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, USA
[2]Otolaryngology – Head and Neck Surgery, University of California, San Francisco, CA 94143, USA
*Correspondence: greg.hickok@uci.edu
DOI 10.1016/j.neuron.2011.01.019

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 audiocentric 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 clin-ical 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, 1990). 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 100 ms (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, 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 nonnoisy, 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., 2010; 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 nonspeech 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., 2010). 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 prelingual 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 that 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 coarticulation 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 homolog to Broca's area, the classic human motor
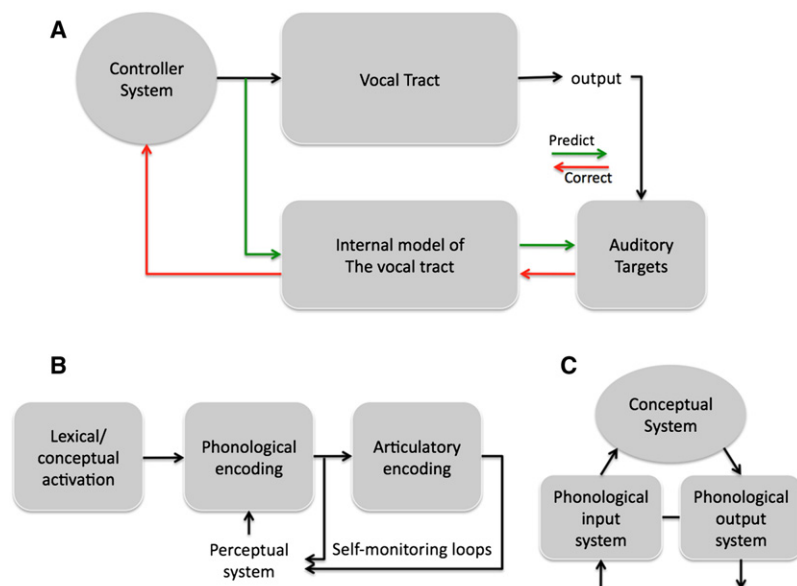
Figure 1. Models of Speech Processing
(A) State feedback control (SFC) model of speech production. The vocal tract is controlled by a motor controller, or set of controllers (Haruno et al., 2001). Motor commands issued to the vocal tract send an efferent copy to an internal model of the vocal tract. 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 versus 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).

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, 2010; 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 version of the 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; Pulvermüller 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; Pulvermüller 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 ventral premotor cortex during the perception of 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) (Sato et al., 2009). 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 (Möttönen 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 only when the speech sounds are partially ambiguous and/or when the behavioral measure is reaction time rather than accuracy. The subtlety of these effects indicates that the motor system's influence on perception is *modulatory* rather than comprising 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 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
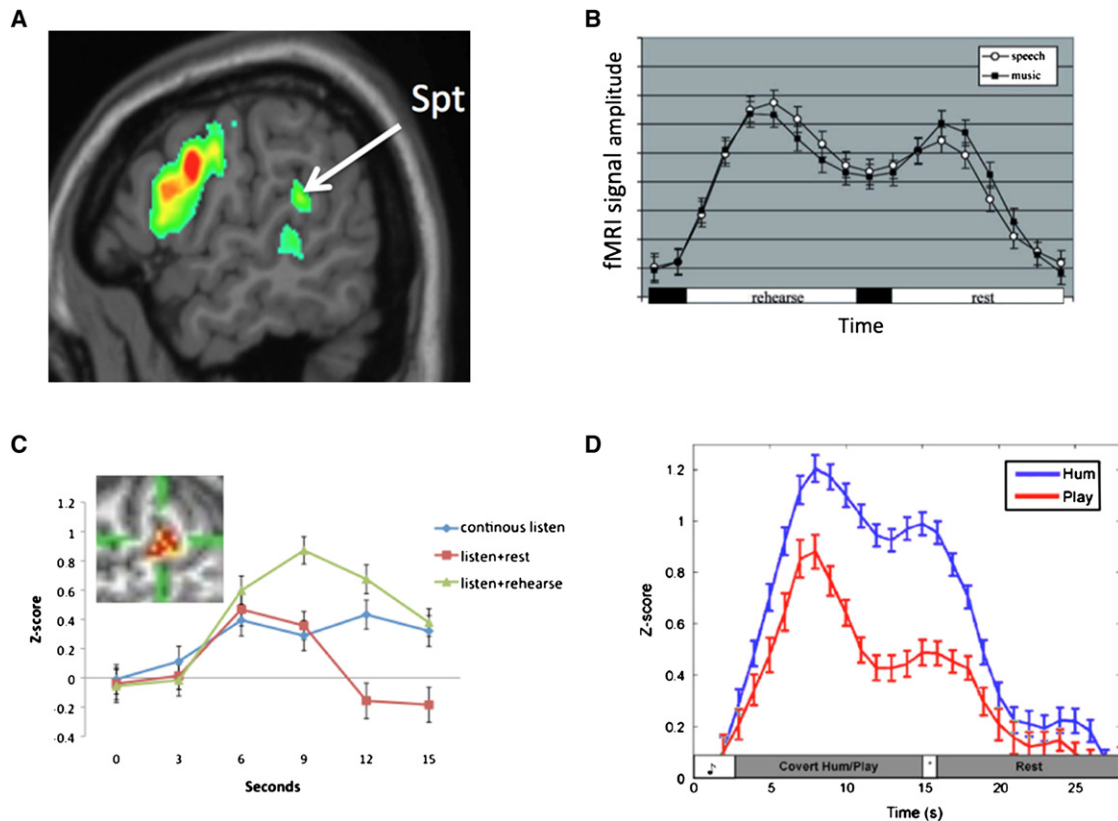
Figure 2. Location and Functional Properties of Area Spt

(A) Activation map for covert speech articulation (rehearsal of a set of nonwords), reprinted with permission from Hickok and Buchsbaum (2003).

(B) Activation time course (fMRI signal amplitude) in Spt during a sensorimotor task for speech and music. A trial is composed of 3 s of auditory stimulation (speech or a piano melody) followed by 15 s covert rehearsal/humming of the heard stimulus followed by 3 s seconds of auditory stimulation followed by 15 s of rest. The two humps represent the sensory responses, the valley between the humps is the motor (covert rehearsal/humming) 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 time course in Spt in three conditions, continuous speech (15 s, blue curve), listen+rest (3 s speech, 12 s rest, red curve), and listen+covert rehearse (3 s speech, 12 s 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., 2009).

(D) Activation time course in Spt in skilled pianists performing a sensorimotor task involving listening to novel melodies and then covertly humming them (blue curve) versus listening to novel melodies and imagine playing them on a keyboard (red curve) (from Pa and Hickok, 2008). This indicates that Spt is relatively selective for vocal tract actions. Reprinted with permission from (Hickok, 2009b).

speech. Early functional imaging studies identified an auditory-related area in the left planum temporale region that was also 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). Further, different subregional *patterns* of activity are apparent during the sensory and motor phases of the task (Hickok et al., 2009), likely reflecting the activation of different

neuronal subpopulations (Dahl et al., 2009) some sensory- and others motor-weighted. Figures 2B–2D show examples of the sensory-motor response properties of Spt and the patchy organization of this region for sensory- versus motor-weighted voxels (Figure 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, 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, 2010). 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 (A.L. Isenberg, K.L. Vaden, K. Saberi, L.T. Muftuler, G.H., unpublished data), raising the possibility that previous work implicating the supramarginal gyrus in speech production may in fact reflect Spt activity.

Area Spt, together with a network of regions including STG, premotor cortex, and the 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. (2008) asked subjects to articulate speech and either fed it back to them 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; Buchsbaum et al., 2011) 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 under discussion. The planum temporale generally has been found to activate under a variety of stimulus conditions. For example, it is sensitive not only 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 composed 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 this more posterior region of the planum temporale, which is consistent with its proposed functional role as an interface

between 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 activations) as well as different patterns of connectivity of the two activation foci as revealed by diffusion tensor imaging (A.L. Isenberg, K.L. Vaden, K. Saberi, L.T. Muftuler, G.H., unpublished data). 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, 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 not only draws 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 as part of the dual stream model of speech processing (e.g., Figure 3) (Hickok and Poeppel, 2000, 2004, 2007; for a similar view, see Rauschecker and Scott, 2009).

As briefly discussed above, Figure 1A depicts 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
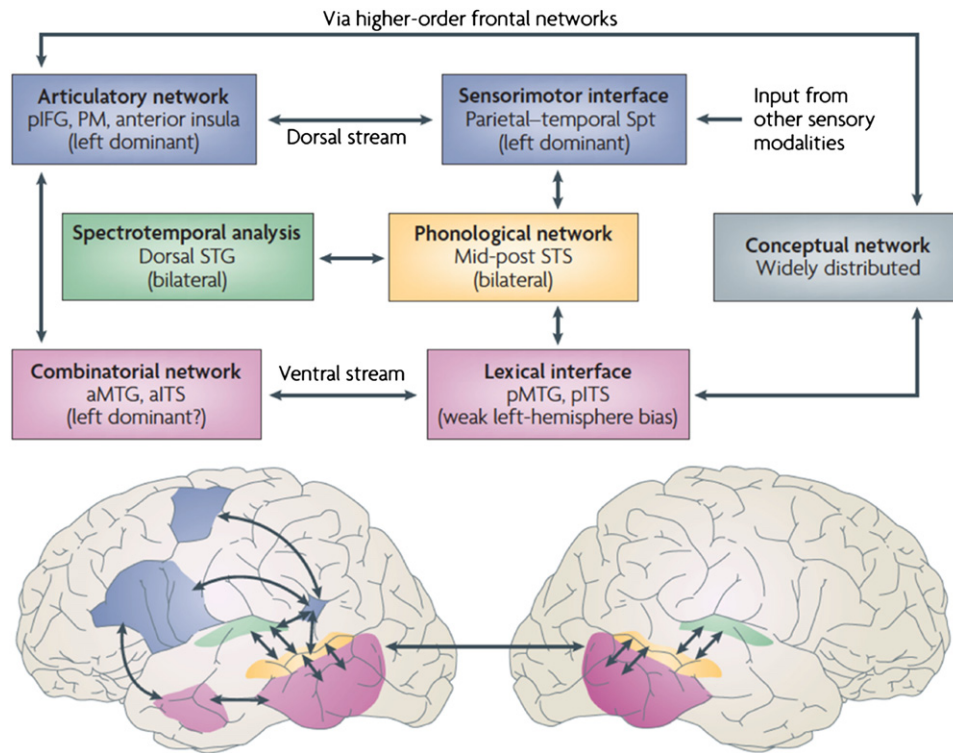
Figure 3. Dual Stream Model of Speech Processing
The dual stream model (Hickok and Poeppel, 2000, 2004, 2007) holds that early stages of speech processing occurs bilaterally in auditory regions on the dorsal STG (spectrotemporal analysis; green) and STS (phonological access/representation; yellow) and 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-semantic 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. Reprinted with permission from (Hickok and Poeppel, 2007).

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 to 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 visuomanual 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 and 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, 1969). Essentially identical views have been promoted by modern theorists (Pulvermüller, 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., 2009). 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 (1) correspond to articulatory or acoustic feature bundles (on the motor and sensory sides, respectively) that define individual speech sounds (phonemes) at the finest grain and/or (2) 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 a 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*, instan-
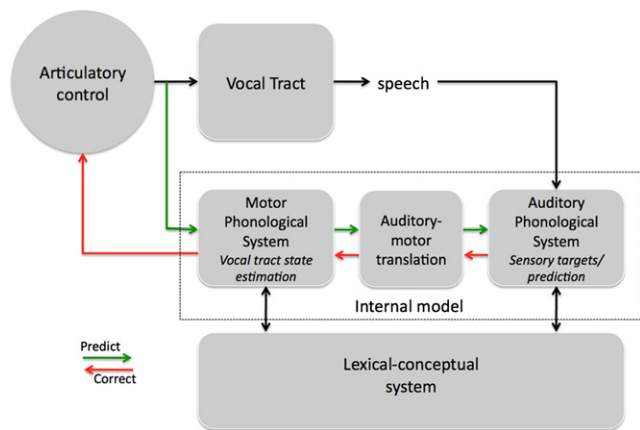

Figure 4. An Integrated State Feedback Control (SFC) Model of Speech Production
Speech models derived from the feedback control, psycholinguistic, and neurolinguistic literatures are integrated into one framework, presented here. The architecture is fundamentally that of a SFC system with a controller, 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.

tiates 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 frontocerebellar 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). His evidence for this claim was that damage to sensory speech systems (1) 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 (2) 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 suggests 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 functional localizations proposed above. Damage to frontal motor-related regions is associated with nonfluent 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. The increase in speech error rate that is observed with damage to the STG/STS is explained by disruption to the 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 disrupts the generation of a forward sensory prediction and therefore 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 (a corrective signal cannot be generated). This is precisely the pattern of deficit in conduction aphasia. In otherwise fluent speech, such patients commit frequent phonemic
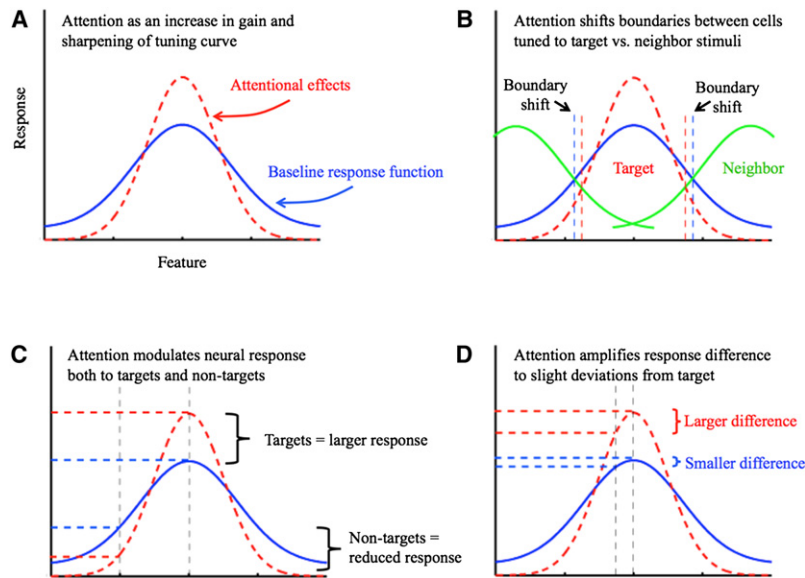
Figure 5. Top-Down Modulation of Perceptual Response Functions
(A) A graph replicated qualitatively from Figure 2 of Boynton (Boynton, 2005) illustrating attentional effects on sensory response functions based on a "feature-similarity gain model" (Martinez-Trujillo and Treue, 2004). The effects include enhancement of the responses to the attended features (red dashed line amplified at the peak) and suppression of the responses to the unattended features (red dashed line suppressed at the tails of the response distribution).
(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 features due to the sharpened response profile in the "attended channel."
(C) Attention enhances coarse discrimination: the response to targets (right vertical grey line) vs. nontargets (left vertical grey line) is larger with attentional modulation (red) than at baseline (blue).
(D) Attention enhances fine discrimination: the response to targets (right vertical gray line) vs. a slight deviation from the target (left vertical gray line) is larger with attentional modulation (red) than at baseline (blue) due to the steeper slope under attentional conditions (see text for additional discussion).

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., 2011). 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 under some circumstances 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 Martínez 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 (1) 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 (2) 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; Buchsbaum et al., 2011). 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 relatively 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
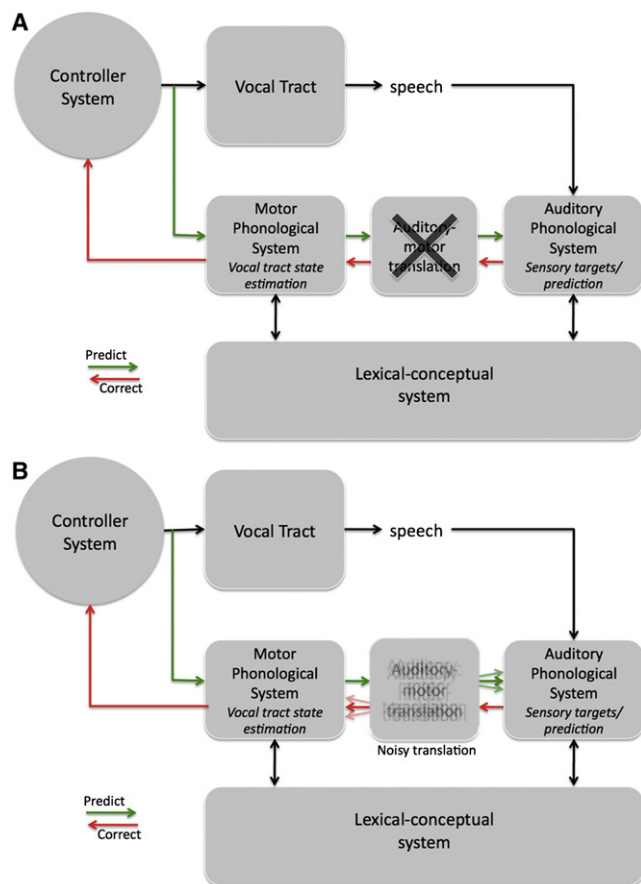
Figure 6. Dysfunctional States of SFC System for Speech

(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.

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 (1) are used to distinguish self- from externally-generated speech, and (2) 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 antipsychotics 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. Our proposed SFC model itself integrates work in psycho- and neurolinguistics with a recently outlined 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., 2010; 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, however, by showing how the model can accommodate motor effects on perception, 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.

REFERENCES

Aliu, S.O., Houde, J.F., and Nagarajan, S.S. (2009). Motor-induced suppression of the auditory cortex. J. Cogn. Neurosci. 21, 791–802.

Andersen, R.A. (1997). Multimodal integration for the representation of space in the posterior parietal cortex. Philos. Trans. R. Soc. Lond. B Biol. Sci. 352, 1421–1428.

Anderson, J.M., Gilmore, R., Roper, S., Crosson, B., Bauer, R.M., Nadeau, S., Beversdorf, D.Q., Cibula, J., Rogish, M., 3rd, Kortencamp, S., et al. (1999). Conduction aphasia and the arcuate fasciculus: A reexamination of the Wernicke-Geschwind model. Brain Lang. 70, 1–12.

Baldo, J.V., Klostermann, E.C., and Dronkers, N.F. (2008). It's either a cook or a baker: Patients with conduction aphasia get the gist but lose the trace. Brain Lang. 105, 134–140.

Benson, D.F., Sheremata, W.A., Bouchard, R., Segarra, J.M., Price, D., and Geschwind, N. (1973). Conduction aphasia. A clinicopathological study. Arch. Neurol. 28, 339–346.

Bishop, D.V., Brown, B.B., and Robson, J. (1990). The relationship between phoneme discrimination, speech production, and language comprehension in cerebral-palsied individuals. J. Speech Hear. Res. 33, 210–219.

Boller, F., and Marcie, P. (1978). Possible role of abnormal auditory feedback in conduction aphasia. Neuropsychologia 16, 521–524.

Boller, F., Vrtunski, P.B., Kim, Y., and Mack, J.L. (1978). Delayed auditory feedback and aphasia. Cortex 14, 212–226.

Boynton, G.M. (2005). Attention and visual perception. Curr. Opin. Neurobiol. 15, 465–469.

Buchsbaum, B., Hickok, G., and Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. Cogn. Sci. 25, 663–678.

Buchsbaum, B.R., Olsen, R.K., Koch, P., and Berman, K.F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. Neuron 48, 687–697.

Buchsbaum, B.R., Baldo, J., D'Esposito, M., Dronkers, N., Okada, K., and Hickok, G. (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory: An aggregate analysis of lesion and fMRI data. Brain Lang., in press. Published online January 21, 2011. 10.1016/j.bandl.2010.12.001.

Burnett, T.A., Freedland, M.B., Larson, C.R., and Hain, T.C. (1998). Voice F0 responses to manipulations in pitch feedback. J. Acoust. Soc. Am. 103, 3153–3161.

Caramazza, A. (1988). Some aspects of language processing revealed through the analysis of acquired aphasia: The lexical system. Annu. Rev. Neurosci. *11*, 395–421.

Christen, H.J., Hanefeld, F., Kruse, E., Imhäuser, S., Ernst, J.P., and Finkenstaedt, M. (2000). Foix-Chavany-Marie (anterior operculum) syndrome in childhood: A reappraisal of Worster-Drought syndrome. Dev. Med. Child Neurol. *42*, 122–132.

Colby, C.L., and Goldberg, M.E. (1999). Space and attention in parietal cortex. Annu. Rev. Neurosci. *22*, 319–349.

Cooper, W.E., and Lauritsen, M.R. (1974). Feature processing in the perception and production of speech. Nature *252*, 121–123.

Corina, D.P., and Knapp, H. (2006). Sign language processing and the mirror neuron system. Cortex *42*, 529–539.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., and Fadiga, L. (2009). The motor somatotopy of speech perception. Curr. Biol. *19*, 381–385.

Dahl, C.D., Logothetis, N.K., and Kayser, C. (2009). Spatial organization of multisensory responses in temporal association cortex. J. Neurosci. *29*, 11924–11932.

Damasio, A.R. (1992). Aphasia. N. Engl. J. Med. *326*, 531–539.

Damasio, H., and Damasio, A.R. (1980). The anatomical basis of conduction aphasia. Brain *103*, 337–350.

de Zubicaray, G.I., and McMahon, K.L. (2009). Auditory context effects in picture naming investigated with event-related fMRI. Cogn. Affect. Behav. Neurosci. *9*, 260–269.

Decety, J., and Michel, F. (1989). Comparative analysis of actual and mental movement times in two graphic tasks. Brain Cogn. *11*, 87–97.

Dell, G.S., Schwartz, M.F., Martin, N., Saffran, E.M., and Gagnon, D.A. (1997). Lexical access in aphasic and nonaphasic speakers. Psychol. Rev. *104*, 801–838.

Delvaux, V., and Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. Phonetica *64*, 145–173.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. Exp. Brain Res. *91*, 176–180.

Dronkers, N., and Baldo, J. (2009). Language: Aphasia. In Encyclopedia of Neuroscience, L.R. Squire, ed. (Oxford: Academic Press), pp. 343–348.

Durisko, C., and Fiez, J.A. (2010). Functional activation in the cerebellum during working memory and simple speech tasks. Cortex *46*, 896–906.

Edwards, E., Nagarajan, S.S., Dalal, S.S., Canolty, R.T., Kirsch, H.E., Barbaro, N.M., and Knight, R.T. (2010). Spatiotemporal imaging of cortical activation during verb generation and picture naming. Neuroimage *50*, 291–301.

Emmorey, K., Xu, J., Gannon, P., Goldin-Meadow, S., and Braun, A. (2010). CNS activation and regional connectivity during pantomime observation: No engagement of the mirror neuron system for deaf signers. Neuroimage *49*, 994–1005.

Fadiga, L., and Craighero, L. (2003). New insights on sensorimotor integration: From hand action to speech perception. Brain Cogn. *53*, 514–524.

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

Fadiga, L., Craighero, L., and D'Ausilio, A. (2009). Broca's area in language, action, and music. Ann. N Y Acad. Sci. *1169*, 448–458.

Fairbanks, G. (1954). Systematic research in experimental phonetics. I. A theory of the speech mechanism as a servosystem. J. Speech Hear. Disord. *19*, 133–139.

Foundas, A.L., Bollich, A.M., Feldman, J., Corey, D.M., Hurley, M., Lemen, L.C., and Heilman, K.M. (2004). Aberrant auditory processing and atypical planum temporale in developmental stuttering. Neurology *63*, 1640–1646.

Franklin, G.F., Powell, J.D., and Emami-Naeini, A. (1991). Feedback control of dynamic systems (Reading, Mass: Addison-Wesley).

Fridriksson, J., Kjartansson, O., Morgan, P.S., Hjaltason, H., Magnusdottir, S., Bonilha, L., and Rorden, C. (2010). Impaired speech repetition and left parietal lobe damage. J. Neurosci. *30*, 11057–11061.

Friederici, A.D. (2009). Pathways to language: Fiber tracts in the human brain. Trends Cogn. Sci. (Regul. Ed.) *13*, 175–181.

Frith, C.D., Blakemore, S., and Wolpert, D.M. (2000). Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. Brain Res. Brain Res. Rev. *31*, 357–363.

Fuster, J.M. (1995). Memory in the cerebral cortex (Cambridge, MA: MIT Press).

Galaburda, A., and Sanides, F. (1980). Cytoarchitectonic organization of the human auditory cortex. J. Comp. Neurol. *190*, 597–610.

Galantucci, B., Fowler, C.A., and Turvey, M.T. (2006). The motor theory of speech perception reviewed. Psychon. Bull. Rev. *13*, 361–377.

Gallese, V., and Lakoff, G. (2005). The Brain's concepts: The role of the Sensory-motor system in conceptual knowledge. Cogn. Neuropsychol. *22*, 455–479.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. Brain *119*, 593–609.

Geschwind, N. (1965). Disconnexion syndromes in animals and man. I. Brain *88*, 237–294, 585–644.

Geschwind, N. (1971). Current concepts: Aphasia. N. Engl. J. Med. *284*, 654–656.

Goldberg, T.E., Gold, J.M., Coppola, R., and Weinberger, D.R. (1997). Unnatural practices, unspeakable actions: A study of delayed auditory feedback in schizophrenia. Am. J. Psychiatry *154*, 858–860.

Golfinopoulos, E., Tourville, J.A., and Guenther, F.H. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. Neuroimage *52*, 862–874.

Goodglass, H. (1992). Diagnosis of conduction aphasia. In Conduction aphasia, S.E. Kohn, ed. (Hillsdale, N.J.: Lawrence Erlbaum Associates), pp. 39–49.

Goodglass, H., Kaplan, E., and Barresi, B. (2001). The assessment of aphasia and related disorders, Third Edition (Philadelphia: Lippincott Williams & Wilkins).

Grafton, S.T., and Hamilton, A.F. (2007). Evidence for a distributed hierarchy of action representation in the brain. Hum. Mov. Sci. *26*, 590–616.

Grafton, S.T., Aziz-Zadeh, L., and Ivry, R.B. (2009). Relative hierarchies and the representation of action. In The cognitive neurosciences, M.S. Gazzaniga, ed. (Cambridge, MA: MIT Press), pp. 641–652.

Graves, W.W., Grabowski, T.J., Mehta, S., and Gordon, J.K. (2007). A neural signature of phonological access: Distinguishing the effects of word frequency from familiarity and length in overt picture naming. J. Cogn. Neurosci. *19*, 617–631.

Graves, W.W., Grabowski, T.J., Mehta, S., and Gupta, P. (2008). The left posterior superior temporal gyrus participates specifically in accessing lexical phonology. J. Cogn. Neurosci. *20*, 1698–1710.

Griffiths, T.D., and Warren, J.D. (2002). The planum temporale as a computational hub. Trends Neurosci. *25*, 348–353.

Guenther, F.H., Hampson, M., and Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. Psychol. Rev. *105*, 611–633.

Guenther, F.H., Ghosh, S.S., and Tourville, J.A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang. *96*, 280–301.

Haruno, M., Wolpert, D.M., and Kawato, M. (2001). Mosaic model for sensorimotor learning and control. Neural Comput. *13*, 2201–2220.

Hauser, M., and Wood, J. (2010). Evolving the capacity to understand actions, intentions, and goals. Annu. Rev. Psychol. 61, 303–324.

Heinks-Maldonado, T.H., Mathalon, D.H., Houde, J.F., Gray, M., Faustman, W.O., and Ford, J.M. (2007). Relationship of imprecise corollary discharge in schizophrenia to auditory hallucinations. Arch. Gen. Psychiatry 64, 286–296.

Heyes, C. (2010). Where do mirror neurons come from? Neurosci. Biobehav. Rev. 34, 575–583.

Hickok, G. (2000). Speech perception, conduction aphasia, and the functional neuroanatomy of language. In Language and the brain, Y. Grodzinsky, L. Shapiro, and D. Swinney, eds. (San Diego: Academic Press), pp. 87–104.

Hickok, G. (2009a). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. J. Cogn. Neurosci. 21, 1229–1243.

Hickok, G. (2009b). The functional neuroanatomy of language. Phys. Life Rev. 6, 121–143.

Hickok, G. (2010a). The role of mirror neurons in speech and language processing. Brain Lang. 112, 1–2.

Hickok, G. (2010b). The role of mirror neurons in speech perception and action word semantics. Lang. Cogn. Process. 25, 749–776.

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

Hickok, G., and Hauser, M. (2010). (Mis)understanding mirror neurons. Curr. Biol. 20, R593–R594.

Hickok, G., and Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. Trends Cogn. Sci. (Regul. Ed.) 4, 131–138.

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

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. Nat. Rev. Neurosci. 8, 393–402.

Hickok, G., Erhard, P., Kassubek, J., Helms-Tillery, A.K., Naeve-Velguth, S., Strupp, J.P., Strick, P.L., and Ugurbil, K. (2000). 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. Neurosci. Lett. 287, 156–160.

Hickok, G., Buchsbaum, B., Humphries, C., and Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. J. Cogn. Neurosci. 15, 673–682.

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

Hickok, G., Okada, K., and Serences, J.T. (2009). Area Spt in the human planum temporale supports sensory-motor integration for speech processing. J. Neurophysiol. 101, 2725–2732.

Hillis, A.E. (2007). Aphasia: Progress in the last quarter of a century. Neurology 69, 200–213.

Holt, L.L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. Psychol. Sci. 16, 305–312.

Holt, L.L., and Lotto, A.J. (2008). Speech perception within an auditory cognitive science framework. Curr. Dir. Psychol. Sci. 17, 42–46.

Indefrey, P., and Levelt, W.J. (2004). The spatial and temporal signatures of word production components. Cognition 92, 101–144.

Jacobs, O.L.R. (1993). Introduction to controll theory (Oxford: Oxford University Press).

Jacquemot, C., Dupoux, E., and Bachoud-Lévi, A.C. (2007). Breaking the mirror: Asymmetrical disconnection between the phonological input and output codes. Cogn. Neuropsychol. 24, 3–22.

Jazayeri, M., and Movshon, J.A. (2006). Optimal representation of sensory information by neural populations. Nat. Neurosci. 9, 690–696.

Jazayeri, M., and Movshon, J.A. (2007). A new perceptual illusion reveals mechanisms of sensory decoding. Nature 446, 912–915.

Kappes, J., Baumgaertner, A., Peschke, C., and Ziegler, W. (2009). Unintended imitation in nonword repetition. Brain Lang. 111, 140–151.

Knapp, H.P., and Corina, D.P. (2010). A human mirror neuron system for language: Perspectives from signed languages of the deaf. Brain Lang. 112, 36–43.

Krigolson, O.E., and Holroyd, C.B. (2007). Hierarchical error processing: Different errors, different systems. Brain Res. 1155, 70–80.

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

Lenneberg, E.H. (1962). Understanding language without ability to speak: A case report. J. Abnorm. Soc. Psychol. 65, 419–425.

Levelt, W.J. (1983). Monitoring and self-repair in speech. Cognition 14, 41–104.

Levelt, W.J.M. (1989). Speaking: From Intention to Articulation (Cambridge, MA: MIT Press).

Levelt, W.J.M., Praamstra, P., Meyer, A.S., Helenius, P., and Salmelin, R. (1998). An MEG study of picture naming. J. Cogn. Neurosci. 10, 553–567.

Levelt, W.J.M., Roelofs, A., and Meyer, A.S. (1999). A theory of lexical access in speech production. Behav. Brain Sci. 22, 1–38, discussion 38–75.

Liberman, A.M. (1957). Some results of research on speech perception. J. Acoust. Soc. Am. 29, 117–123.

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

Liberman, A.M., and Mattingly, I.G. (1989). A specialization for speech perception. Science 243, 489–494.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., and Studdert-Kennedy, M. (1967). Perception of the speech code. Psychol. Rev. 74, 431–461.

Lotto, A.J., Kluender, K.R., and Holt, L.L. (1997). Perceptual compensation for coarticulation by Japanese quail (Coturnix coturnix japonica). J. Acoust. Soc. Am. 102, 1134–1140.

Lotto, A.J., Hickok, G.S., and Holt, L.L. (2009). Reflections on mirror neurons and speech perception. Trends Cogn. Sci. (Regul. Ed.) 13, 110–114.

MacNeilage, P.F., Rootes, T.P., and Chase, R.A. (1967). Speech production and perception in a patient with severe impairment of somesthetic perception and motor control. J. Speech Hear. Res. 10, 449–467.

Maguire, G.A., Yu, B.P., Franklin, D.L., and Riley, G.D. (2004). Alleviating stuttering with pharmacological interventions. Expert Opin. Pharmacother. 5, 1565–1571.

Mahon, B.Z., and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. J. Physiol. Paris 102, 59–70.

Martin, R., and Haroldson, S.K. (1979). Effects of five experimental treatments of stuttering. J. Speech Hear. Res. 22, 132–146.

Martinez-Trujillo, J.C., and Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. Curr. Biol. 14, 744–751.

Massaro, D.W. (1972). Preperceptual images, processing time, and perceptual units in auditory perception. Psychol. Rev. 79, 124–145.

Massaro, D.W., and Chen, T.H. (2008). The motor theory of speech perception revisited. Psychon. Bull. Rev. 15, 453–457, discussion 458–462.

Max, L., Guenther, F.H., Gracco, V.L., Ghosh, S.S., and Wallace, M.E. (2004). Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: A theoretical model of stuttering. Contemporary Issue in Communication Science and Disorders 31, 105–122.

McAdams, C.J., and Maunsell, J.H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. J. Neurosci. *19*, 431–441.

McCarthy, R., and Warrington, E.K. (1984). A two-route model of speech production. Evidence from aphasia. Brain *107*, 463–485.

Meister, I.G., Wilson, S.M., Deblieck, C., Wu, A.D., and Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. Curr. Biol. *17*, 1692–1696.

Moran, J., and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. Science *229*, 782–784.

Möttönen, R., and Watkins, K.E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. J. Neurosci. *29*, 9819–9825.

Mulliken, G.H., and Andersen, R.A. (2009). Forward models and state estimation in posterior parietal cortex. In The cognitive neurosciences, M.S. Gazzaniga, ed. (Cambridge, MA: MIT Press), pp. 599–611.

Murray, S.O., and Wojciulik, E. (2004). Attention increases neural selectivity in the human lateral occipital complex. Nat. Neurosci. *7*, 70–74.

Naeser, M.A., Palumbo, C.L., Helm-Estabrooks, N., Stiassny-Eder, D., and Albert, M.L. (1989). Severe nonfluency in aphasia. Role of the medial subcallosal fasciculus and other white matter pathways in recovery of spontaneous speech. Brain *112*, 1–38.

Navalpakkam, V., and Itti, L. (2007). Search goal tunes visual features optimally. Neuron *53*, 605–617.

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

Oppenheim, G.M., and Dell, G.S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. Cognition *106*, 528–537.

Pa, J., and Hickok, G. (2008). A parietal-temporal sensory-motor integration area for the human vocal tract: Evidence from an fMRI study of skilled musicians. Neuropsychologia *46*, 362–368.

Paus, T., Perry, D.W., Zatorre, R.J., Worsley, K.J., and Evans, A.C. (1996). Modulation of cerebral blood flow in the human auditory cortex during speech: Role of motor-to-sensory discharges. Eur. J. Neurosci. *8*, 2236–2246.

Perenin, M.-T., and Vighetto, A. (1988). Optic ataxia: A specific disruption in visuomotor mechanisms. I. Different aspects of the deficit in reaching for objects. Brain *111*, 643–674.

Price, C.J. (2010). The anatomy of language: A review of 100 fMRI studies published in 2009. Ann. N. Y. Acad. Sci. *1191*, 62–88.

Pulvermüller, F. (1996). Hebb's concept of cell assemblies and the psychophysiology of word processing. Psychophysiology *33*, 317–333.

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., and Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. Proc. Natl. Acad. Sci. USA *103*, 7865–7870.

Purcell, D.W., and Munhall, K.G. (2006). Compensation following real-time manipulation of formants in isolated vowels. J. Acoust. Soc. Am. *119*, 2288–2297.

Rauschecker, J.P., and Scott, S.K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. Nat. Neurosci. *12*, 718–724.

Regan, D., and Beverley, K.I. (1985). Postadaptation orientation discrimination. J. Opt. Soc. Am. A *2*, 147–155.

Reynolds, J.H., and Heeger, D.J. (2009). The normalization model of attention. Neuron *61*, 168–185.

Reynolds, J.H., Chelazzi, L., and Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. J. Neurosci. *19*, 1736–1753.

Rizzolatti, G., and Arbib, M.A. (1998). Language within our grasp. Trends Neurosci. *21*, 188–194.

Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. Annu. Rev. Neurosci. *27*, 169–192.

Rogalsky, C., and Hickok, G. (2010). The role of Broca's area in sentence comprehension. J. Cogn. Neurosci., in press. Published online July 9, 2010. 10.1162/jocn.2010.21530.

Rossetti, Y., Pisella, L., and Vighetto, A. (2003). Optic ataxia revisited: Visually guided action versus immediate visuomotor control. Exp. Brain Res. *153*, 171–179.

Sams, M., Möttönen, R., and Sihvonen, T. (2005). Seeing and hearing others and oneself talk. Brain Res. Cogn. Brain Res. *23*, 429–435.

Sancier, M.L., and Fowler, C.A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. Journal of Phonetics *25*, 421–436.

Sato, M., Tremblay, P., and Gracco, V.L. (2009). A mediating role of the premotor cortex in phoneme segmentation. Brain Lang *111*, 1–7.

Schwartz, J.-L., Basirat, A., Menard, L., and Sato, M. (2010). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. J. Neurolinguist., in press. Published online January 20, 2010. 10.1016/j.jneuroling.2009.12.004.

Scolari, M., and Serences, J.T. (2009). Adaptive allocation of attentional gain. J. Neurosci. *29*, 11933–11942.

Seung, H.S., and Sompolinsky, H. (1993). Simple models for reading neuronal population codes. Proc. Natl. Acad. Sci. USA *90*, 10749–10753.

Shadmehr, R., and Krakauer, J.W. (2008). A computational neuroanatomy for motor control. Exp. Brain Res. *185*, 359–381.

Shelton, J.R., and Caramazza, A. (1999). Deficits in lexical and semantic processing: Implications for models of normal language. Psychon. Bull. Rev. *6*, 5–27.

Shiller, D.M., Sato, M., Gracco, V.L., and Baum, S.R. (2009). Perceptual recalibration of speech sounds following speech motor learning. J. Acoust. Soc. Am. *125*, 1103–1113.

Skipper, J.I., Nusbaum, H.C., and Small, S.L. (2005). Listening to talking faces: Motor cortical activation during speech perception. Neuroimage *25*, 76–89.

Skoyles, J.R. (2010). Mapping of heard speech into articulation information and speech acquisition. Proc. Natl. Acad. Sci. USA *107*, E73–, author reply E74.

Smiley, J.F., Rosoklija, G., Mancevski, B., Mann, J.J., Dwork, A.J., and Javitt, D.C. (2009). Altered volume and hemispheric asymmetry of the superficial cortical layers in the schizophrenia planum temporale. Eur. J. Neurosci. *30*, 449–463.

Spitzer, H., Desimone, R., and Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. Science *240*, 338–340.

Stuart, A., Kalinowski, J., Rastatter, M.P., and Lynch, K. (2002). Effect of delayed auditory feedback on normal speakers at two speech rates. J. Acoust. Soc. Am. *111*, 2237–2241.

Stuart, A., Frazier, C.L., Kalinowski, J., and Vos, P.W. (2008). The effect of frequency altered feedback on stuttering duration and type. J. Speech Lang. Hear. Res. *51*, 889–897.

Sundara, M., Namasivayam, A.K., and Chen, R. (2001). Observation-execution matching system for speech: A magnetic stimulation study. Neuroreport *12*, 1341–1344.

Todorov, E. (2004). Optimality principles in sensorimotor control. Nat. Neurosci. *7*, 907–915.

Tourville, J.A., Reilly, K.J., and Guenther, F.H. (2008). Neural mechanisms underlying auditory feedback control of speech. Neuroimage *39*, 1429–1443.

Treue, S., and Martínez Trujillo, J.C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. Nature *399*, 575–579.

Treue, S., and Maunsell, J.H. (1999). Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. J. Neurosci. *19*, 7591–7602.

van Wassenhove, V., Grant, K.W., and Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. Proc. Natl. Acad. Sci. USA *102*, 1181–1186.

Ventura, M.I., Nagarajan, S.S., and Houde, J.F. (2009). Speech target modulates speaking induced suppression in auditory cortex. BMC Neurosci. *10*, 58.

Waldstein, R.S. (1990). Effects of postlingual deafness on speech production: Implications for the role of auditory feedback. J. Acoust. Soc. Am. *88*, 2099–2114.

Watkins, K., and Paus, T. (2004). Modulation of motor excitability during speech perception: The role of Broca's area. J. Cogn. Neurosci. *16*, 978–987.

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

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

Wernicke, C. (1874). The symptom complex of aphasia: A psychological study on an anatomical basis. Reprinted in Boston Studies in the Philosophy of Science (1969), R.S. Cohen and M.W. Wartofsky, eds. (Dordrecht: D. Reidel Publishing Company), pp. 34–97.

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

Wilson, S.M., Saygin, A.P., Sereno, M.I., and Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. Nat. Neurosci. *7*, 701–702.

Wilson, S.M., Isenberg, A.L., and Hickok, G. (2009). Neural correlates of word production stages delineated by parametric modulation of psycholinguistic variables. Hum. Brain Mapp. *30*, 3596–3608.

Wise, R.J.S., Scott, S.K., Blank, S.C., Mummery, C.J., Murphy, K., and Warburton, E.A. (2001). Separate neural subsystems within 'Wernicke's area'. Brain *124*, 83–95.

Wolpert, D.M., Ghahramani, Z., and Jordan, M.I. (1995). An internal model for sensorimotor integration. Science *269*, 1880–1882.

Yates, A.J. (1963). Delayed auditory feedback. Psychol. Bull. *60*, 213–232.