Abstract and Keywords

Speech production is marked by rapid, coordinated movements of the vocal articulators. This is an impressive feat given the large number of muscles involved in producing even the simplest monosyllable. Yet, fluent speakers meet these demands with relative ease, producing as many as four to seven syllables per second. By 2 years of age, children learning American English have typically mastered the fine articulatory distinctions that differentiate the consonants /b/ and /d/ and the fine timing control that differentiates /b/ and /p/. By grade school they have mastered nearly the full inventory of phonemes, which they can combine in any syllable structure to produce long, complex, intelligible utterances. This article explores how the integration of auditory and somatosensory feedback with motor commands contributes to learning and maintaining these skills.

Keywords: articulation, somatosensory feedback, motor planning, motor learning

Overview: Perception-Production Interactions in Speech Production

Speech production is marked by rapid, coordinated movements of the vocal articulators. This is an impressive feat given the large number of muscles involved in producing even the simplest monosyllable (Zemlin, 1998). Yet, fluent speakers meet these demands with relative ease, producing as many as four to seven syllables per second (Tsao & Weismer, 1997). By 2 years of age, children learning American English have typically mastered the fine articulatory distinctions that differentiate the consonants /b/ and /d/ and the fine timing control that differentiates /b/ and /p/. By grade school they have mastered nearly the full inventory of phonemes, which they can combine in any syllable structure to produce long, complex, intelligible utterances (McLeod & Bleile, 2003). This article explores how the integration of auditory and somatosensory feedback with motor commands contributes to learning and maintaining these skills.

The brain orchestrates the finely tuned pattern of muscle activations needed for speech production using one or both of two broad classes of control. Under feedback control, sensory information guides motor execution. A well-known example of feedback control of speech is the increase in speech intensity in response to increasing
environmental noise, the “Lombard Effect” (Lombard, 1911). Speakers also compensate for perceived errors in their auditory or somatosensory feedback (e.g., Abbs & Gracco, 1984; Jones & Munhall, 2002; Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984; Purcell & Munhall, 2006b). Under feedforward control, task performance is executed from previously learned commands, or speech motor programs, without reliance on incoming task-related sensory information. The ease with which fluent speakers coordinate the rapid movements of multiple articulators is evidence of feedforward control (Neilson & Neilson, 1987). These movements are simply too fast to rely on sensory feedback for guidance.

Learning and maintaining the speech motor programs that support feedforward control, however, relies on sensory feedback. When we produce a speech sound, the motor commands that induce articulator movements are paired with the concomitant sensory feedback. The auditory and somatosensory signals that we expect to result from those motor commands are compared with this feedback. If a difference is encountered, the error is used to drive feedback-based corrective movements and to tune feedforward motor commands (e.g., Aasland, Baum, & McFarland, 2006; Houde & Jordan, 2002; Ito & Ostry, 2010; Jones & Munhall, 2002, 2005; Purcell & Munhall, 2006a; Villacorta, Perkell, & Guenther, 2007). This interplay between feedforward and feedback control systems subserves a preschooler’s initial speech motor learning and the updates necessitated by a growing vocal tract during puberty. At the same time, it supports rapid, online responses to a changing listening environment. This integration of motor and sensory information requires the recruitment and coordination of a diverse set of brain regions. Accordingly, a large portion of the brain is involved in even the simplest speech tasks, such as reading aloud a single word or syllable (e.g., (Fiez & Petersen, 1998; Ghosh, Tourville, & Guenther, 2006; Turkeltaub, Eden, Jones, & Zeffiro, 2002). This is demonstrated in Figure 1, which shows the cortical and subcortical brain regions involved in reading aloud one- or two-syllable words and pseudowords.

When producing a speech sound, motor commands issued from motor and premotor cortex of the lateral frontal lobe induce articulator movements. These areas are highlighted and labeled in Figure 2. The articulator movements result in auditory and somatosensory feedback, which is processed in the superior temporal and lateral parietal cortex, respectively. These cortical regions interact with medial frontal cortex and subcortical structures including the cerebellum, basal ganglia, and brainstem (Figure 1). Together, these regions constitute the neural control system responsible for speech production.

The sections that follow describe how the motor and sensory components of this system interact during speech production and how their interplay enables a speaker to achieve his or her goals. In doing so, behavioral data characterizing the nature of auditory-motor and somatosensory-motor interactions during speech are discussed. Also reviewed is neurophysiological and neuroimaging research that has begun to reveal where and how these interactions are implemented in the brain. These neural mechanisms are described with reference to the Directions Into Velocities of Articulators (DIVA) model of speech acquisition and production (Guenther, Ghosh, & Tourville, 2006; Tourville & Guenther, 2010). DIVA provides a unified quantitative account of a wide range of speech production phenomena and neuroimaging data. The model consists of feedforward and feedback control systems.
with a silent letter viewing baseline task (speech—baseline). The three speech tasks included in the analysis were production of monosyllable nonsense words (11 subjects), bisyllable nonsense words (13 subjects), and monosyllable American English words (10 subjects) for a total of 34 subjects. Thresholded statistical parametric maps (FDR 〈 5 percent; \( t > 3.48 \)) of the pooled speech—baseline contrast are shown superimposed on a representative brain to reveal the brain regions involved in speech production. Far left: A series of coronal slices through the central region of the brain reveals activation in the thalamus and basal ganglia. Numbers to the left of the images denote the anterior-posterior level (y coordinate) of the image in Montreal Neurological Institute (Mazziotta, et al., 2001) stereotactic space (higher y values are more anterior). Center: Activity is superimposed on renderings of the (a) left hemisphere lateral, (b) right lateral, (c) left medial, (d) right medial, (e) ventral, and (f) dorsal surfaces of the brain. The cortical regions associated with speech production include lateral sensorimotor cortex, medial and ventrolateral prefrontal cortex, perisylvian cortex, and superior temporal cortex. A more detailed illustration of the location of these regions on the cerebral surface is provided in Figure 2. Far right: A series of coronal slices through the cerebellum reveals widespread bilateral activation in superior cerebellar cortex and a small cluster of activation in right inferior cerebellar cortex.

![Click to view larger](image)

Figure 2. Lateral cortical regions involved in speech production. Activity associated with motor, somatosensory, and auditory cortical processing during speech production is labeled on the lateral cortical surface. The activation overlays were generated from the pooled analysis of speech production tasks as described in the caption accompanying Figure 1. Sulci that mark relevant anatomic distinctions are marked by dotted lines (see key). The region labeled motor includes motor and premotor cortex on the ventral precentral gyrus and the posterior inferior frontal gyrus. The region labeled somatosensory includes primary and secondary somatosensory cortex on the ventral that are tuned by integrating motor output and sensory feedback and is therefore particularly useful for guiding a discussion of the sensorimotor interactions that underlie speech production.

### Auditory-Motor Interactions in Speech

Many motor acts are aimed at achieving goals in three-dimensional space (e.g., reaching, grasping, walking, and handwriting). For such tasks, visual feedback of task performance plays an important role in monitoring performance and improving skill level (e.g., Held & Freedman, 1963; Helmholtz, 1925; Redding & Wallace, 2006). The primary goal of speech, however, is an acoustic signal that transmits a linguistic message by the listener’s auditory system. Auditory feedback, like visual feedback during spatial task performance, plays an important role in monitoring vocal output and achieving fluency during speech production. We rely on instantaneous auditory feedback to modulate our voices in noisy environments (Lane & Tranel, 1971; Lombard, 1911), increasing the intensity and pitch of our voices and the duration of vowels (Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988; Tartter, Gomes, & Litwin, 1993) as ambient noise increases. This involuntary response improves the clarity of our speech, making it more easily comprehended by a listener (Pittman & Wiley, 2001). Disruption of normal auditory feedback can induce stuttering-like disfluences in normally fluent speakers. This occurs when a short delay (approximately 200 ms) is introduced to a speaker’s auditory feedback (Stuart, Kalinowski, Rastatter, & Lynch, 2002; Yates, 1963). A shorter delay, (approximately 50–75 ms), conversely, results in a transient improvement in the fluency of persons who stutter (e.g., Kalinowski, Stuart, Sark, & Armson, 1996; Ryan & Vankirk, 1974).

Auditory feedback does more than guide ongoing speech production by feedback-based control. In the absence of auditory feedback, development of verbal communication skills is profoundly disrupted. Infants born with severe hearing deficits experience delays in early speech development, including the onset of canonical babbling, which typically...
postcentral and anterior supramarginal gyri. The region labeled auditory includes primary and secondary auditory cortex along the superior temporal gyrus. Arises in the first year (Oller & Eilers, 1988). The intelligibility of speech later produced by such children covaries with residual hearing ability (Smith, 1975). Those born with profound hearing loss may never achieve intelligible verbal communication without interventions that provide some form of hearing ability. Speech development can be significantly improved by cochlear implants (Bouchard, Oullet, & Cohen, 2009), for instance, highlighting the critical role auditory feedback plays in the acquisition of feedforward motor programs that we rely on for fluent verbal communication.

The effects of hearing loss after language acquisition demonstrate another important contribution of auditory feedback to speech production. Although postlingually deaf individuals are able to communicate orally, difficulty controlling vocal pitch and intensity manifests shortly after hearing loss (Cowie & Douglas-Cowie, 1983; Lane & Webster, 1991). The precision of articulator movements and phonetic distinctions also deteriorate, reducing intelligibility (Lane & Webster, 1991; Waldstein, 1990). Auditory feedback is necessary to maintain speech motor skills long after they were first learned. Thus, speech production involves an ongoing, continuous interaction between the speech motor system and sensory systems that allows sensory feedback to modulate ongoing movements of the vocal articulators (feedback control) and to fine tune stored speech motor programs (feedforward control).

Since the Lombard effect was noted in the early 20th century, attempts to characterize auditory-motor interactions during speech have been made by unexpectedly and transiently modifying auditory feedback during speech production in near real-time. Speakers respond to these manipulations by altering their vocal output in the direction opposite the shift. For instance, when the amplitude of a speaker's auditory feedback is shifted down, the speaker responds by increasing his or her vocal intensity (Bauer, Mittal, Larson, & Hain, 2006; Heinks-Maldonado & Houde, 2005). Similarly, when fundamental frequency (F0), which is perceived as vocal pitch, is artificially shifted upward or downward, speakers respond by shifting the F0 in the opposite direction, below or above its normal level (e.g., Burnett & Larson, 2002; Elman, 1981; Jones & Munhall, 2002; Larson, Burnett, Kiran, & Hain, 2000). The speaker effectively corrects for the feedback manipulation, modifying vocal output so that the perceived feedback more closely matches the intended output.

Manipulations of formant frequencies have demonstrated that control of the segmental content of speech is also influenced by online feedback-based control (e.g., Purcell & Munhall, 2006b; Tourville, Reilly, & Guenther, 2008). For instance, shifting the first formant (F1) of a speaker's acoustic output upward makes the word bet sound like bat (the vowel /ɛɛ shifts toward /æ/); a downward shift makes bet sound like bit (/ɛɛ shifts toward /I/). Such a shift alters the phonological and therefore lexical content of the auditory feedback. Speakers compensate for these segmental manipulations, again, by altering the F1 content of their vocal output in the direction opposite the shift, steering output closer to the intended auditory target. The compensatory response can occur within 108 ms of shift onset, fast enough to detect and correct the feedback within the typical duration syllable (Hillenbrand, Clark, & Nearey, 2001; Hillenbrand, Getty, Clark, & Wheeler, 1995) during conversational speech. Likewise, responses to F0 errors are fast enough to correct ongoing speech production. The compensatory response has been noted to persist well beyond the period during which the shift was applied (Donath, Natke, & Kalveram, 2002). Such aftereffects indicate that the change in vocal output is not simply a “reflexive” response to unexpected feedback. Rather, feedback errors and corrective responses are monitored and applied to future movements to preemptively avoid errors.
When a feedback error is repeatedly encountered, speakers adapt to the induced error (Houde, Nagarajan, Sekihara, & Merzenich, 2002; Jones & Munhall, 2000, 2005; Purcell & Munhall, 2006a; Villacorta, et al., 2007). For instance, repeated shifts of F1 in the same direction induce robust compensatory responses in the opposite direction (e.g., Purcell & Munhall, 2006a; Villacorta, et al., 2007). F1 in utterances made immediately after removal or masking of the persistent shift continues to differ from baseline formants in the direction opposite the shift. The “overshoots” or aftereffects that follow removal of the shift indicate a reorganization of the sensory–motor neural mappings that underlie feedforward control in speech. They are evidence of feedforward speech motor commands that have been updated in the presence of persistent auditory error to more accurately achieve auditory goals.

**Neural Correlates**

The interaction of auditory and motor systems during speech production is reflected in the findings of neuroimaging studies. As Figure 2 demonstrates, the motor-related activity of the frontal lobe is accompanied by activity in a large portion of the superior temporal gyrus, the seat of primary and secondary auditory cortices, during speech production. Responses in the more anterior and ventral portions of the superior temporal gyrus have been correlated with increasing intelligibility of speech sounds (Binder, et al., 2000; Giraud & Price, 2001; Scott, Blank, Rosen, & Wise, 2000). This region is thought to support auditory object recognition and speech perception (Rauschecker & Scott, 2009). Activation in the posterior superior temporal gyrus region extends to the inferior parietal cortex including the parietal operculum and anterior supramarginal gyrus, regions implicated in several processes related to speech production including verbal working memory (e.g., Becker, MacAndrew, & Fiez, 1999; Jonides, et al., 1998); phonetic discrimination (Caplan, Gow, & Makris, 1995); and an interface between orthographic, phonological, and lexical-semantic decision-making (Pugh, et al., 2001). Damage to this area results in several language production and perception deficits including Wernicke’s aphasia and conduction aphasia (Damasio & Damasio, 1980; Goodglass, 1993).

It has become clear, however, that activity in the posterior portions of the superior temporal gyrus, including the planum temporale, the portion of the posterior superior temporal gyrus that lies within the Sylvian fissure and behind primary auditory cortex, does not solely reflect the afferent auditory input associated with speech production. Activity in this area changes even when subjects simply imagine speaking without moving the articulators (Hickok, et al., 2000; Numminen & Curio, 1999; Okada & Hickok, 2006; Okada, Smith, Humphries, & Hickok, 2003). Such “covert” speech results in no auditory feedback, yet induces a response in auditory cortex. Moreover, responses to one’s own speech in this region are suppressed compared with responses when listening to recordings of others speaking (Wise, Greene, Buchel, & Scott, 1999) or even recordings of one’s own speech (Curio, Neuloh, Numminen, Jousmaki, & Hari, 2000; Heinks-Maldonado, Mathalon, Gray, & Ford, 2005; Heinks-Maldonado, Nagarajan, & Houde, 2006; Numminen, Salmelin, & Hari, 1999). Single unit recordings from nonhuman primates have shown suppression of auditory cortical neurons immediately prior to self-initiated vocalizations (Eliades & Wang, 2003, 2005). These suppressed neurons become highly active, however, when auditory feedback is altered (Eliades & Wang, 2008). Attenuation of the auditory cortex suppression associated with speech production under altered auditory feedback conditions has also been noted in humans (e.g., Heinks-Maldonado, et al., 2006). The degree of suppression is modulated by how closely feedback matches expectations. Similar modulation of somatosensory responses to self-generated movements (e.g., Blakemore, Wolpert, & Frith, 1998; Wasaka, Hoshiyama, Nakata, Nishihira, & Kakigi, 2003) has been interpreted as evidence that an efference copy of outgoing motor commands sent from motor to sensory cortex encodes the expected
sensory consequences of upcoming movements (Blakemore, Wolpert, & Frith, 2000; Wolpert & Flanagan, 2001). The efference copy attenuates the regions of sensory cortex responsive to the expected sensory feedback. In doing so, the efference copy “cancels” the sensory feedback response resulting from the movement (von Holst & Mittelstaedt, 1950).

Neuroimaging studies of perturbed auditory feedback during speech production have consistently noted increased response in posterior superior temporal gyrus bilaterally as a result of the feedback manipulation. This has been shown for delayed auditory feedback (Hashimoto & Sakai, 2003; Hirano, et al., 1997); noise masking (Christoffels, Formisano, & Schiller, 2007; Zheng, Munhall, & Johnsrude, 2010); F0/pitch shifts (Fu, et al., 2006; McGuire, Silbersweig, & Frith, 1996; Zarate & Zatorre, 2005; see Toyomura et al., 2007, for evidence of increased activity only in the right hemisphere); and formant shifts (Tourville, et al., 2008) during speech production. Greater attenuation of auditory cortical responses was noted when speakers heard normal rather than pitch-shifted auditory feedback. Collectively, these findings indicate that responses in posterior auditory cortex are modulated by input from the speech motor system during speech production. This input selectively suppresses the response to the expected auditory feedback. When feedback does not match this expectation, the response is elevated.

In addition to bilateral superior temporal activity, Tourville et al. (2008) noted increased inferior cerebellum activity in the left hemisphere and right lateralized increased activity in inferior frontal gyrus and lateral premotor cortex when formants were shifted during speech production. Structural equation modeling (McIntosh & Gonzalez-Lima, 1994) revealed greater effective connectivity between the bilateral superior temporal (auditory) areas that responded to the auditory perturbation and these inferior frontal/premotor regions of the right hemisphere. These findings suggest lateral premotor areas of the right hemisphere are selectively involved in the transformation of auditory errors into compensatory articulatory movements that forms the heart of the auditory feedback control network for speech (discussed later). Functional pathways between the lateral premotor areas involved in speech production and posterior auditory cortex have been established in other studies (e.g., Matsumoto, et al., 2004), and more recently white matter tracts between these areas were identified using diffusion tensor imaging. These pathways provide a means for direct communication between the speech motor and auditory systems. Hickok and colleagues have described a region at the posterior end of this pathway, deep within the Sylvian fissure at the junction of the temporal and parietal lobes as a sensorimotor interface for speech (Hickok, Buchsbaum, Humphries, & Muftuler, 2003; Hickok & Poeppel, 2007). The DIVA model, detailed later, offers a mechanistic account of how this interface supports speech motor learning (Guenther, et al., 2006).

**Somatomotor Interactions in Speech**

Mechanoreceptors of the vocal tract provide the brain with information about the configuration of the speech articulators. Like auditory feedback, afferent somatosensory information is carefully monitored during speech production. When somatosensory feedback is unexpected, for instance when the shape of the vocal tract changes (Aasland, et al., 2006; Baum & McFarland, 1997, 2000; Jones & Munhall, 2003; McFarland, Baum, & Chabot, 1996), the speech motor system adapts so that the intended speech target is reached. The influence of somatosensory feedback on speech motor control has been investigated extensively in experiments that have imparted an unexpected, transient perturbation of articulator movements. When speakers encounter disruptions of lip or jaw movements, they adjust movements of unimpeded articulators to compensate for the error (Abbs &
Gracco, 1984; Folkins & Abbs, 1975; Gomi, Honda, Ito, & Murano, 2002; Gracco & Abbs, 1985; Kelso, et al., 1984; Shaiman, 1989). For instance, if the lower lip is displaced downward during production of /aba/, speakers adjust by moving their upper lips lower than normal (Abbs & Gracco, 1984). By doing so, they achieve the lip closure necessary for the /b/.

This compensatory response is fast (within 100 ms of the perturbation) and is functionally specific: compensatory movements are restricted to those that achieve the articulator or acoustic goal. For instance, a downward jaw displacement induces downward movement of the upper lip during the final bilabial closure in /baeb/ but not during the alveolar closure of /baez/ (Kelso, et al., 1984). A modification of the trajectory of the upper lip movement is unnecessary to produce the /z/. The compensatory movements are therefore not simply reflexive (i.e., the result of local stretch receptor dynamics). Rather, they are governed by the articulatory requirements of the intended sensory target, indicating that the compensatory responses are influenced by input from the cerebral cortex where the targets are likely stored. This has been confirmed by Ito and colleagues who demonstrated mediation of compensatory responses to a jaw perturbation by speech motor cortex (Ito, Kimura, & Gomi, 2005).

Although many of the somatosensory perturbations discussed so far were accompanied by a concomitant change in auditory feedback (making it possible that the responses were mediated by auditory feedback error that is transformed into compensatory articulator movements rather than somatosensory errors), other studies have shown correction for somatosensory perturbations in the absence of auditory errors. For example, Lindblom, Lubker, and Gay (1979) showed that subjects produced tongue movements that compensated for a bite block held between the teeth even before they produced acoustic output. More recent studies by Ostry and colleagues have shown that speakers compensate for small jaw perturbations that do not result in detectable changes of auditory feedback (Nasir & Ostry, 2006; Tremblay, Shiller, & Ostry, 2003). Furthermore, when the perturbation is removed, speakers continue to make compensatory movements in anticipation of the perturbation for several trials. This aftereffect suggests that the somatosensory error resulted in modification of stored feedforward motor commands. Similar adaptation to the jaw perturbation was also observed from subjects with profound postlingual hearing loss (Nasir & Ostry, 2008). Together, these results indicate that somatosensory targets are indeed a goal of speech production, and that consistently applied somatosensory perturbations lead to changes in the feedforward motor commands for speech. This latter observation is consistent with the fact that children maintain fluent, intelligible speech in the face of dramatic growth in the vocal tract as they age; although the changes in vocal tract morphology occur relatively slowly, they are akin to the effects of bite blocks, false palates, or lip tubes that induce changes in vocal tract shape in experimental setups.

**Neural Correlates**

Few studies have investigated the brain regions associated with somatosensory feedback contributions to speech motor control. Compensatory responses to a jaw perturbation involve primary motor cortex. This was shown by targeting the left lateral motor cortex with transcranial magnetic stimulation; when the area was stimulated, compensatory responses to jaw perturbations were altered (Ito, et al., 2005). But what of the rest of the brain? Golfinopoulos et al. (2011) used functional magnetic resonance imaging (fMRI) to compare brain responses during speech under normal conditions and when jaw movements were unexpectedly restricted by the inflation of a small balloon. The perturbation resulted in increased activity throughout the brain regions involved in normal speech production (see Figure 1), peaking in the right anterior supramarginal gyrus, with
additional activity in the right inferior frontal gyrus pars triangularis and anterior cingulate gyrus. Activity in several regions, including the supramarginal gyrus, lateral premotor cortex, and inferior frontal gyrus, was significantly greater in the right hemisphere. An earlier fMRI study of a static lip perturbation during speech (Baciu, Abry, & Segebarth, 2000) noted similar results, including a peak in right supramarginal gyrus and a shift to greater input from the right hemisphere during perturbed speech.

The anterior supramarginal gyrus lies immediately behind and is tightly connected with the lateral postcentral gyrus where the primary somatosensory representation of the speech articulators lies (Boling, Reutens, & Olivier, 2002; Fesl, et al., 2003; Lotze, et al., 2000). It is therefore not surprising that this area is active when the speech articulators are perturbed. This area is strongly connected to the speech motor areas of the inferior frontal and lateral premotor cortex (Makris, et al., 2005; Matsumoto, et al., 2004; Rushworth, Behrens, & Johansen-Berg, 2006; Saur, et al., 2008, 2010). An analogous pathway in primates (Luppino, Murata, Govoni, & Matelli, 1999) is believed to contribute to the sensorimotor transformations required to guide movements, such as grasping (see Rizzolatti & Luppino, 2001). Relatedly, this area has been associated with corrective finger movements induced by unexpected load variations during lifting (Jenmalm, Schmitz, Forssberg, & Ehrsson, 2006; Schmitz, Jenmalm, Ehrsson, & Forssberg, 2005). In particular, it is associated with comparing expected sensory feedback and actual feedback from somatosensory afferents during movement. The imaging studies of articulator perturbations cited previously are consistent with this interpretation. It has been argued that speech motor commands and sensory feedback are integrated in the ventral parietal lobe (Guenther, et al., 2006; Hickok & Poeppel, 2004; Rauschecker & Scott, 2009), analogous to the integration of visual and motor signals demonstrated in the dorsal parietal lobe (Andersen, 1997; Rizzolatti, Fogassi, & Gallese, 1997). The next section describes a mechanistic account of this interaction and how it subserves speech motor learning and control.

Brain responses to jaw perturbation (Golfinopoulos, et al., 2011) offer interesting parallels to the auditory feedback perturbation findings discussed in the previous section (Tourville, et al., 2008). Both studies revealed increased bilateral activation of secondary sensory cortices in the modality of the perturbation and right lateralized increases of lateral premotor and inferior frontal activation when feedback was perturbed. Furthermore, both studies revealed increased effective connectivity between the posterior sensory regions and right lateral frontal regions in response to the perturbation, providing further evidence for increased recruitment of right inferior frontal and premotor regions during feedback-based control of speech.

**Neural Computations Underlying Sensorimotor Interactions during Speech**

Since 1992, our laboratory has developed an adaptive neural network model, the DIVA model, which provides a quantitatively explicit description of the sensorimotor interactions involved in speech motor control. The model represents a theory of how feedforward and feedback control systems interact during speech motor learning and production and is presented here to facilitate and guide a unified account of the findings discussed previously. A simplified schematic of the model is shown in Figure 3. DIVA consists of integrated feedforward and feedback control systems that learn to control a simulated vocal tract (a modified version of the synthesizer described by Maeda, 1990). Once trained, the model takes a discrete speech sound (typically a syllable) as input and generates a time varying sequence of articulator positions that command movements of the simulated vocal tract that produce the desired sound. The model provides a unified explanation of a number of speech production
phenomena including motor equivalence, contextual variability, anticipatory and carryover coarticulation, velocity/distance relationships, speaking rate effects, and speaking skill acquisition and retention throughout development (e.g., Callan, Kent, Guenther, & Vorperian, 2000; Guenther, 1994; Guenther, 1995; Guenther, et al., 2006; Guenther, Hampson, & Johnson, 1998; Nieto-Castanon, Guenther, Perkell, & Curtin, 2005). Because it can account for such a wide array of data, the DIVA model has provided the theoretical framework for a number of investigations of normal and disordered speech production. Predictions from the model have guided studies of the role of auditory feedback in normally hearing persons, deaf persons, and persons who have recently regained some hearing through the use of cochlear implants (Lane, et al., 2007; Perkell, et al., 2007; Perkell, et al., 2000; Perkell, Guenther, et al., 2004; Perkell, Matthies, et al., 2004). The model has also aided investigations of the cause of stuttering (Max, Guenther, Gracco, Ghosh, & Wallace, 2004) and apraxia of speech (Robin, et al., 2008; Terband, Maassen, Guenther, & Brumberg, 2009).

The model is comprised of interconnected modules (blocks in Figure 3) that represent maps of neurons in specific regions of the human brain. The association of these modules to cortical, subcortical, and cerebellar regions of the brain is based on human clinical, neuroimaging, and microstimulation studies and nonhuman primate single cell recording and anterograde and retrograde tracing studies (Guenther, et al., 2006; Tourville & Guenther, 2010). Each module corresponds to a set of neurons, or map, which represents a particular type of information. The term mapping is used to refer to a transformation from one neural representation to another. These transformations are represented by arrows in Figure 3 and are analogous to axonal projections in the brain.

A hallmark of the DIVA model is the integration of input from feedforward and feedback control systems to generate the overall motor command that is sent to the articulatory musculature. The “motor programs” for speech sounds are stored in a speech sound map that is hypothesized to lie in lateral premotor cortex in the left hemisphere. Speech production in the model begins with the activation of one of these cells. Signals sent from the speech sound map directly to articulator velocity and position maps in motor cortex to represent feedforward commands (FF pathway in Figure 3). These projections encode a previously learned time series of articulatory movements for producing the desired speech sound; they can thus be thought of as a “gestural score” (e.g., Browman & Goldstein, 1989) for each learned speech sound.

In addition to activating the stored motor program (feedforward command) for the current speech sound, the speech sound map also sends signals to auditory and somatosensory cortex that represent the expected sensory feedback for the associated motor commands. Throughout speech production, sensory feedback is monitored; if a discrepancy between feedback and target is detected, error signals are generated in the auditory and somatosensory error maps, which project to the feedback control map, which in turn sends feedback commands (FB in Figure 3) to motor cortex, where they are added to the feedforward commands to produce the outgoing motor command to the articulatory musculature. When the model produces a highly practiced speech

---

**Figure 3.** A simplified schematic of the DIVA model of the sensorimotor interactions underlying speech production. aSMG = anterior supramarginal gyrus; FF = feedback motor command; mPT = medial planum temporale; pIFG = posterior inferior frontal gyrus; pSTG = posterior superior temporal gyrus; vMC = ventral motor cortex; vPMC = ventral premotor cortex.
sound, the feedforward commands sent from the speech sound map to motor cortex are accurate and the resulting movements are driven primarily by feedforward control. Early in learning, however, speech production is guided primarily by feedback control because poorly tuned feedforward commands result in sensory errors that induce corrective feedback-based motor commands. Corrective feedback commands would similarly arise when the fully tuned system encounters an error caused by an artificial manipulation of sensory feedback (e.g., shifting a formant such that head sounds like had). The model responds to the error in the same way that speakers in auditory perturbation experiments do, producing speech that sounds more like the intended target. Like an adult speaker, the model remains sensitive to errors while maintaining the ability to produce fluent speech.

DIVA’s speech sound map can be likened to the “mental syllabary” described by Levelt and colleagues (Levelt, Roelofs, & Meyer, 1999; Levelt & Wheeldon, 1994). Each cell represents a phoneme or frequently encountered multiphonemic speech sound, with syllables being the most typical sound type represented. This representation is not entirely motor or sensory. Rather, cells of the speech sound map link the sensory goals of articulator movements to the motor commands that generate them. In doing so, these cells embody the core property of what Rizzolati and others have described as “mirror neurons” (Rizzolatti & Craighero, 2004; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). Mirror neurons are so termed because they respond during an action and while viewing (or hearing) that action performed by another animal or person. This describes the behavior of speech sound map cells: production of a speech sound requires activation of the appropriate speech sound map cells; the same cell becomes active when that speech sound is perceived by input from sensory cortex. In the DIVA model, this interaction between the production and perception systems supports speech motor learning. Others have suggested that mirror neurons may also contribute to speech perception (e.g., Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti & Arbib, 1998).

The Role of Sensory Feedback during Feedback Control and Motor Learning

Like an infant, the DIVA model must learn the association between sound and movement of the articulators before it can produce speech sounds. This is done during a training process analogous to infant babbling during which the model produces random movements of the articulators. The resulting auditory signal is then mapped back to that movement. For example, the model learns that the vowel sound /i/, as in heed, can be achieved by moving the tongue toward the hard palate. This sensory-to-motor mapping is a form of inverse model (e.g., Kawato, 1999; Wolpert & Kawato, 1998) because it implements the transformation between the desired sensory outcome and the appropriate motor actions, otherwise known as the inverse kinematic transformation. The auditory-to-motor inverse mapping is represented in Figure 3 by the pathway from the auditory error map to the feedback control map. Once learned, this mapping subserves both the feedback and feedforward controllers: it is the basis for the feedback-based corrective commands that, in turn, tune the feedforward motor programs of the speech sound map.

Following babbling, the DIVA model learns the auditory expectations, or targets, for frequently encountered speech sounds. The model “listens” to the speech sounds (e.g., phoneme, syllable, word) it is presented with, much like an infant listens to his or her mother, and learns the range of time-varying formants that are acceptable for each sound (see Guenther, 1995). As each new target is learned, it becomes associated with a cell in the speech sound map. When a speech sound map cell is activated during a production attempt, the corresponding auditory target becomes active in the auditory target map. Projections such as these, which predict the sensorimotor state resulting from a movement, represent a forward model of the movement (e.g. Davidson &
Wolpert, 2005; Desmurget & Grafton, 2000; Kawato, 1999; Miall & Wolpert, 1996). The auditory target map inhibits the region in the auditory error map hypothesized to lie in the portion of the posterior superior temporal gyrus that has been described as a sensorimotor interface for speech (discussed previously). Direct connections between this region and lateral premotor cortex have recently been established (e.g., Bernal & Altman, 2010; Saur, et al., 2010). The auditory error map also receives excitatory input from primary auditory cortex that encodes afferent auditory feedback.

After a speech sound map cell is associated with an auditory target, the model begins to practice producing the speech sound. With each attempt, the resulting auditory feedback is compared with the target. Any difference between the two results in activation in the auditory error map that induces a feedback-based corrective command by the feedback control map. The feedback-based command improves the ongoing articulator output but also plays a key role in long-term speech motor development. This is because each feedback-based corrective command is added to the feedforward command that generated the articulator movement, thereby improving the feedforward command for the next attempt. With each subsequent attempt to produce the speech sound, the error is further reduced until the feedforward command is able to achieve the desired auditory target without input from the feedback system.

The model also includes a somatosensory feedback control subsystem that is analogous to the auditory feedback control system just described (see Figure 3). Projections from the speech sound map to the somatosensory target map, hypothesized to lie in ventral somatosensory cortex and anterior supramarginal gyrus, encode the expected proprioceptive and tactile feedback associated with each speech sound. According to the model, somatosensory targets are learned with each attempt to produce a speech sound (i.e., in parallel with feedforward learning). The target map inhibits a representation of the expected somatosensory feedback in the somatosensory error map, hypothesized to lie in the same cortical region (discussed previously). The error map also receives excitatory input from the primary somatosensory cortex along the ventral postcentral gyrus, immediately posterior to the motor articulator representations of the precentral gyrus. If somatosensory feedback falls outside the target, cells in the somatosensory error map become active resulting in the generation of a corrective motor command by the feedback control map.

**Lateralized Control Subsystems**

The view that the left hemisphere plays a dominant role in speech production has persisted for more than a century. Lesions to left inferior frontal and premotor cortex are far more likely to result in speech and language production deficits than those in the right hemisphere (see Duffy, 2005; Hillis, et al., 2004 for reviews). For a time, imaging data stood in conflict with this view; speech production was typically associated with bilateral frontal activity (e.g., Bohland & Guenther, 2006; Ozdemir, Norton, & Schlaug, 2006; Soros, et al., 2006; Wise, et al., 1999). There is growing evidence, however, that under normal speaking conditions, lateral premotor activity is indeed significantly greater in the left hemisphere (Ghosh, et al., 2008; Golfinopoulos, et al., 2011; Riecker, Ackermann, Wildgruber, Dogil, & Grodd, 2000; Sidtis, Gomez, Groshong, Strother, & Rottenberg, 2006; Tourville, et al., 2008). When an unanticipated auditory or somatosensory error is encountered, however, increases in premotor activity associated with compensatory responses are significantly greater in the right hemisphere (Golfinopoulos, et al., 2011; Tourville, et al., 2008).

In the DIVA model, these findings are expressed as lateralization of the feedforward and feedback inputs to motor cortex. Feedforward inputs derive from the speech sound map hypothesized to lie in left lateral premotor cortex.
Feedback-based commands, however, are sent to motor cortex from a feedback motor control map in lateral premotor cortex of the right hemisphere. According to the model, under normal feedback conditions the articulators are primarily under the feedforward motor programs stored in left lateral premotor cortex. When a sensory error is encountered, lateral premotor cortex of the right hemisphere is recruited to transform that error into corrective motor commands. The implication is that damage to the left lateral premotor region is more likely to disrupt stored feedforward speech motor commands, and is therefore more likely to result in disordered speech because the feedforward control system is more crucial for fluent speech than the sensory feedback control system (Neilson & Neilson, 1987).

There is additional clinical evidence to support lateralized feedforward and feedback control systems. Apraxia of speech, a disorder frequently associated with damage to left lateral premotor areas or posterior portions of the inferior frontal gyrus (Hillis, et al., 2004; Robin, Jacks, & Ramage, 2007) is typically characterized as involving impaired or missing motor programs for speech production. A speech disorder associated predominantly with right lateral premotor cortex and characterized by impaired feedback-based control has, thus far, not been identified. However, stuttering has been attributed to overreliance on feedback control (Max et al., 2004) and is associated with increased right lateral frontal activity (Brown, Ingham, Ingham, Laird, & Fox, 2005).

**Future Directions**

**Other Regions Involved in Sensorimotor Interactions during Speech**

The DIVA model does not account for all of the brain regions implicated in sensorimotor integration during speech production, especially areas where data are still rather uncertain and contradictory. For example, auditory and somatosensory feedback perturbations result in increased activity in the inferior intermediate cerebellar cortex (lobules VIIb and Villa; Golfinopoulos, et al., 2011; Tourville, et al., 2008). The intermediate cerebellum is structurally (Kelly & Strick, 2003; Schmahmann & Pandya, 1997) and functionally (O’Reilly, Beckmann, Tomassini, Ramnani, & Johansen-Berg, 2009) connected with motor, premotor, and sensory cortices and is known to play a crucial role in sensorimotor learning (Blakemore, Frith, & Wolpert, 2001; Criscimagna-Hemminger, Bastian, & Shadmehr, 2010; Diedrichsen, Hashambhoy, Rane, & Shadmehr, 2005; Smith & Shadmehr, 2005; Tseng, Diedrichsen, Krakauer, Shadmehr, & Bastian, 2007) and it contributes to movements of the vocal articulators (Grod, Hulsmann, Lotze, Wildgruber, & Erb, 2001). The area found active when feedback was perturbed during speech has also been implicated in sensory error processing for reaching movements (Diedrichsen, et al., 2005). It is unknown, however, what specific role this area plays during feedback-based corrective movements during speech. It may contribute to the detection of sensory errors or to the calculation of the corrective motor command. There is evidence for both roles but a clear understanding of the contribution of the inferior cerebellum to sensorimotor integration awaits further study.

Studies have also found activity in the anterior cingulate gyrus and adjacent medial motor areas associated with auditory feedback monitoring during speech (Christoffels, et al., 2007; Fu, et al., 2006) and with somatosensory-based feedback control of speech (Golfinopoulos, et al., 2011). This region of the brain has been associated with a general role in conflict monitoring and error detection (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Fiehler, Ullsperger, & von Cramon, 2004; van Veen, Cohen, Botvinick, Stenger, & Carter, 2001). A straightforward interpretation of its contribution to error correction in speech is hindered by the absence of activity in this area
during auditory feedback perturbation (Tourville, et al., 2008; Toyomura, et al., 2007). Study design may explain the discrepancy (e.g., whether the error is expected or not, whether the error falls above or below conscious awareness) but, as with the cerebellum, further research is required before a detailed account of this region’s contribution to sensorimotor interactions during speech is possible.

Effects of Motor Experience on Speech Perception

So far we have focused on how interactions between the sensory and motor systems influence speech production. Sensory feedback is used to correct ongoing movements and to tune speech motor programs. Recent work has suggested that motor experience can influence speech perception (see Houde, 2009 for review). For example, motor adaptation to a perturbation of auditory feedback during /s/ productions resulted in a concomitant shift in the perception of /s/ (Shiller, Sato, Gracco, & Baum, 2009). The perceptual shift acted to reduce the impact of the perturbation (i.e., the perturbed /s/ sounded more “/s/-like” to the subjects following motor adaptation). Perception was not shifted in a control group that passively listened to recordings of the perturbed utterances, indicating that the perceptual shift was not simply caused by exposure to the perturbed /s/ but was rather mediated by motor adaptation. Indeed, changes in auditory perceptual boundaries can be induced by motor adaptation to a somatosensory perturbation that does not alter auditory feedback (Nasir & Ostry, 2009). Subjects were asked to identify auditory tokens as had or head before and after speakers adapted to a jaw displacement (protrusion). Tokens were drawn from a continuum between had and head that varied only by the first and second formant values of the vowel. Those subjects that adapted to the perturbation were more likely to identify tokens in the middle of the continuum as head following motor learning. Subjects that did not adapt (5 of 23) and those in a control group that performed the experiment without the jaw perturbation did not demonstrate the perceptual shift.

These intriguing findings suggest that persistent errors encountered during adaptation experiments cause updates to the motor programs and the perceptual boundaries for a speech sound. The influence of motor experience on speech perception implies a tighter coupling than any of the previously discussed speech theories provide.

Integration of Somatosensory and Auditory Feedback

For simplicity, the DIVA model currently uses independent auditory and somatosensory channels. Feedback error is calculated within each channel and transformed into corrective commands in the feedback control map in right ventral premotor cortex. There is evidence, however, suggesting that the auditory and somatosensory systems interact during speech, influencing production and perception. A study showing a shift in auditory perception after motor adaptation to a somatosensory perturbation was described in the previous section (Nasir & Ostry, 2009). The influence of somatosensory feedback on auditory perception need not be through motor adaptation. Ito, Tiede, and Ostry (2009) found that perception of tokens on a continuum from had to head shifted when the facial skin of subjects was stretched while listening to the tokens. The direction of the perceptual shift depended on the pattern of the skin stretch: when the skin stretch resembled the pattern associated with production of head, perception was biased toward head; when it resembled had production, perception was biased toward had. Thus, it seems that somatosensory afferents modulate speech perception in normal-hearing individuals.

If somatosensory and auditory feedback systems are not independent, where are they integrated in the brain? A
possible candidate is the medial portion of the planum temporale. This region, traditionally considered a unimodal auditory area, lies adjacent to the secondary somatosensory area of the opercular portion of the supramarginal gyrus, the parietal operculum. It was noted previously that the posterior portion of medial planum temporale is active during passive auditory speech perception and speech production, even if the articulation is silent (e.g., Buchsbaum, et al., 2005; Hickok, et al., 2003). Overlapping activity in this area has also been noted for overt speech and nonspeech movements of the jaw and tongue (Dhanjal, Handunnetthi, Patel, & Wise, 2008).

According to the DIVA model, this area is the location of the auditory target and error maps (i.e., it receives inputs from low-level auditory cortex and from the motor system; Guenther, et al., 2006). Others have proposed a similar role for this region as an auditory-motor interface during speech (e.g., Hickok & Poeppel, 2007; Rauschecker, 2010; Wise, et al., 2001).

There is increasing evidence that somatosensory feedback is also sent to medial planum temporale by inputs from lateral parietal cortex. Projections from cortical and subcortical somatosensory areas to auditory cortex have been shown in rodents (Budinger & Scheich, 2009) and nonhuman primates (Cappe & Barone, 2005; Hackett, et al., 2007; Smiley, et al., 2007). Modulation of auditory cortical neurons in macaques by various forms of somatosensory stimulation has been demonstrated, providing functional evidence of these inputs (Brosch, Selezneva, & Scheich, 2005; Fu, et al., 2003; Lakatos, Chen, O’Connell, Mills, & Schroeder, 2007; Lemus, Hernandez, Luna, Zainos, & Romo, 2010; Schroeder, et al., 2001).

Several studies showing articulator movement-related activity in medial planum temporale have been discussed. As in the macaque, somatosensory stimulation alone, in the absence of movement, can also modulate activity in this region in humans (Beauchamp, Yasar, Frye, & Ro, 2008; Foxe, et al., 2002; Schurmann, Caetano, Hlushchuk, Jousmaki, & Hari, 2006). The anatomic substrates underlying somatosensory inputs to auditory cortex in humans have not been fully characterized. However, recent diffusion tensor imaging studies have shown white matter projections between supramarginal gyrus and posterior superior temporal gyrus (Oishi, et al., 2008) and further indicated that these regions (and lateral premotor cortex) may be connected by the superior longitudinal fasciculus (Saur, et al., 2008; Saur, et al., 2010). Anatomical and physiological studies in macaques (e.g., Cappé & Barone, 2005; Falchier, et al., 2010; Ghazanfar & Lemus, 2010; Kayser, Logothetis, & Panzeri, 2010; Kayser, Petkov, Augath, & Logothetis, 2007; Smiley & Falchier, 2009) and functional imaging studies in humans (Lehmann, et al., 2006; Martuzzi, et al., 2007; Meyer, Baumann, Marchina, & Jancke, 2007; van Atteveldt, Formisano, Goebel, & Blomert, 2004) have also established modulatory visual input to auditory cortex. The “sensory” end of the sensory-motor interface identified in posterior medial planum temporale (Hickok, Okada, & Serences, 2009; Tourville & Guenther, 2010) may therefore be polysensory, an integration of relevant sensory information from the auditory, somatosensory, and visual systems (cf. Dhanjal, et al., 2008).

This work was supported by the National Institute on Deafness and other Communication Disorders (R01 DC02852 and R01 DC07683, Frank Guenther, PI).

References


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource

Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource

Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource

Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource

Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource

Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource


Find This Resource

---

**Jason A. Tourville**

Jason A. Tourville, Department of Speech, Language, and Hearing Sciences, Sargent College of Health and Rehabilitation Sciences, Boston University, and the Center for Cognitive Computation and Neural Technology, Boston University
Maya G. Peeva

Maya G. Peeva, Department of Speech, Language, and Hearing Sciences, Sargent College of Health and Rehabilitation Sciences, Boston University, and the Center for Cognitive Computation and Neural Technology, Boston University

Frank H. Guenther

Frank H. Guenther, Department of Speech, Language, and Hearing Sciences, Sargent College of Health and Rehabilitation Sciences; the Center for Cognitive Computation and Neural Technology; and the Department of Biomedical Engineering, Boston University