically, criteria for all three searches included the common search criteria and the requirement of activity in one of the nine ROIs. First, we searched for general language experiments that had conditions that did not involve oral/facial responses and were in the cognitive behavioural domain of language and phonology, semantics, speech, or syntax. Second, we searched for overt speech production experiments by searching in the behavioural domain of speech execution. Third, we searched for covert speech production experiments with the criteria that experimental paradigm class was covert naming, covert reading, covert recitation/repitition,

**Table 1**

Descriptive counts of all conducted meta-analyses. “Experiments” are “a comparison of two or more brain images that result in a statistical parametric image”. “Conditions” are “the set of experiences or tasks the subjects undergo”. “Locations” is the number of X/Y/Z coordinates given per paper. Quotes and further descriptions can be found at <https://www.brainmap.org/taxonomy/>.

	Meta-analyses search results				
	Papers	Participants	Experiments	Conditions	Locations
<b>Regional overlap analyses</b>					
<i>Natural speech perception</i>					
Words	28	338	96	107	801
Nonwords	31	446	85	103	811
<i>Production</i>					
Articulatory movement	19	255	48	49	291
Speech production	28	343	50	101	654
<i>Network analyses</i>					
<i>Language subprocesses</i>					
BA 6 Lateral	55	759	78	140	1267
IPL	53	741	77	142	1151
Insula	56	832	91	171	1335
BA 6 Medial	93	1379	181	298	2453
BA 44	101	1500	210	337	2432
BA 45	100	1358	177	311	2041
Cerebellum	17	286	21	53	379
3/1/2	16	231	18	48	222
4	19	292	24	46	391
Mean =	56.67	819.78	97.44	171.78	1296.78
<i>Speech production: Overt</i>					
BA 6 Lateral	27	317	40	84	759
IPL	28	329	42	85	926
Insula	19	241	29	57	815
BA 6 Medial	31	397	48	100	1083
BA 44	20	262	36	66	901
BA 45	18	247	28	50	755
Cerebellum	17	330	44	80	1095
3/1/2	31	383	51	102	997
4	29	354	46	92	909
Mean =	24.44	317.78	40.44	79.56	915.56
<i>Speech production: Covert</i>					
BA 6 Lateral	35	474	60	100	996
IPL	24	285	34	65	579
Insula	25	309	41	74	584
BA 6 Medial	48	630	85	140	1346
BA 44	52	740	110	169	1439
BA 45	45	507	82	139	878
Cerebellum	8	104	10	27	138
3/1/2	9	92	12	28	200
4	14	233	21	39	347
Mean =	28.89	374.89	50.56	86.78	723

covert word generation or covert word stem completion. We included experiments using covert production paradigms to further assure that network interactions that involve auditory and somatosensory cortices were not simply involved because participants could hear or feel themselves talking. We then took the overlap of all of the resulting network maps (from searches one, two, and three) while preserving the number of cognitive phonology, semantics, speech, and syntax networks that contributed to each voxel. This map is intended to represent the network variability associated with the many cognitive subprocesses theoretically involved in understanding and producing language in regions that are active in both speech production and natural speech perception.

In addition to these functional connectivity analyses, we also identified the structural connectivity underlying overlapping speech production and perception activity from Hypothesis 1. To do this, we took the overlap of activity in the nine seed regions with the maximum probability tensor maps provided by the International Consortium of Brain Mapping (ICBM), obtained from 81 unimpaired participants (Mazziotta et al., 2001). We then counted the number of contributing voxels for each white matter tract.

### 3.2.4. Hypothesis 3: Dynamic network topology

To test the third hypothesis, we quantified how the topology of networks involved in both speech production and perception changes with context (i.e., stimuli and tasks). To do this, we developed a new data-driven text mining procedure that can be used to classify the networks associated with each seed region as a function of their overall similarity and their dissimilarity with respect to specific stimuli and task related words. Specifically, for each paper contributing to each of the Hypothesis 2 meta-analyses, we automated the retrieval of a description of the methods (from BrainMap) and the title and abstract of that paper from PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>). The returned text was converted to lowercase, numbers and punctuation were removed, common English stop words (e.g., “a” and “the”) and stop words peculiar to neuroimaging (e.g. “analysis” and “mni”) were removed, and the document was stemmed (e.g., “syllables” and “syllable” simply become “syllabl”). We then created a corpus by seed region of all of the terms in common for the phonology, semantics, speech, and syntax studies and the overt and covert speech production co-activation meta-analyses (i.e., excluding terms that do not appear in all papers) and counted the frequency

**Table 2**  
Regional overlap for non-speech related movement of the articulators, speech production, and natural or “passive” listening (see Fig. 3). The overlap of listening and movement of the articulators is only shown where clusters overlap speech production (bold). “Voxels” is the number of significantly overlapping 2 mm<sup>3</sup> voxels. “Production” and “Perception terms” refers to the most significant terms associated with activity at the X, Y and Z centres of mass in an independent set of 3107 meta-analyses z-score maps available from neurosynth.org (individual z-scores can be looked up by entering coordinates at, e.g., [http://neurosynth.org/locations/08\\_04\\_09](http://neurosynth.org/locations/08_04_09)). “Other” refers to cytoarchitectonic region labels, including Brodmann’s Areas. See Fig. 1 for the location of regions.

Overlap of natural speech perception, speech production, and non-speech related movement of the articulators									
MNI centres of mass				Regions		Other	Perception terms	Production terms	
#	Voxels	X	Y	Z	~Increasing distance up to 5 mm		~Decreasing association (average z-scores)		
1	1412 <b>28</b> <b>30</b>	-55 <b>-50</b> <b>-58</b>	-25 <b>-14</b> <b>-32</b>	7 <b>7</b> <b>21</b>	L	Superior temporal gyrus, planum temporale, transverse temporal gyrus, supramarginal gyrus, middle temporal gyrus, subcentral gyrus/sulcus	41, 22, 42, OP 1, TE 1.0, TE 1.1, TE 1.2	Auditory, sounds, audiovisual, listening, speech, acoustic, tones, music, spoken, multisensory (11.62)	Articulatory, motor, sensorimotor, production, sequences, speech production, speaking, tapping, tactile (4.64)
2	1276 <b>125</b>	58 <b>59</b>	-23 <b>-35</b>	6 <b>17</b>	R	Superior temporal gyrus, planum temporale, transverse temporal gyrus, supramarginal gyrus	41, 42, 22, TE 1.0, OP 1	Auditory, sounds, audiovisual, speech, listening, voice, acoustic, vocal, listened, pitch (10.97)	Production, speech production, action, motor network, motor (5.10)
3	1091 <b>323</b> <b>23</b>	-45 <b>-37</b> <b>-42</b>	16 <b>18</b> <b>18</b>	7 <b>4</b> <b>21</b>	L	Pars triangularis, pars opercularis, insula precentral gyrus, insula, inferior frontal gyrus	45, 44, 13, 47	Language, sentences, semantic, linguistic, syntactic, word, verb, lexical, painful, phonological (7.55)	
4	742 <b>335</b>	0 <b>2</b>	12 <b>12</b>	52 <b>50</b>	L/ R	SMA, superior frontal gyrus, medial frontal gyrus, anterior middle cingulate	6, 8, 32	Task, demands, working, working memory, phonological, verbal, word, calculation, maintenance, reading (6.11)	Motor, generation, tapping, execution (4.77)
5	419 <b>544</b>	-51 <b>-50</b>	-5 <b>-1</b>	31 <b>31</b>	L	Precentral gyrus, central sulcus, postcentral gyrus, inferior frontal gyrus	6, 8, 4p, 3a, 4a, 3b, 6, hIP1	Letter, calculation, expertise, phonological, vocal, speech, language (4.12)	Motor, sensorimotor, speech production, overt, production, planning, articulatory, execution, sensory motor, movements (7.34)
6	312 <b>459</b>	50 <b>52</b>	12 <b>13</b>	2 <b>16</b>	R	Rolandic operculum (insula), pars opercularis, superior temporal gyrus,	47, 13, 22, 44, 45	Signal task, stop signal, pain, handed, working memory, working (4.39)	Sequences, motor, execution, motor imagery, motor task, mirror neuron, somatosensory (4.36)
7	290 <b>39</b> <b>37</b>	51 <b>45</b> <b>52</b>	0 <b>3</b> <b>-4</b>	39 <b>30</b> <b>38</b>	R	Precentral gyrus, middle frontal gyrus, inferior frontal gyrus, postcentral gyrus	6, 9, 4a	Visual word, tasks, speech, attention, speech perception, reading, difficulty, auditory, vocal, word form (4.45)	Motor, movements, eye movement, sensorimotor, saccade, execution, sequential, speech production, oral, production (5.08)
8	276	7	-62	-12	R	Cerebellar culmen, culmen of vermis, declive, lingual gyrus	V, VI (hem, vermis)	Rhythm, imagery, externally, sequence, coordination, paced, reorganization, advanced, abilities, reward, anticipation (4.88)	Tapping, motor, movement, force, motor imagery, finger tapping, hand movements, finger movements, sensorimotor (5.99)
9	228	-10	-16	6	L	Thalamus, medial dorsal nucleus, ventral lateral nucleus, mammillary body, ventral posterior lateral nucleus, ventral posterior medial nucleus, parietal, premotor, temporal and prefrontal thalamus		Pain, chronic pain, network, nociceptive, painful, auditory, noxious (4.53)	Motor, movement, finger tapping, sensorimotor, tapping, motor task, paced, movements, somatosensory (4.81)
10	111 <b>21</b>	-24 <b>-21</b>	-63 <b>-62</b>	-26 <b>-27</b>	L	Cerebellum, declive, culmen, uvula	Lobule VI (hem), lobule VIIa crus I (hem)	Music, loop, simultaneously, rhythm, rehearsal, vocal, musical, speech, pitch, rehearsal (4.74)	Motor, speech production, production, finger movements, articulatory, overt, movements, motor control, execution, sequence (6.89)
11	85	-32	-70	-16	L	Fusiform gyrus, lingual gyrus, declive, middle occipital gyrus	18, 19, hoc4v (v4), lobule VI (hem), lobule VIIa crus I (hem)	Visual, array, face, visual field, sighted, objects (4.39)	
12	77	32	21	4	R	Insula (circular sup sulcus, short gyrus), putamen, claustrum, inferior frontal gyrus	13, 47, 45	Task, maintenance, working, working memory, memory wm, pain, wm, demands, load, network (4.70)	
13	75	10	-23	8	R	Thalamus, medial dorsal nucleus, pulvinar, ventral posterior medial nucleus, mammillary body, lateral posterior nucleus, lateral dorsal nucleus, temporal, prefrontal and parietal thalamus		Cross sectional, sectional, pain, stimulation, sparse, electrical, network, noxious (4.44)	Finger movements, movement, movements (4.16)
14	66	14	-17	6	R	Thalamus, ventral posterior lateral nucleus, mammillary body, ventral lateral nucleus, ventral posterior medial nucleus, medial dorsal nucleus, lateral posterior nucleus, premotor, prefrontal,		Pain, paced, sampling, nociceptive, integrate, noxious, coordination, networks, stimulation, painful (4.69)	Motor, sensorimotor, movement, somatosensory, finger tapping, tapping, movements, parkinson disease, disease pd, finger movements (6.22)



Table 2 (continued)

Overlap of natural speech perception, speech production, and non-speech related movement of the articulators							
MNI centres of mass				Regions	Other	Perception terms	Production terms
#	Voxels	X	Y	Z	~Increasing distance up to 5 mm		~Decreasing association (average z-scores)
15	55 58	54 54	-12 -13	40 40	R parietal, motor, and somatosensory thalamus Postcentral gyrus, precentral gyrus, central sulcus	4, 6, 3, 3b, 1, 4a, 4p 6	Vocal, audio, nociceptive, sensory, noxious (4.21) Somatosensory, motor, sensorimotor, coordination, movement, speech production, output, parkinson disease, production, tactile (5.64)
16	43	-41	-14	34	R Postcentral gyrus, central sulcus, precentral gyrus,	6, 3a, 4p, 3b	Vocal, verbal working, speech (4.84) Motor, production, speech production, producing, somatosensory, sensorimotor, execution, produced, tapping, movement (4.72)
17	36	52	9	27	R Pars opercularis, precentral gyrus/sulcus, pars triangularis, middle frontal gyrus	44, 45, 9	Calculation, observation, spatial, difficulty, visually, demands, tasks (4.25) Execution, motor (3.77)
18	32	-24	4	50	L Middle frontal gyrus, superior frontal gyrus, dorsal precentral gyrus, cingulate gyrus, medial frontal gyrus	6, 32, 24	Load, working memory, tasks, paced, spatial, watched, control conditions, memory, difficulty, verbal (5.02) Sequences, planning, execution, motor, movements (3.99)
19	29	4	-31	-10	R Culman, cerebellar vermis	Lobules I–IV (hem)	Pain, expectation, modulatory, distress, painful, noxious, reward (4.76)
20	23	-65	6	8	L Precentral gyrus, postcentral gyrus, pars opercularis, rolandic operculum, subcentral gyrus/sulcus, superior temporal gyrus	6, 44, op 4, 22	Imagery, imagined, words, phonological (4.35) Motor imagery, finger movements, production, speech production, motor execution, tactile, somatosensory, movement (4.27)
21	22	-28	-90	2	L Middle occipital gyrus, inferior occipital gyrus, lingual gyrus, cuneus	18, 19, hoc3v (v3v)	Visual, words, verbs, face (4.30)
22	21	-51	8	38	L Ventral precentral gyrus, middle frontal gyrus, inferior frontal gyrus	44, 6, 9, 8	Task, phonological, bilinguals, reading, processes, language, observation, semantic, speakers, rehearsal (4.69) Motor, mirror, execution (4.32)
	6721	-7	-11	14	L		

of each term. This resulted in one corpus with nine documents (one for each seed region).

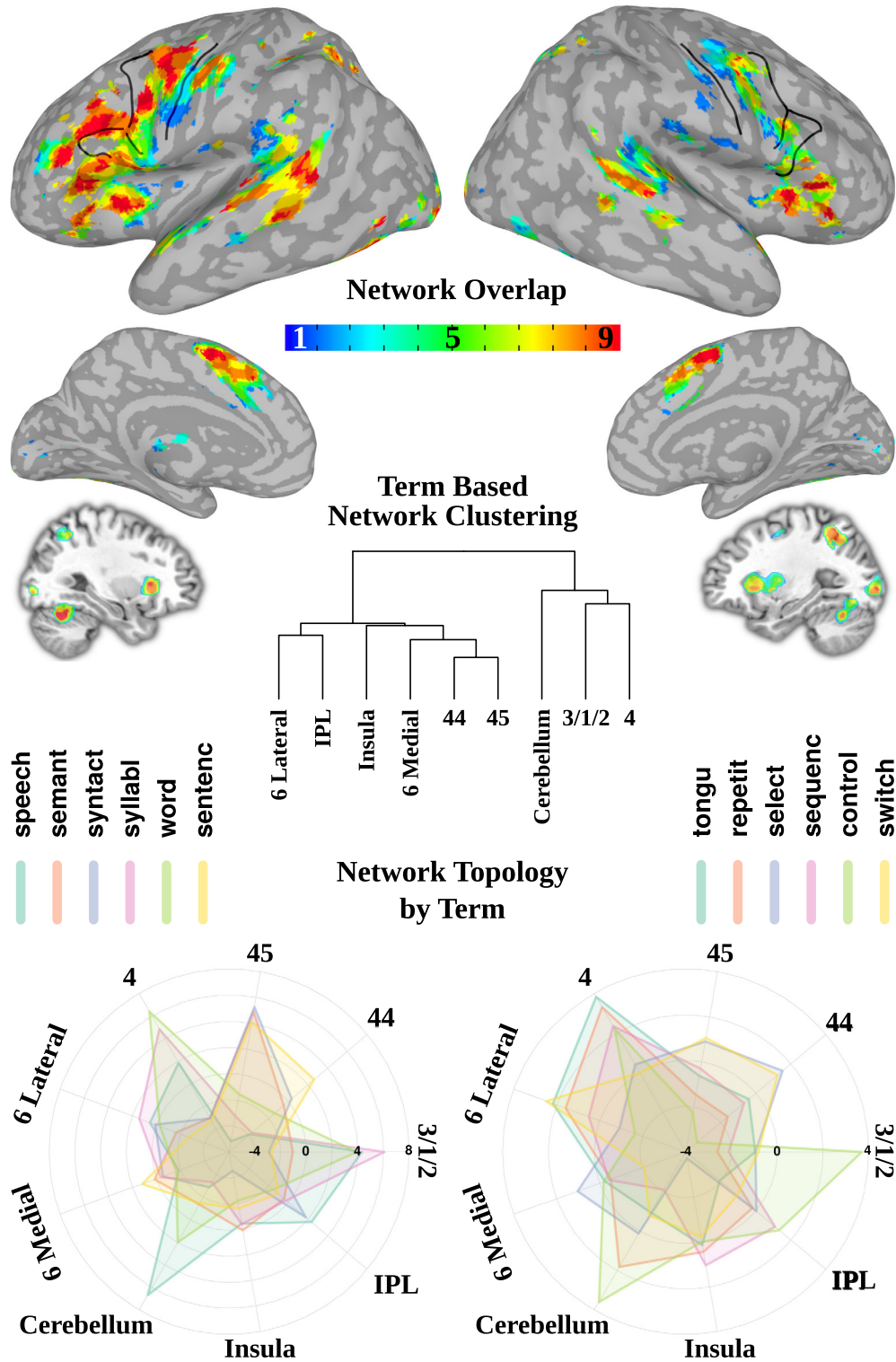
Using (R Core Team, 2014), we then performed hierarchical clustering and some basic statistics on term frequencies, mostly using ‘RcmdrPlugin.temis’ (or ‘R.TeMiS’ for ‘R Text Mining Solution’, Bouchet-Valat & Bastin, 2013). Specifically, regions were hierarchically clustered using Ward’s method based on a  $\chi^2$  distance between documents to reveal the term based network similarity between regions. We also went beyond classifying the relative similarity of nodes in these networks to understand how the relative topology of networks might change as a function of context (i.e., stimuli and tasks), giving us an insight into computations performed by nodes in networks. To do this, we analysed six stemmed terms specific to various units (i.e., “syllabl”, “word” and “sentence”) and corresponding levels of linguistic analysis (i.e., “speech”, “semant” and “syntact”) and six stemmed terms associated with various structures and process level descriptions often associated with speech production (i.e., “tongu”, “repetit”, “select”, “sequenc”, “control”, “switch”). We then graphed the t-value for each term for each region in a radial plot. The t-value is a metric of that terms’ frequency compared to what would be expected given the length of a regions’ associated documents and the global distribution of terms in the corpus (see Bouchet-Valat & Bastin, 2013). We predicted that all terms would load on different nodes differentially, demonstrating that network topology shifts as a function of context.

### 3.3. Results

#### 3.3.1. Hypothesis 1: Regional overlap

A remarkable overlap was observed between the activity patterns associated with naturally listening to both words and non-word sounds with articulatory movements used to overtly produce speech or make movements without producing sound (Fig. 3; Tables 1 and 2). Most of the overlap was in the pars opercularis, ventral precentral sulcus and gyrus, ventral central sulcus, ventral postcentral gyrus, medial superior frontal gyrus (the SMA), anterior insula, and mostly more posterior superior temporal and inferior parietal cortices. Subcortical structures include the thalamus, putamen, and cerebellum. Activity in each of these regions is independently shown to be associated with both high-frequency production and perception related terms mentioned in 11,406 studies (Table 2, two rightmost columns).

Words tended to produce significantly less (and nonwords more) activity in these overlapping regions, particularly in posterior superior temporal regions (yellow in Fig. 3). In the overlap of speech production and word and nonword speech perception, words and nonwords produce similar activity in 58% of all voxels. Of the remaining 42% of voxels, 76% are significantly more active for nonwords. When restricted to the transverse temporal, superior and middle temporal gyri and sulci, and planum temporale and polare, words and nonwords produce similar activity in 53%



**Fig. 4.** Network organization and topologies. Network co-activation meta-analyses for studies of both overt and covert speech production and phonology, semantics, speech, and syntax were conducted using nine seed regions active during both speech production and natural listening (see Fig. 3 and Table 2). The number of networks that each voxel participates in is indicated with a blue-to-red colour spectrum; on one end of the spectrum, regions in blue participate in one network, whereas, on the other end, regions in red participate in nine (top, “Network Overlap”). The black outline serves as a guide to the location of posterior ventral frontal regions (see Fig. 1). The Chi-square distance between seed regions (calculated from a matrix of the term frequencies appearing in all articles contributing to all of the network meta-analyses for each seed) were hierarchically clustered to provide a general description of network relationship between regions (middle, “Term Based Network Clustering”, numbers refer to a seed’s Brodmann’s area, see Fig. 1). Radial plots were used to represent how network topologies change as a function of context (bottom, “Network Topology by Term”). Stemmed terms were analysed by seed region to show how topologies change as a function of different units and associated levels of linguistic analyses (bottom left) and mechanical and mechanistic operations sometimes associated with producing speech (bottom right). The axes of the radial plots are t-values, a metric of how frequently a term was used in combination with a seed given the number of terms and global distribution of those terms across seed regions.

of voxels. Of the remaining 47% of voxels, 99% are significantly more active for nonwords.

### 3.3.2. Hypothesis 2: Network organization

While the prior overlap analyses are suggestive, these results do not mean that regions associated with moving the articulators or speech production form a network nor does it say anything about how fixed or variable those networks are. Network analyses illustrate that a large variety of phonology, semantics, speech, and syntax networks overlap with both overt and covert speech production networks (in regions involved in both speech production and natural speech perception; Fig. 4 Top; Table 1). These networks include nodes in “auditory regions” despite that these were not included as seeds. The networks have different but overlapping distributions, including network connectivity involving regions beyond ventral frontal and posterior superior temporal regions and including connectivity variously involving primary auditory cortex, anterior superior temporal cortices, parietal cortices, the insula, pars opercularis and triangularis, and the SMA (among others). A quantification of the fibre tracts running through the nine seed regions used in the network analysis suggests that, though the bulk of these functional network interactions are mediated by the superior longitudinal fasciculus (including the arcuate fasciculus), these more variable networks can make use of other fibre tracts (Table 2).

### 3.3.3. Hypothesis 3: Dynamic network topology

Hierarchically clustering of term based descriptions of ROI results suggest that lateral BA 6 is most closely related to IPL while the insula is more closely related to medial BA 6 which is itself more closely related to BAs 44 and 45. Finally, the cerebellum, areas 3/1/2 and 4 are closely related (Fig. 4 Middle). These results cohere well with accepted wisdom regarding these regions. For example, BA 4 and BAs 3/1/2 both contain primary motor and somatosensory cortices. Lateral BA 6 and IPL are often described as communicating with one another, e.g., via efference copy (e.g., Skipper et al., 2007). BAs 44 and 45 are often discussed as “higher” phonological and semantic regions (e.g., Poldrack et al., 2001). These results suggest the veracity of this method and also the gross computational similarity of nodes in these networks.

We also used the text-mining approach to understand the relative difference in topologies of networks as a function of specific stimuli and tasks. Consistent with the hierarchical clustering, lateral BA 6, cerebellum, and BAs 3/1/2 play a stronger role in processing associated with smaller levels and units of analysis (“speech”, “syllabl” and “word”) whereas 44 and 45 play a larger role at higher levels and corresponding units (“semant”, “syntact” and “sentence”; Fig. 4 bottom left radial plots). Correspondingly, subprocesses typically associated with speech production load differentially on different regions. For example, “repetit” and “sequenc” load more on BA 4 and lateral 6 whereas “select” loads more on BAs 44 and 45 and medial 6 (Fig. 4 bottom right radial plots). Overall, the topologies of overlapping speech production and perception networks tend to vary dramatically, with different regions in the distributed “core” production network playing more or less of a role in different language subprocesses.

## 4. Discussion

### 4.1. Summary

A vast amount of converging evidence presented in both qualitative and quantitative reviews unambiguously demonstrates that regions and networks of the brain supporting speech production play ubiquitous roles in speech perception (the “if” and “where”

questions). Furthermore, these roles are specific to representations, computations, and mechanisms associated with producing speech (the “how” questions) and are dynamically determined by listening context (the “when” questions).

### 4.1.1. Qualitative review

The qualitative review demonstrates that speech production is a complex process that involves multiple distributed sensorimotor regions and networks. Evidence from behavioural studies arguing that these networks are not involved in speech perception because animals and infants can perform speech tasks despite having never produced speech are not logically sound. In fact, vocal production networks seem to actually play a role in such tasks. Furthermore, there is a great diversity of behavioural findings in humans of all ages that suggests that speech production systems play a role in perception. Likewise, a wide variety of computational models of sound perception perform better when production related information or mechanisms are incorporated into those models. Indeed, vocal production related brain systems play a role in sound perception in various rodent, bird, and primate species. In human primates, speech perception deteriorates with a wide range of damage to speech production systems caused by stroke, focal excision for epilepsy, cerebral palsy, and Parkinson’s disease. Scores of direct cortical, TMS and tDCS studies confirm a causal role for regions involved in producing speech in perception in unimpaired individuals. ECoG and TMS work suggests the specificity of the computations involved (e.g., showing sensorimotor responses and somatotopy) and dynamic nature of these contributions (e.g., results are stronger for speech in noise). Neuroimaging studies confirm this picture without unnatural excitatory or inhibitory stimulation of the brain. Again, results show specific engagement of production systems (e.g., related to place of articulation and somatotopy) and dynamic contextual variability (e.g., being more or less engaged for specific listening tasks). Neuroimaging also suggests an active mechanism by showing that the direction of effects are reverberatory, though most often characterized as top-down or feedback and predictive (e.g., with posterior ventral frontal regions influencing auditory regions).

### 4.1.2. Quantitative review

The qualitative review might have been misleading for a number of reasons. For example, there may be a publication bias (and corresponding file-drawer effect) for demonstrating the engagement of “the motor system” in speech perception. However, unbiased quantitative neuroimaging meta-analyses confirm and extend results of the qualitative review in a number of directions. First, there was a considerable overlap of soundlessly moving the articulators with naturally (i.e., “passively”) listening to speech (Fig. 3). Second, this overlap is not specific to posterior ventral frontal regions (like “Broca’s area”), which many individual studies in the qualitative review fixate on but, rather, includes a large distributed set of regions corresponding to the large set of regions known to be variously involved in speech production (Fig. 3; Table 2). That these regions are each involved in both production and perception was verified using an independent set of large-scale meta-analyses (Table 2, two rightmost columns). Third, though speech production regions appear to be ubiquitously engaged, their contribution is not static: activity increases with the difficulty associated with more natural listening (with nonwords producing more activity than words; Fig. 3). Fourth, the network architecture formed by regions is not fixed: there are many possible functional speech production networks that participate in speech perception through multiple possible fibre tracts (Fig. 4 top; Table 3). Finally, the contribution of any given network varies considerably with context (Fig. 4 polar plots). Indeed, the entire topology of the speech production networks that contribute to

**Table 3**  
White matter tracts connecting regions overlapping speech production and natural or “passive” listening (from Fig. 3).

Tract	White matter tract overlap									
	BA 45	BA 44	BA 6 Lateral	BA 6 Medial	BA 4	Insula	BA 3/1/2	IPL	Cerebellum	
Anterior thalamic radiation	192 (26.41%)	150 (3.78%)	1 (0.05%)	1 (4.76%)					21 (80.77%)	
Corticospinal tract	251 (34.53%)	150 (3.78%)				260 (62.05%)				
Inferior fronto-occipital fasciculus	15 (2.06%)	50 (1.26%)						48 (3.39%)	5 (19.23%)	
Inferior longitudinal fasciculus	156 (21.46%)	3557 (89.66%)	1858 (99.95%)	20 (95.24%)	870 (100%)	80 (19.09%)	1507 (100%)	1370 (96.61%)		
Superior longitudinal fasciculus	113 (15.54%)	60 (1.51%)				79 (18.85%)				
Uncinate fasciculus										
Total	727	3967	1859	21	870	419	1507	1418	26	

speech perception dynamically fluctuate as a function of the linguistic unit and associated level of analysis encountered by the brain (Fig. 4 left polar plots). Likewise, different network topologies during speech perception correspond to different computations associated with subprocess of speech production (Fig. 4 right polar plots).

#### 4.2. Analysis-by-synthesis

As this review began, the question of whether “the motor system” plays a role in speech perception matters because speech scientists, after a great deal of research, have been unable to specify how we achieve perceptual constancy. That is, to provide a theory of how we hear some sounds as speech despite the infinite diversity of acoustic patterns arriving in the cochlea. If this seems esoteric, one simply need recall the experience of hearing an unfamiliar foreign language for the first time. What you hear are not phonemes, syllables, words, or phrases but, rather, a mostly continuous stream of undifferentiated sound. Only after some experience with those sounds do we acquire the illusion that we hear linguistic units. Thus, the brain must make use of knowledge acquired through learning to constrain interpretation of sound. This review strongly suggests that distributed brain systems for producing speech are ubiquitously and dynamically involved in this process. The distributed, sensorimotor and dynamic nature of this contribution implies that any strictly auditory or motor theory of speech perception is wrong. Indeed, the “motor theory of speech perception” (Lieberman & Mattingly, 1985) never received overwhelming support from speech scientists and was criticized from the outset (Galantucci, Fowler, & Turvey, 2006; Lane, 1965; see critiques referenced in McGowan & Faber, 1996). Rather, results of this review suggest the presence of the “analysis-by-synthesis” (AxS) model of speech (Halle & Stevens, 1962; Stevens, 1960; Stevens & Halle, 1967). To see why, the AxS mechanism is next described in more detail in the words of the authors (for more extensive treatments, see Bever & Poeppel, 2010; Poeppel & Monahan, 2011). The AxS model will then be discussed with regard to two fundamental questions: (1) Why might speech production systems be involved in speech perception as opposed to some other mechanism and (2) the necessity of this involvement.

##### 4.2.1. Overview

The AxS model is the conceptual heir of von Helmholtz’s (1867) proposal that vision is necessarily a process of “unconscious inference” due to the poverty of the visual stimulus presented to the eye. Similarly, the AxS model accounts for a poverty of variance in the acoustic makeup of speech signals reaching the cochlea by dynamically appealing to speech production mechanisms for perceptual assistance. Specifically, in the words of the authors, where “V” denotes a “vocal mechanism” for producing speech, “P” is the “abstract representation of an utterance” like phonemes and “A” is “auditory patterns” (Stevens & Halle, 1967, p. 99):

... the auditory patterns that result from an acoustic speech signal at the ears undergo some preliminary analysis, and as a consequence of this preliminary analysis, together with contextual information derived from analysis of adjacent portions of the signal, an hypothesis is made in a control component concerning the abstract representation of the utterance. This hypothesized sequence of units, which we have labeled  $P_{\text{trial}}$ , is then operated on by the phonological rules (the same ones that are used in the generation of speech) to yield a pattern  $V_{\text{trial}}$ . During speech production, this pattern would normally give rise to motor commands that would lead to articulatory activity and sound generation. During speech perception, however, this path is inhibited, and instead an equivalent auditory pattern  $A_{\text{trial}}$  is



derived from  $V_{\text{trial}}$ . ... This computed auditory pattern is compared with the pattern under analysis. If there is agreement between the two patterns, then the trial sequence of elements was correct and this sequence is read out for processing at higher levels. If there is a difference between the patterns, the control component takes note of this error and assembles a new trial sequence of units. This process continues until the trial abstract representation gives a match at the comparator, in which case the correct output sequence is established at P.

That is, when there is substantial phonetic ambiguity or lack of invariance, the non-deterministic relationship is resolved (and perceptual constancy achieved) by using knowledge about how to produce hypothesized phonemes derived from context.

#### 4.2.2. Why speech production systems?

As Fant said, “if auditory analysis in the hearing process has proceeded so far as to allow the proposed articulatory matching [as in the AxS model], the decoding could proceed without an articulatory reference” (as quoted in Lane, 1965). That is, why make reference to speech production systems at all when, for example, some acoustic pattern matching process might suffice? Most of the answers given to this question from various theories centre around two topics: (1) Lack of invariance and (2) Language parity (following explication of these, we add a third). First, it is suggested that the variance problem associated with the acoustic signal is solved by making reference to “the motor system” in part or full. Early versions of the “motor theory of speech perception” claimed that the problem was solved in full by reference to invariant motor commands. Indeed, this position was reasonably posited because of (a) an inability of machines reading an acoustic alphabet to be understood and (b) supporting behavioural research suggesting that speech perception abilities often track aspects of articulation rather than the acoustic signal (Galantucci et al., 2006). However, among other problems, motor commands are likely as or nearly as variable as the acoustic signals themselves. Thus, models that appeal to “the motor system” in part for aid in the form of constraints from speech production systems, as in the AxS model, seem more reasonable while accounting for such empirical evidence.

Second, speech perception abilities evolved, develop, and are used in bi-directional settings in which listeners are producers of speech and vice versa. Thus, a parity must be maintained between acoustic representations and the presumably quite different representations needed for speech production (i.e., the messages received must be like those sent). As such, it was proposed to be more economical to have one shared production based currency as developed more extensively in later versions of the “motor theory of speech perception” (Liberman & Whalen, 2000). Alternately, at minimum, interacting systems are required as in the AxS model. The latter clearly coheres better with neurobiological accounts of sensorimotor systems as reviewed.

A third topic might be subsequently added to the general set of answers given to the question of why speech production systems are involved in perception: (3) Prediction. Specifically, the AxS model was particularly prescient in that “hypotheses” are really predictions of forthcoming speech sounds originating from speech production systems. Indeed, there has been a rash of predictive speech models in which predictions derive from “the motor system” in some manner (Arnal & Giraud, 2012; Callan, Callan, Gamez, Sato, & Kawato, 2010; Gambi & Pickering, 2013; Hickok, Houde, et al., 2011; Kuhl et al., 2014; Rauschecker & Scott, 2009; Skipper, 2015; Skipper et al., 2006). But why would speech production systems implement those predictions? After all, it has more generally been claimed that brains are “essentially prediction machines” (Clark, 2013) and shown that predictions can and do

come from near anywhere in the brain (Bubic, Von Cramon, & Schubotz, 2010).

We have proposed that this is because speech production systems implement a set of processes that can be reused to transform any form of available context (not just acoustic) into smaller units of sound (Skipper, 2014, 2015; Skipper et al., 2006). In particular, when we speak, we must select words (given that there are multiple things we could say), sequence them, and, while producing segments, predict the sensory consequences of individual motor plans. These predictions are an important part of the production process that permit learning and allow for the adjustment of the articulators in real-time when perturbations occur. If these processes are reused during perception, hypotheses deriving from context could be similarly selected, sequenced into individual motor programs, and then used to activate their associated sounds (through feedback or efference copy) to constrain interpretation of information arriving in auditory cortices. For example, an observed iconic flapping gesture made with the hands might activate words like “flapping” and “bird”. The latter might be selected in the context of speech like “Is that a . . .”, sequenced, and used to activate the “b” in bird which can then serve as a constraint. Thus, if the whole set of processes are usurped during speech perception, the brain would have a general purpose mechanism for transforming multiple contextual constraints into sounds that can be used to achieve perceptual constancy (see Skipper, 2014 for empirical support).

#### 4.2.3. Is it necessary?

Many words have been expended to address the question of whether “the motor system” is “necessary” for speech perception. In addressing this question, the authors of the AxS model state that perception *might* be achieved without this production based mechanism but that the described “matching process is always employed as a check”. Consistent with this position, the results reviewed here suggest that brain regions and networks involved in speech production are ubiquitously involved in speech perception. We choose the word “ubiquitous” because it implies the ever presence of production regions but does not require the “necessity” of a one of them. Indeed, as proposed in the AxS model, the results here also suggest that the involvement of speech production regions and networks are dynamically determined by context. For speech perception to even occur, contextual constraints are required to interpret acoustically variable signals. Those constraints might come from one particular dynamic organization of a speech production network or another. At any given moment, that particular network configuration is necessary for speech perception to occur. However, because the contextual constraints available are unique to each listening situation, some other constraint and, therefore, associated speech production network might be necessary for the process of achieving perceptual constancy at some different moment. Dichotomous thinking about necessity mostly derives from reactions to the “motor theory of speech perception” (e.g., “if Broca’s area can be removed and people can still perceive speech than the theory must be wrong”). However, neither that model nor its equally dichotomizing auditory-only siblings is biologically realistic. Thus, “Yes” and “no” answers to the question of necessity should be put to rest for something more continuous and consistent with the sensorimotor organization of the brain and the dynamic and distributed involvement of speech production networks in speech perception as demonstrated here.

Nonetheless, if one assumes the AxS mechanisms are important, one might still ask for a more concrete answer to the question of why some studies seem to show the relative preservation of speech perception after damage to “the motor system” or do not show neuroimaging activity in “the motor system” (e.g., Hickok, Costanzo, et al., 2011). First, the reviewed evidence suggests that these studies are few and far between. Second, the brain is

unambiguously plastic and likely has multiple methods for achieving perceptual constancy, particularly after damage. For example, plasticity might involve supposed hemispheric homologues of regions. Third and related, “the motor system” is clearly not limited to Broca’s area or other nearby posterior ventral frontal regions and, thus, speech perception can rely on other nodes in the production network including subcortical structures (like the basal ganglia) and other cortical regions (like the SMA). Fourth, a lack of activity, e.g., in ventral premotor cortex, following a neuroimaging contrast does not imply that a region is uninvolved in speech perception (given the subtractive logic employed in most studies).

Finally, the predictive brain aspect of the AxS model adds another possible answer to the question of why some neuroimaging studies might not reveal activity in “the motor system” during speech perception. Specifically, if the involvement of speech production systems in perception is predictive, it implies that involvement occurs at a time that precedes onsets that might be used in statistical models for analysis. For example, regression models used in nearly all fMRI experiments assume a “canonical” hemodynamic response function that is convolved with stimulus onset and duration that has a fixed shape and lag of two seconds. However, if prediction is a key aspect of processing speech, these lags and shapes should readily vary as a function of context. This could have the effect of obscuring speech production related activity given that it would a priori occur earlier in time and thus not be well modelled.

#### 4.3. Neurobiological models

The AxS model is not a neurobiological model of speech perception. Nonetheless it seems to account for the data reviewed here better than both textbook or classical models and contemporary models that were intended as neurobiological models. Specifically, the anatomical assertions of classical (Wernicke-Geschwind) models are wrong and they simply cannot account for speech perception or language comprehension (Poeppel & Hickok, 2004; Skipper, 2015; Tremblay & Dick, 2016). The most cited contemporary brain model is the Dual Stream Model of Speech Processing (Hickok & Poeppel, 2007). This model maintains that speech perception and production are supported by two “streams”, presumably corresponding to networks. The “ventral” stream is comprised of structures in the temporal lobes that process sound for the purpose of comprehension. In contrast, the “dorsal” stream, formed of posterior superior temporal regions (around the ascending ramus), the insula, and posterior frontal lobe regions, are a “sensorimotor interface” for transforming speech signals into productions (important in development and perhaps involved in working memory).

For a number of reasons, the dual-stream model, though certainly an advance over classical models, does not account for the roles of speech production systems in perception. First, it encapsulates speech production in mostly posterior ventral frontal regions. This review shows that there are a much larger set of regions that need to be accounted for like the SMA, cerebellum, and basal ganglia. Second, production related mechanisms are relegated to the dorsal stream only. However, this review demonstrates that there are many possible speech production networks involved in perception that encompass both dorsal and ventral stream regions. Third, Hickok (2012) explicitly argues that production regions in the dorsal stream do not contribute to speech recognition (Hickok, Costanzo, et al., 2011; see also Poeppel & Hickok, 2004). The results presented here, however, clearly show that dorsal stream regions do contribute. Fourth, the dual stream model permits no means to account for the dynamic nature of regional and network involvement. For example, as shown in the data reviewed here and elsewhere, auditory, speech, and language networks dynamically reconfigure as a function of context (Andric & Hasson, 2015; Hasson, Nusbaum, & Small, 2009; Skipper, 2014; Skipper, Goldin-

Meadow, Nusbaum, & Small, 2009; Skipper, Goldin-Meadow, et al., 2007; Skipper, 2015). The static architecture of a dual-stream model cannot accommodate such dynamic self-organization. A slew of other dual-stream models have appeared but they also suffer from most if not all of these problems (Rauschecker & Scott, 2009; Skipper et al., 2006).

Thus, we are presently in a position of having a model that was not intended as a neurobiological model that somewhat accounts for the brain data better than existing neurobiological models. What is needed is a better neurobiological model that incorporates an AxS like mechanism – i.e., one that is distributed, specific (sensorimotor), predictive, and dynamic – with an account of the large array of regions and networks involved in production that also participate in perception. In short, we need a model that is more consistent with emerging complex network accounts of the brain (Bullmore & Sporns, 2009; van den Heuvel & Sporns, 2013). By such an account, language comprehension emerges from the reverberatory or oscillatory interaction within and between multiple, dynamically organizing cooperating and competing networks and subnetworks, distributed throughout the entire brain (Skipper, 2015). As this review confirms that regions involved in subprocesses associated with producing speech are unambiguously involved in perceiving speech, we can start asking more nuanced questions along these lines. Specifically, we can start asking more specific “what”, “how” and “when” questions – what roles sensorimotor networks supporting speech production play in speech perception and when. This will require a better conception of the computations performed by each network and node in those networks and how network weightings change as a function of listening context and interaction with other networks. This will ultimately involve abandoning classical, dual-stream, and AxS like models for one that is far more neurobiologically specific, complex network oriented, and that can account for natural language use (Skipper, 2015).

#### 5. Conclusions

This review should be taken as a quietus to the question of whether “the motor system” plays “a role” in speech perception. The very question is malformed because it does not take into account the complexity of speech production as a behaviour and the regions of the brain that support these complex processes. Nor does it take seriously the notion that speech perception is itself a complex contextually determined process. It is uncontroversial that auditory brain regions play a role in speech production. We hope this review makes the converse position, that speech production regions play roles in speech perception, equally uncontroversial. Reviewed results suggest that these roles involve a large distributed set of brain regions, extending well beyond “Broca’s area”, that are dynamically recruited for specific computational reasons associated with speech production. Results like these were predicted by the analysis-by-synthesis model of speech perception but are not supported by popular neurobiological models of language (like the classical and dual-stream models). What is needed now are more dynamic complex network oriented neurobiological models that can predict when speech production networks will be engaged during perception and what computational role those networks are performing as a function of the weighting of the regions in each. Then we can move beyond simplistic speech models to a better understanding of the organization of language and the brain.

#### Acknowledgments

JIS would like to also acknowledge EPSRC EP/M026965/1, “New pathways to hearing: A multisensory noise reducing and palate based sensory substitution device for speech perception”, for

support. JIS would like to thank Howard C Nusbaum for once describing speech perception to him as a grand illusion. JIS has hidden a loving Easter egg for the Rabbit-Worm. DRL was supported by a postdoctoral fellowships from the British Academy and Corpus Christi College, Oxford.

## Appendix A

Search terms for all meta-analyses written in BrainMap's taxonomy (<https://www.brainmap.org/taxonomy/>) using pseudocode, where "&" = "and" and "|" = "or".

### A.1. Common search criteria

All BrainMap database searches were required to meet a set of common search criteria to be included in analyses. These were: [Experiments, Context, is, Normal Mapping] & [Experiments, Activation, is, Activations Only] & [Subjects, Handedness, is, Right] & [Subjects, Age, is more than, 17]

### A.2. Natural ("passive") listening to nonwords and words

In addition to these common search criteria, further search criteria were:

[Conditions, Overt Response, is, All modalities, None] & [Conditions, Stimulus, is, Auditory, Words] & [Experiments, Control, is, High Level]

and

[Conditions, Overt Response, is, All modalities, None] & [Conditions, Stimulus, is, Auditory, Music | Pseudowords | Reversed Speech | Sounds (Environmental) | Syllables] & [Experiments, Control, is, High Level]

### A.3. Speech production and silent articulation

In addition to the common search criteria, further search criteria were:

[[Conditions, Overt Response, is, Oral/Facial, Breath-Hold | Drink | Smile | Swallow] | [Experiments, Paradigm Class, is, Breath-Holding | Chewing/Swallowing | Eating/Drinking | Swallowing | Taste]] & [Conditions, Stimulus, is not, Auditory, All Types] & [Experiments, Control, is, High Level]

and

[Experiments, Behavioural Domain, is, Action, Execution - Speech] & [Experiments, Control, is, High Level]

### A.4. Network analyses

In addition to these common search criteria, further search criteria were:

[Conditions, Overt Response, is not, Oral/Facial, All Types] & [Experiments, Behavioural Domain, is, Cognition, Language -- Phonology | Semantics | Speech | Syntax] & [Locations, MNI Image, is, BA45 | BA44 | Lateral | BA6 | BA6 Medial | BA4 | BA1-3 | IPL | Insula | Cerebellum]

and

[Experiments, Behavioural Domain, is, Action, Execution -- Speech] & [Locations, MNI Image, is, BA45 | BA44 | Lateral | BA6 | BA6 Medial | BA4 | BA1-3 | IPL | Insula | Cerebellum]

and

[Experiments, Paradigm Class, is, Naming (Covert), Reading (Covert), Recitation/Repetition (Covert), Word Generation (Covert),

Word Stem Completion (Covert)] & [Locations, MNI Image, is, BA45 | BA44 | Lateral | BA6 | BA6 Medial | BA4 | BA1-3 | IPL | Insula | Cerebellum]

## References

- Ackermann, H., Mathiak, K., & Riecker, A. (2007). The contribution of the cerebellum to speech production and speech perception: Clinical and functional imaging data. *Cerebellum*, 6(3), 202–213.
- Adank, P. (2012). The neural bases of difficult speech comprehension and speech production: Two Activation Likelihood Estimation (ALE) meta-analyses. *Brain and Language*, 122(1), 42–54.
- Adank, P., & Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *NeuroImage*, 49(1), 1124–1132.
- Agnew, Z. K., McGettigan, C., & Scott, S. K. (2011). Discriminating between auditory and motor cortical responses to speech and nonspeech mouth sounds. *Journal of Cognitive Neuroscience*, 23(12), 4038–4047.
- Alho, J., Lin, F.-H., Sato, M., Tiitinen, H., Sams, M., & Jääskeläinen, I. P. (2014). Enhanced neural synchrony between left auditory and premotor cortex is associated with successful phonetic categorization. *Frontiers in Psychology*, 5, 394.
- Alho, J., Sato, M., Sams, M., Schwartz, J.-L., Tiitinen, H., & Jääskeläinen, I. P. (2012). Enhanced early-latency electromagnetic activity in the left premotor cortex is associated with successful phonetic categorization. *NeuroImage*, 60(4), 1937–1946.
- Allen, P., Larøi, F., McGuire, P. K., & Aleman, A. (2008). The hallucinating brain: A review of structural and functional neuroimaging studies of hallucinations. *Neuroscience and Biobehavioral Reviews*, 32(1), 175–191.
- Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H. B., & Zilles, K. (1999). Broca's region revisited: Cytoarchitecture and intersubject variability. *The Journal of Comparative Neurology*, 412(2), 319–341.
- Anderson, M. L. (2015). Précis of after phonology: neural reuse and the interactive brain. *The Behavioral and Brain Sciences*, 1–22.
- Anderson, L. A., & Linden, J. F. (2016). Mind the gap: Two dissociable mechanisms of temporal processing in the auditory system. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 36(6), 1977–1995.
- Andric, M., & Hasson, U. (2015). Global features of functional brain networks change with contextual disorder. *NeuroImage*, 117, 103–113.
- Appelbaum, I. (1996). The lack of invariance problem and the goal of speech perception. *Proceedings fourth international conference on spoken language*, 1996. ICSLP 96 (Vol. 3), pp. 1541–1544 vol.3.
- Appelbaum, I. (1999). The Dogma of Isomorphism: A Case Study from Speech Perception. *Philosophy of Science*, 66, S250–S259.
- Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, 16(7), 390–398.
- Arsenault, J. S., & Buchsbaum, B. R. (2015). No evidence of somatotopic place of articulation feature mapping in motor cortex during passive speech perception. *Psychonomic Bulletin & Review*. <http://dx.doi.org/10.3758/s13423-015-0988-z>.
- Aziz-Zadeh, L., Sheng, T., & Ghetyanchi, A. (2010). Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. *PLoS One*, 5(1), e8759.
- Badino, L., Canevari, C., Fadiga, L., & Metta, G. (2012). Deep-level acoustic-to-articulatory mapping for DBN-HMM based phone recognition. In *Spoken language technology workshop (SLT), 2012 IEEE* (pp. 370–375). [ieeexplore.ieee.org](http://ieeexplore.ieee.org).
- Baldo, J. V., Wilkins, D. P., Ogar, J., Willock, S., & Dronkers, N. F. (2011). Role of the precentral gyrus of the insula in complex articulation. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, 47(7), 800–807.
- Bamiou, D.-E., Musiek, F. E., & Luxon, L. M. (2003/5). The insula (Island of Reil) and its role in auditory processing: Literature review. *Brain Research Reviews*, 42(2), 143–154.
- Bartoli, E., D'Ausilio, A., Berry, J., Badino, L., Bever, T., & Fadiga, L. (2015). Listener-speaker perceived distance predicts the degree of motor contribution to speech perception. *Cerebral Cortex*, 25(2), 281–288.
- Basirat, A., Schwartz, J.-L., & Sato, M. (2012). Perceptuo-motor interactions in the perceptual organization of speech: Evidence from the verbal transformation effect. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 367(1591), 965–976.
- Basso, A., Casati, G., & Vignolo, L. A. (1977). Phonemic identification defect in aphasia. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, 13(1), 85–95.
- Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronkers, N. F. (2003). Voxel-based lesion-symptom mapping. *Nature Neuroscience*, 6(5), 448–450.
- Bawab, Z. A., Raj, B., & Stern, R. M. (2008). Analysis-by-synthesis features for speech recognition. In *2008 IEEE international conference on acoustics, speech and signal processing* (pp. 4185–4188). [ieeexplore.ieee.org](http://ieeexplore.ieee.org).
- Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., ... Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, 59(s1), 1–26.
- Bever, T. G., & Poeppel, D. (2010). Analysis by synthesis: A (re-) emerging program of research for language and vision. *Biolinguistics*, 4(2–3), 174–200.



- Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, 7(3), 295–301.
- Bishop, D. V., Brown, B. B., & Robson, J. (1990). The relationship between phoneme discrimination, speech production, and language comprehension in cerebral-palsied individuals. *Journal of Speech and Hearing Research*, 33(2), 210–219.
- Blackburn, C. S. (1996). *Articulatory methods for speech production and recognition*. Citeseer.
- Blumstein, S. E. (1994). Impairments of speech production and speech perception in aphasia. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 346(1315), 29–36.
- Blumstein, S. E., Cooper, W. E., Zurif, E. G., & Caramazza, A. (1977). The perception and production of voice-onset time in aphasia. *Neuropsychologia*, 15(3), 371–383.
- Boatman, D. (2004). Cortical bases of speech perception: Evidence from functional lesion studies. *Cognition*, 92(1–2), 47–65.
- Bolhuis, J. J., Okanoya, K., & Scharff, C. (2010). Twitter evolution: Converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience*, 11(11), 747–759.
- Borovsky, A., Saygin, A. P., Bates, E., & Dronkers, N. (2007). Lesion correlates of conversational speech production deficits. *Neuropsychologia*, 45(11), 2525–2533.
- Bouchet-Valat, M., & Bastin, G. (2013). RcmdrPlugin.temis, a Graphical Integrated Text Mining Solution in R. *The R Journal*, 5(1), 188–196.
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, 101(4), 2299–2310.
- Brainard, M. S., & Doupe, A. J. (2013). Translating birdsong: Songbirds as a model for basic and applied medical research. *Annual Review of Neuroscience*, 36, 489–517.
- Brown, S., Ingham, R. J., Ingham, J. C., Laird, A. R., & Fox, P. T. (2005). Stuttered and fluent speech production: An ALE meta-analysis of functional neuroimaging studies. *Human Brain Mapping*, 25(1), 105–117.
- Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, 112(44), 13531–13536.
- Bubic, A., Von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, 4. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2904053/>.
- Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3), 186–198.
- Bybee, J., & McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review*, 22(2–4). <http://dx.doi.org/10.1515/tlir.2005.22.2-4.381>.
- Callan, D., Callan, A., Gamez, M., Sato, M.-A., & Kawato, M. (2010). Premotor cortex mediates perceptual performance. *NeuroImage*, 51(2), 844–858.
- Callan, D., Callan, A., & Jones, J. A. (2014). Speech motor brain regions are differentially recruited during perception of native and foreign-accented phonemes for first and second language listeners. *Frontiers in Neuroscience*, 8, 275.
- Callan, D. E., Jones, J. A., Callan, A. M., & Akahane-Yamada, R. (2004). 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*, 22(3), 1182–1194.
- Canevari, C., Badino, L., D'Ausilio, A., Fadiga, L., & Metta, G. (2013). Modeling speech imitation and ecological learning of auditory-motor maps. *Frontiers in Psychology*, 4, 364.
- Cattaneo, L., & Pavesi, G. (2014). The facial motor system. *Neuroscience and Biobehavioral Reviews*, 38, 135–159.
- Cheung, C., Hamilton, L. S., Johnson, K., & Chang, E. F. (2016). The auditory representation of speech sounds in human motor cortex. *eLife*, 5. <http://dx.doi.org/10.7554/eLife.12577>.
- Chevillet, M. A., Jiang, X., Rauschecker, J. P., & Riesenhuber, M. (2013). Automatic phoneme category selectivity in the dorsal auditory stream. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 33(12), 5208–5215.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences*, 36(3), 181–204.
- Cogan, G. B., Thesen, T., Carlson, C., Doyle, W., Devinsky, O., & Pesaran, B. (2014). Sensory-motor transformations for speech occur bilaterally. *Nature*, 507(7490), 94–98.
- Collard, M. J., Fifer, M. S., Benz, H. L., McMullen, D., Wang, Y., Milsap, G. W., ... Crone, N. E. (2016). Cortical subnetwork dynamics during human language tasks. *NeuroImage*. <http://dx.doi.org/10.1016/j.neuroimage.2016.03.072>.
- Cordemoy, G. de. (1668). *A philosophical discourse concerning speech, conformable to the cartesian principles*. In The Savoy.
- Correia, J. M., Jansma, B. M. B., & Bonte, M. (2015). Decoding articulatory features from fMRI responses in dorsal speech regions. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 35(45), 15015–15025.
- Cummings, J. L., Darkins, A., Mendez, M., Hill, M. A., & Benson, D. F. (1988). Alzheimer's disease and Parkinson's disease: Comparison of speech and language alterations. *Neurology*, 38(5), 680–684.
- D'Ausilio, A., Bufalari, I., Salmas, P., & Fadiga, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 48(7), 882–887.
- D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology: CB*, 19(5), 381–385.
- Damasio, A. R. (1992). Aphasia. *The New England Journal of Medicine*, 326(8), 531–539.
- Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, 56(2), 384–398.
- Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., & Dehaene, S. (2008). How does early brain organization promote language acquisition in humans? *European Review*, 16(4), 399–411.
- Demertzi, A., Gómez, F., Crone, J. S., Vanhauzenhuysse, A., Tshibanda, L., Noirhomme, Q., ... Soddu, A. (2014). Multiple fMRI system-level baseline connectivity is disrupted in patients with consciousness alterations. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, 52, 35–46.
- Démonet, J. F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J. L., Wise, R., ... Frackowiak, R. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain: A Journal of Neurology*, 115(Pt 6), 1753–1768.
- Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, 53(1), 1–15.
- Devlin, J. T., & Watkins, K. E. (2007). Stimulating language: Insights from TMS. *Brain: A Journal of Neurology*, 130(3), 610–622.
- Dick, A. S., Bernal, B., & Tremblay, P. (2014). The language connectome: New pathways, new concepts. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, 20(5), 453–467.
- Diederer, K. M. J., Daalman, K., de Weijer, A. D., Neggers, S. F. W., van Gestel, W., Blom, J. D., ... Sommer, I. E. C. (2012). Auditory hallucinations elicit similar brain activation in psychotic and nonpsychotic individuals. *Schizophrenia Bulletin*, 38(5), 1074–1082.
- Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E., & Ramnani, N. (2009). A probabilistic MR atlas of the human cerebellum. *NeuroImage*, 46(1), 39–46.
- Dierks, T., Linden, D. E., Jandl, M., Formisano, E., Goebel, R., Lanfermann, H., & Singer, W. (1999). Activation of Heschl's gyrus during auditory hallucinations. *Neuron*, 22(3), 615–621.
- Drake, E., & Corley, M. (2015). Articulatory imaging implicates prediction during spoken language comprehension. *Memory & Cognition*, 43(8), 1136–1147.
- Dronkers, N. F., Plaisant, O., Iba-Zizen, M. T., & Cabanis, E. A. (2007). Paul Broca's historic cases: High resolution MR imaging of the brains of Leborgne and Lelong. *Brain: A Journal of Neurology*, 130(Pt 5), 1432–1441.
- Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proceedings of the National Academy of Sciences of the United States of America*, 111(19), 7126–7131.
- Edwards, E., Nagarajan, S. S., Dalal, S. S., Canolty, R. T., Kirsch, H. E., Barbaro, N. M., & Knight, R. T. (2010). Spatiotemporal imaging of cortical activation during verb generation and picture naming. *NeuroImage*, 50(1), 291–301.
- Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., & Fox, P. T. (2012). Activation likelihood estimation meta-analysis revisited. *NeuroImage*, 59(3), 2349–2361.
- Eickhoff, S. B., Bzdok, D., Laird, A. R., Roski, C., Caspers, S., Zilles, K., & Fox, P. T. (2011). Co-activation patterns distinguish cortical modules, their connectivity and functional differentiation. *NeuroImage*, 57(3), 938–949.
- Eickhoff, S. B., Heim, S., Zilles, K., & Amunts, K. (2009). A systems perspective on the effective connectivity of overt speech production. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1896), 2399–2421.
- Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, 25(4), 1325–1335.
- Eidelberg, D. (2009). Metabolic brain networks in neurodegenerative disorders: A functional imaging approach. *Trends in Neurosciences*, 32(10), 548–557.
- Eliades, S. J., & Wang, X. (2002). Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *Journal of Neurophysiology*, 89(4), 2194–2207.
- Eliades, S. J., & Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature*, 453(7198), 1102–1106.
- Eliades, S. J., & Wang, X. (2013). Comparison of auditory-vocal interactions across multiple types of vocalizations in marmoset auditory cortex. *Journal of Neurophysiology*, 109(6), 1638–1657.
- Elorriaga-Santiago, S., Silva-Pereyra, J., Rodríguez-Camacho, M., & Carrasco-Vargas, H. (2013). Phonological processing in Parkinson's disease: A neuropsychological assessment. *NeuroReport*, 24(15), 852–855.
- Evans, S., & Davis, M. H. (2015). Hierarchical organization of auditory and motor representations in speech perception: Evidence from searchlight similarity analysis. *Cerebral Cortex*, 25(12), 4772–4788.
- Ferrari, P. F., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *The European Journal of Neuroscience*, 17(8), 1703–1714.
- Ferri, S., Peeters, R., Nelissen, K., Vanduffel, W., Rizzolatti, G., & Orban, G. A. (2015). A human homologue of monkey F5c. *NeuroImage*, 111, 251–266.



- Fifer, R. C. (1993). Insular stroke causing unilateral auditory processing disorder: Case report. *Journal of the American Academy of Audiology*, 4(6), 364–369.
- Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology*, 61(6), 825–850.
- Fischl, B. (2012). FreeSurfer. *NeuroImage*, 62(2), 774–781.
- Flinker, A., Korzeniewska, A., Sheshyuk, A. Y., Franaszczuk, P. J., Dronkers, N. F., Knight, R. T., & Crone, N. E. (2015). Redefining the role of Broca's area in speech. *Proceedings of the National Academy of Sciences*, 112(9), 2871–2875.
- Fowler, C. A., Brown, J. M., Sabadini, L., & Weising, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, 49(3), 396–413.
- Fox, P. T., Laird, A. R., Fox, S. P., Fox, P. M., Uecker, A. M., Crank, M., ... Lancaster, J. L. (2005). BrainMap taxonomy of experimental design: Description and evaluation. *Human Brain Mapping*, 25(1), 185–198.
- Fox, P. T., & Lancaster, J. L. (2002). Mapping context and content: The BrainMap model. *Nature Reviews Neuroscience*, 3(4), 319–321.
- Fridriksson, J., Fillmore, P., Guo, D., & Rorden, C. (2015). Chronic Broca's aphasia is caused by damage to Broca's and Wernicke's areas. *Cerebral Cortex*, 25(12), 4689–4696.
- Fuertinger, S., Horwitz, B., & Simonyan, K. (2015). The functional connectome of speech control. *PLoS Biology*, 13(7), e1002209.
- Gainotti, G., Miceli, G., Silveri, M. C., & Villa, G. (1982). Some anatomo-clinical aspects of phonemic and semantic comprehension disorders in aphasia. *Acta Neurologica Scandinavica*, 66(6), 652–665.
- Galantucci, B., Fowler, C. A., & Goldstein, L. (2009). Perceptuomotor compatibility effects in speech. *Attention, Perception & Psychophysics*, 71(5), 1138–1149.
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13(3), 361–377.
- Gambi, C., & Pickering, M. J. (2013). Prediction and imitation in speech. *Frontiers in Psychology*, 4, 340.
- Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, 66(3), 363–376.
- Gervain, J., & Werker, J. F. (2008). How infant speech perception contributes to language acquisition. *Language and Linguistics Compass*, 2(6), 1149–1170.
- Geschwind, N. (1970). The organization of language and the brain: Language disorders after brain damage help in elucidating the neural basis of verbal behavior. *Science*, 170(3961), 940–944.
- Ghazanfar, A. A. (2009). The multisensory roles for auditory cortex in primate vocal communication. *Hearing Research*, 258(1–2), 113–120.
- Ghazanfar, A. A., & Eliades, S. J. (2014). The neurobiology of primate vocal communication. *Current Opinion in Neurobiology*, 28, 128–135.
- Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, 462(7272), 502–504.
- Gil-da-Costa, R., Martin, A., Lopes, M. A., Muñoz, M., Fritz, J. B., & Braun, A. R. (2006). Species-specific calls activate homologs of Broca's and Wernicke's areas in the macaque. *Nature Neuroscience*, 9(8), 1064–1070.
- Giraud, A.-L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S. J., & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron*, 56(6), 1127–1134.
- Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics*, 31(3–4), 305–320.
- Golfinopoulos, E., Tourville, J. A., Bohland, J. W., Ghosh, S. S., Nieto-Castanon, A., & Guenther, F. H. (2011). fMRI investigation of unexpected somatosensory feedback perturbation during speech. *NeuroImage*, 55(3), 1324–1338.
- Goodglass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders*. Lea & Febiger.
- Gov, D. W., Jr., & Segawa, J. A. (2009). Articulatory mediation of speech perception: A causal analysis of multi-modal imaging data. *Cognition*, 110(2), 222–236.
- Grabski, K., Lamalle, L., Vilain, C., Schwartz, J.-L., Vallée, N., Tropes, I., ... Sato, M. (2012). Functional MRI assessment of orofacial articulators: Neural correlates of lip, jaw, larynx, and tongue movements. *Human Brain Mapping*, 33(10), 2306–2321.
- Grabski, K., Schwartz, J.-L., Lamalle, L., Vilain, C., Vallée, N., Baci, M., ... Sato, M. (2013). Shared and distinct neural correlates of vowel perception and production. *Journal of Neurolinguistics*, 26(3), 384–408.
- Graziano, M. S. A. (2016). Ethological action maps: A paradigm shift for the motor cortex. *Trends in Cognitive Sciences*, 20(2), 121–132.
- Graziano, M. S., Hu, X. T., & Gross, C. G. (1997). Visuospatial properties of ventral premotor cortex. *Journal of Neurophysiology*, 77(5), 2268–2292.
- Guellai, B., Streri, A., & Yeung, H. H. (2014). The development of sensorimotor influences in the audiovisual speech domain: Some critical questions. *Frontiers in Psychology*, 5, 812.
- Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96(3), 280–301.
- Habib, M., Daquin, G., Milandre, L., Royere, M. L., Rey, M., Lanteri, A., ... Khalil, R. (1995). Mutism and auditory agnosia due to bilateral insular damage—Role of the insula in human communication. *Neuropsychologia*, 33(3), 327–339. 1995/3.
- Hage, S. R., & Nieder, A. (2015). Audio-vocal interaction in single neurons of the monkey ventrolateral prefrontal cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 35(18), 7030–7040.
- Halle, M., & Stevens, K. (1962). Speech recognition: A model and a program for research. *IRE Transactions on Information Theory*, 8(2), 155–159.
- Hasson, U., Nusbaum, H. C., & Small, S. L. (2009). Task-dependent organization of brain regions active during rest. *Proceedings of the National Academy of Sciences*, 106(26), 10841–10846.
- Hickok, G. (2012). The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. *Journal of Communication Disorders*, 45(6), 393–402.
- Hickok, G. (2014). *The myth of mirror neurons: The real neuroscience of communication and cognition*. WW Norton & Company.
- Hickok, G., Costanzo, M., Capasso, R., & Miceli, G. (2011). The role of Broca's area in speech perception: evidence from aphasia revisited. *Brain and Language*, 119(3), 214–220.
- Hickok, G., Holt, L. L., & Lotto, A. J. (2009). Response to Wilson: What does motor cortex contribute to speech perception? *Trends in Cognitive Sciences*, 13(8), 330–331.
- Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, 69(3), 407–422.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402.
- Hillis, A. E. (2007). Aphasia: Progress in the last quarter of a century. *Neurology*, 69(2), 200–213.
- Holmes, C. J., Hoge, R., Collins, L., Woods, R., Toga, A. W., & Evans, A. C. (1998). Enhancement of MR images using registration for signal averaging. *Journal of Computer Assisted Tomography*, 22(2), 324–333.
- Homae, F., Watanabe, H., Nakano, T., & Taga, G. (2011). Large-scale brain networks underlying language acquisition in early infancy. *Frontiers in Psychology*, 2, 93.
- Homae, F., Watanabe, H., & Taga, G. (2014). The neural substrates of infant speech perception. *Language Learning*, 64(s2), 6–26.
- Hope, T. M. H., Seghier, M. L., Leff, A. P., & Price, C. J. (2013). Predicting outcome and recovery after stroke with lesions extracted from MRI images. *NeuroImage Clinical*, 2, 424–433.
- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279(5354), 1213–1216.
- Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience*, 14(8), 1125–1138.
- Howard, I. S., & Messum, P. (2014). Learning to pronounce first words in three languages: An investigation of caregiver and infant behavior using a computational model of an infant. *PLoS One*, 9(10), e110334.
- Hyman, B. T., & Tranel, D. (1989). Hemianesthesia and aphasia. An anatomical and behavioral study. *Archives of Neurology*, 46(7), 816–819.
- Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, 2, 255.
- Indefrey, P., & Levelt, W. J. M. (2000). The neural correlates of language production. In *The new cognitive neurosciences*, pp. 845–865. MIT Press.
- Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1–2), 101–144.
- Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences of the United States of America*, 106(4), 1245–1248.
- Jardri, R., Pouchet, A., Pins, D., & Thomas, P. (2011). Cortical activations during auditory verbal hallucinations in schizophrenia: A coordinate-based meta-analysis. *The American Journal of Psychiatry*, 168(1), 73–81.
- Joly, O., Pallier, C., Ramus, F., Pressnitzer, D., Vanduffel, W., & Orban, G. A. (2012). Processing of vocalizations in humans and monkeys: A comparative fMRI study. *NeuroImage*, 62(3), 1376–1389.
- Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience and Biobehavioral Reviews*, 26(2), 235–258.
- Jürgens, U. (2009). The neural control of vocalization in mammals: A review. *Journal of Voice: Official Journal of the Voice Foundation*, 23(1), 1–10. 2009/1.
- Jusczyk, P. W. (1981). Infant speech perception: A critical appraisal. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kimura, D., & Watson, N. (1989). The relation between oral movement control and speech. *Brain and Language*, 37(4), 565–590.
- King, S., Frankel, J., Livescu, K., McDermott, E., Richmond, K., & Wester, M. (2007). Speech production knowledge in automatic speech recognition. *The Journal of the Acoustical Society of America*, 121(2), 723–742.
- Kittredge, A. K., & Dell, G. S. (2016). Learning to speak by listening: Transfer of phonotactics from perception to production. *Journal of Memory and Language*, 89, 8–22.
- Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science*, 237(4819), 1195–1197.
- Knight, R. T. (2007). Neuroscience. Neural networks debunk phrenology. *Science*, 316(5831), 1578–1579.
- Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*, 297(5582), 846–848.
- Kreisler, A., Godefroy, O., Delmaire, C., Debachy, B., Leclercq, M., Pruvo, J. P., & Leys, D. (2000). The anatomy of aphasia revisited. *Neurology*, 54(5), 1117–1123.
- Kröger, B. J., & Cao, M. (2015). The emergence of phonetic–phonological features in a biologically inspired model of speech processing. *Journal of Phonetics*, 53, 88–100.
- Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron*, 67(5), 713–727.

- Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, *190*(4209), 69–72.
- Kuhl, P. K., Ramírez, R. R., Bosseler, A., Lin, J.-F. L., & Imada, T. (2014). Infants' brain responses to speech suggest analysis by synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(31), 11238–11245.
- Kumar, V., Croxson, P. L., & Simonyan, K. (2016). Structural organization of the laryngeal motor cortical network and its implication for evolution of speech production. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *36*(15), 4170–4181.
- Laird, A. R., Lancaster, J. L., & Fox, P. T. (2005). BrainMap: The social evolution of a human brain mapping database. *Neuroinformatics*, *3*(1), 65–78.
- Laird, A. R., Robinson, J. L., McMillan, K. M., Tordesillas-Gutiérrez, D., Moran, S. T., Gonzales, S. M., ... Lancaster, J. L. (2010). Comparison of the disparity between Talairach and MNI coordinates in functional neuroimaging data: Validation of the Lancaster transform. *NeuroImage*, *51*(2), 677–683.
- Lametti, D. R., Krol, S. A., Shiller, D. M., & Ostry, D. J. (2014). Brief periods of auditory perceptual training can determine the sensory targets of speech motor learning. *Psychological Science*, *25*(7), 1325–1336.
- Lametti, D. R., Oostwoud Wijdenes, L., Bonaiuto, J., Bestmann, S., & Rothwell, J. C. (2016). Cerebellar tDCS dissociates the timing of perceptual decisions from perceptual change in speech. *Journal of Neurophysiology*. jn.00433.2016.
- Lametti, D. R., Rochet-Capellan, A., Neufeld, E., Shiller, D. M., & Ostry, D. J. (2014). Plasticity in the human speech motor system drives changes in speech perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(31), 10339–10346.
- Lametti, D. R., & Watkins, K. E. (2016). Cognitive Neuroscience: The Neural Basis of Motor Learning by Observing. *Current Biology: CB*, *26*(7), R288–R290.
- Lancaster, J. L., Tordesillas-Gutiérrez, D., Martínez, M., Salinas, F., Evans, A., Zilles, K., ... Fox, P. T. (2007). Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Human Brain Mapping*, *28*(11), 1194–1205.
- Lane, H. (1965). The motor theory of speech perception: a critical review. *Psychological Review*, *72*, 275–309.
- Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, *14*(4), 677–709.
- Lazar, R. M., & Antoniollo, D. (2008). Variability in recovery from aphasia. *Current Neurology and Neuroscience Reports*, *8*(6), 497–502.
- Lee, J. D., Park, H.-J., Park, E. S., Oh, M.-K., Park, B., Rha, D.-W., ... Park, C. I. (2011). Motor pathway injury in patients with periventricular leukomalacia and spastic diplegia. *Brain: A Journal of Neurology*, *134*(Pt 4), 1199–1210.
- Leeper, H. A., Jr, Shewan, C. M., & Booth, J. C. (1986). Altered acoustic cue discrimination in Broca's and conduction aphasics. *Journal of Communication Disorders*, *19*(2), 83–103.
- Leroy, F., Glasel, H., Dubois, J., Hertz-Pannier, L., Thirion, B., Mangin, J.-F., & Dehaene-Lambertz, G. (2011). Early maturation of the linguistic dorsal pathway in human infants. *Journal of Neuroscience*, *31*(4), 1500–1506.
- Lesser, R. P., Lueders, H., Dinner, D. S., Hahn, J., & Cohen, L. (1984). The location of speech and writing functions in the frontal language area. Results of extraoperative cortical stimulation. *A Journal of Neurology*, *107*(Pt 1), 275–291.
- Li, M., Kim, J., Lammert, A., Ghosh, P. K., Ramanarayanan, V., & Narayanan, S. (2016). Speaker verification based on the fusion of speech acoustics and inverted articulatory signals. *Computer Speech & Language*, *36*, 196–211.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36.
- Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, *4*(5), 187–196.
- Liebenthal, E., Sabri, M., Beardsley, S. A., Mangalathu-Arumana, J., & Desai, A. (2013). Neural dynamics of phonological processing in the dorsal auditory stream. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(39), 15414–15424.
- Linden, D. E. J., Thornton, K., Kuswanto, C. N., Johnston, S. J., van de Ven, V., & Jackson, M. C. (2011). The brain's voices: Comparing nonclinical auditory hallucinations and imagery. *Cerebral Cortex*, *21*(2), 330–337.
- Londei, A., D'Ausilio, A., Basso, D., Sestieri, C., Del Gratta, C., Romani, G. L., & Olivetti Belardinelli, M. (2007). Brain network for passive word listening as evaluated with ICA and Granger causality. *Brain Research Bulletin*, *72*(4–6), 284–292.
- Londei, A., D'Ausilio, A., Basso, D., Sestieri, C., Gratta, C. D., Romani, G.-L., & Belardinelli, M. O. (2010). Sensory-motor brain network connectivity for speech comprehension. *Human Brain Mapping*, *31*(4), 567–580.
- Long, M. A., Katlowitz, K. A., Svirsky, M. A., Clary, R. C., Byun, T. M., Majaj, N., ... Greenlee, J. D. (2016). Functional segregation of cortical regions underlying speech timing and articulation. *Neuron*, *89*(6), 1187–1193.
- Lotte, F., Brumberg, J. S., Brunner, P., Gunduz, A., Ritaccio, A. L., Guan, C., & Schalk, G. (2015). Electroencephalographic representations of segmental features in continuous speech. *Frontiers in Human Neuroscience*, *9*, 97.
- Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, *13*(3), 110–114.
- Lotto, A. J., & Holt, L. L. (2000). The illusion of the phoneme. In S. J. Billings, J. P. Boyle, & A. M. Griffith (Eds.), *Presented at the Chicago Linguistic Society, Chicago* (pp. 191–204).
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *The Behavioral and Brain Sciences*, *21*(4), 499–511. discussion 511–46.
- Makris, N., Goldstein, J. M., Kennedy, D., Hodge, S. M., Caviness, V. S., Faraone, S. V., ... Seidman, L. (2006). Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophrenia Research*, *83*(2–3), 155–171.
- Martin, N., & Saffran, E. M. (2002). The relationship of input and output phonological processing: An evaluation of models and evidence to support them. *Aphasiology*, *16*(1–2), 107–150.
- Matchin, W., Groulx, K., & Hickok, G. (2014). Audiovisual speech integration does not rely on the motor system: Evidence from articulatory suppression, the McGurk effect, and fMRI. *Journal of Cognitive Neuroscience*, *26*(3), 606–620.
- Matsumoto, R., Nair, D. R., LaPresto, E., Najm, I., Bingaman, W., Shibusaki, H., & Lüders, H. O. (2004). Functional connectivity in the human language system: A cortico-cortical evoked potential study. *Brain: A Journal of Neurology*, *127*(Pt 10), 2316–2330.
- Maudoux, A., Lefebvre, P., Cabay, J.-E., Demertzi, A., Vanhauudenhuysse, A., Laureys, S., & Soddu, A. (2012). Auditory resting-state network connectivity in tinnitus: A functional MRI study. *PLoS One*, *7*(5), e36222.
- Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., ... Mazoyer, B. (2001). A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *356*(1412), 1293–1322.
- McGettigan, C., Agnew, Z. K., & Scott, S. K. (2010). Are articulatory commands automatically and involuntarily activated during speech perception? *Proceedings of the National Academy of Sciences of the United States of America*, *107*(12), E42. author reply E43.
- McGowan, R. S., & Faber, A. (1996). Introduction to papers on speech recognition and perception from an articulatory point of view. *The Journal of the Acoustical Society of America*, *99*(3), 1680–1682.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264* (5588), 746–748.
- Meier, J. D., Affalo, T. N., Kastner, S., & Graziano, M. S. A. (2008). complex organization of human primary motor cortex: A high-resolution fMRI study. *Journal of Neurophysiology*, *100*(4), 1800–1812.
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology: CB*, *17* (19), 1692–1696.
- Melinger, A., Branigan, H. P., & Pickering, M. J. (2014). Parallel processing in language production. *Language, Cognition and Neuroscience*, *29*(6), 663–683.
- Messum, P., & Howard, I. S. (2015). Creating the cognitive form of phonological units: The speech sound correspondence problem in infancy could be solved by mirrored vocal interactions rather than by imitation. *Journal of Phonetics*, *53*, 125–140.
- Miceli, G., Gainotti, G., Caltagirone, C., & Masullo, C. (1980). Some aspects of phonological impairment in aphasia. *Brain and Language*, *11*(1), 159–169.
- Miller, C. T., Dimauro, A., Pistorio, A., Hendry, S., & Wang, X. (2010). Vocalization induced CFos expression in marmoset cortex. *Frontiers in Integrative Neuroscience*, *4*, 128.
- Miller, C. T., Thomas, A. W., Nummela, S. U., & de la Mothe, L. A. (2015). Responses of primate frontal cortex neurons during natural vocal communication. *Journal of Neurophysiology*, *114*(2), 1158–1171.
- Mirman, D., Chen, Q., Zhang, Y., Wang, Z., Faseyitan, O. K., Coslett, H. B., & Schwartz, M. F. (2015). Neural organization of spoken language revealed by lesion-symptom mapping. *Nature Communications*, *6*, 6762.
- Mitra, V., Sivaraman, G., Nam, H., Espy-Wilson, C., & Saltzman, E. (2014). Articulatory features from deep neural networks and their role in speech recognition. In *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 3017–3021). [ieeexplore.ieee.org](http://ieeexplore.ieee.org).
- Mochida, T., Kimura, T., Hiroya, S., Kitagawa, N., Gomi, H., & Kondo, T. (2013). Speech misperception: Speaking and seeing interfere differently with hearing. *PLoS One*, *8*(7), e68619.
- Mohr, J. P., Pessin, M. S., Finkelnstein, S., Funkenstein, H. H., Duncan, G. W., & Davis, K. R. (1978). Broca aphasia: Pathology and clinical. *Neurology*, *28*(4), 311–324.
- Moineau, S., Dronkers, N. F., & Bates, E. (2005). Exploring the processing continuum of single-word comprehension in aphasia. *Journal of Speech, Language, and Hearing Research: JSLHR*, *48*(4), 884.
- Mooney, R. (2014). Auditory-vocal mirroring in songbirds. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369* (1644), 20130179.
- Möttönen, R., Dutton, R., & Watkins, K. E. (2013). Auditory-motor processing of speech sounds. *Cerebral Cortex*, *23*(5), 1190–1197.
- Möttönen, R., van de Ven, G. M., & Watkins, K. E. (2014). Attention fine-tunes auditory-motor processing of speech sounds. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(11), 4064–4069.
- Mottron, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *Journal of Neuroscience*, *29*(31), 9819–9825.
- Moulin-Frier, C., Diard, J., Schwartz, J.-L., & Bessière, P. (2015). COSMO (“Communicating about Objects using Sensory-Motor Operations”): A Bayesian modeling framework for studying speech communication and the emergence of phonological systems. *Journal of Phonetics*, *53*, 5–41.
- Moulin-Frier, C., Laurent, R., Bessière, P., Schwartz, J. L., & Diard, J. (2012). Adverse conditions improve distinguishability of auditory, motor, and perceptuo-motor theories of speech perception: An exploratory Bayesian modelling study. *Language and Cognitive Processes*, *27*(7–8), 1240–1263.
- Munding, D., Dubarry, A.-S., & Alario, F.-X. (2016). On the cortical dynamics of word production: A review of the MEG evidence. *Language, Cognition and Neuroscience*, *31*(4), 441–462.
- Murakami, T., Kell, C. A., Restle, J., Ugawa, Y., & Ziemann, U. (2015). Left dorsal speech stream components and their contribution to phonological processing.



- The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 35(4), 1411–1422.
- Naeser, M. A., Palumbo, C. L., Helm-Estabrooks, N., Stiassny-Eder, D., & Albert, M. L. (1989). Severe nonfluency in aphasia. Role of the medial subcallosal fasciculus and other white matter pathways in recovery of spontaneous speech. *Brain: A Journal of Neurology*, 112(Pt 1), 1–38.
- Nasir, S. M., & Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proceedings of the National Academy of Sciences of the United States of America*, 106(48), 20470–20475.
- Nelissen, K., Luppino, G., Vanduffel, W., Rizzolatti, G., & Orban, G. A. (2005). Observing others: Multiple action representation in the frontal lobe. *Science*, 310(5746), 332–336.
- Nelson, A., & Mooney, R. (2016). The basal forebrain and motor cortex provide convergent yet distinct movement-related inputs to the auditory cortex. *Neuron*, 90(3), 635–648.
- Nelson, A., Schneider, D. M., Takatoh, J., Sakurai, K., Wang, F., & Mooney, R. (2013). A circuit for motor cortical modulation of auditory cortical activity. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 33(36), 14342–14353.
- Nusbaum, N. S., & Henly, A. S. (1992). Listening to speech through an adaptive window of analysis. *The Auditory Processing of Speech: From Sounds to Words*, 339.
- Nuttall, H. E., Kennedy-Higgins, D., Hogan, J., Devlin, J. T., & Adank, P. (2016). The effect of speech distortion on the excitability of articulatory motor cortex. *NeuroImage*, 128, 218–226.
- Nye, P. W., & Fowler, C. A. (2003/1). Shadowing latency and imitation: The effect of familiarity with the phonetic patterning of English. *Journal of Phonetics*, 31(1), 63–79.
- Ojemann, G. A. (1983). Brain organization for language from the perspective of electrical stimulation mapping. *The Behavioral and Brain Sciences*, 6(02), 189–206.
- Ojemann, G. A. (1991). Cortical organization of language. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 11(8), 2281–2287.
- Ojemann, G., & Mateer, C. (1979). Human language cortex: Localization of memory, syntax, and sequential motor-phoneme identification systems. *Science*, 205(4413), 1401–1403.
- Onnis, L., & Spivey, M. J. (2012). Toward a new scientific visualization for the language sciences. *Information. An International Interdisciplinary Journal*, 3(1), 124–150.
- Osnes, B., Hugdahl, K., & Specht, K. (2011). Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. *NeuroImage*, 54(3), 2437–2445.
- Ostry, D. J., & Gribble, P. L. (2016). Sensory plasticity in human motor learning. *Trends in Neurosciences*, 39(2), 114–123.
- Oudeyer, P.-Y. (2005). The self-organization of speech sounds. *Journal of Theoretical Biology*, 233(3), 435–449.
- Paulesu, E., Perani, D., Blasi, V., Silani, G., Borghese, N. A., De Giovanni, U., ... Fazio, F. (2003). A functional-anatomical model for lipreading. *Journal of Neurophysiology*, 90(3), 2005–2013.
- Paus, T. (2005). Inferring causality in brain images: A perturbation approach. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 360(1457), 1109–1114.
- Peeters, M., Verhoeven, L., de Moor, J., & van Balkom, H. (2009). Importance of speech production for phonological awareness and word decoding: The case of children with cerebral palsy. *Research in Developmental Disabilities*, 30(4), 712–726.
- Peeters, M., Verhoeven, L., van Balkom, H., & de Moor, J. (2008). Foundations of phonological awareness in pre-school children with cerebral palsy: The impact of intellectual disability. *Journal of Intellectual Disability Research: JIDR*, 52(Pt 1), 68–78.
- Peh, W. Y. X., Roberts, T. F., & Mooney, R. (2015). Imaging auditory representations of song and syllables in populations of sensorimotor neurons essential to vocal communication. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 35(14), 5589–5605.
- Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain: A Journal of Neurology*. <http://dx.doi.org/10.1093/brain/60.4.38900>.
- Perani, D., Saccuman, M. C., Scifo, P., Anwander, A., Anwander, A., Spada, D., ... Friederici, A. D. (2011). Neural language networks at birth. *Proceedings of the National Academy of Sciences of the United States of America*, 108(38), 16056–16061.
- Perkell, J., Guenther, F., Lane, H., Matthies, M., Payan, Y., Perrier, P., ... Zandipour, M. (2001). The sensorimotor control of speech production. In *Proceedings of the First International Symposium on Measurement, Analysis and Modeling of Human Functions* (pp. 21–23).
- Petkov, C. I., & Jarvis, E. D. (2012). Birds, primates, and spoken language origins: Behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience*, 4, 12.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *The Behavioral and Brain Sciences*, 36(4), 329–347.
- Poehpel, D., & Hickok, G. (2004). Towards a new functional anatomy of language. *Cognition*, 92(1–2), 1–12.
- Poehpel, D., & Monahan, P. J. (2011). Feedforward and feedback in speech perception: Revisiting analysis by synthesis. *Language and Cognitive Processes*, 26(7), 935–951.
- Poldrack, R. A., Temple, E., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., & Gabrieli, J. (2001). Relations between the neural bases of dynamic auditory processing and phonological processing: Evidence from fMRI. *Journal of Cognitive Neuroscience*, 13(5), 687–697.
- Port, R. F. (2010). Rich memory and distributed phonology. *Language Sciences*, 32(1), 43–55. 2010/1.
- Prather, J. F., Nowicki, S., Anderson, R. C., Peters, S., & Mooney, R. (2009). Neural correlates of categorical perception in learned vocal communication. *Nature Neuroscience*, 12(2), 221–228.
- Prather, J. F., Peters, S., Nowicki, S., & Mooney, R. (2008). Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature*, 451(7176), 305–310.
- Price, C. J., Crinion, J., & Macsweeney, M. (2011). A generative model of speech production in Broca's and Wernicke's areas. *Frontiers in Psychology*, 2, 237.
- Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews. Neuroscience*, 11(5), 351–360.
- Pulvermüller, F., Huss, M., Kherif, F., del Prado, Moscoso, Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, 103(20), 7865–7870.
- Pytte, C. L., & Suthers, R. A. (1999). A bird's own song contributes to conspecific song perception. *NeuroReport*, 10(8), 1773–1778.
- R Core Team (2014). R: A language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria. URL: <http://www.R-project.org/>.
- Rastle, K., Davis, M. H., & Brysbaert, M. (2010). Response to McGettigan et al.: Task-based accounts are not sufficiently coherent to explain articulatory effects in speech perception. *Proceedings of the National Academy of Sciences*, 107(12), E43. E43–E43.
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724.
- Rhone, A. E., Nourski, K. V., Oya, H., Kawasaki, H., Howard, M. A., 3rd, & McMurray, B. (2016). Can you hear me yet? An intracranial investigation of speech and non-speech audiovisual interactions in human cortex. *Language, Cognition and Neuroscience*, 31(2), 284–302.
- Richard, F. D., Bond, C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology: Journal of Division 1, of the American Psychological Association*, 7(4), 331–363.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169–192.
- Robbins, J. A., Logemann, J. A., & Kirshner, H. S. (1986). Swallowing and speech production in Parkinson's disease. *Annals of Neurology*, 19(3), 283–287.
- Rogalsky, C., Love, T., Driscoll, D., Anderson, S. W., & Hickok, G. (2011). Are mirror neurons the basis of speech perception? Evidence from five cases with damage to the purported human mirror system. *Neurocase*, 17(2), 178–187.
- Romanski, L. M., & Goldman-Rakic, P. S. (2002). An auditory domain in primate prefrontal cortex. *Nature Neuroscience*, 5(1), 15–16.
- Romero, L., Walsh, V., & Papagno, C. (2006). The neural correlates of phonological short-term memory: A repetitive transcranial magnetic stimulation study. *Journal of Cognitive Neuroscience*, 18(7), 1147–1155.
- Roon, K. D., & Gafos, A. I. (2015). Perceptuo-motor effects of response-distractor compatibility in speech: Beyond phonemic identity. *Psychonomic Bulletin & Review*, 22(1), 242–250.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638.
- Rowe, J. B., & Siebner, H. R. (2012). The motor system and its disorders. *NeuroImage*, 61(2), 464–477.
- Roy, A., & Mooney, R. (2007). Auditory plasticity in a basal ganglia-forebrain pathway during decrystallization of adult birdsong. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 27(24), 6374–6387.
- Saad, Z. S., Reynolds, R. C., Argall, B., Japee, S., & Cox, R. W. (2004). SUMA: An interface for surface-based intra- and inter-subject analysis with AFNI. In *Proceedings of the IEEE international symposium on biomedical imaging, Arlington, VA* (pp. 1510–1513).
- Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Brain Research. Cognitive Brain Research*, 23(2–3), 429–435.
- Sato, M., Tremblay, P., & Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language*, 111(1), 1–7.
- Saur, D., Lange, R., Baumgaertner, A., Schraknepper, V., Willmes, K., Rijntjes, M., & Weiller, C. (2006). Dynamics of language reorganization after stroke. *Brain: A Journal of Neurology*, 129(Pt 6), 1371–1384.
- Scarbel, L., Beaudet, D., Schwartz, J.-L., & Sato, M. (2014). The shadow of a doubt? Evidence for perceptuo-motor linkage during auditory and audiovisual close-shadowing. *Frontiers in Psychology*, 5, 568.
- Scheck, S. M., Boyd, R. N., & Rose, S. E. (2012). New insights into the pathology of white matter tracts in cerebral palsy from diffusion magnetic resonance imaging: A systematic review. *Developmental Medicine and Child Neurology*, 54(8), 684–696.
- Schirmer, A. (2004). Timing speech: A review of lesion and neuroimaging findings. *Brain Research. Cognitive Brain Research*, 21(2), 269–287.
- Schneider, D. M., & Mooney, R. (2015). Motor-related signals in the auditory system for listening and learning. *Current Opinion in Neurobiology*, 33, 78–84.
- Schneider, D. M., Nelson, A., & Mooney, R. (2014). A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature*, 513(7517), 189–194.

- Schomers, M. R., Kirilina, E., Weigand, A., Bajbouj, M., & Pulvermüller, F. (2015). Causal influence of articulatory motor cortex on comprehending single spoken words: TMS evidence. *Cerebral Cortex*, 25(10), 3894–3902.
- Schomers, M. R., & Pulvermüller, F. (2016). Is the sensorimotor cortex relevant for speech perception and understanding? An integrative review. *Frontiers in Human Neuroscience*, 10. <http://dx.doi.org/10.3389/fnhum.2016.00435>.
- Schouten, B., Gerrits, E., & van Hoesen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, 41(1), 71–80.
- Schubotz, R. I. (2007). Prediction of external events with our motor system: Towards a new framework. *Trends in Cognitive Sciences*, 11(5), 211–218.
- Schwartz, J.-L., Basirat, A., Ménard, L., & Sato, M. (2012/9). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, 25(5), 336–354.
- Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, 10(4), 295–302.
- Scott, S. K., Rosen, S., Wickham, L., & Wise, R. J. S. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *The Journal of the Acoustical Society of America*, 115(2), 813–821.
- Sehm, B., Schnitzler, T., Obleser, J., Groba, A., Ragert, P., Villringer, A., & Obrig, H. (2013). Facilitation of inferior frontal cortex by transcranial direct current stimulation induces perceptual learning of severely degraded speech. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 33(40), 15868–15878.
- Shergill, S. S., Brammer, M. J., Fukuda, R., Bullmore, E., Amaro, E., Jr., Murray, R. M., & McGuire, P. K. (2002). Modulation of activity in temporal cortex during generation of inner speech. *Human Brain Mapping*, 16(4), 219–227.
- Shergill, S. S., Brammer, M. J., Williams, S. C., Murray, R. M., & McGuire, P. K. (2000). Mapping auditory hallucinations in schizophrenia using functional magnetic resonance imaging. *Archives of General Psychiatry*, 57(11), 1033–1038.
- Shewan, C. M. (1980). Phonological processing in Broca's aphasics. *Brain and Language*, 10(1), 71–88.
- Shiller, D. M., & Rochon, M.-L. (2014). Auditory-perceptual learning improves speech motor adaptation in children. *Journal of Experimental Psychology: Human Perception and Performance*, 40(4), 1308–1315.
- Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America*, 125(2), 1103–1113.
- Shultz, S., Vouloumanos, A., Bennett, R. H., & Pelphrey, K. (2014). Neural specialization for speech in the first months of life. *Developmental Science*, 17(5), 766–774.
- Skipper, J. I. (2014). Echoes of the spoken past: How auditory cortex hears context during speech perception. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369(1651), 20130297.
- Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain and Language*, 101(3), 260–277.
- Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2009). Gestures orchestrate brain networks for language understanding. *Current Biology: CB*, 19(8), 661–667.
- Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: Motor cortical activation during speech perception. *NeuroImage*, 25(1), 76–89.
- Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2006). Lending a helping hand to hearing: Another motor theory of speech perception. *Action to Language via the Mirror*, 250–285.
- Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, 17(10), 2387–2399.
- Skipper, J. I. (2015). The NOLB model: A model of the natural organization of language and the brain. In R. M. Willems (Ed.), *Cognitive neuroscience of natural language use*. Cambridge University Press.
- Smalle, E. H. M., Rogers, J., & Möttönen, R. (2015). Dissociating contributions of the motor cortex to speech perception and response bias by using transcranial magnetic stimulation. *Cerebral Cortex*, 25(10), 3690–3698.
- Smith, M. M. (2001). Simply a speech impairment? Literacy challenges for individuals with severe congenital speech impairments. *International Journal of Disability, Development and Education*, 48(4), 331–353.
- Solis, M. M., & Doupe, A. J. (2000). Compromised neural selectivity for song in birds with impaired sensorimotor learning. *Neuron*, 25(1), 109–121.
- Stasenko, A., Bonn, C., Teghipco, A., Garcea, F. E., Sweet, C., Dombrov, M., ... Mahon, B. Z. (2015). A causal test of the motor theory of speech perception: A case of impaired speech production and spared speech perception. *Cognitive Neuropsychology*, 32(2), 38–57.
- Stevens, K. N. (1960). Toward a model for speech recognition. *The Journal of the Acoustical Society of America*, 32(1), 47–55.
- Stevens, K. N., & Halle, M. (1967). Remarks on analysis by synthesis and distinctive features. In W. Walther-Dunn (Ed.), *Models for the perception of speech and visual form* (pp. 88–102).
- Szenkovits, G., Peelle, J. E., Norris, D., & Davis, M. H. (2012). Individual differences in premotor and motor recruitment during speech perception. *Neuropsychologia*, 50(7), 1380–1392.
- Tan, Y.-W., Liu, W.-J., Jiang, W., & Zheng, H. (2015). Integration of articulatory knowledge and voicing features based on DNN/HMM for Mandarin speech recognition. In *2015 international joint conference on neural networks (IJCNN)* (pp. 1–8).
- Taylor, L. B. (1979). Psychological assessment of neurosurgical patients. In T. Rasmussen & R. Marino (Eds.), *Functional neurosurgery* (pp. 165–180). Raven Press.
- Thulborn, K. R., Carpenter, P. A., & Just, M. A. (1999). Plasticity of language-related brain function during recovery from stroke. *Stroke: A Journal of Cerebral Circulation*, 30(4), 749–754.
- Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science*, 21(10), 1532–1540.
- Tremblay, P., & Dick, A. S. (2016). Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain and Language*, 162, 60–71.
- Tremblay, P., & Small, S. L. (2011). On the context-dependent nature of the contribution of the ventral premotor cortex to speech perception. *NeuroImage*, 57(4), 1561–1571.
- Turkeltaub, P. E., Eden, G. F., Jones, K. M., & Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *NeuroImage*, 16(3), 765–780.
- Turkeltaub, P. E., Eickhoff, S. B., Laird, A. R., Fox, M., Wiener, M., & Fox, P. (2012). Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Human Brain Mapping*, 33(1), 1–13.
- van de Ven, V., Esposito, F., & Christoffels, I. K. (2009). Neural network of speech monitoring overlaps with overt speech production and comprehension networks: A sequential spatial and temporal ICA study. *NeuroImage*, 47(4), 1982–1991.
- van den Heuvel, M. P., & Sporns, O. (2013). Network hubs in the human brain. *Trends in Cognitive Sciences*, 17(12), 683–696.
- Venezia, J. H., Saberi, K., Chubb, C., & Hickok, G. (2012). Response bias modulates the speech motor system during syllable discrimination. *Frontiers in Psychology*, 3, 157.
- Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., ... Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage*, 30(4), 1414–1432.
- Vitale, C., Marcelli, V., Abate, T., Pianese, A., Allocca, R., Moccia, M., ... Cavaliere, M. (2016). Speech discrimination is impaired in parkinsonian patients: Expanding the audiologic findings of Parkinson's disease. *Parkinsonism & Related Disorders*, 22(Suppl. 1), S138–S143.
- von Helmholtz, H. (1867). *Treatise on physiological optics, Vol. III*. Dover.
- Wand, M., Koutník, J., & Schmidhuber, J. (2016, January 29). *Lipreading with Long Short-Term Memory*. arXiv [cs.CV]. <<http://arxiv.org/abs/1601.08188>>.
- Wang, J., Samal, A., Green, J. R., & Rudzicz, F. (2012a). Sentence recognition from articulatory movements for silent speech interfaces. In *2012 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 4985–4988). Dover. [ieeexplore.ieee.org](http://ieeexplore.ieee.org).
- Wang, J., Samal, A., Green, J. R., & Rudzicz, F. (2012b). Whole-word recognition from articulatory movements for silent speech interfaces. In *Interspeech* (pp. 1327–1330).
- Warren, R. M. (1968). Verbal transformation effect and auditory perceptual mechanisms. *Psychological Bulletin*, 70(4), 261–270.
- Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41(8), 989–994.
- Weller, M. (1993). Anterior opercular cortex lesions cause dissociated lower cranial nerve palsies and anarthria but no aphasia: Foix-Chavany-Marie syndrome and "automatic voluntary dissociation" revisited. *Journal of Neurology*, 240(4), 199–208.
- Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology*, 50, 509–535.
- Westermann, G., & Reck Miranda, E. (2004). A new model of sensorimotor coupling in the development of speech. *Brain and Language*, 89(2), 393–400.
- Whitehill, T. L., Francis, A. L., & Ching, C. K. Y. (2003). Perception of place of articulation by children with cleft palate and posterior placement. *Journal of Speech, Language, and Hearing Research: JSLHR*, 46(2), 451–461.
- Williams, H., & Nottebohm, F. (1985). Auditory responses in avian vocal motor neurons: A motor theory for song perception in birds. *Science*, 229(4710), 279–282.
- Wilson, S. M. (2009). Speech perception when the motor system is compromised. *Trends in Cognitive Sciences*, 13(8), 329–330. author reply 330–1.
- Wilson, S. M., Henry, M. L., Besbris, M., Ogar, J. M., Dronkers, N. F., Jarrold, W., ... Gorno-Tempini, M. L. (2010). Connected speech production in three variants of primary progressive aphasia. *Brain: A Journal of Neurology*, 133(Pt 7), 2069–2088.
- Wilson, S. M., & Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *NeuroImage*, 33(1), 316–325.
- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7(7), 701–702.
- Yakunina, N., Tae, W. S., Lee, K. U., Kim, S. S., & Nam, E.-C. (2013). Spatiotemporal segregation of neural response to auditory stimulation: An fMRI study using independent component analysis and frequency-domain analysis. *PLoS One*, 8(6), e66424.
- Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, 8(8), 665–670.



- Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., ... Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165.
- Yeung, H. H., & Werker, J. F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychological Science*, *24*(5), 603–612.
- Yildiz, I. B., & Kiebel, S. J. (2011). A hierarchical neuronal model for generation and online recognition of birdsongs. *PLoS Computational Biology*, *7*(12), e1002303.
- Yuen, I., Davis, M. H., Brysbaert, M., & Rastle, K. (2010). Activation of articulatory information in speech perception. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(2), 592–597.
- Zanini, S., Melatini, A., Capus, L., Gioulis, M., Vassallo, A., & Bava, A. (2003). Language recovery following subthalamic nucleus stimulation in Parkinson's disease. *NeuroReport*, 511–516.
- Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, *256*(5058), 846–849.
- Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top-down and bottom-up processes in speech comprehension. *NeuroImage*, *32*(4), 1826–1836.