

Review Article

Modeling the Role of Sensory Feedback in Speech Motor Control and Learning

Benjamin Parrell^a and John Houde^b

Purpose: While the speech motor system is sensitive to feedback perturbations, sensory feedback does not seem to be critical to speech motor production. How the speech motor system is able to be so flexible in its use of sensory feedback remains an open question.

Method: We draw on evidence from a variety of disciplines to summarize current understanding of the sensory systems' role in speech motor control, including both online control and motor learning. We focus particularly on computational models of speech motor control that incorporate sensory feedback, as these models provide clear encapsulations of different theories of sensory systems' function in speech production. These computational models include the well-established directions into velocities of articulators model and computational models that we have been developing in our labs based on the domain-general theory of state feedback control (feedback aware control of tasks in speech model).

Results: After establishing the architecture of the models, we show that both the directions into velocities of articulators

and state feedback control/feedback aware control of tasks models can replicate key behaviors related to sensory feedback in the speech motor system. Although the models agree on many points, the underlying architecture of the 2 models differs in a few key ways, leading to different predictions in certain areas. We cover key disagreements between the models to show the limits of our current understanding and point toward areas where future experimental studies can resolve these questions.

Conclusions: Understanding the role of sensory information in the speech motor system is critical to understanding speech motor production and sensorimotor learning in healthy speakers as well as in disordered populations. Computational models, with their concrete implementations and testable predictions, are an important tool to understand this process. Comparison of different models can highlight areas of agreement and disagreement in the field and point toward future experiments to resolve important outstanding questions about the speech motor control system.

Speech production is one of the most complex human motor behaviors, with roughly 100 muscles coordinated precisely in space and time to produce rapid speech movements at a high rate. A major question in speech science is how such a complex system can be controlled to produce fluent speech. One approach has been to develop computational models of speech motor control. Computational models provide an important tool to understand the speech motor system: These models provide a concrete hypothesis of how the system (or a subpart of that system) may function, generating output that can be compared

against real speech data to test these hypotheses. Computational models can also provide the impetus for new experiments with human speakers to test the predictions of that model or to distinguish between competing computational approaches.

For these reasons, developing a model of how the complex speech articulatory system is controlled to produce fluent speech has been a goal of speech scientists for many decades. Many models have been developed over the years, focusing on a number of different aspects of speech motor control: coordination of multiple articulators in time and space (Birkholz, Kroger, & Neuschaefer-Rube, 2011; Guenther, 1995a; Saltzman, 1986; Saltzman & Munhall, 1989; Sanguineti, Laboissière, & Ostry, 1998), the role of muscle dynamics (Perrier, 2005; Perrier, Løevenbruck, & Payan, 1996), the role of somatosensory and auditory feedback (Guenther, 2016; J. F. Houde & Nagarajan, 2011; Parrell, Ramanarayanan, Nagarajan, & Houde, 2018; Ramanarayanan, Parrell, Goldstein, Nagarajan, & Houde, 2016; Tourville & Guenther, 2011), and the neural substrate of speech production (Guenther, 2016; J. F. Houde &

^aDepartment of Communication Sciences and Disorders, University of Wisconsin–Madison

^bDepartment of Otolaryngology—Head and Neck Surgery, University of California, San Francisco

Correspondence to Benjamin Parrell: bparrell@wisc.edu

Editor-in-Chief: Ben Maassen

Editor: Hayo Terband

Received April 5, 2018

Revision received September 8, 2018

Accepted February 22, 2019

https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0127

Publisher Note: This article is part of the Special Issue: Select Papers From the 7th International Conference on Speech Motor Control.

Disclosure: The authors have declared that no competing interests existed at the time of publication.

Chang, 2015; J. F. Houde & Nagarajan, 2011; Tourville & Guenther, 2011).

In the current article, we focus on the role of the sensory system in speaking. We examine basic computational architectures underlying different models of speech motor control, experimental evidence for the role of the sensory system in speech production at various time scales, and previous attempts to model speech as a sensorimotor system. We track recent progress in our labs on developing a model of speech production as a state feedback control (SFC) system (Diedrichsen, Shadmehr, & Ivry, 2010; J. F. Houde & Nagarajan, 2011; Parrell et al., 2018; Parrell, Ramanarayanan, et al., 2019; Ramanarayanan et al., 2016; Shadmehr & Krakauer, 2008; Todorov, 2004; Todorov & Jordan, 2002). In an SFC system, an estimate of the current state of the system (e.g., positions and velocities of the lips, tongue, jaw, etc.) is generated from a combination of sensory information and internal predictions. This state estimate is then used to generate state-dependent motor commands to drive changes in the system. We compare where the predictions of our state feedback approach differ from the directions into velocities of articulators (DIVA) model (Guenther, 2016), a prominent model of speech motor control, with the aim of pointing toward open questions about sensory function in the speech motor system and some avenues for further research to elucidate the control system underlying the complex act of speaking.

Control Mechanisms and Sensory Feedback

Computational approaches to both speech and non-speech motor control typically rely on basic concepts developed first in engineering control theory (Jacobs, 1993; Stengel, 1994; Parrell, Lammert, Ciccarelli, & Quatieri, 2019). The two basic control systems, which form the foundation for the more complex control architectures typically considered in current approaches to human motor control, are *feedforward* and *feedback* control. Although no current theories rely only on one of these components, it is useful to first describe them separately before describing how they are currently employed.

In a feedforward control system, motor commands are instantiated as preplanned trajectories that are executed by the articulatory system in a time-locked manner. In the most basic forms of feedforward control, these time-varying motor commands are executed without any reference to the outcomes of these actions. That is to say, sensory feedback from self-produced actions plays no role in this type of control. In feedback control, motor commands are calculated and issued online during movement rather than being preplanned. In these systems, sensory feedback is compared to the desired end state at each time point. This error, or difference between feedback and goal, is used to generate a motor command to change the state of the system. Feedback control systems in speech have used both somatosensory and auditory feedbacks (Fairbanks, 1954; Guenther, 2016; Saltzman, 1986; Saltzman & Munhall, 1989).

Neither pure feedforward or sensory feedback control systems are likely to occur in human speech. Pure feedforward control is precluded by the ample evidence (reviewed below) that sensory feedback plays a role in the online control of speech. A pure sensory feedback system is equally unlikely, as it supposes accurate, real-time sensory information from the somatosensory (e.g., the positions and velocities of the speech articulators) or auditory (e.g., vowel formants) systems. While this type of information is likely available to the central nervous system (CNS) through afferent information originating in muscle spindles and the auditory system, real neural systems provide information that is both noisy and delayed in time. Thus, any architecture that relies critically on accurate, real-time sensory information to generate motor commands will produce erroneous movements as the state error used to generate motor commands will be calculated using inaccurate and outdated information.

A third type of controller—*predictive control*—uses the basic architecture of a feedback controller, substituting sensory afferent signals with an internal prediction or estimate of the feedback (Wolpert & Miall, 1996). This eliminates the delays inherent to sensory feedback process in the nervous system, which would otherwise preclude feedback control of fast, ballistic movements, such as those produced during speech (Hollerbach, 1982), allowing for fast comparison of the estimated feedback with the action goal, as in a stereotypical feedback controller. As long as the internal estimate is accurate, such a short-latency feedback control system based on this internal estimate is functionally equivalent to a feedforward control system (Shadmehr & Krakauer, 2008), though still without pre-specification of a time-dependent motor plan.

With these approaches to control in mind, in the sections below, we review the role of the sensory system in speech motor control. This includes how sensory information is used for online control (as in a feedback control system) as well as how it is used for learning and updating of the speech motor control system (such as would be used in a feedforward or predictive control scheme).

Sensory Feedback Use in Online Control

Both auditory and somatosensory feedback have been implicated in the control of speech. There is some evidence that auditory feedback plays a role in the online control of speech in addition to its role in long-term calibration. One area of evidence comes from studies that use loud noise to mask participants' ability to hear their own speech. These studies typically find differences between masked and clear speech in a wide variety of parameters: speech volume, pitch, duration, voice quality, and formant frequencies (Ladefoged, 1967; Lane & Tranel, 1971; Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). Interestingly, speech in noise has been found to be more intelligible than clear speech (Summers et al., 1988), potentially suggesting that removing online auditory feedback can enhance the accuracy of the speech motor system. However, these results are complicated by the fact that the use of loud

masking noise introduces a new stimulus to the speech production system, and speaking in the presence of masking noise may cause a qualitative change in the role of the auditory system in speech (Lane & Tranel, 1971). An alternative account of the role of auditory feedback in online control comes from a study with cochlear implant users with postlingual deafness (Svirsky, Lane, Perkell, & Wozniak, 1992). This study, though limited to three individuals, found that turning off the cochlear implant (and thus eliminating auditory feedback) caused changes in speech volume, pitch, voice quality, and vowel formants, though the pattern of changes varied across individuals. These results suggest some role for auditory feedback in the online control of speech.

As is true for auditory feedback, somatosensory feedback has also been shown to play an important role in speech production. A very limited number of studies have examined the speech of patients with loss of somatosensation in the oral cavity, generally finding substantial articulatory deficits (MacNeilage, Rootes, & Chase, 1967). However, the congenital nature of the sensory deficit in these patients makes it difficult to separate problems in online versus predictive/feedforward control. The single study that has examined patients with acquired sensory disorders suggests speech may be essentially normal (after a recovery period), with large deficits only present when the system is perturbed, as with a bite block (Hoole, 1987). It should be noted that this patient showed essentially absent tactile sensation, but proprioception was not directly assessed.

Other studies have addressed this question by attempting to mask oral sensation in healthy speakers. However, while topical anesthetics may mask tactile sensation in the oral cavity (Casserly, Rowley, Marino, & Pollack, 2016), proprioceptive information from muscles, tendons, and skin is difficult to mask. Even the few studies that have employed subcutaneous nerve injections have blocked primarily tactile sensation, likely leaving proprioception intact (Niemi, Laaksonen, Ojala, Aaltonen, & Happonen, 2006; C. M. Scott & Ringel, 1971). Since proprioceptive sensory afferent information for the tongue is likely carried along the same nerves as outgoing efferent motor commands (e.g., hypoglossal nerve; Fitzgerald & Sachithanandan, 1979), pharmacological nerve blocks administered to these pathways will have effects on both sensory and motor systems. Despite the inability to completely block somatosensation, blocking of tactile sensation leads to substantial imprecision in speech articulation, suggesting an important role for the somatosensory system in the precise control necessary for speech (Putnam & Ringel, 1976; Ringel & Steer, 1963; C. M. Scott & Ringel, 1971). However, speech in these experiments remains largely intelligible (though distorted), suggesting that some type of predictive or feedforward controller operates independently of sensory feedback. This conclusion must be taken cautiously, however, as proprioceptive feedback, as well as tactile feedback from the posterior tongue, was still available in these studies.

Additional evidence for the role of somatosensory information comes from studies on nonspeech motor

behavior in nonhuman primates. There is some evidence that deafferented primates show severely abnormal movements (Desmurget & Grafton, 2003; Lassek, 1955; Mott & Sherrington, 1895). However, these severe abnormalities are not permanent, with at least some individuals recovering somewhat (though not fully) normal motor function (Knapp, Taub, & Berman, 1963). Moreover, monkeys trained to make particular single-joint pointing movements retained the ability to make accurate movements after deafferentation (Polit & Bizzi, 1979), which was replicated in single finger movements in humans with temporary deafferentation applied by an inflatable arm cuff (Kelso & Holt, 1980). However, deafferented patients struggle with more complex movements (Rothwell et al., 1982). While the evidence is mixed, it seems that while some simple movements can be performed accurately even in the absence of somatosensory feedback, this feedback is critical for more complex movements (such as speech).

Another way to examine the role of sensory feedback in online control of speech is to examine the response of the speech motor system to externally imposed perturbations of sensory feedback. There is substantial evidence from studies of this type that the speech production system does make use of both auditory and somatosensory feedback for online control.

The earliest studies examining the role of auditory feedback in speech motor control used delayed auditory feedback (DAF) systems, where a participant's speech is recorded, delayed by some amount (typically 100–200 ms), and then played back to the participant over headphones. DAF has been shown to severely impact the fluency of speech (Fairbanks, 1954; Lee, 1950; Yates, 1963), indicating that the speech motor system monitors the auditory afferent signal. Similarly, loud masking noise causes a number of changes, including increased loudness (Lane & Tranel, 1971; Lombard, 1911) and changes in vowel formants (Summers, Johnson, Pisoni, & Bernacki, 1989). Conversely, increasing the amplitude of feedback causes a decrease in speech loudness (Chang-Yit, Pick, & Siegel, 1975). These results are consistent with the role for auditory feedback, at least in regulating speech volume.

Further evidence in favor of a role for auditory feedback in speech motor control comes from altered auditory feedback paradigms that impose external perturbations on vowel formants (Niziolek & Guenther, 2013; Purcell & Munhall, 2006b; Tourville, Reilly, & Guenther, 2008) or vocal pitch (Elman, 1981; Jones & Munhall, 2000). In these cases, participants respond to external perturbations by altering their speech to oppose the perturbation during the duration of a single production. For example, when a vowel formant or vocal pitch is lowered, speakers respond by raising that formant or pitch real time. Similarly, speakers will lower their vowel formants or pitch when these are raised.

Analogous results have been found for somatosensory perturbations. When unexpected loads are applied to the lower lip during a bilabial closure (e.g., for /p/), the upper lip lowers to a greater degree than normal to compensate

for this perturbation, allowing the lips to achieve closure (Abbs & Gracco, 1984; Gracco & Abbs, 1985). Critically, the response of the lower lip to an upper lip perturbation is found only when producing bilabial sounds like /p/ or /b/, produced with both upper and lower lips, but not labiodental /f/, produced with the lower lip and upper teeth (Shaiman & Gracco, 2002). This indicates that compensatory reactions to imposed loads are task dependent and produced only when needed to accomplish the current production goal. Similar results are found when the load is applied to the jaw—loads imposed during production of /p/ induce compensatory lowering of the upper lip but no compensatory response from the tongue, while loads imposed during production of /s/ show the opposite pattern (Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984).

Together, these results indicate that both auditory and somatosensory feedback are used, when available, for the online control of speech movements. Somatosensory feedback, in particular, may be used to control online movement at both cortical and spinal levels. This use of somatosensory feedback at multiple levels is known to be the case for arm movements (S. H. Scott, 2012). In speech, a role for somatosensory feedback at both spinal/peripheral control (Perrier, 2005; Perrier, Løevenbruck, et al., 1996; Perrier, Ostry, & Laboissière, 1996; Sanguineti, et al. 1998) and cortically based control (Guenther, 2016; J. F. Houde & Nagarajan, 2011; Saltzman & Kelso, 1987) have been suggested. While it seems likely that peripheral control does make use of sensory feedback in a similar manner as in limb control, there is, to our knowledge, currently no direct evidence of this process. There is, however, ample evidence for a cortical use of somatosensory feedback. In speech, the latencies of responses to sensory feedback perturbations in human speakers suggest that these responses have a cortical, rather than reflexive (spinal or brainstem), origin. The first spinal reflex responses appear only 15–25 ms after the onset of a somatosensory perturbation (Desmurget & Grafton, 2003; Gottlieb & Agarwal, 1979). Though these estimates come from studies of limb control, they likely overestimate the speed of reflex responses in speech given the much shorter distance between the speech articulators and brainstem compared to the limb extremities and spine (Guenther, 2006). However, electromyographic latencies to mechanical perturbations are reported to be around 30–85 ms (Abbs & Gracco, 1984). Moreover, transcranial magnetic stimulation results have confirmed a cortical role in motor responses to jaw-load perturbations (Ito, Kimura, & Gomi, 2005). The previously described task specificity of feedback responses (Kelso et al., 1984; Shaiman & Gracco, 2002) also suggests a cortical origin of the response, as lower level (reflex) responses are identical in form regardless of the current task or action goal (Desmurget & Grafton, 2003, though cf. Weiler, Gribble, & Pruszynski, 2019).

A similar cortical role for auditory feedback is evident. Response latencies (measured behaviorally) to auditory responses are even longer than for somatosensory feedback,

roughly 150 ms (Cai et al., 2012; Parrell, Agnew, Nagarajan, Houde, & Ivry, 2017; Tourville et al., 2008), much longer than the 20-ms perioral auditory reflex (McClellan & Sapir, 1981). Auditory perturbations also elicit task-specific responses; for example, the sensitivity of the speech system to auditory feedback is influenced by linguistic stress and emphasis (Kalveram & Jäncke, 1989; H. Liu, Zhang, Xu, & Larson, 2007; Natke, Grosser, & Kalveram, 2001; Natke & Kalveram, 2001). To date, the extent of subcortical contributions to auditory feedback control pathways remains unclear.

Sensory Feedback Use in Long-Term Calibration and Tuning

In addition to its role in the online control of speech actions, sensory feedback also plays an important role in the long-term updating and maintenance of feedforward or predictive speech motor control. The speech motor system undergoes changes at both long- and short-term time scales, but the speech motor system is able to deal with both types of change to maintain accurate, fluent speech. For example, the vocal tract changes shape and configuration during childhood, and the shape of the vocal tract may be artificially altered by orthodontics or surgical procedures. At shorter time scales, the motor system must deal with muscle fatigue throughout the course of the day. The motor system is able to adapt relatively quickly to these changes, and there is strong evidence that sensory feedback plays a critical role in this process.

There is convincing evidence that auditory feedback is important for long-term calibration of the speech motor system. While some individuals who acquire speech normally but become deaf postlingually typically maintain the ability to produce mostly fluent speech, the speech of other individuals is highly impaired (Cowie & Douglas-Cowie, 1983). These impairments affect both prosodic (pitch, voicing quality, intersegment articulatory timing) and articulatory control of both consonants and vowels (Cowie & Douglas-Cowie, 1983; Lane & Webster, 1991). Moreover, these speakers typically develop decreased contractiveness between different speech sounds over time (Cowie & Douglas-Cowie, 1992; Lane & Webster, 1991). These deficits indicate that auditory feedback likely plays an important role in the maintenance and fine-tuning of feedforward control over longer time scales (Perkell, 2012).

The role of sensory feedback in adaptation of feedforward or predictive control has also been investigated using a variety of manipulations that perturb auditory or somatosensory feedback consistently over a large number of trials. Typically, the speech motor system changes over time to counteract the imposed perturbation, indicating that sensory feedback plays an important role in maintaining accurate speech production. Such adaptive responses have been seen in response to auditory perturbation of vowel formants (J. F. Houde & Jordan, 1998, 2002; Purcell & Munhall, 2006a), vocal pitch (Jones & Munhall, 2000),

and the spectral characteristics of fricatives (Shiller, Sato, Gracco, & Baum, 2009), and these results have been replicated many times. Similar changes have been seen in speech production when physical perturbations are introduced that create both somatosensory and auditory perturbations. When participants produce sibilant consonants after being fitted with a prosthesis that changes the shape of the hard palate (creating unexpected tactile feedback and alterations to the spectral center of the fricative), they alter their production such that the fricative centroid returns to near baseline values (Baum & McFarland, 1997, 2000). Similarly, when participants produce the vowel /u/ after being fitted with a lip tube of a fixed diameter (creating a proprioceptive error and a change in vowel formants), they can adapt their productions over time to return their production to near baseline values (Savariaux, Perrier, & Orliaguet, 1995; Savariaux, Perrier, Orliaguet, & Schwartz, 1999). Adaptive changes are seen when the skin is mechanically stretched laterally during speech, which leads to increases in lip rounding to compensate for the lateral pull of the skin (Ito & Ostry, 2010).

Adaptation to somatosensory perturbations has also been shown to drive adaptive changes in speech production even when they do not induce auditory errors, indicating that somatosensory feedback itself is also used to update feedforward/predictive control. Experiments with dental prosthetics, similar to the palatal prosthetics discussed earlier, have shown that speakers adapt to the alterations in the shape of the vocal tract even when auditory feedback is masked by loud auditory noise (Jones & Munhall, 2003). Participants also adapt to a velocity-dependent force field artificially applied to the jaw by a robot during jaw opening and closing movements for speech, even though these perturbations have no acoustic consequences (Tremblay, Shiller, & Ostry, 2003). Interestingly, participants do not adapt to the same perturbation applied to nonspeech jaw movements, suggesting that motor adaptation in the oral motor control system is task specific in a similar way to online compensations for sensory or mechanical perturbations.

While both somatosensory and auditory feedback play a role in adaptation of speech motor control, the relative contribution of each domain is unclear. Although speakers are able to adapt to changes in vocal tract shape in the presence of masking noise, adaptation is more complete when auditory feedback is available (Jones & Munhall, 2003). This suggests that auditory and somatosensory feedback have complimentary roles. However, when incompatible auditory and somatosensory perturbations are introduced concurrently (pulling the jaw upward while raising F1), participants compensated for the auditory, rather than the somatosensory, perturbations (Feng, Gracco, & Max, 2011), suggesting a primary role for auditory feedback, at least for vowels. On the other hand, there is substantial interspeaker variability in adaptation to both somatosensory (Baum & McFarland, 2000; Lametti, Nasir, & Ostry, 2012; Ménard, Perrier, Aubin,

Savariaux, & Thibeault, 2008; Savariaux et al., 1995) and auditory (Lametti et al., 2012; Munhall, MacDonald, Byrne, & Johnsrude, 2009; Purcell & Munhall, 2006a) perturbations, suggesting that the use of sensory feedback for adaptive learning may vary substantially between individuals. In fact, the relative weight of auditory versus somatosensory feedback has been shown to vary across individuals, with some speakers showing larger adaptations to somatosensory perturbations and others to auditory perturbations (Lametti et al., 2012), even when the perturbations are applied concurrently (in this experiment, the somatosensory perturbation did not affect speech acoustics).

Neural Evidence for Sensory Feedback Use in Speech Production

Neuroanatomical Pathways Supporting Sensory Feedback

There is a great deal of neuroanatomical evidence that the sensory systems of the brain make contact with the motor systems governing speaking. Certainly, there are local afferent/efferent connections in the brainstem and spinal cord (Jürgens, 2002; Webster, 1992). This is most well known to be true for somatosensation, but it is also the case for audition, with ascending auditory information in the inferior colliculus making contact with descending vocal motor pathways in the periaqueductal gray. In cortex, there are many more ways that sensory processing areas, including even primary auditory cortex, link up with the complex of speech motor areas, including primary motor cortex (Huang, Liu, Yang, Mu, & Hsia, 1991; Skipper, Devlin, & Lametti, 2017; Yeterian & Pandya, 1999). Of these, what is considered one of the most important connections is the arcuate fasciculus and, more generally, the superior longitudinal fasciculus, which links high-level auditory processing areas (posterior superior temporal sulcus/superior temporal gyrus [STG], posterior planum temporale, ventral supra-marginal gyrus [vSMG], sylvio partial temporal [SPT]) to the premotor areas (ventral premotor cortex [PMv] and dorsal premotor cortex) and frontal areas (pars triangularis, ventral operculum, dorsal operculum—all fields of Broca's region) implicated in speech production (Friederici, 2011; Glasser & Rilling, 2008; Schmahmann et al., 2007; Upadhyay, Hallock, Ducros, Kim, & Ronen, 2008). This pathway has been referred to as the “dorsal stream” speech feedback processing pathway (Hickok & Poeppel, 2007).

When this pathway is disrupted, a number of speech sensorimotor disorders appear to result (Hickok, Houde, & Rong, 2011). One of these is conduction aphasia, which results from a stroke along this pathway and is characterized by a state in which production and comprehension of speech are preserved but the ability to repeat speech sound sequences just heard is impaired (Geschwind, 1965). Temporary disruptions of the dorsal stream have also been shown to rather directly affect auditory feedback processing for speech production. Shum et al. used repetitive

transcranial magnetic stimulation of the inferior parietal lobe (near the SPT region) to inhibit its activity as subjects spoke while exposed to altered formant feedback (Shum, Shiller, Baum, & Gracco, 2011). The authors found that repetitive transcranial magnetic stimulation of this region did in fact reduce the degree that subjects altered their behavior in response to the altered feedback.

Functional Imaging Studies of Sensory Feedback Processing During Speaking

Speaking-Induced Suppression

Given the neuroanatomical substrates linking sensory and motor areas, it is not surprising that functional imaging studies have shown evidence of sensorimotor interactions. One of the most basic is an effect we refer to as *speaking-induced suppression* (SIS), where the response of a subject's auditory cortices to his or her self-produced speech is significantly smaller than their response to similar but externally produced speech (e.g., tape playback of the subject's previous self-productions). This effect has been seen using positron emission tomography (Hirano et al., 1996; Hirano, Kojima, et al., 1997; Hirano, Naito, et al., 1997), electroencephalography (Ford & Mathalon, 2004; Ford et al., 2001), magnetoencephalography (MEG; Curio, Neuloh, Numminen, Jousmäki, & Hari, 2000; Heinks-Maldonado, Nagarajan, & Houde, 2006; J. F. Houde, Nagarajan, Sekihara, & Merzenich, 2002; Numminen & Curio, 1999; Numminen, Salmelin, & Hari, 1999; Ventura, Nagarajan, & Houde, 2009), and electrocorticography (ECoG; Chang, Niziolek, Knight, Nagarajan, & Houde, 2013; Greenlee et al., 2011). An analog of the SIS effect has also been seen in nonhuman primates (Eliades & Wang, 2003, 2005, 2008). MEG experiments have shown that the SIS effect is only minimally explained by a general suppression of auditory cortex during speaking and that this suppression is not happening in the more peripheral parts of the CNS (J. F. Houde et al., 2002). The observed suppression goes away if the subject's feedback is altered to mismatch his or her expectations (Heinks-Maldonado et al., 2006; J. F. Houde et al., 2002), as is consistent with some of the positron emission tomography study findings.

These features of the SIS effect suggest that the motor actions generating speech allow auditory cortex to anticipate the auditory consequences of speaking. Speech, however, is highly variable, as can be seen in repeated productions of a vowel, and so a further question is whether the feedback prediction mechanisms supporting SIS can account for this variability. Niziolek, Nagarajan, and Houde (2013), examining how SIS varies as a function of variation in the formants (F1/F2) of a vowel's production, found that SIS for outlying productions of a vowel was significantly less pronounced than was SIS for productions closer to the vowel's mean. The results suggest that not all production variability is tracked by auditory feedback predictions. Whether this is due to a limitation in the feedback prediction process (e.g., the prediction only represents the mean expected production) or is instead due to production

variability arising downstream from cortical sources remains to be resolved. However, the results do further strengthen the case for SIS being a neural marker for a prediction comparison process that underlies the auditory feedback processing during speaking.

Neural Responses to Altered Sensory Feedback During Speaking

Another approach to exhibiting the sensory prediction comparison process is to contrast neural responses during speaking with normal feedback with responses observed during speaking when sensory feedback is altered. Using this approach, functional imaging studies examining speaking with audio feedback alterations have identified a number of cortical areas that are potentially involved in processing auditory feedback. A variety of feedback alterations have been investigated, including pitch shifts (Fu et al., 2006; McGuire, Silbersweig, & Frith, 1996; Parkinson et al., 2012; Toyomura et al., 2007; Zarate, Wood, & Zatorre, 2010; Zarate & Zatorre, 2008), DAF (Hashimoto & Sakai, 2003; Hirano, Kojima, et al., 1997), formant shifts (Tourville et al., 2008), and replacement of the voiced feedback with noise modulated by the amplitude envelope of the voiced feedback (Zheng, Munhall, & Johnsrude, 2009). All studies found the STG to be more active when feedback was altered, with some studies localizing activity to the mid-STG (Fu et al., 2006; Hirano, Kojima, et al., 1997; McGuire et al., 1996; Tourville et al., 2008) or mid- to post-STG (Hashimoto & Sakai, 2003; Parkinson et al., 2012) and most studies finding the posterior STG (pSTG) region particularly responsive to altered feedback (Hashimoto & Sakai, 2003; Takaso, Eisner, Wise, & Scott, 2010; Tourville et al., 2008; Zarate et al., 2010; Zarate & Zatorre, 2008; Zheng et al., 2009). Nearby the pSTG, many studies also found the vSMG more active in altered feedback trials (Hashimoto & Sakai, 2003; Tourville et al., 2008; Toyomura et al., 2007). Regions of right anterior and posterior cerebellum (anterior vermis and Lobule VIII) have also been shown to be more active in trials with altered auditory feedback (Tourville et al., 2008).

The altered feedback approach has also been used to identify regions involved in somatosensory feedback processing during speaking. Golfinopoulos et al. (2011) used functional magnetic resonance imaging to contrast neural responses during speaking with normal feedback with speaking when a jaw movement was unexpectedly blocked via rapid inflation of a balloon placed between the molars. The contrast between normal and perturbed speech showed that the somatosensory perturbation increased activation in ventral motor cortex, supramarginal gyrus, inferior frontal gyrus, premotor cortex, and inferior cerebellum.

Speech Perturbation Response Enhancement

Studies of altered feedback responses like those described above use the "whole trial" alteration method, where contrasts are made between trials with and without altered feedback. Another way to gauge responses to altered feedback is to use transient, midutterance perturbations. In this case, the baseline for judging neural responses to the altered

feedback can be the time period immediately adjacent to the onset of the feedback perturbation. This allows for a potentially more sensitive contrast to isolate responses to the altered feedback relative to the actual present state of the speech motor system, not just a similar state on a different trial. Of course, the functional imaging method needed to make such adjacent time interval comparisons must have high time resolution, so studies based on this approach have been limited to electroencephalography, ECoG, and MEG.

Studies using this approach have examined responses to transient, midutterance perturbations of the pitch of the audio feedback heard by speakers in their ongoing phonation. Speakers are observed to compensate within several hundred milliseconds of such pitch feedback perturbations—the so-called pitch perturbation reflex (Burnett, Freedland, Larson, & Hain, 1998). This clear causal link between an applied feedback perturbation and a compensatory response makes this experimental paradigm ideal for studying the role of sensory feedback in speaking. Neural studies of the pitch perturbation reflex found a result that, in some sense, is the opposite of the SIS effect described above. In response to altered feedback, some auditory cortical areas have an enhanced response, compared to their response to hearing the same altered feedback when passively listening (Behroozmand, Karvelis, Liu, & Larson, 2009; Chang et al., 2013; Kort, Cuesta, Houde, & Nagarajan, 2016; Kort, Nagarajan, & Houde, 2014). This effect has been called *speech perturbation response enhancement* (SPRE) and has been localized to the posterior auditory cortical regions (left pSTG, left vSMG, right mSTG) by a functional magnetic resonance imaging study comparing auditory responses to noise-altered feedback during speaking with passive listening. In our own studies of SPRE using MEG, we localized the effect to right parietal and premotor cortex and left posterior temporal cortex (Kort et al., 2016). In our ECoG study, we found that both pSTG and ventral temporal areas exhibited SPRE, and we also found that responses in these auditory areas were significant predictors of the trial-to-trial variability of subjects' compensatory responses. Studies of SPRE have therefore produced particularly strong evidence for a role of sensory feedback in ongoing speaking.

Models of Speech

As outlined above, there is strong evidence that sensory feedback is used for online speech control at a cortical or central control level. This indicates that the speech system is not a purely feedforward controller or even a feedforward controller with sensory feedback being used at a lower, reflexive level. However, the latencies of feedback responses mean a pure feedback control is also not plausible. How can we reconcile this information, and what does that reconciliation tell us about the speech motor control system?

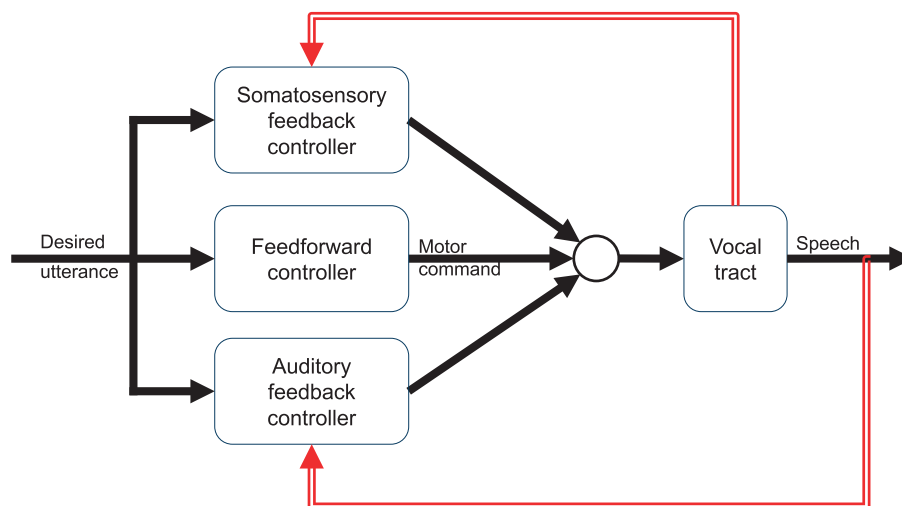
Here, we focus principally on two models of speech motor control: the DIVA model (Guenther, 2016) and the SFC/feedback aware control of tasks in speech (FACTS) model, which we have been developing in our labs (J. F. Houde & Nagarajan, 2011; J. F. Houde, Niziolek, Kort,

Agnew, & Nagarajan, 2014; Parrell et al., 2018; Parrell, Ramanarayanan, et al., 2019; Ramanarayanan et al., 2016). These models represent two different approaches to modeling speech motor control. DIVA combines a feedforward controller specifying preplanned articulatory trajectories with feedback controllers that rely on sensory reafference to correct ongoing movement. FACTS, instead, combines an internal estimate of sensory feedback with actual sensory reafference in a SFC system without preplanned articulatory trajectories. Many other models can be viewed as alternative instantiations of these two basic approaches: the ACTION-based model of speech production (Kröger, Kannampuzha, & Neuschaefer-Rube, 2009) shares much of its system architecture with DIVA, whereas the Tian and Poeppel (Tian & Poeppel, 2010) and Hickok (Hickok, 2014; Hickok et al., 2011) models are alternative implementations of an SFC architecture. While there may be subtle differences between these alternative models and DIVA/FACTS, we believe that most of the predictions made by the DIVA and FACTS models would likely hold for these other models, given their similar architectures. We do not discuss models of speech motor control that do not directly address the role of sensory feedback. These include models based on the equilibrium point hypothesis (Perrier, 2005; Perrier, Løevenbruck, et al., 1996; Perrier, Ostry, et al., 1996; Sanguineti et al., 1998) and the task dynamics model (Saltzman & Kelso, 1987; Saltzman & Munhall, 1989). While these models have proven useful in explaining some important phenomena in speech production (including the role of mechanics and vocal tract dynamics in control, coarticulation, and speech unit sequencing), they do not speak to how sensory feedback is used at the level of central control.

DIVA

One approach to modeling speech motor control is to combine a feedback controller with a feedforward controller specifying preplanned, time-varying trajectories. This approach has a long history in nonspeech motor control models (Arbib, 1981; Kawato, Furukawa, & Suzuki, 1987). In speech, this combined feedforward/feedback architecture is used in the DIVA model (Guenther, 2016; Tourville & Guenther, 2011). In terms of its control architecture, the DIVA model is a variant of the feedback error learning model of control developed to explain nonspeech motor behavior (Kawato & Gomi, 1992). The key features of the feedback error learning architecture are (a) motor commands are generated as the sum of the outputs of a feedforward controller and a feedback controller and (b) the output of the feedback controller teaches the feedforward controller. In this model, the control system initially relies completely on sensory feedback and the feedback controller to generate appropriate motor commands. Gradually, however, the feedforward controller is adjusted to minimize the output of the feedback controller, and the control system becomes less reliant (and ultimately nonreliant) on sensory feedback.

Figure 1. The directions into velocities of articulators (DIVA) model, after Guenther (2016). The DIVA model combines a feedforward controller (middle) with separate feedback controllers for auditory and somatosensory feedbacks.



These features of the feedback error learning model are recapitulated in the DIVA model for the control of speech, which builds on the Kawato and Gomi (1992) model with substantial speech-specific innovations in terms of proposed neural bases and auditory control space. DIVA combines a feedforward controller, which specifies preplanned trajectories in auditory, somatosensory, and articulatory spaces, with two feedback controllers processing auditory and somatosensory feedback, respectively (see Figure 1). The output of these three controllers are summed to generate the final motor commands issued to the speech system (in DIVA, which uses a kinematic model of the vocal tract, these motor commands are desired positions and velocities of the speech articulators rather than muscle activations).

While the articulatory trajectory (or feedforward motor plan) is specified as a time-locked series of discrete positions, the auditory and somatosensory signals are defined as time-varying regions rather than single trajectories. Sensory errors and subsequent feedback-based motor commands are only generated when sensory feedback falls outside these target regions. This means that, effectively, any sensory feedback that falls within these boundaries is considered to be equally acceptable.¹

Because of its hybrid architecture, the DIVA model is able to reproduce some of the key findings regarding the role of the sensory system in speech motor control. In the fully trained model, equivalent to a typical adult speaker, the planned trajectory in articulatory space used by the

feedforward controller is accurate enough to maintain the auditory and somatosensory signals generated by the speech system within the sensory target boundaries. Thus, for well-learned speech, the system operates in a purely feedforward manner. In fact, because the feedforward controller is separate from the two feedback controllers, the system is able to produce speech even in the absence of sensory feedback by relying only on the feedforward controller. However, the presence of the feedback controllers means that the system is responsive to deviations from expected sensory feedback, whether these are caused by inaccurate feedforward motor plans or external perturbations of either auditory (Tourville et al., 2008) or somatosensory feedback (Golfinopoulos et al., 2011).

DIVA also models the use of sensory feedback for adaptation seen in human speech. In the feedback error learning model described above, the output of the feedback controller is used to learn the feedforward controller. This is also the case in DIVA, where the feedforward controller is updated using the output of the feedback controllers to update the articulatory trajectory used as the feedforward motor command. This iterative process is used to initially train the model (i.e., learn the correct feedforward commands for a given auditory target) and remains online to update the motor commands based on any deviations from the sensory targets. Because learning relies on feedback corrections generated during online control (though these corrections need not necessarily be executed), adaptation only occurs when sensory feedback falls outside the time-varying regions that define sensory targets.

DIVA is the most fully developed model of speech motor control to date. Developed over the past 25 years, the DIVA has computationally implemented a number of important aspects of the speech production system beyond the basic control architecture described above (Guenther, 2016). DIVA has a fully developed implementation of speech

¹In one article, Gaussian distributions (where the width of the distribution is related to auditory acuity) were proposed as an alternative to the categorical nature of auditory targets (Villacorta et al., 2007). However, the standard description of DIVA includes all-or-nothing targets, and how Gaussian distributions would play out with other aspects of the model, such as target learning and coarticulation, have not been detailed.

motor development and later tuning of feedforward commands. DIVA has also explicitly localized all of its proposed computational processes to different cortical and subcortical structures and is capable of producing simulated neural activity that can be compared to functional magnetic resonance imaging data from human speakers. Extensions to the model have also addressed syllable and phoneme sequencing and gating (Bohland, Bullock, & Guenther, 2010).

SFC

There is another way to create a control system that, like speaking, is both responsive to and yet not dependent on sensory feedback. Instead of explicitly separating feedforward from feedback control, this control system is entirely based on feedback control. However, instead of directly relying on only reafferent sensory information as the input to the feedback controller, feedback also comes from an internally generated source: an estimate of the current state of the system being controlled. These two sources of feedback, internal and reafferent, are combined and used in a single comparison with the production target. This approach is called *state feedback control* or *SFC*, and it has long been a province of engineering control theory, partly because of its ability to generate controls even when sensor feedback is only intermittently available and partly because it forms the basis for optimal control theory (Todorov, 2004, 2006; Todorov & Jordan, 2002). Indeed, because the compatibility of the SFC architecture with current theories of optimal motor control, SFC has gained wide acceptance as a model of how the CNS controls movements across a variety of motor domains (Diedrichsen et al., 2010; Franklin & Wolpert, 2011; S. H. Scott, 2004, 2012; Shadmehr & Krakauer, 2008). Our goal in developing an SFC-based model has been to attempt to test the applicability of this widely accepted framework to the speech motor system, working from the hypothesis that the same principles and control structures that underlie other human motor systems are shared with speech.

The key concept in SFC is that the system being controlled (e.g., an arm, a vocal tract) has a dynamic state. This state could be (and, in practice, is often assumed to be) the positions and velocities of the articulators comprising the system, but the important thing is that the state is an instantaneous description of the system sufficient to predict its future behavior. A controller in the CNS trying to accomplish some task (e.g., generate a desired movement, speak some utterance) could use this state as the input to a *state feedback control law*, which can generate controls specific to the current state that would move the system closer to the state that would accomplish the desired task.

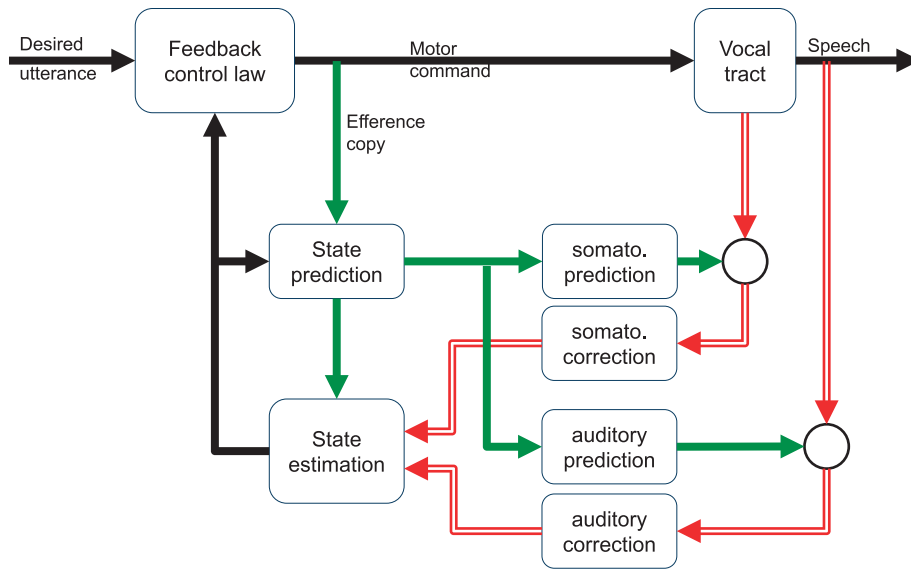
In general, however, it is assumed that the CNS does not actually have direct access to this state due to noise and delays in the neural system. Instead, the CNS must rely on an internal estimate of the current state. This estimate is generated by a *state estimator* based on a number of different inputs. First, the estimator relies on a *prediction* of the current state of the system generated from a copy of previously

issued motor commands, or *effference copy*. The system that converts this effference copy to a state prediction is known as a *forward model*. The other input to the state estimator, along with the state prediction, is information from the sensory system or *sensory feedback*. The combined system (state feedback control law and state estimator) is thus responsive to sensory feedback if it is present (the state estimator uses sensory feedback to update and correct the state prediction), but the system is not dependent on sensory feedback (the state estimator can still provide a state estimate based on the output of the forward model even in the absence of sensory feedback). The forward model learns to predict the system state by comparing the predicted output with measurements of the system's response as relayed by sensory feedback.

In our SFC model of the control of speaking (see Figure 2), we have implemented an articulatory state feedback control law driving the vocal tract to produce desired utterances. Critically, this type of control system generates controls online at every time step based on a state estimate provided by an articulatory state estimator. The state estimate is generated by a recurrent prediction/correction process that runs continually during speaking. At every time step, a prediction of the current articulatory state is first made based on the previous state estimate and an effference copy of the current articulatory controls. This state prediction, in turn, is used to make predictions about what somatosensory and auditory feedbacks should currently be expected from the vocal tract. Incoming sensory feedback is compared with those predictions, and any mismatches (feedback prediction errors) are converted into corrections that are applied to the predicted state. The result is an updated estimate of the current articulatory state that is used by the articulatory state feedback control law, along with a desired end state of the movement, to generate further controls sent to the vocal tract articulators.

In the SFC model, therefore, motor programs for desired utterances are not represented directly as time-dependent sequences of articulatory controls or desired articulatory positions (trajectories) but indirectly as one or more desired end states (or target states) that the vocal tract articulators should achieve in order for the desired utterance to be produced. The articulatory control law, at every time step, creates controls to move from the currently estimated articulatory state toward the currently desired target state. This arrangement has two consequences. First, the state feedback control law generating controls is a mapping between states and controls that is potentially sharable between utterance motor programs. Second, since utterance motor programs are represented as desired end states rather than desired trajectories, they are not necessarily fixed temporal output sequences but potentially could be adaptive time courses, with sequencing of the next desired state possibly dependent on whether the previous desired state has been achieved. While our SFC model, like most SFC or optimal control models in other domains, uses target end states (or temporally varying patterns of target end states) as the input to the controller rather than time-dependent trajectories, the general architecture of SFC is

Figure 2. The state feedback control model, after J. F. Houde and Nagarajan (2011). There is no separate feedforward controller in this model. Instead, there is an internal state prediction and estimation process that is able to operate with delayed, noisy, or even absent sensory feedback.

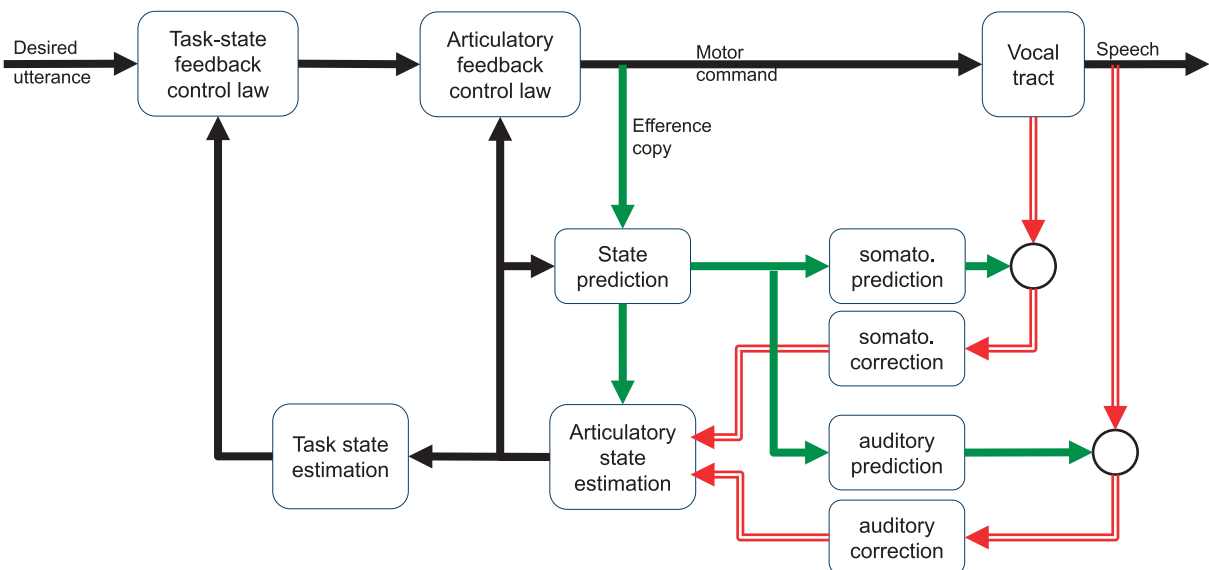


also compatible with cases where explicit trajectory control may be necessary, such as in slow reaching tasks (Cluff & Scott, 2015). Such a model would be substantially similar to the feedforward controller in DIVA, though with a more complex state estimation component.

We have recently expanded our SFC model of articulatory control to incorporate control of higher level speech

tasks, which we call the *feedback aware control of tasks in speech* or *FACTS* model (Parrell et al., 2018; Parrell et al., 2019; Ramanarayanan et al., 2016), shown in Figure 3. There is substantial evidence that the goals of speech production, like the goals of nonspeech motor actions (Diedrichsen et al., 2010), are not particular articulatory states or trajectories. For example, although the jaw is typically mobile in

Figure 3. The feedback aware control of tasks model. This model is an extension of the state feedback control approach to the control of high-level speech tasks, rather than simply at the articulatory level. The model combines a low-level articulatory state feedback controller (cf. state feedback control model in Figure 2) with a high-level task-based controller (Saltzman & Munhall, 1989).



speech, speakers produce immediate and full compensation when their jaw is fixed with a bite block (Fowler & Turvey, 1981), suggesting that consistent positions of the individual speech articulators are not active goals for speech production. Moreover, this compensation is immediate, suggesting that the speech motor system does not need time to learn or adapt to the perturbation (as it does for auditory perturbations) and that the speech motor control system is flexible enough to allow for variation in the positions of individual speech articulators without adversely affecting speech output. This is similar to studies that have applied unexpected mechanical loads to the lower lip or jaw (Abbs & Gracco, 1984; Gracco & Abbs, 1985; Kelso et al., 1984; Shaiman & Gracco, 2002), which consistently show compensatory movements in other, unperturbed articulators involved in producing the current speech segment. Further evidence comes from studies that have examined patterns of variability in articulator movement. For example, when the movements of the upper lip, lower lip, and jaw were measured during repeated production of the bilabial stop /p/, the movement variability of each individual articulator was greater than their summed movement (Gracco & Abbs, 1986), suggesting that the most relevant dimension for control is the joint movement of all three articulators (i.e., lip closure) rather than the individual articulator positions.

Together, this evidence strongly suggests that the control of the speech motor system is not organized at the level of individual articulator movements but at some high-order level. In FACTS, we currently take these high-level tasks to be constrictions in the vocal tract, following the task dynamic model of speech motor control (Saltzman, 1986; Saltzman & Kelso, 1987; Saltzman & Munhall, 1989). Tasks include lip aperture, lip protrusion, the location and degree of constriction between the tongue tip and palate, the location and degree of constriction between the tongue body and palate/rear pharyngeal wall, and degree of velum opening, among others. However, there is evidence that other types of targets exist, such as auditory targets for American English /r/ (Guenther et al., 1999; Nieto-Castanon, Guenther, Perkell, & Curtin, 2005). The separation of tasks from articulation in FACTS allows for the use of both constriction targets and targets in other reference frames, which we are currently exploring. The overall architecture of such a system would be substantially similar to the current model.

These tasks are controlled in an analogous manner to the articulatory feedback control system in standard SFC models: a task-level feedback control law generates task commands to move the system toward a task-level goal based on the current task state (the “position” and “velocity” of the tasks, which is, to say, their current value and current rate of change). The task-level control law is taken from the dynamical systems approach in task dynamics, where each gesture is modeled as a dynamic system that outputs changes in task acceleration based on the current task position, task velocity, and task goal. Each gesture is active during a particular time window for a given utterance, and more than one gesture may be active at any given time. The gestures for a particular utterance can be arranged into a “gestural score”

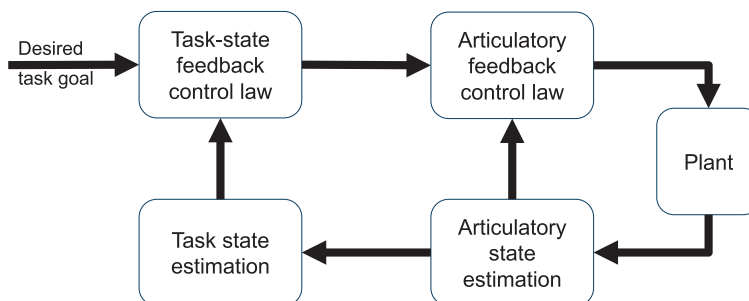
in the framework of articulatory phonology (Browman & Goldstein, 1992, 1995), which is nothing more than a specification of which gestures are active in the control law at each time step. Effectively, each gestural score defines a time-varying task state feedback control law that will direct the system toward the attainment of the gestures that are active at a given point in time. These gestural scores serve as the input to the task-level feedback controller in FACTS.

The output of the task-level control law is used as the input to a lower-level articulatory state feedback controller. As in our SFC model, this articulatory controller generates articulator accelerations (equivalent to motor commands in the model) to move a model of the vocal tract, generating somatosensory and auditory output.² This sensory feedback is compared with an internal prediction of the state generated from an efference copy of the motor command to estimate the current articulatory state, which is passed back to the articulatory control law. This articulatory state is additionally used to estimate the current task state needed as an input to the task-level controller. The FACTS model thus has a hierarchical architecture, with a high-level task state feedback controller operating on a low-level articulatory state feedback controller, which controls the speech production mechanism itself. A similar hierarchical approach has been proposed in the optimal control literature as a domain-general approach for controlling abstract, high-level tasks in redundant motor systems like speech, as shown in Figure 4 (Haar & Donchin, 2019; Todorov, Li, & Pan, 2005). The principle benefit of such a hierarchical control structure is to simplify the control of tasks by separating this feedback law from the high dimensionality of the articulatory system while maintaining the link between low-dimensional, high-level task control and the high-dimensional articulatory (or muscle) synergies used to accomplish these goals. Interestingly, such a hierarchical structure emerges in optimal feedback control of redundant tasks (like speech), even when they are not explicitly designed to be hierarchical (Todorov et al., 2005). This suggests that an explicitly hierarchical architecture is reasonable for redundant systems such as speech production.

In the current implementation of FACTS, sensory feedback is a combination of both auditory (F1, F2, F3) and somatosensory information (the position and velocities of the model articulators). Neural noise is modeled by adding scaled Gaussian white noise to the system with separate standard deviations for the auditory and somatosensory channels. Because of the use of sensory feedback in generating the final articulatory state estimate in the lower-level state feedback controller, this system is able to make compensatory responses to both auditory and somatosensory perturbations, which have been implemented as alterations to the first formant in the auditory feedback channel and

²While the output of the low-level controller is accelerations in the current model, the flexibility of the SFC concept would allow us to output desired forces, individual muscle activations, or any combination of them as well, given a more realistic model of the plant (S. H. Scott, 2004).

Figure 4. Generic hierarchical state feedback control architecture, as proposed in Todorov et al. (2005). The feedback aware control of tasks model can be viewed as a speech-specific implementation of this domain-general architecture.



forces applied to the jaw during consonant closure, respectively (Parrell et al., 2018; Parrell et al., 2019).

The behavior of the model provides a qualitative match to human behavior. For auditory perturbations, this means the model produces only partial compensation. In the model, this is because final state estimate is a weighted combination of the internal prediction, somatosensory feedback, and auditory feedback. The induced auditory error is somewhat balanced out by the fact that sensory feedback matches the internal predictions, which results in the partial compensation seen in the model output. On the other hand, compensations to jaw perturbations are complete, but task specific, such that a perturbation to the jaw causes the upper lip to lower when producing a bilabial stop (to achieve the task of bilabial closure) while the same perturbation during production of a coronal stop causes the tongue tip to raise (to achieve the task of closing the tongue tip against the palate). At the same time as the model is able to produce correct responses to sensory perturbations, the model is also able to produce relatively stable speech even in the presence of noisy sensory feedback and to produce speech when the sensory feedback is removed entirely by relying on the efference copy-based state prediction process, though the speech produced in the absence of sensory feedback is more variable than when sensory feedback is present.

Comparing Model Approaches

We have presented two approaches to modeling the role of sensory feedback in motor control and learning in the speech production system: DIVA and FACTS. In many areas, there is broad agreement across these and other models of speech motor control. From an algorithmic point of view, it is quite clear that speech motor control, like non-speech motor control, involves a combination of feedback and feedforward or predictive control. It is also clear that sensory feedback is used both for online feedback control and for long-term updating of the predictive or feedforward control system. Also, from an anatomical point of view, it is also quite evident which areas of the brain are active during the speech production process. The basic involvement of these areas is primarily motivated by the results of empirical

studies revealing what cortical and subcortical areas are active during speech production. So, any model of speech production attempting to explain this neural activity must posit roles for these areas.

While both models are thus broadly similar, the different architectures and approaches used in the two models create a number of differences, many of which make different predictions about various speech phenomena. Here, we present some of these differences to highlight where the models disagree and to point out areas where our understanding of the speech motor system can be improved. It should be noted that some of the discussion below regards differences between DIVA versus the SFC approach in general and not direct comparisons with the FACTS model. Given the long development period of DIVA compared to FACTS, it is unsurprising that many aspects of the DIVA model, such as learning, are not currently implemented in FACTS. We indicate these cases in the sections below where appropriate.

Sensorimotor Learning

One of the outstanding questions in speech motor control and, indeed, in motor control, in general, is how the feedforward or predictive control system is learned and updated. While there are multiple types of learning present in motor systems (e.g., reinforcement, use dependent, strategic/instruction; Haith & Krakauer, 2015; Krakauer, 2015), the most studied form of learning in speech motor control is sensorimotor learning. Sensorimotor learning (also known as *sensorimotor adaptation*) refers to the process of changing behavior to reduce errors in response to unexpected sensory feedback. This type of learning has been studied extensively in speech by measuring the behavioral responses to sustained perturbations of vowel formants (J. F. Houde & Jordan, 1998, 2002; Katseff, Houde, & Johnson, 2012; Purcell & Munhall, 2006a) or pitch (Jones & Munhall, 2000, 2005).

There are two basic theories of how the (speech) motor system adapts to sensory perturbation. In the first theory, previous behavior serves as a teaching signal for altering the feedforward control system. As one performs

a movement under sensory alterations, they will experience sensory errors and have the opportunity to correct for those errors. This corrective behavior can be seen in speech in the online compensatory response to unexpected auditory perturbations (Burnett et al., 1998; Purcell & Munhall, 2006b). In this theory, those corrective motor commands serve as a teaching signal and are added to the feedforward motor commands for future movements with some shifting in time (Kawato & Gomi, 1992). This is the way sensorimotor learning is included in DIVA, where feedforward commands are updated by adding previously issued commands from the feedback control pathways.

The second theory of sensorimotor learning hypothesizes that the predictive control system is updated directly from sensory errors and not from previously executed motor commands. In this theory, the difference between expected and perceived sensory signals (also known as the *sensory prediction error*) serve as a teaching signal for the internal forward model, which predicts the motor and sensory consequences of motor commands in an SFC system (Haith & Krakauer, 2013; Shadmehr, Smith, & Krakauer, 2010). By adjusting the output of the forward model, the system will start to produce behavior that will minimize the sensory error. While we have not yet implemented this type of learning in SFC/FACTS, the architecture of the models and the critical role of the forward model in predictive control is consistent with this theory of sensorimotor learning.

One key difference between these two theories of sensorimotor learning is whether corrective movements are necessary for learning or whether the experience of sensory prediction errors alone, even in the absence of corrective motor commands, is sufficient. While this issue has not been addressed in speech to date, a number of experiments in saccade adaptation (Noto & Robinson, 2001; Wallman & Fuchs, 1998) and reaching (Tseng, Diedrichsen, Krakauer, Shadmehr, & Bastian, 2007) have explored this question. In these studies, participants made movements and experienced sensory prediction errors under conditions that eliminated or greatly reduced the ability to make motor corrections for those errors. In the reaching tasks (Tseng et al., 2007), for example, this was done by having subjects make rapid reaching movements that limit the time available for corrective movements. Other studies provide feedback at only the end point of the reaching motion instead of continuously throughout the reach (Izawa & Shadmehr, 2011). These studies have consistently shown that adaptation is unaltered by the restricted ability to make corrective movements, suggesting that sensory prediction error alone is sufficient to drive sensorimotor adaptation, though there is some evidence that feedback motor commands may play an additional role in learning (Albert & Shadmehr, 2016). This evidence suggests that learning from feedback commands (as in DIVA, where feedback motor commands can update feedforward command whether or not they are actually executed) and from sensory errors (as in SFC models) may both play a role in sensorimotor learning in speech. However, adaption in the speech motor system remains to be thoroughly tested.

Incorporation of Plant Dynamics in Central Control

Because the vocal tract articulators have dynamic properties (e.g., they have mass), ultimately, forces must be generated to move the articulators. However, the degree to which the high-level motor control system incorporates the dynamic properties of the body into the generation of motor commands has been a matter of debate (Feldman & Levin, 2009; Loeb, 2012; Perrier, Ostry, et al., 1996; S. H. Scott, 2012). One possibility is that central control is essentially kinematic, with generation of muscle force arising at a lower level from the combination of descending commands and afferent proprioceptive signals from the muscles. Alternatively, the dynamics of the physical system may be incorporated into the central control mechanism itself.

On one hand, simulations have shown that a system that controls speech via setting the equilibrium point of muscles and relegates control of force to the peripheral nervous system is able to reproduce many of the kinematic patterns of speech (Perrier, Ostry, et al., 1996; Sanguineti et al., 1998). Moreover, short-latency stretch reflexes are consistent with a system that regulates the activation threshold of individual muscles (S. H. Scott, 2012). At first pass, this suggests that such a system may be a plausible candidate for human speech motor control.

However, there is evidence that the dynamic properties of the body play an additional role in high-level/central control. In both speech and nonspeech domains, subjects will learn to adjust their control of the limb and jaw to counter the effects of externally applied dynamic fields (Lametti et al., 2012; Nasir & Ostry, 2006; Shadmehr & Mussa-Ivaldi, 1994; Tremblay & Ostry, 2006; Tremblay et al., 2003). There is also evidence that subjects account for the inertia and dynamics of a task (e.g., inertia, gravity, elasticity) when executing that task in both reaching (D. Liu & Todorov, 2007; Sabes, Jordan, & Wolpert, 1998) and speech (Derrick, Stavness, & Gick, 2015; Hoole, 1998; R. A. Houde, 1968; Ostry, Gribble, & Gracco, 1996; Perrier, Payan, Zandipour, & Perkell, 2003; Shiller, Ostry, & Gribble, 1999; Shiller, Ostry, Gribble, & Laboisière, 2001). Lastly, while short-term stretch reflexes are consistent with an equilibrium point model of muscle control, long-term stretch reflexes clearly incorporate more complex dynamic control and are cortically generated (Gribble & Ostry, 1999; Kurtzer, Pruszynski, & Scott, 2008; Pruszynski et al., 2011).

As currently implemented, both the DIVA and FACTS models operate on a purely kinematic articulatory system, meaning that they operate without regard for the mass of the plant nor the actual forces that would be needed to move it. The outputs of both systems are in terms of kinematics (desired velocities in DIVA or accelerations in FACTS), not forces. Both models are, however, compatible with a lower level system, which could generate forces from these kinematic controls, such as an equilibrium point control model, where the high-level motor control system supplies changes in the equilibrium points of vocal tract muscles (Perrier, Ostry, et al., 1996). In fact, such a model has been implemented computationally in an earlier version of DIVA

that used time-invariant acoustic regions as targets rather than the current time-varying acoustic/somatosensory/motor trajectories (Zandipour et al., 2004).

In theory, both DIVA and FACTS could be modified to account for central control of dynamics. Given that we are not the developers of DIVA, we hesitate to speculate in detail on how such a model would work, but it would conceivably be accomplished by altering the feedforward controller to take into account the dynamics of the plant, as in the “*Gestures Shaped by the Physics and by a Perceptually Oriented Targets Optimization*” model (Perrier, Ma, & Payan, 2005). This may also entail changing the output of the feedforward and feedback controllers (potentially to muscle activations instead of articulatory velocities) and may also affect how feedforward control is learned. Incorporating dynamic control in FACTS is relatively straightforward in principle, as the SFC architecture on which this model is based is compatible with incorporation of plant dynamics into central control (S. H. Scott, 2012). FACTS could implement dynamic control in the articulatory feedback control law by (a) taking into account the dynamic properties of the plan and (b) outputting forces rather than (or in addition to) articulatory accelerations. More evidence is needed to resolve the importance of plant dynamics in central control for speech (e.g., the nature and source of short- and long-latency speech reflex responses), but it seems likely that the physical dynamics of the speech production apparatus are incorporated into neural control both peripherally and centrally. Importantly, nonspeech motor control studies have shown the descending outputs of neurons in primary motor cortex appear to represent a range of abstractions of motor tasks from desired movement directions, to desired joint angle changes, to desired changes in individual muscle activations (Takei, Hoffman, & Strick, 1999, 2003). Such a complex control signal is likely also present in the speech system, and both DIVA and FACTS will need to be modified to incorporate such control.³

Sources of Sensory Predictions

The process of matching incoming sensory feedback with a prediction of that feedback is central to the state estimation process that lies at the heart of the SFC and FACTS models. This matching is also critical to the function of the sensory feedback controllers in the DIVA model. However, much remains unknown about how this process works and the nature of feedback predictions. One key issue to resolve is to determine what the sources are within the CNS that are used to make these sensory predictions.

The idea that the efference copy of motor actions could be a source of information to generate sensory predictions has been around for quite some time. The idea was originally proposed to explain how reafference of self-generated

sensations can be filtered out of the sensory stream to reveal only externally generated sensations. To accomplish such sensory filtering, it was hypothesized that internal feedback of the neural commands generating motor actions might also generate expectations of their sensory consequences (Jeannerod, 1988; von Holst, 1954; von Holst & Mittelstaedt, 1950). These expectations would be “subtracted” from the incoming sensory information, thus highlighting any remaining unpredicted sensations, which were assumed to arise from external sources. In this way, efference copy was hypothesized to enable the CNS to distinguish “self” from “nonself” sources of sensory input.

A more general version of this idea can be seen in the later-developed concept of predictive coding as an explanation for how the CNS processes incoming sensory information (Rao & Ballard, 1999). In the predictive coding model, the main job of sensory information arriving from the periphery (bottom-up information) is to confirm or correct predictions of that sensory input that have been made from inferences about context (both spatial and temporal) and from information from other sensory and even motor sources in the CNS (top-down) expectations.

The model was first proposed to explain visual processing but has been generalized to explain other sensory processing and even cognitive processing. Examples in speech perception that can be explained by the predictive coding model include the phonetic restoration effect, where listeners perceive an illusory speech sound that matches phonetic context at point in running speech where the sound has been replaced by a noise burst (Warren, 1970). Another well-known example that could be explained by predictive coding is the McGurk effect, where listeners hear the audio of the production of one speech sound and simultaneously see the video of the face of a speaker producing another speech sound and perceive a speech sound that is intermediate between the two (McGurk & MacDonald, 1976).

Current models of speech production take the concept of predictive coding a step further, suggesting that not only is incoming sensory feedback processed by comparison with predictions but also that the result of the comparison—the sensory prediction error—is used to correct ongoing speech motor output. How exactly this motor output correction process proceeds, however, differs between models, as has been discussed above, and these differences have consequences for what the models postulate to be the sources of the sensory predictions used in the speech motor control process. In the DIVA model, the sensory prediction error is used as a direct correction of the current motor output, just as in a pure feedback control system. As a result, in DIVA, the source of the sensory prediction is the currently intended sensory target (or sensory goal) associated with the desired speech utterance. In SFC, the sensory prediction error instead updates the estimate of the current articulatory state. As a result, in SFC, the sensory prediction is derived from the current articulatory state estimate. Early versions of DIVA also compared sensory feedback with sensory predictions generated from an efference copy of outgoing motor commands (Guenther, 1995b; Guenther, Hampson, &

³Some other models, especially the “*Gestures Shaped by the Physics and by a Perceptually Oriented Targets Optimization*” model (Patri, Diard, & Perrier, 2015; Perrier et al., 2005), do incorporate dynamics in central control.

Johnson, 1998), though the details of how efference copy was used were somewhat different from their use in SFC. However, the use of efference copy has been replaced in the current version of the model with comparison with intended sensory targets.

Given the difference in how sensory predictions are postulated to originate in different models of speaking and the ramifications of this on how sensory prediction errors are used, it is surprising how little research has been done addressing the nature of the sensory predictions and their possible sources. Two studies by Niziolek et al. are relevant to this issue, though they do not fully resolve the question. One study looked at the effect of speakers' vowel category boundaries on how much they compensated within trials for unpredictable perturbations of the auditory feedback of their vowel formants (Niziolek & Guenther, 2013). It was found that formant perturbations that caused the vowel in a particular utterance to cross or approach a speaker-specific category boundary into a neighboring vowel space generated greater compensatory responses than did perturbations that kept speakers' productions within the same vowel category. This result suggests that the auditory feedback response increases when the perturbation crosses some perceptual boundary, consistent with the DIVA hypothesis. However, compensation was still seen even for within-category perturbations (even when these perturbations moved the production closer to the center of the vowel's distribution), suggesting that auditory predictions have some precision above the categorical level, consistent with predictions of the SFC model. Another study looked at the limits of the resolution of the sources of auditory feedback predictions during speaking (Niziolek et al., 2013). As discussed above, existence of the SIS effect supports the idea that incoming auditory feedback is compared with a prediction of that feedback. The study by Niziolek et al. showed that the degree of SIS varies as a function of nearness to a vowel's average formant values, suggesting that not all the variability in speech output is anticipated in the feedback prediction. However, it is unclear if this variability is not predicted because the auditory predictions come from learned associations or goals (as in DIVA) or because it arises from unpredictable variability in motor execution due to low-level motor noise (as in SFC/FACTS).

One possible way to separate the predictions of the two models is through mechanical, rather than auditory, perturbations. In particular, DIVA predicts that the nature of the current speech task will affect neural responses to somatosensory feedback perturbations, while SFC predicts it will not. In DIVA, somatosensory targets are implemented as tolerable ranges for the incoming sensory feedback feature to take on without generating a sensory mismatch error signal. So, for example, if the current task is to bring the lips together to make a bilabial stop consonant, then the somatosensory target would be represented in high-level terms as a target lip opening of zero, and any perturbations of the upper or lower lip would create deviations from this target lip opening. Such somatosensory perturbations would therefore generate large somatosensory mismatch

responses. On the other hand, if the current task was to produce a /t/, the critical constriction is between the tongue tip and palate, with the lips largely unconstrained. In this case, the tolerable range for the lower lip would be quite large, and so any mechanical perturbations of it would therefore generate no (or small) somatosensory mismatch responses. Thus, DIVA predicts that the neural somatosensory response to an upper lip perturbation would vary depending on the speech task. In SFC, however, both the speech target specification and the task-dependent flexibility are implemented in the state feedback control law in frontal/motor cortex and are independent from the articulatory state estimation process distributed between frontal cortex and the sensory cortices, with sensory feedback predictions made solely from the current articulatory state prediction. Crucially, this state prediction is not dependent on the current speech task goal. So, in the above described example, SFC predicts that a perturbation of the lower lip will generate a feedback prediction error reflected in the response of somatosensory cortex, but that response will be largely the same regardless of whether the current speech task is to produce a bilabial stop or coronal stop. Thus, SFC predicts that the neural somatosensory response to an upper lip perturbation would not vary across speech tasks. We are currently working toward implementing such a test in our labs.

An alternative test would be to examine the response of human speakers to very small auditory or somatosensory perturbations. The use of sensory target regions in DIVA makes the prediction that compensation for externally imposed alterations of sensory feedback should only occur for perturbations that push the feedback outside the target range, though the precise size of the target is hypothesized to vary between speakers (Villacorta, Perkell, & Guenther, 2007). SFC/FACTS, however, predicts that even very small errors will lead to behavioral changes by altering the estimated system state.

Neural Substrates of Sensorimotor Control

The areas of the CNS that are seen to be active during speech production are many and vary depending on the exact nature of the speaking task being considered (Skipper et al., 2017). Full discussion of this complicated pattern of activity is beyond the scope of this review article and has yet to be fully accounted for by any model of speech motor control. Here, we focus on what the DIVA and SFC models say about the neural substrate of the sensorimotor interactions that appear to take place during speaking. While the hypothesized neural basis of the DIVA model is well established, with well-defined links between computational components and neural regions and implemented neural simulations, the hypotheses of the SFC approach are, at this point, mainly speculative.

Several areas' roles in the speech production process are largely agreed upon across speech motor control models. Primary motor cortex must ultimately be the source of descending influence on lower motor systems to actually move the vocal tract articulators to make speech. Stimulation

of the speech areas of primary motor cortex can cause movements of the vocal tract articulators and vocalizations (Penfield & Boldrey, 1937), and lesions of these areas cause mutism (Jürgens, 2002). Furthermore, earlier in this review article, we discussed results that localized the sites of sensory feedback processing during speaking, with the area of the supramarginal gyrus/inferior parietal lobe for somatosensory feedback processing and the area of the SPT/pSTG for auditory feedback processing. Also as discussed above, both the DIVA and SFC/FACTS models propose that, in these areas, sensory feedback is processed by comparing it with a prediction of that feedback, resulting in sensory prediction errors if there are mismatches. However, the two models differ considerably in their assumptions about what processes give rise to the feedback predictions and for what processes the sensory prediction errors serve as input.

Sensorimotor Interactions in DIVA

In the DIVA model, the left ventral premotor cortex (LvPMC) is a feedforward source of time-varying speech sound target specifications output to primary motor cortex and higher level somatosensory and auditory cortices. In response to this target input, these receiving areas then have their own internal dynamics: Primary motor cortex plays out the desired articulatory trajectory, while the somatosensory and auditory areas play out the trajectories/time courses of the somatosensory and auditory feedback predictions expected from the production of the speech sound target. There are no pathways back to the LvPMC during the production of a speech sound. Primary motor cortex has its own recurrent connections for generating the desired articulatory trajectory that the lower motor system should follow, and if incoming sensory feedback matches expectations, there is no information sent back to it from the sensory areas. If sensory feedback mismatches are detected, this mismatch information propagates from the sensory areas through the right ventral premotor cortex (RvPMC) to primary motor cortex. In DIVA, the RvPMC is proposed to be the site of the “feedback control mapping”—the mapping from sensory prediction/target errors to the associated articulatory trajectory corrections. This role was postulated for the RvPMC based on studies suggesting it plays a central role in responding to feedback perturbations, something that has been seen in studies of responses to formant feedback perturbations (Tourville et al., 2008), pitch feedback perturbations (Kort et al., 2016, 2014), and somatosensory feedback perturbations (Golfinopoulos et al., 2011).

Sensorimotor Interactions in SFC/FACTS

The sensorimotor interactions in DIVA are based on the feedback control concept: The difference between sensory target (feedback predictions) and sensory feedback drives corrections to the ongoing motor output. In SFC models, sensorimotor interactions are based on the state estimation concept: The estimate of the current articulatory state gives rise to feedback predictions, and if these differ from actual feedback, the difference drives corrections to the current state estimate. Thus, in SFC models, all sensorimotor

interactions are encapsulated in the state estimation process. The actual process of generating motor controls is still based on the feedback control concept, but because the feedback for this process (the feedback control law) is the internally maintained state estimate, the control process is isolated from any sensorimotor interactions. Thus, when considering the neural substrate of sensorimotor interactions in SFC models, the key question is: Where is the state estimate maintained? This maintenance process involves predicting the next state and integrating feedback prediction errors to update the prediction into the current state estimate.

There are a number of candidate areas in the CNS that could be the neural substrates of these processes. Many studies suggest that the cerebellum is crucial for making predictions of the outcomes of motor actions (Herzfeld & Shadmehr, 2014; Wolpert & Miall, 1996). Parietal cortex is also well situated to integrate different sources of sensory feedback information. Damage to parietal cortex and cerebellum is associated with an impaired ability to update motor output in response to sensory feedback perturbations in speech and other motor domains (Parrell et al., 2017; Shadmehr & Krakauer, 2008; Shum et al., 2011). However, especially for speech, the premotor cortices are also well situated for integrating sensory information with motor efference copy. The arcuate fasciculus and the superior longitudinal fasciculus link the PMv to the high-level auditory and somatosensory processing areas that, as discussed above, appear to be involved in sensory feedback processing in speech (Makris et al., 2005; Saur et al., 2008). The PMv is also well placed to receive efference copy from and provide state estimates to primary motor cortex. As discussed above, the RvPMC has been associated with responses to speech auditory feedback perturbations.

As our working hypothesis, we currently postulate that the LvPMC is where the articulatory state estimate is maintained during running speech. The LvPMC is the source of a dynamically changing state estimate that is continuously passed to the primary motor, somatosensory, and auditory cortices as production of the currently desired speech sound progresses. Primary motor cortex implements a state feedback control law that outputs the next articulatory controls based on the desired next change in task state and the currently estimated articulatory state. It also sends efference copy of its output controls back to the LvPMC. The LvPMC, through interaction with the cerebellum, uses this efference copy and the current articulatory state to predict the next articulatory state. As with DIVA, if incoming sensory feedback matches expectations, there is no information sent back from the sensory areas. However, if sensory feedback mismatches are detected, this mismatch information propagates from the sensory areas through the RvPMC to the LvPMC, where it is used to correct the current articulatory state estimate.

Conclusion

In this review article, we have presented a summary of the role of sensory information in the speech motor system.

Drawing on evidence from speech perturbation experiments, development, speech disorders, and neuroimaging, we have shown that the sensory system is integrally linked to the speech motor system. Both auditory and somatosensory information is monitored and used in speech production. At the same time, this sensory information (at least for the auditory system) does not seem to be critical for maintaining relatively accurate supralaryngeal speech function, precluding the idea that the speech motor system is a sensory servo controller. Moreover, the latencies and noisiness of the neural sensory system are nontrivial and would render a pure feedback control system unstable.

We have also reviewed the role of sensory feedback in learning in the speech motor system. While we have focused our discussion primarily on the adaptability of the mature speech system, rather than in speech development, the sensory system presumably plays a role in development as well (Guenther, 2016). We have shown how the speech motor system will adapt to external perturbation of both auditory and somatosensory feedback, consistent with the idea that the production system uses sensory feedback to modulate internal predictions about the motor and sensory outcomes of speech actions.

We have presented two computational models of the speech motor system that have sought to explain the somewhat contradictory notions that sensory feedback is highly relevant to speech motor control but, at the same time, does not form a critical component of that process. We showed how the DIVA model handles this dissociation by separating feedforward control, which proceeds invariantly regardless of sensory feedback and so can produce accurate speech even when feedback is removed, from separate auditory and somatosensory feedback pathways. We have also presented an alternative view of motor control based on the concept of SFC, where internal predictions about the outcomes of motor actions are corrected by the sensory system and fed back to a single feedback control law. We have reviewed the broad applicability of the SFC approach to nonspeech motor domains and presented recent work from our labs on implementing this approach in the speech motor system in the SFC and FACTS models. We showed how these models can replicate the role of sensory feedback in speech, being sensitive to sensory feedback (and particularly to sensory prediction errors) when it is available but being able to operate in absence due to the model-predictive component. Importantly, these models are only examples of two general approaches to speech motor control, and other models of both types (Kröger, 2013; Tian & Poeppel, 2010) and models that do not address the role of sensory feedback in central control (Patri, Perrier, & Diard, 2016; Perrier et al., 2005; Sanguineti et al., 1998) may make somewhat different predictions in some of these areas.

Lastly, we have reviewed areas of agreement between SFC/FACTS and DIVA, as well as areas where the predictions of models may differ. These models (as well as other speech motor control models) agree that speech production relies on both feedback and feedforward or predictive control, that sensory feedback is used at multiple timescales,

and that a certain subset of neural regions are involved in speech production. However, these models make different predictions in a number of different areas: sensorimotor learning, the origin of sensory predictions, and some aspects of the neural substrates of the speech motor system.

It is highly likely that no model, in its current form, accurately reflects the complex structure of the speech motor control system. In some cases, it is possible that both models reflect some true component of the system. For example, it is possible that both feedforward and predictive control systems exist, rather than only one or the other. This seems particularly likely regarding the source of dynamic control, where evidence suggests both central and peripheral control systems reflect the physical dynamics of the body. Our goal in identifying differences between current models has not been to suggest the superiority of one model over the other but to suggest fruitful avenues for future research that can resolve these different predictions and bring us closer to understanding the complex human speech motor system.

Acknowledgments

This work was supported by National Institutes of Health Grants R01 DC017696 (awarded to John F. Houde), U01 NS098971 (awarded to Edward F. Chang), R01 DC013979 (awarded to Srikantan S. Nagarajan), R01 DC017091 (awarded to John F. Houde), and F32 DC014211 (awarded to Benjamin Parrell).

References

- Abbs, J. H., & Gracco, V. L. (1984). Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech. *Journal of Neurophysiology*, *51*(4), 705–723.
- Albert, S. T., & Shadmehr, R. (2016). The neural feedback response to error as a teaching signal for the motor learning system. *The Journal of Neuroscience*, *36*(17), 4832–4845. <https://doi.org/10.1523/JNEUROSCI.0159-16.2016>
- Arbib, M. A. (1981). Perceptual structures and distributed motor control. In J. M. Brookhart, V. B. Mountcastle, & V. Brooks (Eds.), *Handbook of physiology, Supplement 2: Handbook of physiology, the nervous system, motor control*. Hoboken, NJ: Wiley. <https://doi.org/10.1002/cphy.cp010233>
- Baum, S. R., & McFarland, D. H. (1997). The development of speech adaptation to an artificial palate. *The Journal of the Acoustical Society of America*, *102*(4), 2353–2359.
- Baum, S. R., & McFarland, D. H. (2000). Individual differences in speech adaptation to an artificial palate. *The Journal of the Acoustical Society of America*, *107*(6), 3572–3575.
- Behroozmand, R., Karvelis, L., Liu, H., & Larson, C. R. (2009). Vocalization-induced enhancement of the auditory cortex responsiveness during voice f0 feedback perturbation. *Clinical Neurophysiology*, *120*(7), 1303–1312. <https://doi.org/10.1016/j.clinph.2009.04.022>
- Birkholz, P., Kroger, B. J., & Neuschaefer-Rube, C. (2011). Model-based reproduction of articulatory trajectories for consonant–vowel sequences. *IEEE Transactions on Audio, Speech, and Language Processing*, *19*(5), 1422–1433.
- Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural representations and mechanisms for the performance of simple

- speech sequences. *Journal of Cognitive Neuroscience*, 22(7), 1504–1529. <https://doi.org/10.1162/jocn.2009.21306>
- Browman, C. P., & Goldstein, L.** (1992). Articulatory phonology: An overview. *Phonetica*, 49(3–4), 155–180.
- Browman, C. P., & Goldstein, L.** (1995). Dynamics and articulatory phonology. In R. Port & T. van Gelder (Eds.), *Mind as motion: Dynamics, behavior, and cognition* (pp. 175–194). Cambridge, MA: MIT Press.
- Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C.** (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America*, 103(6), 3153–3161.
- Cai, S., Beal, D. S., Ghosh, S. S., Tiede, M. K., Guenther, F. H., & Perkell, J. S.** (2012). Weak responses to auditory feedback perturbation during articulation in persons who stutter: Evidence for abnormal auditory–motor transformation. *PLOS ONE*, 7(7), e41830. <https://doi.org/10.1371/journal.pone.0041830>
- Casserly, E. D., Rowley, M. C., Marino, F. R., & Pollack, E. A.** (2016). Acoustic analysis of speech produced with degradation of acoustic and somatosensory feedback. *Proceedings of Meetings on Acoustics*, 29(1), 060016. <https://doi.org/10.1121/2.0000651>
- Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F.** (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences*, 110(7), 2653–2658. <https://doi.org/10.1073/pnas.1216827110>
- Chang-Yit, R., Pick, H. L., & Siegel, G. M.** (1975). Reliability of sidetone amplification effect in vocal intensity. *Journal of Communication Disorders*, 8(4), 317–324.
- Cluff, T., & Scott, S. H.** (2015). Apparent and actual trajectory control depend on the behavioral context in upper limb motor tasks. *The Journal of Neuroscience*, 35(36), 12465–12476. <https://doi.org/10.1523/JNEUROSCI.0902-15.2015>
- Cowie, R. I. D., & Douglas-Cowie, E.** (1983). 6-Speech production in profound postlingual deafness. In M. E. Lutman & M. P. Haggard (Eds.), *Hearing science and hearing disorders* (pp. 183–230). London, United Kingdom: Academic Press. <https://doi.org/10.1016/B978-0-12-460440-7.50011-9>
- Cowie, R. I. D., & Douglas-Cowie, E.** (1992). *Postlingually acquired deafness: Speech deterioration and the wider consequences*. Berlin, NY: Mouton de Gruyter.
- Curio, G., Neuloh, G., Numminen, J., Jousmäki, V., & Hari, R.** (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Human Brain Mapping*, 9(4), 183–191.
- Derrick, D., Stavness, I., & Gick, B.** (2015). Three speech sounds, one motor action: Evidence for speech-motor disparity from English flap production. *The Journal of the Acoustical Society of America*, 137(3), 1493–1502.
- Desmurget, M., & Grafton, S.** (2003). Feedback or feedforward control: End of a dichotomy. In S. H. Johnson-Frey (Ed.), *Taking action: Cognitive neuroscience perspectives on intentional acts* (pp. 289–338). Cambridge, MA: MIT Press.
- Diedrichsen, J., Shadmehr, R., & Ivry, R. B.** (2010). The coordination of movement: Optimal feedback control and beyond. *Trends in Cognitive Sciences*, 14(1), 31–39. <https://doi.org/10.1016/j.tics.2009.11.004>
- Eliades, S. J., & Wang, X.** (2003). Sensory–motor interaction in the primate auditory cortex during self-initiated vocalizations. *Journal of Neurophysiology*, 89(4), 2194–2207. <https://doi.org/10.1152/jn.00627.2002>
- Eliades, S. J., & Wang, X.** (2005). Dynamics of auditory–vocal interaction in monkey auditory cortex. *Cerebral Cortex*, 15(10), 1510–1523. <https://doi.org/10.1093/cercor/bhi030>
- Eliades, S. J., & Wang, X.** (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature*, 453(7198), 1102–1106. <https://doi.org/10.1038/nature06910>
- Elman, J. L.** (1981). Effects of frequency-shifted feedback on the pitch of vocal productions. *The Journal of the Acoustical Society of America*, 70(1), 45–50. <https://doi.org/10.1121/1.386580>
- Fairbanks, G.** (1954). Systematic research in experimental phonetics. 1. A theory of the speech mechanism as a servosystem. *Journal of Speech and Hearing Disorders*, 19(2), 133–139. <https://doi.org/10.1044/jshd.1902.133>
- Feldman, A. G., & Levin, M. F.** (2009). The equilibrium-point hypothesis—Past, present and future. In D. Sternad (Ed.), *Progress in motor control* (pp. 699–726). Boston, MA: Springer. https://doi.org/10.1007/978-0-387-77064-2_38
- Feng, Y., Gracco, V. L., & Max, L.** (2011). Integration of auditory and somatosensory error signals in the neural control of speech movements. *Journal of Neurophysiology*, 106(2), 667–679. <https://doi.org/10.1152/jn.00638.2010>
- Fitzgerald, M. J., & Sachithanandan, S. R.** (1979). The structure and source of lingual proprioceptors in the monkey. *Journal of Anatomy*, 128(Pt. 3), 523–552.
- Ford, J. M., & Mathalon, D. H.** (2004). Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking. *Journal of Psychiatric Research*, 38(1), 37–46. [https://doi.org/10.1016/S0022-3956\(03\)00095-5](https://doi.org/10.1016/S0022-3956(03)00095-5)
- Ford, J. M., Mathalon, D. H., Heinks, T., Kalba, S., Faustman, W. O., & Roth, W. T.** (2001). Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. *The American Journal of Psychiatry*, 158(12), 2069–2071. <https://doi.org/10.1176/appi.ajp.158.12.2069>
- Fowler, C. A., & Turvey, M. T.** (1981). Immediate compensation in bite-block speech. *Phonetica*, 37(5–6), 306–326. <https://doi.org/10.1159/000260000>
- Franklin, D. W., & Wolpert, D. M.** (2011). Computational mechanisms of sensorimotor control. *Neuron*, 72(3), 425–442. <https://doi.org/10.1016/j.neuron.2011.10.006>
- Friederici, A. D.** (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*, 91(4), 1357–1392. <https://doi.org/10.1152/physrev.00006.2011>
- Fu, C. H. Y., Vythelingum, G. N., Brammer, M. J., Williams, S. C. R., Amaro, E., Andrew, C. M., ... McGuire, P. K.** (2006). An fMRI study of verbal self-monitoring: Neural correlates of auditory verbal feedback. *Cerebral Cortex*, 16(7), 969–977. <https://doi.org/10.1093/cercor/bhj039>
- Geschwind, N.** (1965). Disconnexion syndromes in animals and man. II. *Brain: A Journal of Neurology*, 88(3), 585–644.
- Glasser, M. F., & Rilling, J. K.** (2008). DTI tractography of the human brain’s language pathways. *Cerebral Cortex*, 18(11), 2471–2482. <https://doi.org/10.1093/cercor/bhn011>
- Golfinopoulos, E., Tourville, J. A., Bohland, J. W., Ghosh, S. S., Nieto-Castanon, A., & Guenther, F. H.** (2011). fMRI investigation of unexpected somatosensory feedback perturbation during speech. *NeuroImage*, 55(3), 1324–1338. <https://doi.org/10.1016/j.neuroimage.2010.12.065>
- Gottlieb, G. L., & Agarwal, G. C.** (1979). Response to sudden torques about ankle in man: Myotatic reflex. *Journal of Neurophysiology*, 42(1 Pt. 1), 91–106. <https://doi.org/10.1152/jn.1979.42.1.91>
- Gracco, V. L., & Abbs, H. J.** (1985). Dynamic control of the perioral system during speech: Kinematic analyses of autogenic and nonautogenic sensorimotor processes. *Journal of Neurophysiology*, 54, 418–432.

- Gracco, V. L., & Abbs, J. H. (1986). Variant and invariant characteristics of speech movements. *Experimental Brain Research*, 65(1), 156–166.
- Greenlee, J. D. W., Jackson, A. W., Chen, F., Larson, C. R., Oya, H., Kawasaki, H., ... Iii, M. A. H. (2011). Human auditory cortical activation during self-vocalization. *PLOS ONE*, 6(3), e14744. <https://doi.org/10.1371/journal.pone.0014744>
- Gribble, P. L., & Ostry, D. J. (1999). Compensation for interaction torques during single- and multijoint limb movement. *Journal of Neurophysiology*, 82(5), 2310–2326. <https://doi.org/10.1152/jn.1999.82.5.2310>
- Guenther, F. H. (1995a). A modeling framework for speech motor development and kinematic articulator control. In K. Elenius & P. Branderud (Eds.), *Proceedings of the 13th International Congress of Phonetic Sciences* (Vol. 2, pp. 92–99). Stockholm, Sweden: Arne Strömbergs Grafiska.
- Guenther, F. H. (1995b). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, 102, 594–621.
- Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, 39(5), 350–365. <https://doi.org/10.1016/j.jcomdis.2006.06.013>
- Guenther, F. H. (2016). *Neural control of speech*. Cambridge, MA: MIT Press.
- Guenther, F. H., Espy-Wilson, C., Boyce, S., Matthies, M., Zandipour, M., & Perkell, S. J. (1999). Articulatory trade-offs reduce acoustic variability during American English /r/ production. *The Journal of the Acoustical Society of America*, 105, 2854–2865.
- Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, 105, 611–633.
- Haar, S., & Donchin, O. (2019, February 2). A revised computational neuroanatomy for motor control. <https://doi.org/10.31219/osf.io/t6prj>
- Haith, A. M., & Krakauer, J. W. (2013). Model-based and model-free mechanisms of human motor learning. *Advances in Experimental Medicine and Biology*, 782, 1–21. https://doi.org/10.1007/978-1-4614-5465-6_1
- Haith, A. M., & Krakauer, J. W. (2015). Theoretical models of motor control and motor learning. In A. Gollhofer, W. Taube, & J. B. Nielsen (Eds.), *Routledge handbook of motor control and motor learning* (pp. 7–28). London, United Kingdom: Routledge.
- Hashimoto, Y., & Sakai, K. L. (2003). Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: An fMRI study. *Human Brain Mapping*, 20(1), 22–28. <https://doi.org/10.1002/hbm.10119>
- Heinks-Maldonado, T. H., Nagarajan, S. S., & Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. *NeuroReport*, 17(13), 1375–1379. <https://doi.org/10.1097/01.wnr.0000233102.43526.e9>
- Herzfeld, D. J., & Shadmehr, R. (2014). Cerebellum estimates the sensory state of the body. *Trends in Cognitive Sciences*, 18(2), 66–67. <https://doi.org/10.1016/j.tics.2013.10.015>
- Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language and Cognitive Processes*, 29(1), 2–20. <https://doi.org/10.1080/01690965.2013.834370>
- Hickok, G., Houde, J. F., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, 69(3), 407–422. <https://doi.org/10.1016/j.neuron.2011.01.019>
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402. <https://doi.org/10.1038/nrn2113>
- Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., ... Konishi, J. (1996). Cortical speech processing mechanisms while vocalizing visually presented languages. *NeuroReport*, 8(1), 363–367.
- Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., ... Konishi, J. (1997). Cortical processing mechanism for vocalization with auditory verbal feedback. *NeuroReport*, 8(9–10), 2379–2382.
- Hirano, S., Naito, Y., Okazawa, H., Kojima, H., Honjo, I., Ishizu, K., ... Konishi, J. (1997). Cortical activation by monaural speech sound stimulation demonstrated by positron emission tomography. *Experimental Brain Research*, 113(1), 75–80.
- Hollerbach, J. M. (1982). Computers, brains and the control of movement. *Trends in Neurosciences*, 5, 189–192. [https://doi.org/10.1016/0166-2236\(82\)90111-4](https://doi.org/10.1016/0166-2236(82)90111-4)
- Hoole, P. (1987). Bite-block speech in the absence of oral sensibility. *Proceedings of the 11th International Congress on Phonetic Sciences*, 4, 16–19.
- Hoole, P. (1998). Do airstream mechanisms influence tongue movement paths. *Phonetica*, 55(3), 131–146. <https://doi.org/10.1159/000028428>
- Houde, J. F., & Chang, E. F. (2015). The cortical computations underlying feedback control in vocal production. *Current Opinion in Neurobiology*, 33, 174–181. <https://doi.org/10.1016/j.conb.2015.04.006>
- Houde, J. F., & Jordan, I. M. (1998). Sensorimotor adaptation in speech production. *Science*, 279, 1213–1216.
- Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research*, 45(2), 295–310.
- Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, 5(82). <https://doi.org/10.3389/fnhum.2011.00082>
- Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience*, 14(8), 1125–1138. <https://doi.org/10.1162/089892902760807140>
- Houde, J. F., Niziolek, C., Kort, N., Agnew, Z., & Nagarajan, S. S. (2014). Simulating a state feedback model of speaking. *Proceedings of the 10th International Seminar on Speech Production, 2014*, 202–205.
- Houde, R. A. (1968). *A study of tongue body motion during selected speech sounds* (No. SCRL-Monograph-2). Santa Barbara, CA: Speech Communications Research Lab. Retrieved from <http://www.dtic.mil/docs/citations/AD0682966>
- Huang, C., Liu, G., Yang, B.-Y., Mu, H., & Hsia, C.-F. (1991). Auditory receptive area in the cerebellar hemisphere is surrounded by somatosensory areas. *Brain Research*, 541(2), 252–256. [https://doi.org/10.1016/0006-8993\(91\)91025-V](https://doi.org/10.1016/0006-8993(91)91025-V)
- Ito, T., Kimura, T., & Gomi, H. (2005). The motor cortex is involved in reflexive compensatory adjustment of speech articulation. *NeuroReport*, 16(16), 1791–1794.
- Ito, T., & Ostry, D. J. (2010). Somatosensory contribution to motor learning due to facial skin deformation. *Journal of Neurophysiology*, 104(3), 1230–1238. <https://doi.org/10.1152/jn.00199.2010>
- Izawa, J., & Shadmehr, R. (2011). Learning from sensory and reward prediction errors during motor adaptation. *PLoS Computational Biology*, 7(3), e1002012. <https://doi.org/10.1371/journal.pcbi.1002012>

- Jacobs, O. L. R.** (1993). *Introduction to control theory*. Oxford, NY: Oxford University Press.
- Jeannerod, M.** (1988). *The neural and behavioural organization of goal-directed movements*. Oxford, United Kingdom: Oxford University Press.
- Jones, J. A., & Munhall, K. G.** (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, 108(3), 1246–1251.
- Jones, J. A., & Munhall, K. G.** (2003). Learning to produce speech with an altered vocal tract: The role of auditory feedback. *The Journal of the Acoustical Society of America*, 113(1), 532–543.
- Jones, J. A., & Munhall, K. G.** (2005). Remapping auditory–motor representations in voice production. *Current Biology*, 15(19), 1768–1772. <https://doi.org/10.1016/j.cub.2005.08.063>
- Jürgens, U.** (2002). Neural pathways underlying vocal control. *Neuroscience & Biobehavioral Reviews*, 26(2), 235–258. [https://doi.org/10.1016/S0149-7634\(01\)00068-9](https://doi.org/10.1016/S0149-7634(01)00068-9)
- Kakei, S., Hoffman, D. S., & Strick, P. L.** (1999). Muscle and movement representations in the primary motor cortex. *Science*, 285(5436), 2136–2139.
- Kakei, S., Hoffman, D. S., & Strick, P. L.** (2003). Sensorimotor transformations in cortical motor areas. *Neuroscience Research*, 46(1), 1–10. [https://doi.org/10.1016/S0168-0102\(03\)00031-2](https://doi.org/10.1016/S0168-0102(03)00031-2)
- Kalveram, K. T., & Jäncke, L.** (1989). Vowel duration and voice onset time for stressed and nonstressed syllables in stutterers under delayed auditory feedback condition. *Folia Phoniatrica et Logopaedica*, 41(1), 30–42. <https://doi.org/10.1159/000265930>
- Katseff, S., Houde, J. F., & Johnson, K.** (2012). Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback? *Language and Speech*, 55(2), 295–308. <https://doi.org/10.1177/0023830911417802>
- Kawato, M., Furukawa, K., & Suzuki, R.** (1987). A hierarchical neural-network model for control and learning of voluntary movement. *Biological Cybernetics*, 57(3), 169–185. <https://doi.org/10.1007/BF00364149>
- Kawato, M., & Gomi, H.** (1992). A computational model of four regions of the cerebellum based on feedback-error learning. *Biological Cybernetics*, 68(2), 95–103. <https://doi.org/10.1007/BF00201431>
- Kelso, J. A. S., & Holt, K. G.** (1980). Exploring a vibratory systems analysis of human movement production. *Journal of Neurophysiology*, 43(5), 1183–1196. <https://doi.org/10.1152/jn.1980.43.5.1183>
- Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A.** (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, 10(6), 812–832.
- Knapp, H. D., Taub, E., & Berman, A. J.** (1963). Movements in monkeys with deafferented forelimbs. *Experimental Neurology*, 7(4), 305–315. [https://doi.org/10.1016/0014-4886\(63\)90077-3](https://doi.org/10.1016/0014-4886(63)90077-3)
- Kort, N. S., Cuesta, P., Houde, J. F., & Nagarajan, S. S.** (2016). Bihemispheric network dynamics coordinating vocal feedback control. *Human Brain Mapping*, 37(4), 1474–1485. <https://doi.org/10.1002/hbm.23114>
- Kort, N. S., Nagarajan, S. S., & Houde, J. F.** (2014). A bilateral cortical network responds to pitch perturbations in speech feedback. *NeuroImage*, 86, 525–535. <https://doi.org/10.1016/j.neuroimage.2013.09.042>
- Krakauer, J. W.** (2015). The applicability of motor learning to neurorehabilitation. In V. Dietz & N. Ward (Eds.), *Oxford textbook of neurorehabilitation* (pp. 55–64). Oxford, United Kingdom: Oxford University Press. <https://doi.org/10.1093/med/9780199673711.003.0007>
- Kröger, B. J.** (2013). Modeling speech production from the perspective of neuroscience. In D. Mehnert, D. Kordon, & M. Wolff (Eds.), *Systemtheorie, signalverarbeitung und sprachtechnologie. Rüdiger Hoffmann zum 65. Geburtstag* (pp. 218–225). Dresden, Germany: TUDpress.
- Kröger, B. J., Kannampuzha, J., & Neuschaefer-Rube, C.** (2009). Towards a neurocomputational model of speech production and perception. *Speech Communication*, 51(9), 793–809. <https://doi.org/10.1016/j.specom.2008.08.002>
- Kurtzer, I. L., Pruszynski, J. A., & Scott, S. H.** (2008). Long-latency reflexes of the human arm reflect an internal model of limb dynamics. *Current Biology*, 18(6), 449–453. <https://doi.org/10.1016/j.cub.2008.02.053>
- Ladefoged, P.** (1967). *Three areas of experimental phonetics*. Oxford, United Kingdom: Oxford University Press.
- Lametti, D. R., Nasir, S. M., & Ostry, D. J.** (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *The Journal of Neuroscience*, 32(27), 9351–9358. <https://doi.org/10.1523/JNEUROSCI.0404-12.2012>
- Lane, H., & Tranel, B.** (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, 14(4), 677–709. <https://doi.org/10.1044/jshr.1404.677>
- Lane, H., & Webster, J. W.** (1991). Speech deterioration in postlingually deafened adults. *The Journal of the Acoustical Society of America*, 89(2), 859–866. <https://doi.org/10.1121/1.1894647>
- Lassek, A. M.** (1955). Effect of combined afferent lesions on motor function. *Neurology*, 5(4), 269. <https://doi.org/10.1212/WNL.5.4.269>
- Lee, B. S.** (1950). Some effects of side-tone delay. *The Journal of the Acoustical Society of America*, 22(5), 639–640. <https://doi.org/10.1121/1.1906665>
- Liu, D., & Todorov, E.** (2007). Evidence for the flexible sensorimotor strategies predicted by optimal feedback control. *The Journal of Neuroscience*, 27(35), 9354–9368. <https://doi.org/10.1523/JNEUROSCI.1110-06.2007>
- Liu, H., Zhang, Q., Xu, Y., & Larson, C. R.** (2007). Compensatory responses to loudness-shifted voice feedback during production of Mandarin speech. *The Journal of the Acoustical Society of America*, 122(4), 2405–2412. <https://doi.org/10.1121/1.2773955>
- Loeb, G. E.** (2012). Optimal isn't good enough. *Biological Cybernetics*, 106(11–12), 757–765. <https://doi.org/10.1007/s00422-012-0514-6>
- Lombard, E.** (1911). Le signe de l'élevation de la voix. *Annales des Maladies de L'Oreille et du Larynx*, 37, 101–119.
- MacNeilage, P. F., Rootes, T. P., & Chase, R. A.** (1967). Speech production and perception in a patient with severe impairment of somesthetic perception and motor control. *Journal of Speech and Hearing Research*, 10(3), 449–467. <https://doi.org/10.1044/jshr.1003.449>
- Makris, N., Kennedy, D. N., McInerney, S., Sorensen, A. G., Wang, R., Caviness, V. S., Jr., & Pandya, D. N.** (2005). Segmentation of subcomponents within the superior longitudinal fascicle in humans: A quantitative, *in vivo*, DT-MRI study. *Cerebral Cortex*, 15(6), 854–869. <https://doi.org/10.1093/cercor/bhh186>

- McClean, M. D., & Sapir, S.** (1981). Some effects of auditory stimulation on perioral motor unit discharge and their implications for speech production. *The Journal of the Acoustical Society of America*, *69*(5), 1452–1457. <https://doi.org/10.1121/1.385777>
- McGuire, P. K., Silbersweig, D. A., & Frith, C. D.** (1996). Functional neuroanatomy of verbal self-monitoring. *Brain: A Journal of Neurology*, *119*(Pt. 3), 907–917.
- McGurk, H., & MacDonald, J.** (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.
- Ménard, L., Perrier, P., Aubin, J., Savariaux, C., & Thibeault, M.** (2008). Compensation strategies for a lip-tube perturbation of French [u]: An acoustic and perceptual study of 4-year-old children. *The Journal of the Acoustical Society of America*, *124*(2), 1192–1206. <https://doi.org/10.1121/1.2945704>
- Mott, F. W., & Sherrington, C. S.** (1895). Experiments upon the influence of sensory nerves upon movement and nutrition of the limbs. Preliminary communication. *Proceedings of the Royal Society of London*, *57*, 481–488.
- Munhall, G. K., MacDonald, E. N., Byrne, S. K., & Johnsrude, I.** (2009). Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *The Journal of the Acoustical Society of America*, *125*(1), 384–390.
- Nasir, S. M., & Ostry, D. J.** (2006). Somatosensory precision in speech production. *Current Biology*, *16*(19), 1918–1923. <https://doi.org/10.1016/j.cub.2006.07.069>
- Natke, U., Grosser, J., & Kalveram, K. T.** (2001). Fluency, fundamental frequency, and speech rate under frequency-shifted auditory feedback in stuttering and nonstuttering persons. *Journal of Fluency Disorders*, *26*(3), 227–241. [https://doi.org/10.1016/S0094-730X\(01\)00099-7](https://doi.org/10.1016/S0094-730X(01)00099-7)
- Natke, U., & Kalveram, K. T.** (2001). Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables. *Journal of Speech, Language, and Hearing Research*, *44*(3), 577–584.
- Niemi, M., Laaksonen, J.-P., Ojala, S., Aaltonen, O., & Happonen, R.-P.** (2006). Effects of transitory lingual nerve impairment on speech: An acoustic study of sibilant sound /s/. *International Journal of Oral and Maxillofacial Surgery*, *35*(10), 920–923. <https://doi.org/10.1016/j.ijom.2006.06.002>
- Nieto-Castanon, A., Guenther, F. H., Perkell, J., & Curtin, H.** (2005). A modeling investigation of articulatory variability and acoustic stability during American English /r/ production. *The Journal of the Acoustical Society of America*, *117*(5), 3196–3212.
- Niziolek, C. A., & Guenther, F. H.** (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *The Journal of Neuroscience*, *33*(29), 12090–12098. <https://doi.org/10.1523/JNEUROSCI.1008-13.2013>
- Niziolek, C. A., Nagarajan, S. S., & Houde, J. F.** (2013). What does motor efference copy represent? Evidence from speech production. *The Journal of Neuroscience*, *33*(41), 16110–16116. <https://doi.org/10.1523/JNEUROSCI.2137-13.2013>
- Noto, C. T., & Robinson, F. R.** (2001). Visual error is the stimulus for saccade gain adaptation. *Cognitive Brain Research*, *12*(2), 301–305. [https://doi.org/10.1016/S0926-6410\(01\)00062-3](https://doi.org/10.1016/S0926-6410(01)00062-3)
- Numminen, J., & Curio, G.** (1999). Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neuroscience Letters*, *272*(1), 29–32. [https://doi.org/10.1016/S0304-3940\(99\)00573-X](https://doi.org/10.1016/S0304-3940(99)00573-X)
- Numminen, J., Salmelin, R., & Hari, R.** (1999). Subject's own speech reduces reactivity of the human auditory cortex. *Neuroscience Letters*, *265*(2), 119–122. [https://doi.org/10.1016/S0304-3940\(99\)00218-9](https://doi.org/10.1016/S0304-3940(99)00218-9)
- Ostry, D., Gribble, P. L., & Gracco, V. L.** (1996). Coarticulation of jaw movements in speech production: Is context sensitivity in speech kinematics centrally planned? *The Journal of Neuroscience*, *16*(4), 1570–1579.
- Parkinson, A. L., Flagmeier, S. G., Manes, J. L., Larson, C. R., Rogers, B., & Robin, D. A.** (2012). Understanding the neural mechanisms involved in sensory control of voice production. *NeuroImage*, *61*(1), 314–322. <https://doi.org/10.1016/j.neuroimage.2012.02.068>
- Parrell, B., Agnew, Z., Nagarajan, S., Houde, J., & Ivry, R. B.** (2017). Impaired feedforward control and enhanced feedback control of speech in patients with cerebellar degeneration. *The Journal of Neuroscience*, *37*(38), 9249–9258. <https://doi.org/10.1523/JNEUROSCI.3363-16.2017>
- Parrell, B., Lammert, A. C., Ciccirelli, G., & Quatieri, T. F.** (2019). Current models of speech motor control: A control-theoretic overview of architectures and properties. *The Journal of the Acoustical Society of America*, *145*(3), 1456–1481. <https://doi.org/10.1121/1.5092807>
- Parrell, B., Ramanarayanan, V., Nagarajan, S., & Houde, J. F.** (2018). FACTS: A hierarchical task-based control model of speech incorporating sensory feedback. *Interspeech 2018*. 2018, 1497–1501.
- Parrell, B., Ramanarayanan, V., Nagarajan, S., & Houde, J. F.** (2019). The FACTS model of speech motor control: Fusing state estimation and task-based control. *BioRxiv*, 543728. <https://doi.org/10.1101/543728>
- Patri, J.-F., Diard, J., & Perrier, P.** (2015). Optimal speech motor control and token-to-token variability: A Bayesian modeling approach. *Biological Cybernetics*, *109*(6), 611–626. <https://doi.org/10.1007/s00422-015-0664-4>
- Patri, J.-F., Perrier, P., & Diard, J.** (2016). Bayesian modeling in speech motor control: A principled structure for the integration of various constraints. *17th Annual Conference of the International Speech Communication Association (Interspeech 2016)*, 2016, 3588–3592.
- Penfield, W., & Boldrey, E.** (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain*, *60*(4), 389–443. <https://doi.org/10.1093/brain/60.4.389>
- Perkell, J. S.** (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neuro-linguistics*, *25*(5), 382–407. <https://doi.org/10.1016/j.jneuroling.2010.02.011>
- Perrier, P.** (2005). Control and representations in speech production. *ZAS Papers in Linguistics*, *40*, 109–132.
- Perrier, P., Løevenbruck, H., & Payan, Y.** (1996). Control of tongue movements in speech: The equilibrium point hypothesis perspective. *Journal of Phonetics*, *24*, 53–75.
- Perrier, P., Ma, L., & Payan, Y.** (2005). Modeling the production of VCV sequences via the inversion of a biomechanical model of the tongue. *Proceedings of the INTERSPEECH: Interspeech 2005–Eurospeech, 9th European Conference on Speech Communication and Technology, Lisbon, Portugal, 2005*, 1041–1044.
- Perrier, P., Ostry, D., & Laboisière, R.** (1996). The equilibrium point hypothesis and its application to speech motor control. *Journal of Speech and Hearing Research*, *39*, 365–378.
- Perrier, P., Payan, Y., Zandipour, M., & Perkell, J.** (2003). Influence of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. *The Journal of the Acoustical Society of America*, *114*(3), 1582–1599.
- Polit, A., & Bizzi, E.** (1979). Characteristics of motor programs underlying arm movements in monkeys. *Journal of Neurophysiology*, *42*, 183–194.

- Pruszynski, J. A., Kurtzer, I., Nashed, J. Y., Omrani, M., Brouwer, B., & Scott, S. H. (2011). Primary motor cortex underlies multi-joint integration for fast feedback control. *Nature*, 478(7369), 387–390. <https://doi.org/10.1038/nature10436>
- Purcell, D. W., & Munhall, K. G. (2006a). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, 120(2), 966–977.
- Purcell, D. W., & Munhall, K. G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, 119(4), 2288–2297.
- Putnam, A. H. B., & Ringel, R. L. (1976). A cineradiographic study of articulation in two talkers with temporarily induced oral sensory deprivation. *Journal of Speech and Hearing Research*, 19(2), 247–266. <https://doi.org/10.1044/jshr.1902.247>
- Ramanarayanan, V., Parrell, B., Goldstein, L., Nagarajan, S., & Houde, J. F. (2016). A new model of speech motor control based on task dynamics and state feedback. *Interspeech 2016*, 2016, 3564–3568.
- Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87. <https://doi.org/10.1038/4580>
- Ringel, R. L., & Steer, M. D. (1963). Some effects of tactile and auditory alterations on speech output. *Journal of Speech and Hearing Research*, 6(4), 369–378. <https://doi.org/10.1044/jshr.0604.369>
- Rothwell, J. C., Traub, M. M., Day, B. L., Obeso, J. A., Thomas, P. K., & Marsden, C. D. (1982). Manual motor performance in a deafferented man. *Brain*, 105(3), 515–542. <https://doi.org/10.1093/brain/105.3.515>
- Sabes, P. N., Jordan, M. I., & Wolpert, D. M. (1998). The role of inertial sensitivity in motor planning. *The Journal of Neuroscience*, 18(15), 5948–5957. <https://doi.org/10.1523/JNEUROSCI.18-15-05948.1998>
- Saltzman, E. (1986). Task dynamic coordination of the speech articulators: A preliminary model. *Experimental Brain Research Series*, 15, 129–144.
- Saltzman, E., & Kelso, J. A. S. (1987). Skilled actions: A task dynamic approach. *Psychological Review*, 94, 84–106.
- Saltzman, E., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1(4), 333–382.
- Sanguineti, V., Laboissière, R., & Ostry, D. J. (1998). A dynamic biomechanical model for neural control of speech production. *The Journal of the Acoustical Society of America*, 103(3), 1615–1627. <https://doi.org/10.1121/1.421296>
- Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., . . . Weiller, C. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences*, 105(46), 18035–18040. <https://doi.org/10.1073/pnas.0805234105>
- Savariaux, C., Perrier, P., & Orliaguet, J. P. (1995). Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production. *The Journal of the Acoustical Society of America*, 98(5), 2428–2442.
- Savariaux, C., Perrier, P., Orliaguet, J.-P., & Schwartz, J.-L. (1999). Compensation strategies for the perturbation of French [u] using a lip tube. II. Perceptual analysis. *The Journal of the Acoustical Society of America*, 106(1), 381–393. <https://doi.org/10.1121/1.427063>
- Schmahmann, J. D., Pandya, D. N., Wang, R., Dai, G., D'Arceuil, H. E., de Crespigny, A. J., & Wedeen, V. J. (2007). Association fibre pathways of the brain: Parallel observations from diffusion spectrum imaging and autoradiography. *Brain*, 130(3), 630–653. <https://doi.org/10.1093/brain/awl359>
- Scott, C. M., & Ringel, R. L. (1971). Articulation without oral sensory control. *Journal of Speech and Hearing Research*, 14(4), 804–818.
- Scott, S. H. (2004). Optimal feedback control and the neural basis of volitional motor control. *Nature Reviews Neuroscience*, 5(7), 532–546. <https://doi.org/10.1038/nrn1427>
- Scott, S. H. (2012). The computational and neural basis of voluntary motor control and planning. *Trends in Cognitive Sciences*, 16(11), 541–549. <https://doi.org/10.1016/j.tics.2012.09.008>
- Shadmehr, R., & Krakauer, J. W. (2008). A computational neuroanatomy for motor control. *Experimental Brain Research*, 185(3), 359–381. <https://doi.org/10.1007/s00221-008-1280-5>
- Shadmehr, R., & Mussa-Ivaldi, F. A. (1994). Adaptive representation of dynamics during learning of a motor task. *The Journal of Neuroscience*, 14(5 Pt. 2), 3208–3224.
- Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annual Review of Neuroscience*, 33, 89–108. <https://doi.org/10.1146/annurev-neuro-060909-153135>
- Shaiman, S., & Gracco, V. L. (2002). Task-specific sensorimotor interactions in speech production. *Experimental Brain Research*, 146(4), 411–418. <https://doi.org/10.1007/s00221-002-1195-5>
- Shiller, D. M., Ostry, D. J., & Gribble, P. L. (1999). Effects of gravitational load on jaw movements in speech. *The Journal of Neuroscience*, 19(20), 9073–9080. <https://doi.org/10.1523/JNEUROSCI.19-20-09073.1999>
- Shiller, D. M., Ostry, D. J., Gribble, P. L., & Laboissière, R. (2001). Compensation for the effects of head acceleration on jaw movement in speech. *The Journal of Neuroscience*, 21(16), 6447–6456.
- Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America*, 125(2), 1103–1113. <https://doi.org/10.1121/1.3058638>
- Shum, M., Shiller, D. M., Baum, S. R., & Gracco, V. L. (2011). Sensorimotor integration for speech motor learning involves the inferior parietal cortex. *The European Journal of Neuroscience*, 34(11), 1817–1822. <https://doi.org/10.1111/j.1460-9568.2011.07889.x>
- Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain and Language*, 164, 77–105. <https://doi.org/10.1016/j.bandl.2016.10.004>
- Stengel, R. F. (1994). *Optimal control and estimation*. New York, NY: Dover Publications.
- Summers, W. V., Johnson, K., Pisoni, D. B., & Bernacki, R. H. (1989). An addendum to “Effects of noise on speech production: Acoustic and perceptual analyses” [J. Acoust. Soc. Am. 84, 917–928 (1988)]. *The Journal of the Acoustical Society of America*, 86(5), 1717–1721.
- Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America*, 84(3), 917–928.
- Svirsky, M. A., Lane, H., Perkell, J. S., & Wozniak, J. (1992). Effects of short-term auditory deprivation on speech production in adult cochlear implant users. *The Journal of the Acoustical Society of America*, 92(3), 1284–1300. <https://doi.org/10.1121/1.403923>
- Takaso, H., Eisner, F., Wise, R. J. S., & Scott, S. K. (2010). The effect of delayed auditory feedback on activity in the temporal lobe while speaking: A positron emission tomography study.

- Journal of Speech, Language, and Hearing Research*, 53(2), 226–236. [https://doi.org/10.1044/1092-4388\(2009/09-0009\)](https://doi.org/10.1044/1092-4388(2009/09-0009))
- Tian, X., & Poeppel, D.** (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology*, 1, 166. <https://doi.org/10.3389/fpsyg.2010.00166>
- Todorov, E.** (2004). Optimality principles in sensorimotor control. *Nature Neuroscience*, 7(9), 907–915. <https://doi.org/10.1038/nn1309>
- Todorov, E.** (2006). Optimal control theory. In K. Doya, S. Ishii, & A. Pouget (Eds.), *Bayesian brain: Probabilistic approaches to neural coding* (pp. 269–298). Cambridge, MA: MIT Press.
- Todorov, E., & Jordan, M. I.** (2002). Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, 5(11), 1226–1235. <https://doi.org/10.1038/nn963>
- Todorov, E., Li, W., & Pan, X.** (2005). From task parameters to motor synergies: A hierarchical framework for approximately optimal control of redundant manipulators. *Journal of Robotic Systems*, 22(11), 691–710. <https://doi.org/10.1002/rob.20093>
- Tourville, J. A., & Guenther, F. H.** (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26(7), 952–981. <https://doi.org/10.1080/01690960903498424>
- Tourville, J. A., Reilly, K. J., & Guenther, F. H.** (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, 39(3), 1429–1443. <https://doi.org/10.1016/j.neuroimage.2007.09.054>
- Toyomura, A., Koyama, S., Miyamaoto, T., Terao, A., Omori, T., Murohashi, H., & Kuriki, S.** (2007). Neural correlates of auditory feedback control in human. *Neuroscience*, 146(2), 499–503. <https://doi.org/10.1016/j.neuroscience.2007.02.023>
- Tremblay, S., & Ostry, D.** (2006). The achievement of somatosensory targets as an independent goal of speech production—Special status of vowel-to-vowel transitions. In P. Divenyi, S. Greenberg, & G. Meyer (Eds.), *Dynamics of speech production and perception* (pp. 33–43). Amsterdam, the Netherlands: IOS Press.
- Tremblay, S., Shiller, D., & Ostry, D.** (2003). Somatosensory basis of speech production. *Nature*, 423, 866–869.
- Tseng, Y.-W., Diedrichsen, J., Krakauer, J. W., Shadmehr, R., & Bastian, A. J.** (2007). Sensory prediction errors drive cerebellum-dependent adaptation of reaching. *Journal of Neurophysiology*, 98(1), 54–62. <https://doi.org/10.1152/jn.00266.2007>
- Upadhyay, J., Hallock, K., Ducros, M., Kim, D.-S., & Ronen, I.** (2008). Diffusion tensor spectroscopy and imaging of the arcuate fasciculus. *NeuroImage*, 39(1), 1–9. <https://doi.org/10.1016/j.neuroimage.2007.08.046>
- Ventura, M. I., Nagarajan, S. S., & Houde, J. F.** (2009). Speech target modulates speaking induced suppression in auditory cortex. *BMC Neuroscience*, 10, 58. <https://doi.org/10.1186/1471-2202-10-58>
- Villacorta, V. M., Perkell, J. S., & Guenther, F. H.** (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America*, 122(4), 2306–2319. <https://doi.org/10.1121/1.2773966>
- von Holst, E.** (1954). Relations between the central nervous system and the peripheral organs. *The British Journal of Animal Behaviour*, 2(3), 89–94. [https://doi.org/10.1016/S0950-5601\(54\)80044-X](https://doi.org/10.1016/S0950-5601(54)80044-X)
- von Holst, E., & Mittelstaedt, H.** (1950). Das reafferenzprinzip. *Naturwissenschaften*, 37, 464–476.
- Wallman, J., & Fuchs, A. F.** (1998). Saccadic gain modification: Visual error drives motor adaptation. *Journal of Neurophysiology*, 80(5), 2405–2416. <https://doi.org/10.1152/jn.1998.80.5.2405>
- Warren, R. M.** (1970). Perceptual restoration of missing speech sounds. *Science*, 167(3917), 392–393.
- Webster, D. B.** (1992). An overview of mammalian auditory pathways with an emphasis on humans. In D. B. Webster, A. N. Popper, & R. R. Fay (Eds.), *The mammalian auditory pathway: Neuroanatomy* (pp. 1–22). New York, NY: Springer. https://doi.org/10.1007/978-1-4612-4416-5_1
- Weiler, J., Gribble, P. L., & Pruszynski, J. A.** (2019). Spinal stretch reflexes support efficient hand control. *Nature Neuroscience*, 22, 1. <https://doi.org/10.1038/s41593-019-0336-0>
- Wolpert, D. M., & Miall, R. C.** (1996). Forward models for physiological motor control. *Neural Networks*, 9(8), 1265–1279.
- Yates, A. J.** (1963). Delayed auditory feedback. *Psychological Bulletin*, 60(3), 213–232. <https://doi.org/10.1037/h0044155>
- Yeterian, E. H., & Pandya, D. N.** (1999). Corticostriatal connections of the superior temporal region in rhesus monkeys. *Journal of Comparative Neurology*, 399(3), 384–402. [https://doi.org/10.1002/\(SICI\)1096-9861\(19980928\)399:3<384::AID-CNE7>3.0.CO;2-X](https://doi.org/10.1002/(SICI)1096-9861(19980928)399:3<384::AID-CNE7>3.0.CO;2-X)
- Zandipour, M., Guenther, F., Perkell, J., Perrier, P., Payan, Y., & Badin, P.** (2004). Vowel–vowel planning in acoustic and muscle space. *Proceedings of “From sound to sense: 50+ years of discoveries in speech communication,” 2014*, C103–C108.
- Zarate, J. M., Wood, S., & Zatorre, R. J.** (2010). Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. *Neuropsychologia*, 48(2), 607–618. <https://doi.org/10.1016/j.neuropsychologia.2009.10.025>
- Zarate, J. M., & Zatorre, R. J.** (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing. *NeuroImage*, 40(4), 1871–1887. <https://doi.org/10.1016/j.neuroimage.2008.01.026>
- Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S.** (2009). Functional overlap between regions involved in speech perception and in monitoring one’s own voice during speech production. *Journal of Cognitive Neuroscience*, 22(8), 1770–1781. <https://doi.org/10.1162/jocn.2009.21324>