Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization

Gregory Hickok[1], **John Houde**[2], and **Feng Rong**[1]

[1]Center for Cognitive Neuroscience, Center for Hearing Research, Department of Cognitive Sciences, University of California, Irvine, CA 92697

[2]Otolaryngology – Head and Neck Surgery, University of California, San Francisco, CA 94143

Abstract

Sensorimotor integration is an active domain of speech research and is characterized by two main ideas, that the auditory system is critically involved in speech *production*, and that the motor system is critically involved in speech *perception*. Despite the complementarity of these ideas, there is little crosstalk between these literatures. We propose an integrative model of the speech-related "dorsal stream" in which sensorimotor interaction primarily supports speech production, in the form of a state feedback control architecture. A critical component of this control system is forward sensory prediction, which affords a natural mechanism for limited motor influence on perception, as recent perceptual research has suggested. Evidence shows that this influence is *modulatory* but not necessary for speech perception. The neuroanatomy of the proposed circuit is discussed as well as some probable clinical correlates including conduction aphasia, stuttering, and aspects of schizophrenia.

Introduction

Sensorimotor integration in the domain of speech processing is an exceptionally active area of research and can be summarized by two main ideas: (1) the auditory system is critically involved in the *production* of speech and (2) the motor system is critically involved in the *perception* of speech. Both ideas address the need for "parity", as Liberman and Mattingly put it (Liberman and Mattingly, 1985, 1989), between the auditory and motor speech systems, but emphasize opposite directions of influence and situate the point of contact in a different place. The audio-centric view suggests that the goal of speech *production* is to generate a target *sound*, thus the common currency is acoustic in nature. The motor-centric view suggests that the goal of speech *perception* is to recover the *motor* gesture that generated a perceptual speech event, thus the common currency is motoric in nature. Somewhat paradoxically, it is the researchers studying speech production who promote an audio-centric view and the researchers studying speech perception who promote a motor-centric view. Even more paradoxically, despite the obvious complementarity between these lines of investigation, there is virtually no theoretical interaction between them.

A major goal of this review is to consider the relation between these two ideas regarding sensorimotor interaction in speech and whether they might be integrated into a single

functional anatomic framework. To this end, we will review evidence for the role of the auditory system in speech production, evidence for the role of the motor system in speech perception, and recent progress in mapping an auditory-motor integration circuit for speech and related functions (vocal music). We will then consider a unified framework based on a state feedback control architecture, in which sensorimotor integration functions primarily in support of speech production, but can also subserve top-down motor modulation of the auditory system during speech perception. Finally, we discuss a range of possible clinical correlates of dysfunction of this sensorimotor integration circuit.

The role of the auditory system in speech production

All it takes is one bad telephone connection, in which one's own voice echoes in the earpiece with a slight delay, to be convinced that input to the auditory system affects speech production. The disruptive effect of delayed auditory feedback is well-established (Stuart et al., 2002; Yates, 1963) but is just one source of evidence for the acoustic influence on speech output. Adult-onset deafness is another: individuals who become deaf after becoming proficient with a language nonetheless suffer speech articulation declines as a result of the lack of auditory feedback which is critical to maintain phonetic precision over the long term (Waldstein, 1989). Other forms of altered auditory feedback, such as digitally shifting the voice pitch or the frequency of a speech formant (frequency band) has been shown experimentally to lead to automatic compensatory adjustments on the part of the speaker within approximately 100msec (Burnett et al., 1998; Purcell and Munhall, 2006). At a higher level of analysis, research on speech error patterns at the phonetic, lexical, and syntactic levels shows that the perceptual system plays a critical role in self-monitoring of speech output (both overt and inner speech) and that this self-perception provides feedback signals that guide repair processes in speech production (Levelt, 1983; Levelt, 1989).

It is not just acoustic perception of one's own voice that affects speech production. The common anecdotal observation that speakers can pick up accents as a result of spending extended periods in a different linguistic community, so-called `gestural drift', has been established quantitatively (Sancier and Fowler, 1997). In the laboratory setting, it has been shown that phonetic patterns such voice pitch and vowel features introduced into "ambient speech" of the experimental setting is unintentionally (i.e., automatically) reproduced in the subjects' speech (Cooper and Lauritsen, 1974; Delvaux and Soquet, 2007; Kappes et al., 2009). This body of work demonstrates that perception of others' speech patterns influence the listener's speech patterns. Nowhere is this more evident than in development, where the acoustic input to a prelingual child determines the speech patterns s/he acquires.

Thus, it is uncontroversial that the auditory system plays an important role in speech production; without it, speech cannot be learned or maintained with normal precision. Computationally, auditory-motor interaction in the context of speech production has been characterized in terms of feedback control models. Such models can trace their lineage back to Fairbanks (Fairbanks, 1954), who adapted Wiener's (Wiener, 1948) feedback control theory to speech motor control. Fairbanks proposed that speech goals were represented in terms of a sequence of desired sensory outcomes, and that the articulators were driven to produce speech by a system that minimized the error between desired and actual sensory feedback. This idea, that sensory feedback could be the basis of online control of speech output, has persisted into more recent computational models (Guenther et al., 1998), but runs into several practical problems: stable feedback control requires non-noisy, undelayed feedback (Franklin et al., 1991), but real sensory feedback is noisy (e.g. due to background noise), delayed (due to synaptic and processing delays), and, especially in the case of auditory feedback, intermittently absent (e.g., due to loud masking noise). To address these problems, some feedback-based models have been hybridized by including a feedforward

controller that ignores sensory feedback (Golfinopoulos et al., 2009; Guenther et al., 2006). However, a more principled approach is taken by newer models of motor control derived from state feedback control (SFC) theory (Jacobs, 1993). Of late, SFC models have been highly successful at explaining the role of the CNS in non-speech motor phenomena (Shadmehr and Krakauer, 2008; Todorov, 2004) and an SFC model of speech motor control has recently been proposed (Ventura et al., 2009).

Like the Fairbanks model, in the SFC model, online articulatory control is based on feedback, but in this case not on direct sensory feedback. Instead, online feedback control comes from an internally maintained representation, an internal model estimate of the current dynamical state of the vocal tract. The internal estimate is based on previously learned associations between issued motor commands and actual sensory outcomes. Once these associations are learned, the internal system can then predict likely sensory consequences of a motor command prior to the arrival of actual sensory feedback and can use these predictions to provide rapid corrective feedback to the motor controllers if the likely sensory outcome differs from the intended outcome (Figure 1A). Thus, in the SFC framework online feedback control is achieved primarily via internal forward model predictions whereas actual feedback is used to train and update the internal model. Of course, actual feedback can also be used to correct overt prediction/feedback mismatch errors. It should be clear that this approach has much in common with self-monitoring notions developed within the context of psycholinguistic research (Levelt, 1983).

The role of the motor system in speech perception

The idea that speech perception relies critically on the motor speech system was put forward as a possible solution to the observation that there is not a one-to-one relation between acoustic patterns and perceived speech sounds (Liberman, 1957; Liberman et al., 1967). Rather, the acoustic patterns associated with individual speech sounds are context-dependent. For example, a /d/ sound has a different acoustic pattern in the context of /di/ versus /du/. This is because articulation of the following vowel is already commenced during articulation of /d/ (coarticulation). The intuition behind Liberman and colleagues' motor theory of speech perception was that while the auditory signal associated with speech sounds can be variable, the motor gestures that produce them are not: a /d/ is always produced by stopping and releasing airflow using the tip of the tongue on the roof of the mouth (although this assumption is questionable (Schwartz et al., in press)). Therefore, Liberman et al. reasoned, the goal of speech perception must be to recover the invariant motor gestures that produce speech sounds rather than to decode the acoustic patterns themselves; however, no mechanism was proposed to explain *how* the gestures were recovered (for a recent discussion see (Galantucci et al., 2006; Massaro and Chen, 2008)).

Although the motor theory represents an intriguing possible solution to a vexing problem, it turned out to be empirically incorrect in its strong form. Subsequent research has shown convincingly that the motor speech system is not necessary for solving the context-dependency problem (Lotto et al., 2009; Massaro and Chen, 2008). For example, the ability to perceive speech sounds has been demonstrated in patients who have severely impaired speech production due to chronic stroke (Naeser et al., 1989; Weller, 1993), in individuals who have acute and complete deactivation of speech production due to left carotid artery injection of sodium amobarbital (Wada procedure) (Hickok et al., 2008), in individuals who never acquired the ability to speak due to congenital disease or pre-lingual brain damage (Bishop et al., 1990; Christen et al., 2000; Lenneberg, 1962; MacNeilage et al., 1967), and even in nonhuman mammals (chinchilla) and birds (quail) (Kuhl and Miller, 1975; Lotto et al., 1997) which don't have the biological capacity to speak. Further, contextual dependence in speech perception has been demonstrated in the purely acoustic domain: perception of

syllables along a da-ga continuum -- syllables which differ in the onset frequencies of their 3rd formant -- is modulated by listening to a preceding sequence of tones with an average frequency aligned with the onset frequency of one syllable versus the other (Holt, 2005). This shows that the auditory system maintains a running estimate of the acoustic context and uses this information in the encoding of incoming sounds. Such a mechanism provides a means for dealing with acoustic variability due to co-articulation that does not rely on reconstructing motor gestures but rather uses the broader acoustic context (Holt and Lotto, 2008; Massaro, 1972). In sum, the motor system is not necessary for solving the contextual dependence problem in speech perception and the auditory system appears to have a mechanism for solving it.

The discovery of mirror neurons in macaque area F5 -- a presumed homologue to Broca's area, the classic human motor speech area -- has resurrected motor theories of perception in general (Gallese and Lakoff, 2005; Rizzolatti and Craighero, 2004) and the motor theory of speech perception in particular (Fadiga and Craighero, 2003; Fadiga et al., 2009; Rizzolatti and Arbib, 1998). Mirror neurons fire both during the execution and observation of actions and are widely promoted as supporting the "understanding" of actions via motor simulation (di Pellegrino et al., 1992; Gallese et al., 1996; Rizzolatti and Craighero, 2004), although this view has been challenged on several fronts (Corina and Knapp, 2006; Emmorey et al., 2010; Hauser and Wood, 2010; Heyes, 2010; Hickok, 2009a; Hickok and Hauser, in press; Knapp and Corina, 2010; Mahon and Caramazza, 2008). It is important to recognize that the discovery of mirror neurons, while interesting, does not negate the empirical evidence against a strong motor theory of speech perception (Hickok, 2010b; Lotto et al., 2009) and any theory of speech perception will have to take previous evidence into account (Hickok, 2010a). Unfortunately, mirror neuron inspired discussions of speech perception (Fadiga et al., 2009; Pulvermuller et al., 2006) have not taken this broader literature into account (Skoyles, 2010).

This renewed interest in the motor theory has generated a flurry of studies that have suggested a *limited* role for the motor system in speech perception. Several transcranial magnetic stimulation (TMS) and functional imaging experiments have found that the perception of speech, with no explicit motor task, is sufficient to activate (or potentiate) the motor speech system in a highly specific, i.e., somatotopic, fashion (Fadiga et al., 2002; Pulvermuller et al., 2006; Skipper et al., 2005; Sundara et al., 2001; Watkins and Paus, 2004; Watkins et al., 2003; Wilson et al., 2004). But it is unclear whether such activations are causally related to speech recognition or rather are epiphenomenal, reflecting spreading activation between associated networks. For this reason, more recent studies have attempted to modulate perceptual responses via motor-speech stimulation, with some success. One study showed that stimulation of premotor cortex resulted in a decline in the ability to identify syllables in noise (Meister et al., 2007), while another stimulated the same region using clear speech stimuli and found no effect on accuracy across several measures of speech perception but reported that response times in one task were slowed (namely, a phoneme discrimination task in which subjects judged whether pairs of syllables start with the same sound or not). A third study found that stimulation of motor lip or tongue areas resulted in a facilitation (faster reaction times) in identification of lip- or tongue-related speech sounds (D'Ausilio et al., 2009), and a fourth found that stimulation of motor lip areas resulted in decreased ability to discriminate lip-related speech sounds (Mottonen and Watkins, 2009). Still other work has found that motor learning can also modulate the perception of speech (Shiller et al., 2009). It is relevant that these effects emerge either only when the speech sounds are partially ambiguous and/or in reaction time measures only indicating that the effects are quite subtle. Although suggestive, there is reason to question all of this work as none of them used methods that separate perceptual discrimination effects from response bias effects (Hickok, 2010b). But for present purposes we will assume their

validity for the sake of argument. Assuming that there is a real effect on perception, the totality of the evidence clearly indicates a *modulatory* effect rather than a necessary component of speech sound recognition.

In sum, there is unequivocal neuropsychological evidence that a strong version of the motor theory of speech perception, one in which the motor system is a necessary component, is untenable. However, there is suggestive evidence that the motor system is capable of *modulating* the perceptual system to some degree. Models of speech perception will need to account for both sets of observations.

The functional anatomy of sensorimotor integration for speech

During the last decade a great deal of progress has been made in mapping the neural organization of sensorimotor integration for speech. Early functional imaging studies identified an auditory-related area in the left planum temporale region as involved in speech *production* (Hickok et al., 2000; Wise et al., 2001). Subsequent studies showed that this left dominant region, dubbed Spt for its location in the Sylvian fissure at the parietal-temporal boundary (Figure 2A) (Hickok et al., 2003), exhibited a number of properties characteristic of sensorimotor integration areas such as those found in macaque parietal cortex (Andersen, 1997; Colby and Goldberg, 1999). Most fundamentally, Spt exhibits sensorimotor response properties, activating both during the passive perception of speech and during covert (subvocal) speech articulation (covert speech was used to ensure that overt auditory feedback was not driving the activation) (Buchsbaum et al., 2001; Buchsbaum et al., 2005; Hickok et al., 2003), and further that different subregional *patterns* of activity are apparent during the sensory and motor phases of the task (Hickok et al., 2009b), likely reflecting the activation of different neuronal subpopulations (Dahl et al., 2009) some sensory- and others motor-weighted. Figure 2B–D show examples of the sensory-motor response properties of Spt and the patchy organization of this region for sensory- versus motor-weighted voxels (2C, inset). Spt is not speech specific; its sensorimotor responses are equally robust when the sensory stimulus is tonal melodies and (covert) humming is the motor task (see the two curves in Figure 2B) (Hickok et al., 2003). Activity in Spt is highly correlated with activity in the pars opercularis (Buchsbaum et al., 2001; Buchsbaum et al., 2005), which is the posterior sector of Broca's region. White matter tracts identified via diffusion tensor imaging suggest that Spt and the pars opercularis are densely connected anatomically (for review see (Friederici, 2009; Rogalsky and Hickok, in press)). Finally, consistent with some sensorimotor integration areas in the monkey parietal lobe (Andersen, 1997; Colby and Goldberg, 1999), Spt appears to be motor-effector selective, responding more robustly when the motor task involves the vocal tract than the manual effectors (Figure 2D) (Pa and Hickok, 2008). More broadly, Spt is situated in the middle of a network of auditory (superior temporal sulcus) and motor (pars opercularis, premotor cortex) regions (Buchsbaum et al., 2001; Buchsbaum et al., 2005; Hickok et al., 2003), perfectly positioned both functionally and anatomically to support sensorimotor integration for speech and related vocal-tract functions. It is worth noting that the supramarginal gyrus, a region just dorsal to Spt in the inferior parietal lobe, has been implicated in aspects of speech production (for a recent review see Price, 2010). In group-averaged analyses using standard brain anatomy normalization, area Spt can mis-localize to the supramarginal gyrus (Isenberg et al., in press) raising the possibility that previous work implicating the this structure in speech production may in fact reflect Spt activity.

Area Spt, together with a network of regions including STG, premotor, and cerebellum, has been implicated in auditory feedback control of speech production, suggesting that Spt is part of the SFC system. In an fMRI study, Tourville et al. (Tourville et al., 2008) asked subjects to articulate speech and either fed it back altered (up or down shift of the first

formant frequency) or unaltered. Shifted compared to unshifted speech feedback resulted in activation of area Spt, as well as bilateral superior temporal areas, right motor and somatosensory-related regions, and right cerebellum. Interestingly, damage to the vicinity of Spt has been associated with conduction aphasia, a syndrome in which sound-based errors in speech production is the dominant symptom (Baldo et al., 2008; Goodglass, 1992) and these patients have a decreased sensitivity to the normally disruptive effects of delayed auditory feedback (Boller and Marcie, 1978; Boller et al., 1978). These observations are in line with the view that Spt plays a role in auditory feedback control of speech production.

A brief digression is in order at this point regarding the functional organization of the planum temporale in relation to Spt and the mechanisms we've been discussing. The planum temporale generally has been found to activate under a variety of stimulus conditions. For example, not only is it sensitive to speech-related acoustic features as discussed above, but also to auditory spatial features (Griffiths and Warren, 2002; Rauschecker and Scott, 2009). This has led some authors to propose that the planum temporale functions as a "computational hub" (Griffiths and Warren, 2002) and/or supports a "common computational mechanism" (Rauschecker and Scott, 2009) that applies to a variety of stimulus events. An alternative is that the planum temporale is functionally segregated with, for example, one sector supporting sensorimotor integration and other sectors supporting other functions, such as spatial-related processes (Hickok, 2009b).

Several lines of evidence are consistent with this view. First, the planum temporale is comprised of several cytoarchitectonic fields, the most posterior of which, area Tpt, is outside of auditory cortex proper (Galaburda and Sanides, 1980). This suggests a multifunctional organization with a major division between auditory cortex (anterior sectors) and auditory-related cortices (posterior sectors). Second, Spt is located within the most posterior region of the planum temporale, which is consistent with its proposed functional role as at the interface of auditory and motor systems. Finally, a recent experiment that directly compared sensorimotor and spatial activations within subjects found spatially distinct patterns of activation within the planum temporale (sensorimotor activations were posterior to spatial activation) and different patterns of connectivity of the two activation foci (Isenberg et al., in press). Thus, it seems that the sensorimotor functions of Spt in the posterior planum temporale region are distinguishable from the less-well characterized auditory functions of the more anterior region(s).

Synthesis: a unified framework for sensorimotor integration for speech

The foregoing review of the literature points to several conclusions regarding sensorimotor processes in speech. On the output side it is clear that auditory information plays an important role in feedback control of speech production. On the input side it, while the motor speech system is not necessary for speech perception, it is activated during passive listening to speech and may provide a *modulatory* influence on perception of speech sounds. Finally, the neural network supporting sensorimotor functions in speech includes premotor cortex, area Spt, STG (auditory cortex), and the cerebellum. We propose a unified view of these observations within the framework of a SFC model of speech production. We suggest that such a circuit also explains, as a consequence of its feedback control computations both the activation of motor cortex during perception and the top-down modulatory influence the motor system may have on speech perception.

A state feedback control (SFC) model of speech production

As noted above, feedback control architectures for speech production have been developed previously. Here we propose a model that draws not only on recent developments in SFC theory but also seeks to integrate models of the speech processing derived from

psycholinguistic and neurolinguistic research. The model can be viewed as a spelling out of the computations involved in the "dorsal" auditory/speech stream proposed previously (Hickok and Poeppel, 2000, 2004, 2007; Rauschecker and Scott, 2009). Figure 2 shows the functional and anatomical components of one of these models with the dorsal sensory-motor stream indicated in blue shading (see figure legend for details).

As briefly discussed above, Figure 1A presents a SFC architecture presented in the context of motor control for speech production as adapted from (Ventura et al., 2009). In this framework, a motor command issued by the motor controller to the vocal tract articulators is accompanied by a corollary discharge to an internal model of the vocal tract, which represents an estimate of the dynamical state the vocal tract given the recent history of the system and incoming (corollary) motor commands. This state estimate is transformed into a forward prediction of the acoustic consequences of the motor command. We also assume that a forward prediction of the somatosensory consequences of the motor command is generated, although we will not discuss the role of this system here. The forward auditory prediction, in turn, supports two functions as noted above. One is a rapid internal monitoring function, which calculates whether the current motor commands are likely to hit their intended sensory targets (this implies that the targets are known independently of the forward predictions, see below) and provides corrective feedback if necessary. Needless to say, the usefulness of this internal feedback depends on how accurate the internal model/ forward prediction is. Therefore it is important to use actual sensory feedback to update and tune the internal model to ensure it is making accurate predictions. This is the second (slower, external monitoring) function of the forward predictions: to compare predicted with actual sensory consequences and use prediction error to generate a corrective signal to update the internal model, which in turn provides input the motor controller. Of course, if internal feedback monitoring fails to catch an error in time, external feedback can be used to correct movements as well.

As noted above, an internal feedback loop that generates a forward prediction of the sensory consequence of an action is useless if the intended sensory target is not known. This raises an interesting issue because unlike in typical visuo-manual paradigms where actions are often directed at external sensory targets, in most speech acts there is no immediate externally provided sensory target (unless one is repeating heard speech). Instead the sensory goal of a speech act is an internal representation (e.g., a sequence of speech sounds) called up from memory on the basis of a higher-level goal, namely, to express a concept via a word or phrase that corresponds to that concept. This, in turn, implies that speech production involves the activation of not only motor speech representations but also internal representations of sensory speech targets that can be used to compare against both predicted and actual consequences of motor speech acts.

Psycholinguistic models of speech production typically assume an architecture that is consistent with the idea that speech production involves the activation of a sensory target. For example, major stages of such models include: the activation of a lexical-conceptual representation, access to the corresponding phonological representation followed by articulatory coding (Dell et al., 1997; Levelt et al., 1999); both external and internal monitoring loops have been proposed (Levelt, 1983) (Figure 1B). Although the content of the phonological stage is not typically associated specifically with a sensory or motor representation in these models, several studies have suggested that the neural correlates of phonological access involve (but are not necessarily limited to) auditory-related cortices in the posterior superior temporal sulcus/gyrus (de Zubicaray and McMahon, 2009; Edwards et al., 2010; Graves et al., 2007; Graves et al., 2008; Indefrey and Levelt, 2004; Levelt et al., 1998; Okada and Hickok, 2006; Wilson et al., 2009).

Research in the neuropsychological tradition has generated additional information regarding the phonological level of processing, suggesting in fact two components of a phonological system, one corresponding to sensory input processes and another to motor output systems (Figure 1C) (Caramazza, 1988; Jacquemot et al., 2007; Shelton and Caramazza, 1999). Briefly, the motivation for this claim comes from observations that brain damage can cause a disruption of the ability to articulate words without affecting the perceptual recognition of words and in other instances can cause a disruption of word recognition without affecting speech fluency (speech output is agile, although often error prone). This viewpoint is consistent with Wernicke's early model in which he argued that the representation of speech, e.g., a word, has two components, one sensory (what the word sounds like) and one motor (what sequence of movements will generate that sequence of sounds) (Wernicke, 1874/1969). Essentially identical views have been promoted by modern theorists (Pulvermuller, 1996).

An integrated model of the speech production system can be derived by merging the three models in Figure 1. This integrated model is depicted in Figure 4. The basic architecture is that of a SFC system with motor commands generating a corollary discharge to an internal model that is used for feedback control. Input to the system comes from a lexical-conceptual network as assumed by both the psycholinguistic and neurolinguistic frameworks and the output of the system is controlled by a low-level articulatory controller as in the psycholinguistic and SFC models. In between the input/output system is a phonological system that is split into two components, corresponding to sensory input and motor output subsystems, as in the neuropsychological model. We have also added a sensorimotor translation component. Sensorimotor translation is assumed to occur in the neurolinguistic models (Jacquemot et al., 2007), and as reviewed above, Spt is a likely neural correlate of this translation system (Buchsbaum et al., 2001;Hickok et al., 2003;Hickok et al., 2009b). Similar translation networks have been identified in the primate visuomotor system (Andersen, 1997). As assumed by both the psycho- and neurolinguistic models, the phonological representations we have in mind here are relatively high-level in that they correspond to articulatory or acoustic feature bundles (on the motor and sensory sides, respectively) which define individual speech sounds (phonemes) at the finest grain and/or at a courser-grained level, correspond to sequences of sounds/movements (~syllables), a concept consistent with Levelt's mental "syllabary" (Levelt, 1989). The model also incorporates both external and internal feedback loops as in the SFC framework and in Levelt's psycholinguistic model (Levelt, 1983).

In the context of an SFC framework, two kinds of internal forward models are maintained, one that makes forward predictions regarding the state of the motor effectors and one that makes forward predictions regarding the sensory consequences of these motor effector states (Wolpert et al., 1995). Deviations between the predicted sensory consequences and the sensory targets generate an error signal that can be used to update the internal motor model and provide corrective feedback to the controller. We suggest that neuronal ensembles coding learned motor sequences, such as those stored in the hypothesized "motor phonological system" form an internal forward model of the vocal tract in the sense that activation of a code for a speech sequence, say that for articulating the word *cat*, instantiates a prediction of future states of the vocal tract, namely those corresponding to the articulation of that particular sequence of sounds. Thus, activation of the high-level motor ensemble coding for the word *cat* drives the execution of that sequence in the controller. Corollary discharge from the motor controller back to the higher-level motor phonological system can provide information (predictions) about where in the sequence of movements the vocal tract is at a given time point. Alternatively, or perhaps in conjunction, lower levels of the motor system, such as a fronto-cerebellar circuit, may fill in the details of where the vocal tract is in the predicted sequence given the particulars of the articulation, taking into account

velocities, fatigue, etc. A hierarchically organized feedback control system, with internal models and feedback loops operating at different grains of analysis, is in line with recent hypotheses (Grafton et al., 2009; Grafton and Hamilton, 2007; Krigolson and Holroyd, 2007) and makes sense in the context of speech where the motor system must hit sensory targets corresponding to features (formant frequency), sound categories (phonemes), sequences of sound categories (syllables/words), and even phrasal structures (syntax) (Levelt, 1983). Given that the concepts of sensory hierarchies and motor hierarchies are both firmly established, the idea of sensorimotor hierarchies would seem to follow (Fuster, 1995). Thus while we discuss this system at a fairly course grain of analysis, the phonological level, we are open to the possibility that both finer-grained and more coarse-grained SFC systems exist.

Activation of a motor phonological representation not only makes forward predictions regarding the state of the vocal tract, but when transformed into a sensory representation, also makes forward predictions about the sensory consequences of the movements: if the system activates the motor program for generating the word *cat*, the sensory system can expect to hear the acoustic correlates of the word. Thus, activation of the motor phonological system can generate predictions about the expected sensory consequences in the auditory phonological system. In our model, forward models of sensory events are instantiated within the sensory system. Direct evidence for this view comes from the motor-induced suppression effect: the response to hearing one's own speech is attenuated compared to hearing the same acoustic event in the absence of the motor act of speaking (e.g., when the subject's own speech is recorded and played back) (Aliu et al., 2009; Paus et al., 1996). This is expected if producing speech generates corollary discharges that propagate to the auditory system.

Wernicke proposed that speaking a word involves parallel inputs to both the motor and auditory speech systems, or in our terminology, the motor and auditory phonological systems (Wernicke, 1874/1969). His evidence for this claim was that damage to sensory speech systems (i) did not interrupt fluency, showing that it was possible to activate motor programs for speech in the absence of an intact sensory speech system, but (ii) caused errors in otherwise fluent speech, showing that the sensory system played a critical role. His clinical observations have since been confirmed: patients with left posterior temporal lobe damage produce fluent but error prone speech (Damasio, 1992; Goodglass et al., 2001; Hillis, 2007), and his theoretical conclusions are still valid. More recent work has also argued for a dual-route architecture for speech production (McCarthy and Warrington, 1984). Accordingly, we also assume that activation of the speech production network involves parallel inputs to the motor and auditory phonological systems. Activation of the auditory component comprises the sensory targets of the action, whereas activation of motor phonological system defines the initial motor plan that, via internal feedback loops can be compared against the sensory targets. In an SFC framework, damage to the auditory phonological speech system results in speech errors because the internal feedback mechanism that would normally detect and correct errors is no longer functioning. An alternative to the idea of parallel inputs to sensory and motor phonological speech systems is a model in which the initial input is to the motor component only, with sensory involvement coming only via internal feedback (Edwards et al., 2010). However, as noted above, an internal feedback signal is not useful if there is no target to reference it against. Additional evidence for parallel activation of motor and auditory phonological systems comes from conduction aphasia, which we discuss in a later section.

Neural correlates of the SFC system

Can this SFC system be localized in the brain? In broad sketch, yes. One approach to localizing the network is to use imagined speech. It has been found that imagined

movements closely parallel the timing of real movements (Decety and Michel, 1989), and research on imagined speech suggest that it shares properties with real speech, for example subjects report inner "slips of the tongue" that show a lexical bias (slips tend to form words rather than nonwords) just as in overt speech (Oppenheim and Dell, 2008). In the context of a SFC framework the ability to generate accurate estimates of the timing of a movement based on mental simulation has been attributed to the use of an internal model (Mulliken and Andersen, 2009; Shadmehr and Krakauer, 2008). Following this logic, the distribution of activity in the brain during imagined speech should provide at least a first-pass estimate of the neural correlates of the SFC network. Several studies of imagined speech (covert rehearsal) have been carried out (Buchsbaum et al., 2001; Buchsbaum et al., 2005; Hickok et al., 2003), which identified a network including the STS/STG, Spt, and premotor cortex, including both ventral and more dorsolateral regions (Figure 2A), as well as the cerebellum (Durisko and Fiez, 2010; Tourville et al., 2008). We suggest that the STS/STG corresponds to the auditory phonological system, Spt corresponds to the sensorimotor translation system, and the premotor regions correspond to the motor phonological system, consistent with previous models of these functions (Hickok, 2009b; Hickok and Poeppel, 2007). The role of the cerebellum is less clear, although it may support internal model predictions at a finer-grained level of motor control.

Lesion evidence supports the above functional localization proposal. Damage to frontal motor-related regions is associated with non-fluent speech output (classical Broca's aphasia) (Damasio, 1992; Dronkers and Baldo, 2009; Hillis, 2007) as one would expect if motor phonological representations could not be activated. Damage to the STG/STS and surrounding tissue results in fluent speech output that is characterized by speech errors (as in Wernicke's or conduction aphasia) (Damasio, 1992; Dronkers and Baldo, 2009; Hillis, 2007). Preserved fluency with such a lesion is explained on the basis of an intact motor phonological system that can be innervated directly from the lexical conceptual system. Errors follow from damage to the sensory system that codes the sensory targets of speech production: without the ability to evaluate the sensory consequences of coded movements, potential errors cannot be prevented and the error rate is therefore expected to rise. Wernicke's aphasics are typically unaware of their speech errors indicating that external feedback is also ineffective, presumably because the sensory targets cannot be activated effectively. Damage to the sensorimotor interface system should result in fluent speech with an increase in error rate because the internal feedback system is not able to detect and correct errors prior to execution. However, errors should be detectable via external feedback because the sensory targets are normally activated (assuming parallel input to motor and sensory systems). Once detected however, the error should not be correctable because of the dysfunctional sensorimotor interface. This is precisely the pattern of deficit in conduction aphasia. In otherwise fluent speech, such patients commit frequent phonemic speech errors, which they can detect and attempt to correct upon hearing their own overt speech. However, correction attempts are often unsuccessful, leading to the characteristic *conduite d'approche* behavior (repeated self-correction attempts) of these patients (Goodglass, 1992). The lesions associated with conduction aphasia have been found to overlap area Spt (Buchsbaum et al., in press). See below for further discussion of conduction aphasia.

Explaining motor effects on speech perception

As is clear from Figure 4, a sensory feedback control model of speech production includes pathways both for the activation of motor speech systems from sensory input (the feedback correction pathway) and for the activation of auditory speech systems from motor activation (the forward prediction pathway). Given this architecture, the activation of motor speech systems from passive speech listening is straightforwardly explained on the assumption that others' speech can excite the same sensory-to-motor feedback circuit. This is an empirically

defensible assumption given the necessary role of others' speech in language development and the observation that acoustic-phonetic features of ambient speech modulates the speech output of listeners (Cooper and Lauritsen, 1974;Delvaux and Soquet, 2007;Kappes et al., 2009;Sancier and Fowler, 1997). Put differently, just as it is necessary to use one's own speech feedback to generate corrective signals for motor speech acts, we also use others' speech to learn (or tune) new motor speech patterns. Thus according to this view, motor-speech networks in the frontal cortex are activated during passive speech listening not because they are critical for analyzing phonemic information for perception but rather because auditory speech information, both self and others', is relevant for production.

On the perceptual side, we suggest, following other authors (Rauschecker and Scott, 2009; Sams et al., 2005; van Wassenhove et al., 2005), that forward predictions from the motor speech system can modulate the perception of others' speech. We propose further that this is a kind of "exaptation" of a process that developed for internal-feedback control. Specifically, forward predictions are necessary for motor control because they allow the system to calculate deviation between predicted and target/observed sensory consequences. Note that a forward prediction generates a sensory expectation, or in the terminology of the attentional literature, a selective attentional gain applied to the expected sensory features (and/or suppression of irrelevant features). Thus, forward predictions generated via motor commands can function as a top-down attentional modulation of sensory systems. Such attentional modulation may be important for sensory feedback control because it sharpens the perceptual acuity of the sensory system to the relevant range of expected inputs (see below). This "attentional" mechanism might then be easily co-opted for motor-directed modulation of the perception of others' speech, which would be especially useful under noisy listening conditions, thus explaining the motor speech-induced effects of perception as summarized above.

On the face of it, there seems to be a tension between error correction and selective attention. One the one hand, selective attention increases perceptual detectability to attended features and decreases detectability to unattended features. On the other hand, for error correction the system needs to be able to detect deviations from the expected (attended) pattern. However, these two computational effects are not mutually exclusive. Suppose selective attention in this context both increases the gain of the response in networks tuned to the attended units and sharpens the tuning selectivity for the relevant features (Figure 5). The increased gain will result in facilitation of detection of the presence of expected (attended) features, whereas the sharpened tuning curve may make deviations from the expectation more salient. The idea that attention can modulate gain is well established (Boynton, 2005;McAdams and Maunsell, 1999;Moran and Desimone, 1985;Reynolds et al., 1999;Reynolds and Heeger, 2009;Treue and Martinez Trujillo, 1999;Treue and Maunsell, 1999). Whether attention can sharpen the tuning properties of neurons is less well established although limited evidence exists (Murray and Wojciulik, 2004;Spitzer et al., 1988).

An alternative approach to explaining how selective attention could both enhance detection of deviation from an expected target and enhance detection of the presence of the expected target comes from recent work on the nature of the gain modulation induced by selective attention. The traditional view is that attention to a given feature increases the gain of neurons that are selective for that feature, and this model works well for detecting the presence of a stimulus or for making coarse discriminations. However, recent theoretical and experimental work has suggested that for fine discriminations, gain is applied not to neurons coding the target feature, but to neurons tuned slightly away from the target (Jazayeri and Movshon, 2006, 2007; Navalpakkam and Itti, 2007; Regan and Beverley, 1985; Scolari and Serences, 2009; Seung and Sompolinsky, 1993). These flanking cells are maximally

informative in that their response varies the most with small changes in the stimulus feature because the stimuli fall on a steeper portion of the tuning curve compared to units tuned to the target. This work has further shown that gain is adaptively applied depending on the task to optimize performance (Jazayeri and Movshon, 2007; Scolari and Serences, 2009). For example, if a fine discrimination is required, gain is applied to the flanking units, which are maximally informative for fine discriminations, whereas if a coarse discrimination is required, gain is applied to target-tuned units, which are maximally informative for coarse discriminations.

For present purposes, we can conceptualize forward predictions as attentional gain signals that are applied adaptively depending on the task; indeed a forward prediction may be implemented via a gain allocation mechanism. If the task is to detect relatively fine deviations from the intended target during speech production, gain may be applied to neurons tuned to flanking values of a target feature thus maximizing error detection. If, on the other hand, the task is to identify, say, which syllable is being spoken by someone else, gain may be applied to cells tuned to the target features themselves, thereby facilitating identification or coarse discrimination.

No matter the details of the mechanism, the above discussion is intended to highlight (i) that a plausible mechanism exists for motor-induced modulation of speech perception within the framework of a sensory feedback control model of speech production and (ii) that error detection in one's own speech and attentional facilitation of perception of others' speech are not conflicting computational tasks. An interesting by-product of this line of thinking is that it suggests a point of contact between or even integration of research on aspects of motor control and selective attention.

Clinical correlates

Developmental or acquired dysfunction of the sensorimotor integration circuit for speech should result in clinically relevant speech disorders. Here we consider some clinical correlates of dysfunction in a SFC system for speech.

Conduction aphasia

In the visuomotor domain, damage to sensorimotor areas in the parietal lobe is associated with optic ataxia, a disorder in which patients can recognize objects but have difficulty reaching for them accurately and tend to grope for visual targets (Perenin and Vighetto, 1988; Rossetti et al., 2003). Conduction aphasia is a linguistic analog to optic ataxia in that affected patients can comprehend speech but have great difficulty repeating it verbatim (i.e., achieving auditory targets that are presented to them), often verbally "groping" for the appropriate sound sequence in their frequent phonemic errors and repeated self-correction attempts (Benson et al., 1973; Damasio and Damasio, 1980; Dronkers and Baldo, 2009; Goodglass, 1992). Classically, damage to the arcuate fasciculus, the white matter fiber bundle connecting the posterior and anterior language areas, was thought to cause conduction aphasia (Geschwind, 1965, 1971), but modern data suggest a cortical lesion centered around the temporal-parietal junction, overlapping area Spt, is the source of the deficits (Anderson et al., 1999; Baldo et al., 2008; Fridriksson et al., 2010; Hickok, 2000; Hickok et al., 2000). Interestingly, there is evidence that patients with conduction aphasia have a decreased sensitivity to the disruptive effects of delayed auditory feedback (Boller and Marcie, 1978; Boller et al., 1978) as one would expect with damage to a circuit that supports auditory feedback control of speech production.

In terms of our SFC model, and as noted above, a lesion to Spt would disrupt the ability to generate forward predictions in auditory cortex and thereby the ability to perform internal

feedback monitoring, making errors more frequent than in an unimpaired system (Figure 6A). However, this would not disrupt the activation of high-level auditory targets in the STS via the lexical semantic system, thus leaving the patient capable of detecting errors in their own speech, a characteristic of conduction aphasia. Once an error is detected however, the correction signal will not be accurately translated to the internal model of the vocal tract due to disruption of Spt. The ability to detect but not accurately correct speech errors should result in repeated unsuccessful self-correction attempts, again a characteristic of conduction aphasia. Complete disruption of the predictive/corrective mechanisms in a SFC system might be expected to result in a progressive deterioration of the speech output as noise- or drift-related errors accumulate in the system with no way of correcting them, yet conduction aphasics do not develop this kind of hopeless deterioration. This may be because sensory feedback from the somatosensory system is still intact and is sufficient to keep the system reasonably tuned, or because other, albeit less efficient pathways exist for auditory feedback control. Conduction aphasia is a relative rare chronic disorder, being more often observed in the acute disease state. Perhaps many patients learn to rely more effectively on these other mechanisms as part of the recovery process.

Developmental stuttering

Developmental stuttering is a disorder affecting speech fluency in which sounds, syllables, or words may be repeated or prolonged during speech production. Behavioral, anatomical, and computational modeling work suggests that developmental stuttering is related to dysfunction of sensorimotor integration circuits. Behaviorally, it is well-documented that delayed auditory feedback can result in a paradoxical improvement in fluency among people who stutter (Martin and Haroldson, 1979; Stuart et al., 2008). Anatomically, it has been found that this paradoxical effect is correlated with asymmetry of the planum temporale, which contains area Spt: stutterers who show the delayed auditory feedback effect were found to have a reversed planum temporale asymmetry (right > left) compared to controls (Foundas et al., 2004). Computationally, recent modeling work has led to the proposal that stuttering can be caused by dysfunction of internal models involved in motor control of speech (Max et al., 2004). Broadly consistent with this previous account we argue that in people who stutter, the internal model of the vocal tract is intact as is the sensory system/ error calculation mechanism in auditory cortex (targets are accurately coded), but the mapping between the internal model of the vocal tract and the sensory system, mediated by Spt, is *noisy* (Figure 6B). A noisy mapping between sensory and motor systems still allows the internal model to be trained because statistically it will converge on an accurate model as long as there is sufficient sampling. However, for a given utterance, the forward sensory prediction of a speech gesture will tend to generate incorrect predictions because of the increased variance of the mapping function. These incorrect predictions in turn will trigger an *invalid* error signal when compared to the (accurately represented) sensory target. This results in a sensory-to-motor "error" correction signal, which itself is noisy and inaccurate. In this way, the system ends up in an inaccurate, iterative predict-correct loop that results in stuttering (this is similar to Max et al.'s 2004 claim although the details differ somewhat). Producing speech in chorus (while others are speaking the same utterance) dramatically improves fluency in people who stutter. This may be because the sensory system (which is coding the inaccurate prediction) is bombarded with external acoustic input that matches the sensory target and thus washes out and overrides the inaccurate prediction allowing for fluent speech. The degree of noise in the sensorimotor mappings may be proportional to the load on the system, which could be realized in terms of temporal demands (speech rate) or neuromodulatory systems (e.g., stress-induced factors). Many details need to be worked out, but it there is a significant amount of circumstantial evidence implicating some aspect of the feedback control systems in developmental stuttering (Max et al., 2004).

Auditory hallucinations in schizophrenia

Although seemingly unrelated to conduction aphasia and stuttering, schizophrenia is another disorder that appears to involve an auditory feedback control dysfunction. A prominent positive symptom of schizophrenia is auditory hallucinations, typically involving perceived voices. It has recently been suggested that this symptom results from dysfunction in generating forward predictions of motor speech acts (Heinks-Maldonado et al., 2007) see also (Frith et al., 2000). The reasoning for this claim is as follows. An important additional function of motor-to-sensory corollary discharges is to distinguish self- from externally-generated action. For example, self-generated eye movements do not result in the percept of visual motion even though an image moves across the retina. If corollary discharges associated with speech acts (i) are used to distinguish self- from externally-generated speech, and (ii) if this system is imprecise in schizophrenia, self-generated speech (perhaps even subvocal speech) may be perceived as externally generated, i.e., hallucinations. Consistent with this hypothesis, a recent study found that hallucinating patients do not show the normal suppression of auditory response to self-generated speech and the degree of abnormality correlated both with severity of hallucinations and misattributions of self-generated speech (Heinks-Maldonado et al., 2007). Schizophrenics also have anatomical abnormalities of the plaunum temporale, particularly in the upper cortical layers (I–III, the cortico-cortical layers) of the caudal region (likely corresponding to the location of Spt) in the left hemisphere which show a reduced fractional volume relative to controls (Smiley et al., 2009). Thus, in schizophrenia the nature of the behavioral and physiological effects (implicating sensorimotor integration), the location of anatomical abnormalities (left posterior PT), and the level of cortical processing implicated (cortico-cortical) are all consistent with dysfunction involving area Spt. As with stuttering, a research emphasis on this functional circuit is warranted in understanding aspects of schizophrenia.

Diversity and similarity of symptoms associated with SFC dysfunction

One would not have expected a connection between disorders as apparently varied as conduction aphasia, stuttering, and schizophrenia, yet they all seem to involve, in part, dysfunction of the same region and functional circuit. A closer look at these syndromes reveals other similarities. For example, all three disorders show atypical responses to delayed auditory feedback. Fluency of speech in both stutterers and conduction aphasics is not negatively affected by delayed auditory feedback and may show paradoxical improvement (Boller et al., 1978; Martin and Haroldson, 1979; Stuart et al., 2008), whereas in schizophrenia delayed auditory feedback induces the reverse effect: greater than normal speech dysfluency (Goldberg et al., 1997). Further, both stuttering and schizophrenia appear to be associated with dopamine abnormalities: dopamine antagonists such as risperidone and olanzapine (atypical anti-psychotics commonly used to treat schizophrenia) have recently been shown to improve stuttering (Maguire et al., 2004). Although on first consideration it seems problematic to have such varied symptoms associated with disruption of the same circuit, having the opportunity to study a variety of breakdown scenarios may prove to be particularly instructive in working out the details of the circuit.

Summary and Conclusions

Research on sensorimotor integration in speech is largely fractionated, with one camp seeking to understand the role of sensory systems in production and the other camp seeking to understand the role of the motor system in perception. Little effort has been put into trying to integrate these lines of investigation. We intended here to show that such an integration is possible -- that both lines of research are studying two sides of the same coin -- and indeed potentially fruitful in that it leads to new hypotheses regarding the nature of the sensorimotor system as well as the basis for some clinical disorders.

In short, we propose that sensorimotor integration exists to support speech production, that is, the capacity to learn how to articulate the sounds of one's language, keep motor control processes tuned, and support online error detection and correction. This is achieved, we suggest, via a state feedback control mechanism. Once in place, the computational properties of the system afford the ability to modulate perceptual processes somewhat, and it is this aspect of the system that recent studies of motor involvement in perception have tapped into.

The ideas we have outlined build on previous work. The SFC model itself integrates work in psycho- and neurolinguistics with a recently proposed SFC model of speech production (Ventura et al., 2009), which itself derives from recent work on SFC systems in the visuo-manual domain (Shadmehr and Krakauer, 2008). In addition our SFC model is closely related to previous sensory feedback models of speech production (Golfinopoulos et al., 2009; Guenther et al., 1998). Neuroanatomically, our model can be viewed as an elaboration of previously proposed models of the dorsal speech stream (Hickok and Poeppel, 2000, 2004, 2007; Rauschecker and Scott, 2009). The present proposal goes beyond previous work, not only in showing how the model can accommodate motor effects on perception, by showing how state feedback control models might relate to psycholinguistic and neurolinguistic models of speech processes, and how forward predictions might be related to attentional mechanisms. We submit these as hypotheses that can provide a framework for future work in sensorimotor integration for speech processing.

Acknowledgments

References

Aliu SO, Houde JF, Nagarajan SS. Motor-induced suppression of the auditory cortex. J Cogn Neurosci 2009;21:791–802. [PubMed: 18593265]

Andersen R. Multimodal integration for the representation of space in the posterior parietal cortex. Philos Trans R Soc Lond B Biol Sci 1997;352:1421–1428. [PubMed: 9368930]

Anderson JM, Gilmore R, Roper S, Crosson B, Bauer RM, Nadeau S, Beversdorf DQ, Cibula J, Rogish M III, Kortencamp S, et al. Conduction aphasia and the arcuate fasciculus: A reexamination of the Wernicke-Geschwind model. Brain and Language 1999;70:1–12. [PubMed: 10534369]

Baldo JV, Klostermann EC, Dronkers NF. It's either a cook or a baker: patients with conduction aphasia get the gist but lose the trace. Brain Lang 2008;105:134–140. [PubMed: 18243294]

Benson DF, Sheremata WA, Bouchard R, Segarra JM, Price D, Geschwind N. Conduction aphasia: A clincopathological study. Archives of Neurology 1973;28:339–346. [PubMed: 4696016]

Bishop DV, Brown BB, Robson J. The relationship between phoneme discrimination, speech production, and language comprehension in cerebral-palsied individuals. Journal of Speech and Hearing Research 1990;33:210–219. [PubMed: 2359262]

Boller F, Marcie P. Possible role of abnormal auditory feedback in conduction aphasia. Neuropsychologia 1978;16:521–524. [PubMed: 692865]

Boller F, Vrtunski PB, Kim Y, Mack JL. Delayed auditory feedback and aphasia. Cortex 1978;14:212–226. [PubMed: 679702]

Boynton GM. Attention and visual perception. Curr Opin Neurobiol 2005;15:465–469. [PubMed: 16023853]

Buchsbaum B, Hickok G, Humphries C. Role of Left Posterior Superior Temporal Gyrus in Phonological Processing for Speech Perception and Production. Cognitive Science 2001;25:663–678.

Buchsbaum BR, Baldo J, D'Esposito M, Dronkers N, Okada K, Hickok G. Conduction Aphasia and Phonological Short-term Memory: A Meta-Analysis of Lesion and fMRI data. Brain and Language. in press.

Buchsbaum BR, Olsen RK, Koch P, Berman KF. Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. Neuron 2005;48:687–697. [PubMed: 16301183]

Burnett TA, Freedland MB, Larson CR, Hain TC. Voice F0 responses to manipulations in pitch feedback. J Acoust Soc Am 1998;103:3153–3161. [PubMed: 9637026]

Caramazza A. Some aspects of language processing revealed through the analysis of acquired aphasia: the lexical system. Annu Rev Neurosci 1988;11:395–421. [PubMed: 2452596]

Christen HJ, Hanefeld F, Kruse E, Imhauser S, Ernst JP, Finkenstaedt M. Foix-Chavany-Marie (anterior operculum) syndrome in childhood: a reappraisal of Worster-Drought syndrome. Dev Med Child Neurol 2000;42:122–132. [PubMed: 10698330]

Colby CL, Goldberg ME. Space and attention in parietal cortex. Annual Review of Neuroscience 1999;22:319–349.

Cooper WE, Lauritsen MR. Feature processing in the perception and production of speech. Nature 1974;252:121–123. [PubMed: 4425217]

Corina DP, Knapp H. Sign language processing and the mirror neuron system. Cortex 2006;42:529–539. [PubMed: 16881265]

D'Ausilio A, Pulvermuller F, Salmas P, Bufalari I, Begliomini C, Fadiga L. The motor somatotopy of speech perception. Curr Biol 2009;19:381–385. [PubMed: 19217297]

Dahl CD, Logothetis NK, Kayser C. Spatial organization of multisensory responses in temporal association cortex. J Neurosci 2009;29:11924–11932. [PubMed: 19776278]

Damasio AR. Aphasia. New England Journal of Medicine 1992;326:531–539. [PubMed: 1732792]

Damasio H, Damasio AR. The anatomical basis of conduction aphasia. Brain 1980;103:337–350. [PubMed: 7397481]

de Zubicaray GI, McMahon KL. Auditory context effects in picture naming investigated with event-related fMRI. Cogn Affect Behav Neurosci 2009;9:260–269. [PubMed: 19679762]

Decety J, Michel F. Comparative analysis of actual and mental movement times in two graphic tasks. Brain Cogn 1989;11:87–97. [PubMed: 2789819]

Dell GS, Schwartz MF, Martin N, Saffran EM, Gagnon DA. Lexical access in aphasic and nonaphasic speakers. Psychological Review 1997;104:801–838. [PubMed: 9337631]

Delvaux V, Soquet A. The influence of ambient speech on adult speech productions through unintentional imitation. Phonetica 2007;64:145–173. [PubMed: 17914281]

di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G. Understanding motor events: a neurophysiological study. Exp Brain Res 1992;91:176–180. [PubMed: 1301372]

Dronkers, N.; Baldo, J. Language: Aphasia. In: Squire, LR., editor. Encyclopedia of Neuroscience. Academic Press; Oxford: 2009. p. 343-348.

Durisko C, Fiez JA. Functional activation in the cerebellum during working memory and simple speech tasks. Cortex 2010;46:896–906. [PubMed: 19853247]

Edwards E, Nagarajan SS, Dalal SS, Canolty RT, Kirsch HE, Barbaro NM, Knight RT. Spatiotemporal imaging of cortical activation during verb generation and picture naming. Neuroimage 2010;50:291–301. [PubMed: 20026224]

Emmorey K, Xu J, Gannon P, Goldin-Meadow S, Braun A. CNS activation and regional connectivity during pantomime observation: no engagement of the mirror neuron system for deaf signers. Neuroimage 2010;49:994–1005. [PubMed: 19679192]

Fadiga L, Craighero L. New insights on sensorimotor integration: from hand action to speech perception. Brain Cogn 2003;53:514–524. [PubMed: 14642302]

Fadiga L, Craighero L, Buccino G, Rizzolatti G. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. Eur J Neurosci 2002;15:399–402. [PubMed: 11849307]

Fadiga L, Craighero L, D'Ausilio A. Broca's area in language, action, and music. Ann N Y Acad Sci 2009;1169:448–458. [PubMed: 19673823]

Fairbanks G. Systematic research in experimental phonetics: 1. A theory of the speech mechanism as a servosystem. Journal of Speech and Heearing Disorders 1954;19:133–139.

Foundas AL, Bollich AM, Feldman J, Corey DM, Hurley M, Lemen LC, Heilman KM. Aberrant auditory processing and atypical planum temporale in developmental stuttering. Neurology 2004;63:1640–1646. [PubMed: 15534249]

Franklin, GF.; Powell, JD.; Emami-Naeini, A. Feedback control of dynamic systems. Reading, Mass; Addison-Wesley: 1991.

Fridriksson J, Kjartansson O, Morgan PS, Hjaltason H, Magnusdottir S, Bonilha L, Rorden C. Impaired speech repetition and left parietal lobe damage. J Neurosci 2010;30:11057–11061. [PubMed: 20720112]

Friederici AD. Pathways to language: fiber tracts in the human brain. Trends Cogn Sci 2009;13:175–181. [PubMed: 19223226]

Frith CD, Blakemore S, Wolpert DM. Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. Brain Res Brain Res Rev 2000;31:357–363. [PubMed: 10719163]

Fuster, JM. Memory in the cerebral cortex. MIT Press; Cambridge, MA: 1995.

Galaburda A, Sanides F. Cytoarchitectonic organization of the human auditory cortex. Journal of Comparative Neurology 1980;190:597–610. [PubMed: 6771305]

Galantucci B, Fowler CA, Turvey MT. The motor theory of speech perception reviewed. Psychon Bull Rev 2006;13:361–377. [PubMed: 17048719]

Gallese V, Fadiga L, Fogassi L, Rizzolatti G. Action recognition in the premotor cortex. Brain 1996;119(Pt 2):593–609. [PubMed: 8800951]

Gallese V, Lakoff G. The brain's concepts: The role of the sensory-motor system in conceptual knowledge. Cogn Neuropsychol 2005;22:455–479. [PubMed: 21038261]

Geschwind N. Disconnexion syndromes in animals and man. Brain 1965;88:237–294. 585–644. [PubMed: 5318481]

Geschwind N. Aphasia. New England Journal of Medicine 1971;284:654–656. [PubMed: 5545606]

Goldberg TE, Gold JM, Coppola R, Weinberger DR. Unnatural practices, unspeakable actions: a study of delayed auditory feedback in schizophrenia. Am J Psychiatry 1997;154:858–860. [PubMed: 9167517]

Golfinopoulos E, Tourville JA, Guenther FH. The integration of large-scale neural network modeling and functional brain imaging in speech motor control. Neuroimage. 2009

Goodglass, H. Diagnosis of conduction aphasia. In: Kohn, SE., editor. Conduction aphasia. Lawrence Erlbaum Associates; Hillsdale, N.J.: 1992. p. 39-49.

Goodglass, H.; Kaplan, E.; Barresi, B. The assessment of aphasia and related disorders. 3rd ed. Lippincott Williams & Wilkins; Philadelphia: 2001.

Grafton, ST.; Aziz-Zadeh, L.; Ivry, RB. Relative hierarchies and the representation of action. In: Gazzaniga, MS., editor. The cognitive neurosciences. MIT Press; Cambridge, MA: 2009. p. 641-652.

Grafton ST, Hamilton AF. Evidence for a distributed hierarchy of action representation in the brain. Hum Mov Sci 2007;26:590–616. [PubMed: 17706312]

Graves WW, Grabowski TJ, Mahta S, Gordon JK. A neural signature of phonological access: distinguishing the effects of word frequency from familiarity and length in overt picture naming. Journal of Cognitive Neuroscience 2007;19:617–631. [PubMed: 17381253]

Graves WW, Grabowski TJ, Mehta S, Gupta P. The left posterior superior temporal gyrus participates specifically in accessing lexical phonology. J Cogn Neurosci 2008;20:1698–1710. [PubMed: 18345989]

Griffiths TD, Warren JD. The planum temporale as a computational hub. Trends Neurosci 2002;25:348–353. [PubMed: 12079762]

Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang 2006;96:280–301. [PubMed: 16040108]

Guenther FH, Hampson M, Johnson D. A theoretical investigation of reference frames for the planning of speech movements. Psychological Review 1998;105:611–633. [PubMed: 9830375]

Haruno M, Wolpert DM, Kawato M. Mosaic model for sensorimotor learning and control. Neural Comput 2001;13:2201–2220. [PubMed: 11570996]

Hauser M, Wood J. Evolving the capacity to understand actions, intentions, and goals. Annu Rev Psychol 2010;61:303–324. C301. [PubMed: 19575605]

Heinks-Maldonado TH, Mathalon DH, Houde JF, Gray M, Faustman WO, Ford JM. Relationship of imprecise corollary discharge in schizophrenia to auditory hallucinations. Arch Gen Psychiatry 2007;64:286–296. [PubMed: 17339517]

Heyes C. Where do mirror neurons come from? Neurosci Biobehav Rev 2010;34:575–583. [PubMed: 19914284]

Hickok, G. Speech perception, conduction aphasia, and the functional neuroanatomy of language. In: Grodzinsky, Y.; Shapiro, L.; Swinney, D., editors. Language and the brain. Academic Press; San Diego: 2000. p. 87-104.

Hickok G. Eight problems for the mirror neuron theory of action understanding in monkeys and humans. J Cogn Neurosci 2009a;21:1229–1243. [PubMed: 19199415]

Hickok G. The functional neuroanatomy of language. Physics of Life Reviews 2009b;6:121–143. [PubMed: 20161054]

Hickok G. The role of mirror neurons in speech and language processing. Brain Lang 2010a;112:1–2. [PubMed: 19948355]

Hickok G. The role of mirror neurons in speech perception and action word semantics. Language and Cognitive Processes 2010b;25:749–776.

Hickok G, Buchsbaum B. Temporal lobe speech perception systems are part of the verbal working memory circuit: evidence from two recent fMRI studies. Behav Brain Sci 2003;26:740–741.

Hickok G, Buchsbaum B, Humphries C, Muftuler T. Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. Journal of Cognitive Neuroscience 2003;15:673–682. [PubMed: 12965041]

Hickok G, Erhard P, Kassubek J, Helms-Tillery AK, Naeve-Velguth S, Strupp JP, Strick PL, Ugurbil K. A functional magnetic resonance imaging study of the role of left posterior superior temporal gyrus in speech production: implications for the explanation of conduction aphasia. Neuroscience Letters 2000;287:156–160. [PubMed: 10854735]

Hickok G, Hauser M. (Mis)understanding mirror neurons. Current Biology. in press.

Hickok G, Okada K, Barr W, Pa J, Rogalsky C, Donnelly K, Barde L, Grant A. Bilateral capacity for speech sound processing in auditory comprehension: evidence from Wada procedures. Brain Lang 2008;107:179–184. [PubMed: 18976806]

Hickok G, Okada K, Serences JT. Area Spt in the Human Planum Temporale Supports Sensory-Motor Integration for Speech Processing. J Neurophysiol. 2009a

Hickok G, Okada K, Serences JT. Area Spt in the human planum temporale supports sensory-motor integration for speech processing. J Neurophysiol 2009b;101:2725–2732. [PubMed: 19225172]

Hickok G, Poeppel D. Towards a functional neuroanatomy of speech perception. Trends in Cognitive Sciences 2000;4:131–138. [PubMed: 10740277]

Hickok G, Poeppel D. Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. Cognition 2004;92:67–99. [PubMed: 15037127]

Hickok G, Poeppel D. The cortical organization of speech processing. Nat Rev Neurosci 2007;8:393–402. [PubMed: 17431404]

Hillis AE. Aphasia: progress in the last quarter of a century. Neurology 2007;69:200–213. [PubMed: 17620554]

Holt LL. Temporally nonadjacent nonlinguistic sounds affect speech categorization. Psychol Sci 2005;16:305–312. [PubMed: 15828978]

Holt LL, Lotto AJ. Speech perception within an auditory cognitive science framework. Current Directions in Psychological Science 2008;17:42–46. [PubMed: 19060961]

Indefrey P, Levelt WJ. The spatial and temporal signatures of word production components. Cognition 2004;92:101–144. [PubMed: 15037128]

Isenberg AL, Vaden KL, Saberi K, Muftuler LT, Hickok G. Functionally Distinct Regions for Spatial Processing and Sensory Motor Integration in the Planum Temporale. Human Brain Mapping. in press.

Jacobs, OLR. Introduction to controll theory. Oxford University Press; Oxford: 1993.

Jacquemot C, Dupoux E, Bachoud-Levi AC. Breaking the mirror: Asymmetrical disconnection between the phonological input and output codes. Cogn Neuropsychol 2007;24:3–22. [PubMed: 18416481]

Jazayeri M, Movshon JA. Optimal representation of sensory information by neural populations. Nat Neurosci 2006;9:690–696. [PubMed: 16617339]

Jazayeri M, Movshon JA. A new perceptual illusion reveals mechanisms of sensory decoding. Nature 2007;446:912–915. [PubMed: 17410125]

Kappes J, Baumgaertner A, Peschke C, Ziegler W. Unintended imitation in nonword repetition. Brain Lang 2009;111:140–151. [PubMed: 19811813]

Knapp HP, Corina DP. A human mirror neuron system for language: Perspectives from signed languages of the deaf. Brain Lang 2010;112:36–43. [PubMed: 19576628]

Krigolson OE, Holroyd CB. Hierarchical error processing: different errors, different systems. Brain Res 2007;1155:70–80. [PubMed: 17498670]

Kuhl PK, Miller JD. Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. Science 1975;190:69–72. [PubMed: 1166301]

Lenneberg EH. Understanding language without ability to speak: a case report. Journal of Abnormal and Social Psychology 1962;65:419–425. [PubMed: 13929636]

Levelt WJ. Monitoring and self-repair in speech. Cognition 1983;14:41–104. [PubMed: 6685011]

Levelt, WJM. Speaking: From intention to articulation. MIT Press; Cambridge, MA: 1989.

Levelt WJM, Praamstra P, Meyer AS, Helenius P, Salmelin R. An MEG study of picture naming. Journal of Cognitive Neuroscience 1998;10:553–567. [PubMed: 9802989]

Levelt WJM, Roelofs A, Meyer AS. A theory of lexical access in speech production. Behavioral & Brain Sciences 1999;22:1–75. [PubMed: 11301520]

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

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. Psychol Rev 1967;74:431–461. [PubMed: 4170865]

Liberman AM, Mattingly IG. The motor theory of speech perception revised. Cognition 1985;21:1–36. [PubMed: 4075760]

Liberman AM, Mattingly IG. A specialization for speech perception. Science 1989;243:489–494. [PubMed: 2643163]

Lotto AJ, Hickok GS, Holt LL. Reflections on Mirror Neurons and Speech Perception. Trends Cogn Sci 2009;13:110–114. [PubMed: 19223222]

Lotto AJ, Kluender KR, Holt LL. Perceptual compensation for coarticulation by Japanese quail (Coturnix coturnix japonica). J Acoust Soc Am 1997;102:1134–1140. [PubMed: 9265760]

MacNeilage PF, Rootes TP, Chase RA. Speech production and perception in a patient with severe impairment of somesthetic perception and motor control. J Speech Hear Res 1967;10:449–467. [PubMed: 6081929]

Maguire GA, Yu BP, Franklin DL, Riley GD. Alleviating stuttering with pharmacological interventions. Expert Opin Pharmacother 2004;5:1565–1571. [PubMed: 15212606]

Mahon BZ, Caramazza A. A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. J Physiol Paris 2008;102:59–70. [PubMed: 18448316]

Martin R, Haroldson SK. Effects of five experimental treatments of stuttering. J Speech Hear Res 1979;22:132–146. [PubMed: 502494]

Martinez-Trujillo JC, Treue S. Feature-based attention increases the selectivity of population responses in primate visual cortex. Curr Biol 2004;14:744–751. [PubMed: 15120065]

Massaro DW. Preperceptual images, processing time, and perceptual units in auditory perception. Psychol Rev 1972;79:124–145. [PubMed: 5024158]

Massaro DW, Chen TH. The motor theory of speech perception revisited. Psychon Bull Rev 2008;15:453–457. discussion 458–462. [PubMed: 18488668]

Max L, Guenther FH, Gracco VL, Ghosh SS, Wallace ME. Unstable or insufficiently activated internal models and feedback-biased motor control as sournces of dysfluency: A theoretical model of stuttering. Contemporary Issue in Communication Science and Disorders 2004;31:105–122.

McAdams CJ, Maunsell JH. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. J Neurosci 1999;19:431–441. [PubMed: 9870971]

McCarthy R, Warrington EK. A two-route model of speech production. Evidence from aphasia. Brain 1984;107:463–485. [PubMed: 6722512]

Meister IG, Wilson SM, Deblieck C, Wu AD, Iacoboni M. The essential role of premotor cortex in speech perception. Curr Biol 2007;17:1692–1696. [PubMed: 17900904]

Moran J, Desimone R. Selective attention gates visual processing in the extrastriate cortex. Science 1985;229:782–784. [PubMed: 4023713]

Mottonen R, Watkins KE. Motor representations of articulators contribute to categorical perception of speech sounds. J Neurosci 2009;29:9819–9825. [PubMed: 19657034]

Mulliken, GH.; Andersen, RA. Forward models and state estimation in posterior parietal cortex. In: Gazzaniga, MS., editor. The cognitive neurosciences. MIT Press; Cambridge, MA: 2009. p. 599-611.

Murray SO, Wojciulik E. Attention increases neural selectivity in the human lateral occipital complex. Nat Neurosci 2004;7:70–74. [PubMed: 14647291]

Naeser MA, Palumbo CL, Helm-Estabrooks N, Stiassny-Eder D, Albert ML. Severe nonfluency in aphasia: Role of the medical subcallosal fasciculus and other white matter pathways in recovery of spontaneous speech. Brain 1989;112:1–38. [PubMed: 2917272]

Navalpakkam V, Itti L. Search goal tunes visual features optimally. Neuron 2007;53:605–617. [PubMed: 17296560]

Okada K, Hickok G. Left posterior auditory-related cortices participate both in speech perception and speech production: Neural overlap revealed by fMRI. Brain Lang 2006;98:112–117. [PubMed: 16716388]

Oppenheim GM, Dell GS. Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. Cognition 2008;106:528–537. [PubMed: 17407776]

Pa J, Hickok G. A parietal-temporal sensory-motor integration area for the human vocal tract: Evidence from an fMRI study of skilled musicians. Neuropsychologia 2008;46:362–368. [PubMed: 17709121]

Paus T, Perry DW, Zatorre RJ, Worsley KJ, Evans AC. Modulation of cerebral blood flow in the human auditory cortex during speech: Role of motor-to-sensory discharges. European Journal of Neuroscience 1996;8:2236–2246. [PubMed: 8950088]

Perenin M-T, Vighetto A. Optic ataxia: A specific disruption in visuomotor mechanisms. I. Different aspects of the deficit in reaching for objects. Brain 1988;111:643–674. [PubMed: 3382915]

Pulvermuller F. Hebb's concept of cell assemblies and the psychophysiology of word processing. Psychophysiology 1996;33:317–333. [PubMed: 8753932]

Pulvermuller F, Huss M, Kherif F, Moscoso del Prado Martin F, Hauk O, Shtyrov Y. Motor cortex maps articulatory features of speech sounds. Proc Natl Acad Sci U S A 2006;103:7865–7870. [PubMed: 16682637]

Purcell DW, Munhall KG. Compensation following real-time manipulation of formants in isolated vowels. J Acoust Soc Am 2006;119:2288–2297. [PubMed: 16642842]

Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 2009;12:718–724. [PubMed: 19471271]

Regan D, Beverley KI. Postadaptation orientation discrimination. J Opt Soc Am A 1985;2:147–155. [PubMed: 3973752]

Reynolds JH, Chelazzi L, Desimone R. Competitive mechanisms subserve attention in macaque areas V2 and V4. J Neurosci 1999;19:1736–1753. [PubMed: 10024360]

Reynolds JH, Heeger DJ. The normalization model of attention. Neuron 2009;61:168–185. [PubMed: 19186161]

Rizzolatti G, Arbib M. Language within our grasp. Trends in Neurosciences 1998;21:188–194. [PubMed: 9610880]

Rizzolatti G, Craighero L. The mirror-neuron system. Annu Rev Neurosci 2004;27:169–192. [PubMed: 15217330]

Rogalsky C, Hickok G. The role of Broca's area in sentence comprehension. Journal of Cognitive Neuroscience. in press.

Rossetti Y, Pisella L, Vighetto A. Optic ataxia revisited: visually guided action versus immediate visuomotor control. Exp Brain Res 2003;153:171–179. [PubMed: 13680044]

Sams M, Mottonen R, Sihvonen T. Seeing and hearing others and oneself talk. Brain Res Cogn Brain Res 2005;23:429–435. [PubMed: 15820649]

Sancier ML, Fowler CA. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. Journal of Phonetics 1997;25:421–436.

Schwartz J-L, Basirat A, Menard L, Sato M. The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. Journal of Neurolinguistics. in press.

Scolari M, Serences JT. Adaptive allocation of attentional gain. J Neurosci 2009;29:11933–11942. [PubMed: 19776279]

Seung HS, Sompolinsky H. Simple models for reading neuronal population codes. Proc Natl Acad Sci U S A 1993;90:10749–10753. [PubMed: 8248166]

Shadmehr R, Krakauer JW. A computational neuroanatomy for motor control. Exp Brain Res 2008;185:359–381. [PubMed: 18251019]

Shelton JR, Caramazza A. Deficits in lexical and semantic processing: Implications for models of normal language. Psychon B Rev 1999;6:5–27.

Shiller DM, Sato M, Gracco VL, Baum SR. Perceptual recalibration of speech sounds following speech motor learning. J Acoust Soc Am 2009;125:1103–1113. [PubMed: 19206885]

Skipper JI, Nusbaum HC, Small SL. Listening to talking faces: motor cortical activation during speech perception. Neuroimage 2005;25:76–89. [PubMed: 15734345]

Skoyles JR. Mapping of heard speech into articulation information and speech acquisition. Proc Natl Acad Sci U S A 2010;107:E73. author reply E74. [PubMed: 20427741]

Smiley JF, Rosoklija G, Mancevski B, Mann JJ, Dwork AJ, Javitt DC. Altered volume and hemispheric asymmetry of the superficial cortical layers in the schizophrenia planum temporale. Eur J Neurosci 2009;30:449–463. [PubMed: 19656176]

Spitzer H, Desimone R, Moran J. Increased attention enhances both behavioral and neuronal performance. Science 1988;240:338–340. [PubMed: 3353728]

Stuart A, Frazier CL, Kalinowski J, Vos PW. The effect of frequency altered feedback on stuttering duration and type. J Speech Lang Hear Res 2008;51:889–897. [PubMed: 18658059]

Stuart A, Kalinowski J, Rastatter MP, Lynch K. Effect of delayed auditory feedback on normal speakers at two speech rates. J Acoust Soc Am 2002;111:2237–2241. [PubMed: 12051443]

Sundara M, Namasivayam AK, Chen R. Observation-execution matching system for speech: a magnetic stimulation study. Neuroreport 2001;12:1341–1344. [PubMed: 11388407]

Todorov E. Optimality principles in sensorimotor control. Nature Neuroscience 2004;7:907–915.

Tourville JA, Reilly KJ, Guenther FH. Neural mechanisms underlying auditory feedback control of speech. Neuroimage 2008;39:1429–1443. [PubMed: 18035557]

Treue S, Martinez Trujillo JC. Feature-based attention influences motion processing gain in macaque visual cortex. Nature 1999;399:575–579. [PubMed: 10376597]

Treue S, Maunsell JH. Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. J Neurosci 1999;19:7591–7602. [PubMed: 10460265]

van Wassenhove V, Grant KW, Poeppel D. Visual speech speeds up the neural processing of auditory speech. Proc Natl Acad Sci U S A 2005;102:1181–1186. [PubMed: 15647358]

Ventura MI, Nagarajan SS, Houde JF. Speech target modulates speaking induced suppression in auditory cortex. BMC Neurosci 2009;10:58. [PubMed: 19523234]

Waldstein RS. Effects of postlingual deafness on speech production: Implications for the role of auditory feedback. Journal of the Acoustical Society of America 1989;88:2099–2144. [PubMed: 2269726]

Watkins K, Paus T. Modulation of motor excitability during speech perception: the role of Broca's area. J Cogn Neurosci 2004;16:978–987. [PubMed: 15298785]

Watkins KE, Strafella AP, Paus T. Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia 2003;41:989–994. [PubMed: 12667534]

Weller M. Anterior opercular cortex lesions cause dissociated lower cranial nerve palsies and anarthria but no aphasia: Foix-Chavany-Marie syndrome and "automatic voluntary dissociation" revisited. J Neurol 1993;240:199–208. [PubMed: 7684439]

Wernicke, C. The symptom complex of aphasia: A psychological study on an anatomical basis. In: Cohen, RS.; Wartofsky, MW., editors. Boston studies in the philosophy of science. D. Reidel Publishing Company; Dordrecht: 1874/1969. p. 34-97.

Wiener, N. Cybernetics: Control and Communication in the Animal and the Machine. John Wiley & Sons, Inc.; New York: 1948.

Wilson SM, Isenberg AL, Hickok G. Neural correlates of word production stages delineated by parametric modulation of psychollinguistic variables. Human Brain Mapping. 2009

Wilson SM, Saygin AP, Sereno MI, Iacoboni M. Listening to speech activates motor areas involved in speech production. Nat Neurosci 2004;7:701–702. [PubMed: 15184903]

Wise RJS, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA. Separate neural sub-systems within "Wernicke's area". Brain 2001;124:83–95. [PubMed: 11133789]

Wolpert DM, Ghahramani Z, Jordan MI. An internal model for sensorimotor integration. Science 1995;269:1880–1882. [PubMed: 7569931]

Yates AJ. Delayed auditory feedback. Psychological Bulletin 1963;60:213–251. [PubMed: 14002534]
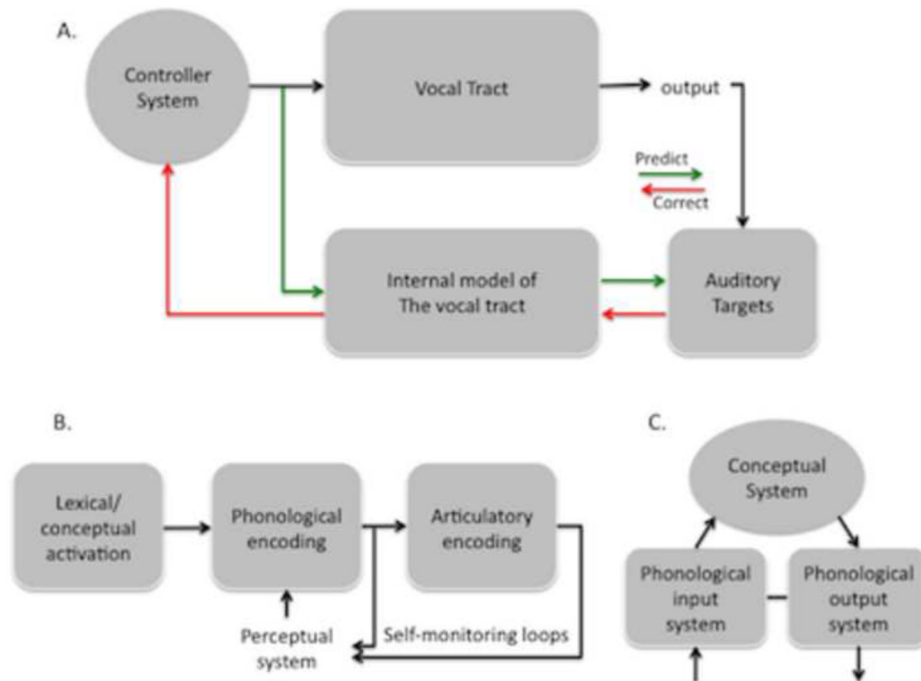
Figure 1. Models of speech processing
A. State feedback control (SFC) model of speech production. The vocal tract is controlled by a motor controller, or set of controller (Haruno et al., 2001). Motor commands issued to the vocal tract send an efferent copy to an internal model of the vocal tract. The internal model of the vocal tract maintains an estimate of the current dynamic state of vocal tract. The sensory consequences of an issued motor command are predicted by a function that translates the current dynamic state estimate of the vocal tract into an auditory representation. Predicted auditory consequences can be compared against both the intended and actual auditory targets of a motor command. Deviation between the predicted vs. intended/actual targets result in the generation of an error correction signal that feeds back into the internal model and ultimately to the motor controllers. See text for details. B. A psycholinguistic model of speech production. Although details vary, psycholinguistic models of speech production agree on a multistage process that includes minimally a lexical/ conceptual system, a phonological system, and an articulatory process that generates the motor code to produce speech (Dell et al., 1997; Levelt et al., 1999). C. A neurolinguistic model of speech processing. Research from patients with language disorders have documented dissociations in the ability to access phonological codes for receptive and expressive speech, leading to the idea that phonological processes have separable but linked motor and sensory components (Jacquemot et al., 2007).
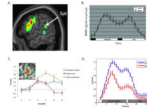
Figure 2. Location and functional properties of area Spt
A. Activation map for covert speech articulation (rehearsal of a set of nonwords), from (Hickok and Buchsbaum, 2003). B. Activation timecourse (fMRI signal amplitude) in Spt during a sensorimotor task for speech and music. A trial is composed of 3s of auditory stimulation followed by 15s covert rehearsal/humming of the heard stimulus followed by 3s seconds of auditory stimulation followed by 15s of rest. The two humps represent the sensory responses, the valley between the humps is the motor (covert rehearsal) response, and the baseline values at the onset and offset of the trial reflect resting activity levels. Note similar response to both speech and music. Adapted from (Hickok et al., 2003)C. Activation timecourse in Spt in three conditions, continuous speech (15s, blue curve), listen+rest (3s speech, 12s rest, red curve), and listen+covert rehearse (3s speech, 12s rehearse, green curve). The pattern of activity within Spt (inset) was found to be different for listening to speech compared to rehearsing speech assessed at the end of the continuous listen versus listen+rehearse conditions despite the lack of a significant signal amplitude difference at that time point. Adapted from (Hickok et al., 2009a). D. Activation timecourse in Spt in skilled pianists performing a sensorimotor task involving listening to novel melodies and then covertly humming them (blue curve) vs. listening to novel melodies and imagine playing them on a keyboard (red curve). This indicates that Spt is relatively selective for vocal tract actions. Reprinted with permission from (Hickok, 2009b).
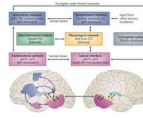
Figure 3. Dual stream model of speech processing
The dual stream model (Hickok and Poeppel, 2000, 2004, 2007) holds that early stages of speech processing occurs bilaterally in auditory regions on the dorsal STG (spectrotemporal analysis; green) and STS (phonological access/representation; yellow), and then diverges into two broad streams: a temporal lobe ventral stream supports speech comprehension (lexical access and combinatorial processes; pink) whereas a strongly left dominant dorsal stream supports sensory-motor integration and involves structures at the parietal-temporal junction (Spt) and frontal lobe. The conceptual network (gray box) is assumed to be widely distributed throughout cortex. IFG, inferior frontal gyrus; ITS, inferior temporal sulcus; MTG, middle temporal gyrus; PM, premotor; Spt, Sylvian parietal-temporal; STG, superior temporal gyrus; STS, superior temporal sulcus Figure reprinted with permission from (Hickok and Poeppel, 2007).
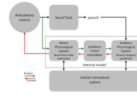
Figure 4. An integrated state feedback control (SFC) model of speech production
Speech models derived from the feedback control, psycholinguistic, and neurolinguistic literatures are integrated into one framework, presented here. The architecture is fundamentally that of a SFC system with a controller, or set of controllers (Haruno et al., 2001), localized to primary motor cortex, which generates motor commands to the vocal tract and sends a corollary discharge to an internal model which makes forward predictions about both the dynamic state of the vocal tract and about the sensory consequences of those states. Deviations between predicted auditory states and the intended targets or actual sensory feedback generates an error signal that is used to correct and update the internal model of the vocal tract. The internal model of the vocal tract is instantiated as a "motor phonological system", which corresponds to the neurolinguistically elucidated *phonological output lexicon*, and is localized to premotor cortex. Auditory targets and forward predictions of sensory consequences are encoded in the same network, namely the "auditory phonological system", which corresponds to the neurolinguistically elucidated *phonological input lexicon*, and is localized to the STG/STS. Motor and auditory phonological systems are linked via an auditory-motor translation system, localized to area Spt. The system is activated via parallel inputs from the lexical-conceptual system to the motor and auditory phonological systems.
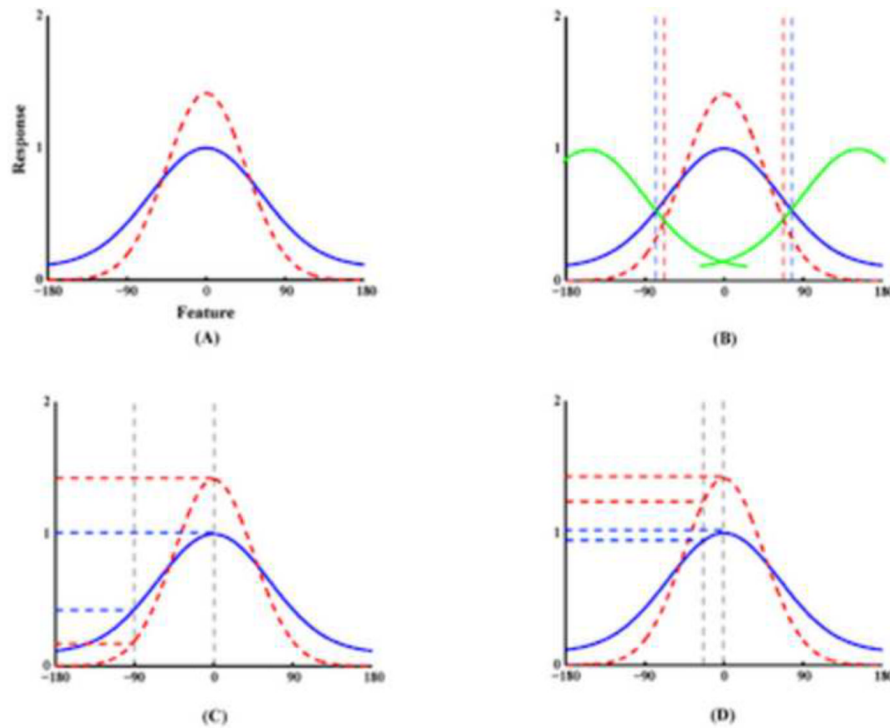
Figure 5.
Top-down modulation of perceptual response functions. **(A)** A graph replicated qualitatively from figure 2 of Boynton (Boynton, 2005) illustrating attentional effects on sensory response functions based on a `feature-similarity gain model' (Martinez-Trujillo and Treue, 2004). The effects include enhancement of the responses to the attended features and suppression of the responses to the unattended features (red dash line curve vs. blue solid line curve as modulated vs. baseline). **(B)** Increased discrimination capacity. Inward-shift of the boundaries (vertical dashed lines) makes it more likely for other perceptual `channels' (green solid curves) to respond to stimuli with features different from the attended due to the sharpened response profile in the `attended channel'. **(C)** Enhancement of the perceptual selectivity between different features achieved by increases of the response to the attended and decreases of the response to the unattended when the features are significantly different from each other, and **(D)** For features similar to the focus of attention, the contrast between responses to attended and unattended features is also increased though both responses to attended and unattended are increased.
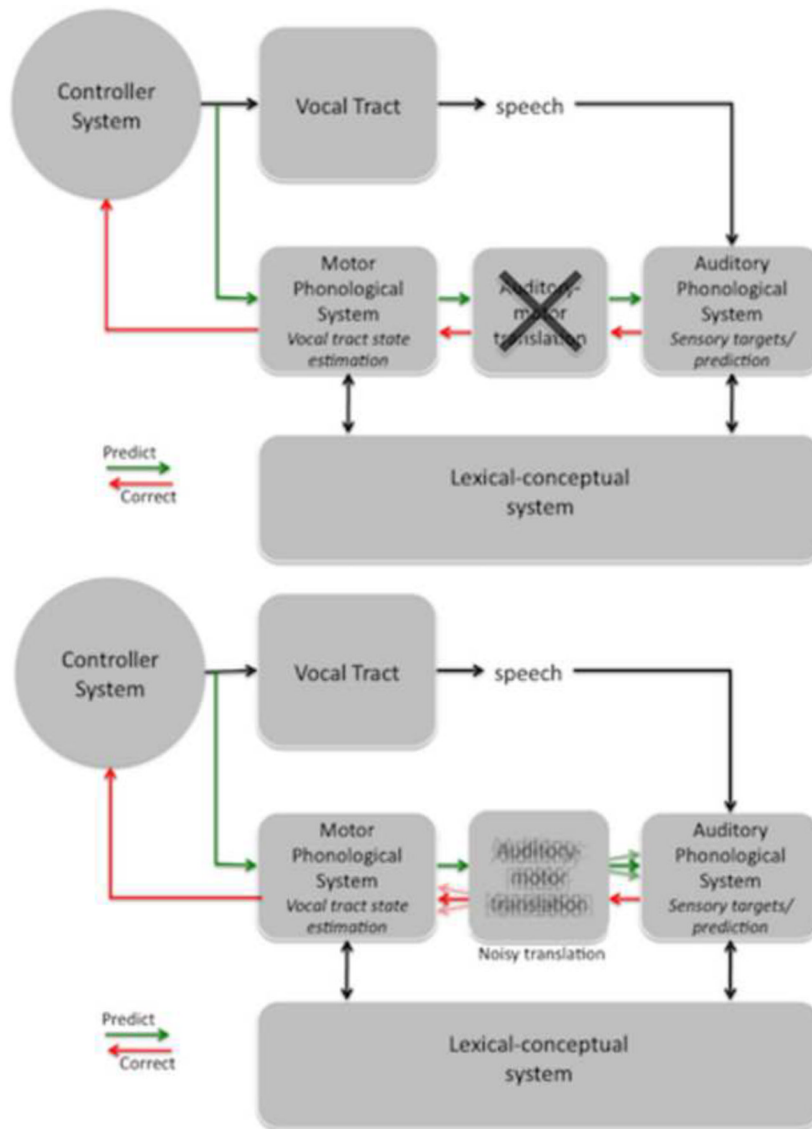
Figure 6.
Dysfunctional states of SFC system for speech. A. Proposed source of the deficit in conduction aphasia: damage to the auditory-motor translation system. Input from the lexical conceptual system to motor and auditory phonological systems are unaffected allowing for fluent output and accurate activation of sensory targets. However, internal forward sensory predictions are not possible leading to an increase in error rate. Further, errors detected as a consequence of mismatches between sensory targets and actual sensory feedback cannot be used to correct motor commands. B. Proposed source of the dysfunction in stuttering: noisy auditory-motor translation. Motor commands result in sometimes inaccurate sensory predictions due to the noisy sensorimotor mapping which trigger error correction signals that are themselves noisy, further exacerbating the problem and resulting in stuttering.