# Articulating: The Neural Mechanisms of Speech Production

Elaine Kearney[1] and Frank H. Guenther[1,2,3,4]

[1]*Department of Speech, Language, and Hearing Sciences, Boston University, 635 Commonwealth Avenue, Boston, MA 02215*

[2]*Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, MA 02215*

[3]*The Picower Institute for Learning and Memory, Massachusetts Institute of Technology, 43 Vassar Street, Cambridge, MA 02139*

[4]*Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, 149 13th Street, Charlestown, MA  02129*

**Corresponding Author:** Prof. Frank H. Guenther, Boston University, 677 Beacon Street, Boston, MA 02215. Phone: 617-353-5765. Fax: 617-353-7755. Email: guenther@bu.edu.

1

Word Count (including references): 8859

**Abstract**

Speech production is a highly complex sensorimotor task involving tightly coordinated processing across large expanses of the cerebral cortex. Historically, the study of the neural underpinnings of speech suffered from the lack of an animal model. The development of non-invasive structural and functional neuroimaging techniques in the late $20_{th}$ century has dramatically improved our understanding of the speech network. Techniques for measuring regional cerebral blood flow have illuminated the neural regions involved in various aspects of speech, including feedforward and feedback control mechanisms. In parallel, we have designed, experimentally tested, and refined a neural network model detailing the neural computations performed by specific neuroanatomical regions during speech. Computer simulations of the model account for a wide range of experimental findings, including data on articulatory kinematics and brain activity during normal and perturbed speech. Furthermore, the model is being used to investigate a wide range of communication disorders.

## 1. Introduction

Speech production is a highly complex motor act involving respiratory, laryngeal, and supraglottal vocal tract articulators working together in a highly coordinated fashion. Nearly every speech gesture involves several articulators – even an isolated vowel such as "ee" involves coordination of the jaw, tongue, lips, larynx, and respiratory system. Underlying this complex motor act is the *speech motor control system* that readily integrates auditory, somatosensory, and motor information represented in the temporal, parietal, and frontal cortex, respectively, along with associated sub-cortical structures, to produce fluent and intelligible speech – whether the speech task is producing a simple nonsense syllable or a single real word (Ghosh, Tourville, & Guenther, 2008; Petersen, Fox, Posner, Mintun, & Raichle, 1988; Sörös et al., 2006; Turkeltaub, Eden, Jones, & Zeffiro, 2002).

In *Speaking*, Levelt (1989) laid out a broad theoretical framework of language production from the conceptualization of an idea to the articulation of speech sounds. In comparison to linguistic processes, speech motor control mechanisms differ in a number of ways. They are closer to the neural periphery, more similar to neural substrates in non-human primates, and better understood in terms of neural substrates and computations. These characteristics have shaped the study of the speech motor control system from early work with non-human primates and more recently with neural modelling and experimental testing.

The present article takes a historical perspective in describing the neural mechanisms of speech motor control. We begin in the first section with a review of models and theories of speech production, outlining the state of the field in 1989 and introducing the DIVA model – a computational neural network that describes the sensorimotor interactions involved in articulator control during speech production (Guenther, 1995; Guenther, Ghosh, & Tourville, 2006). Taking a similar approach in the next section, we review the key empirical findings regarding the neural bases of speech production prior to 1989 and highlight the primary developments in cognitive neuroimaging that followed and transformed our ability to conduct non-invasive speech research in humans. The neural correlates of speech production are discussed in the context of the DIVA model; as a neural network, the model's components correspond to neural populations and are given specific anatomical regions that can then be tested against neuroimaging data. Data from experiments that investigated the neural mechanisms of auditory feedback control are presented to illustrate how the model quantitatively fits to both behavioural and neural data. In the final section, we demonstrate the utility of neurocomputational models in furthering the scientific understanding of motor speech disorders and informing the development of novel, targeted treatments for those who struggle to translate their message "from intention to articulation".

## 2.  Models and Theories of Speech Production

In summarizing his review of the models and theories of speech production, Levelt (1989, p. 452) notes that "There is no lack of theories, but there is a great need of convergence." This section first briefly reviews a number of the theoretical proposals that led to this conclusion, culminating with the influential task dynamic model of speech production, which appeared in print the same year as *Speaking*. We then introduce the DIVA model of speech production, which incorporates many prior proposals in providing a unified account of the neural mechanisms responsible for speech motor control.

### *State of the Field Prior to 1989*

One of the simplest accounts for speech motor control is the idea that each phoneme is associated with an articulatory target (e.g., MacNeilage, 1970) or a muscle length target (*e.g.* Fel'dman, 1966a, 1966b) such that production of the phoneme can be carried out simply by moving the articulators to that muscle/articulatory configuration. By 1989, substantial evidence against such a simple articulatory target view was already available, including studies indicating that, unlike some "higher-level" articulatory targets such as lip aperture, individual articulator positions often vary widely for the same phoneme depending on things like phonetic context (Daniloff & Moll, 1968; Kent, 1977; Recasens, 1989), external loads applied to the jaw or lip (Abbs & Gracco, 1984; Folkins & Abbs, 1975; Gracco & Abbs, 1985),  and simple trial-to-trial variability (Abbs, 1986).

The lack of invariance in the articulator positions used to produce phonemes prompted researchers to search for a different kind of phonemic "target" that could

account for speech articulations. An attractive possibility is that the targets are *acoustic* or *auditory* (e.g., Fairbanks, 1954). As Levelt and others note, however, such a view is left with the difficult problem of accounting for how the brain's neural control system for speech can achieve these targets, given that they must ultimately be achieved through muscle activations and articulator positions whose relationship to the acoustic signal is rather complex and incompletely understood. For example, if my second formant frequency is too low, how does my brain know that I need to move my tongue forward?

To overcome this problem, several models postulate that auditory targets may be equated to targets in a somatosensory reference frame that is more closely related to the articulators than an acoustic reference frame (e.g., Lindblom, Lubker, & Gay, 1979; Perkell, 1981). For example, Lindblom et al. (1979) proposed that the area function of the vocal tract (i.e., the 3D shape of the vocal tract "tube"), which largely determines its acoustic properties, acts as a proxy for the auditory target that can be sensed through somatic sensation. Furthermore, they posit that the brain utilises an *internal model* that can estimate the area function based on somatosensory feedback of articulator positions and generate corrective movements if the estimated area function mismatches the target area function.

Published in the same year as Levelt's landmark book, the *task dynamic* model (Saltzman & Munhall, 1989) provided a fleshed-out treatment of vocal tract shape targets. According to this model, the primary targets of speech are the locations and degrees of key constrictions of the vocal tract (which dominate the acoustic signal

compared to less-constricted parts of the vocal tract), specified within a time-varying *gestural score*. The model was mathematically specified and simulated on a computer to verify its ability to achieve constriction targets using different combinations of articulators in different speaking conditions.

The task dynamic model constitutes an important milestone in speech modelling and continues to be highly influential today. However, it does not account for several key aspects of speech: for example, the model is not neurally specified, it does not account for development of speaking skills (all parameters are provided by the modeller rather than learned), and it does not account for auditory feedback control mechanisms such as those responsible for compensatory responses to purely auditory feedback manipulations. The DIVA model introduced in the following subsection addresses these issues by integrating past proposals such as auditory targets, somatosensory targets, and internal models into a relatively straightforward, unified account of both behavioural and neural findings regarding speech production.

### *The DIVA Model*

Since 1992, our laboratory has developed, tested, and refined an adaptive neural network model of the brain computations underlying speech production called the *Directions Into Velocities of Articulators* (DIVA) model (e.g., Guenther, 1994; Guenther, 1995, 2016; Guenther et al., 2006; Guenther, Hampson, & Johnson, 1998). This model combines a control theory account of speech motor control processes with a neurocomputational description of the roles played by the various cortical and subcortical regions involved in

speech production. In the current subsection, we briefly relate the DIVA model to the stages of word production proposed in the model of Levelt and colleagues, followed by a description of the control structure of the model. A description of the model's neural substrates is provided in the following section.

The model of word production proposed by Levelt (1989) begins at *conceptual preparation*, in which the intended meaning of an utterance is initially formulated. This is followed by a *lexical selection* stage, in which candidate items in the lexicon, or lemmas, are identified. The chosen lemmas must then be translated into morphemes (*morphological encoding*) and then into sound units for production (*phonological encoding*). The output of the phonological encoding stage is a set of syllables chosen from a *mental syllabary*. The DIVA model's input is approximately equivalent to the output of the phonological encoding stage. The DIVA model then provides a detailed neural and computational account of Levelt's *phonetic encoding* and *articulation* stages, as detailed in the following paragraphs.

The control scheme utilised by the DIVA model is depicted in Figure 1. The DIVA controller utilises information represented in three different reference frames: a motor reference frame, an auditory reference frame, and a somatosensory reference frame. Mathematical treatments of the model are included elsewhere (e.g., Guenther, 2016; Guenther et al., 2006); here we present a qualitative treatment for brevity.

Production of a speech sound (which can be a frequently produced phoneme, syllable, or word) starts with activation of the sound's neural representation in a *speech*

*sound map* hypothesised to reside in the left ventral premotor cortex. It is useful to think of the output of the phonological encoding stage of Levelt's (1989) framework as the input to the speech sound map in DIVA. In Levelt's framework, these inputs take the form of syllables from the mental syllabary. Although DIVA similarly assumes that the typical form of the inputs is syllabic, the model also allows for larger multi-syllabic chunks in frequently produced words (which can be reconciled with Levelt's view by assuming that the motor system recognises when it has a motor program for producing consecutive syllables specified by the phonological encoding stage) as well as individual phonemes in the speech sound map that are necessary for producing novel syllables.

Activation of a speech sound map node leads to the readout of a learned set of motor commands for producing the sound, or *motor target*, along with *auditory* and *somatosensory targets* which represent the desired states of the auditory and somatosensory systems for producing the current sound. The motor target can be thought of as the sound's "motor program" and consists of a sequence of articulatory movements that have been learned to produce. The *feedforward controller* compares this motor target to an internal estimate of the current motor state to generate a time series of articulator velocities (labelled $\dot{M}_{FF}$ in Figure 1) which move the speech articulators to produce the appropriate acoustic signal for the sound. The feedforward command is summed with sensory feedback-based commands arising from *auditory* and *somatosensory feedback controllers* to generate the overall motor command ($\dot{M}$ in the figure) to the vocal tract musculature.

The auditory and somatosensory feedback controllers act to correct any production errors sensed via auditory or somatosensory feedback sent to the cerebral cortex by comparing this feedback with the auditory and somatosensory targets for the sound. These targets represent learned sensory expectations that accompany successful productions of the sound. If the current sensory feedback mismatches the sensory target, the appropriate feedback controller transforms this sensory error into motor corrective commands (labelled $\dot{M}_S$ and $\dot{M}_A$ in Figure 1) that act to decrease the sensory error within the current production. These corrective commands also act to update the motor target for future productions (indicated by dashed arrows in Figure 1) to (partially) incorporate the corrective movements.

The auditory feedback controller in the DIVA model is, in essence, an instantiation of Fairbanks' (1954) model of speech production as an auditory feedback control process. Although both the Levelt framework and DIVA model utilise auditory feedback for error monitoring, the Levelt framework focuses on the detection of phonological/phonetic errors at the level of discrete phonological entities such as phonemes, whereas DIVA focuses on its use for lower-level tuning of speech motor programs with little regard for their higher-level phonological structure.

A number of researchers have noted that the delays inherent in the processing of auditory feedback preclude the use of purely feedback control mechanisms for speech motor control; this motivates the DIVA model's use of a feedforward mechanism that generates learned articulatory gestures which are similar to the *gestural score* of the task

11

dynamic model (Saltzman & Munhall, 1989). DIVA's somatosensory feedback controller is essentially an implementation of the proposal of Perkell (1981) and Lindblom et al. (1979) that speech motor control involves desired somatosensory patterns that correspond to the auditory signals generated for a speech sound. DIVA unifies these prior proposals in the form of a straightforward control mechanism that explicates the interactions between auditory, somatosensory, and motor representations, while at the same time characterizing the neural substrates that implement these interactions, as described in the next section.

### 3.  Neural Bases of Speech Production

In this section, we begin by reviewing studies prior to 1989 that informed our early knowledge of the neural bases of speech production, specifically focusing on the control of vocalizations in nonhuman primates and early human work based on lesion and electrical stimulation studies. Following this review, we highlight key developments in cognitive neuroimaging and its subsequent application in identifying neural correlates and quantitative fits of the DIVA model.

### *State of the Field Prior to 1989*

When *Speaking* was released in 1989, our knowledge of the neural bases of speech relied heavily on work with nonhuman primates as well as a limited number of human studies that reported the effects of brain lesions and electrical stimulation on speech production. From an evolutionary perspective, the neural bases of speech production have developed from the production of learned voluntary vocalizations in our evolutionary predecessors, namely nonhuman primates. The study of these neural mechanisms in nonhuman primates has benefitted from a longer history than speech production in humans due to the ability to use more invasive methods, such as single-unit electrophysiology, focal lesioning, and axonal trackers. From works conducted throughout the twentieth century, a model for the control of learned primate vocalization was developed (Jürgens & Ploog, 1970; Müller-Preuss & Jürgens, 1976; Thoms & Jürgens, 1987). Figure 2, adapted from Jürgens (2009), illustrates the brain regions and axonal tracts involved in the control of learning primate vocalization, primarily based on studies of squirrel monkeys.

Based on this model, two hierarchically organised pathways converge onto the reticular formation (RF) of the pons and medulla oblongata and subsequently the motoneurons that control muscles involved in respiration, vocalization, and articulation (Jürgens & Richter, 1986; Thoms & Jürgens, 1987). The first pathway (limbic) follows a route from the anterior cingulate cortex to the RF via the periaqueductal grey matter and controls motivation or readiness to vocalise by means of a gating function, allowing commands from the cerebral cortex to reach the motor periphery through the RF. As such, this pathway controls the initiation and intensity of vocalization, but not the control of muscle patterning or the specific acoustic signature (Düsterhöft, Häusler, & Jürgens, 2004; Larson, 1991). The second pathway (motor cortical) maps from the motor cortex to the RF and generates the final motor command for the production of learned vocalizations. In the motor cortex, distinct areas represent the oral and laryngeal muscles (Jürgens & Ploog, 1970), and these areas integrate with two feedback loops involving subcortical structures that preprocess the motor commands: a loop through the pons, cerebellum, and thalamus, and a loop through the putamen, pallidum, and thalamus. Together, the components of the motor cortical pathway control the specific pattern of vocalization.

Many cortical areas beyond the primary motor cortex, however, are involved in speech production, and before neuroimaging little was known about their role due to an inability to measure brain activity non-invasively in humans. Some of the earliest evidence of localization of speech and language function stemmed from cortical lesion studies of patients with aphasias or language difficulties. This work, pioneered by Paul

14

Broca and Carl Wernicke in the nineteenth century, associated different brain regions to a loss of language function. Two seminal papers by Broca suggested that damage to the inferior frontal gyrus of the cerebral cortex was related to impaired speech output and that lesions of the left hemisphere, but typically not the right, interfered with speech (Broca, 1861, 1865). The patients described in these papers primarily had a loss of speech output but had a relatively spared ability to perceive speech; a pattern of impairment that was termed Broca's aphasia. Shortly thereafter, Wernicke identified a second brain region associated with another type of aphasia, sensory or Wernicke's aphasia, which is characterised by poor speech comprehension and relatively preserved and fluent speech output (Wernicke, 1874). This sensory aphasia was associated with lesions to the posterior portion of the superior temporal gyrus in the left cerebral hemisphere.

Lesion studies, however, are not without their limitations, and as a result can be very difficult to interpret. First, it is uncommon for different patients to share precisely the same lesion location in the cortex. Second, lesions often span multiple cortical areas affecting different neural systems making it challenging to match a brain region to a specific functional task. Third, it's possible for spared areas of the cortex to compensate for lesioned areas, potentially masking the original function of the lesion site. Finally, a large amount of variation may occur in the location of a particular brain function across individuals, especially for higher-level regions of the cortex, as is evident for syntactic processing (Caplan, 2001).

More direct evidence for the localization of speech function in the brain came from electrical stimulation studies conducted with patients who were undergoing excision surgery for focal epilepsy in the 1930s to 1950s. Wilder Penfield and colleagues at the Montreal Neurological Institute delivered short bursts of electrical stimulation via an electrode to specific locations on the cerebral cortex while patients were conscious and then recorded their behavioural responses and sensations (Penfield & Rasmussen, 1950; Penfield & Roberts, 1959). By doing so, they uncovered fundamental properties of the functional organization of the cerebral cortex. Specifically, they showed evidence of somatotopic organization of the body surface in the primary somatosensory and motor cortices, and these representations included those of the vocal tract. Stimulation of the postcentral gyrus (primary somatosensory cortex) was found to elicit tingling, numbness or pulling sensations in various body parts, sometimes accompanied by movement, while stimulation of the postcentral gyrus (primary motor cortex) elicited simple movements of the body parts. Using this method, the lips, tongue, jaw and laryngeal system were localised to the ventral half of the lateral surface of the postcentral and precentral gyri.

At this point in history, we had some idea of which cortical areas were involved in the control of speech production. As reviewed above, studies of nonhuman primates and human lesion and electrical stimulation studies provided evidence for the role of the motor cortex, the inferior frontal gyrus, the superior temporal gyrus, and the somatosensory cortex. Little evidence, however, had yet to emerge regarding the differentiated functions of these cortical areas.

16

*Cognitive Neuroimaging*

The advent of cognitive neuroimaging in the late 1980s changed the landscape of speech research and for the first time, it was possible to conduct neurophysiological investigations of the uniquely human capacity to speak in a large number of healthy individuals. The first technology harnessed for the purpose of assessing brain activity during a speech task was positron emission tomography (PET) (Petersen et al., 1988). PET detects gamma rays emitted from radioactive tracers injected into the body and can be used to measure changes in regional cerebral blood flow, which is indicative of local neural activity. An increase in blood flow to a region, or the hemodynamic response, is associated with that region's involvement in the task. Petersen et al. (1988) examined the hemodynamic response during a single word production task with the words presented auditorily or visually, and showed increased activity in motor and somatosensory areas along the ventral portion of the central sulcus, the superior temporal gyrus, and the supplementary motor area; thus, replicating earlier cortical stimulation studies (e.g., Penfield & Roberts, 1959).

Further advancements in the 1990s led to magnetic resonance imaging (MRI) technology being employed to measure the hemodynamic response, a method known as functional MRI (fMRI). Compared to PET, fMRI has some key advantages: (1) fMRI does not require the use of radioactive tracer injections, and (2) fMRI facilitates the collection of structural data for localization in the same scan as functional data. As a result, a large number of fMRI studies of speech and language have been performed in the last two decades. It is the widespread availability of neuroimaging technology that has

17

made it possible to develop neurocomputational models that explicitly make hypotheses that can be tested and refined if necessary based on the experimental results. This scientific approach leads to a much more mechanistic understanding of the functions of different brain regions involved in speech.

Figure 3 illustrates cortical activity during simple speech production tasks such as reading single words aloud as measured using fMRI. High areas of cortical activity are observed in anatomically and functionally distinct areas. These include the precentral gyrus (known functionally as the motor and premotor cortex), inferior frontal gyrus, anterior insula, postcentral gyrus (somatosensory cortex), Heschl's gyrus (primary auditory cortex), and superior temporal gyrus (higher-order auditory cortex).

In the early 2000s, MRI technology was again harnessed for the study of speech production, with advanced methods being developed for voxel-based lesion-symptom mapping (VLSM; Bates et al., 2003). VLSM analyses the relationship between tissue damage and behavioural performance on a voxel-by-voxel basis in order to identify the functional architecture of the brain. In contrast to fMRI studies conducted with healthy individuals that highlight areas of brain activity during a particular behaviour, VLSM may identify brain areas that are critical to that behaviour. Using this approach, a number of studies have provided insights into the neural correlates of disorders of speech production. For example, lesions in the left precentral gyrus are associated with apraxia of speech (Baldo, Wilkins, Ogar, Willock, & Dronkers, 2011; Itabashi et al., 2016), damage to the paravermal and hemispheric lobules V and V1 in the cerebellum is

associated with ataxic dysarthria (Schoch, Dimitrova, Gizewski, & Timmann, 2006), and the cortico-basal ganglia-thalamo-cortical loop is implicated in neurogenic stuttering (Theys, De Nil, Thijs, Van Wieringen, & Sunaert, 2013). Speech disorders are considered in further detail in Section 5.

The following subsections describe the computations performed by the cortical and subcortical areas involved in speech production according to the DIVA model, including quantitative fits of the model to relevant behavioural and neuroimaging experimental results.

### *Neural Correlates of the DIVA Model*

A distinctive feature of the DIVA model is that all of the model components have been associated with specific anatomical locations and localised in Montreal Neurological Institute space, allowing for direct comparisons between model simulations and experimental results. These locations are based on synthesised findings from neurophysiological, neuroanatomical, and lesion studies of speech production (see Guenther, 2016; Guenther et al., 2006). Figure 4 shows the neural correlates of the DIVA model. Each box represents a set of model nodes that together form a neural map that is associated with a specific type of information in the model. Cortical regions are indicated by large boxes and subcortical regions by small boxes. Excitatory and inhibitory axonal projections are denoted by arrows and lines terminating in circles, respectively. These projections transform neural information from one reference frame into another.

Each speech sound map node, representing an individual speech sound, is hypothesised to correspond to a cluster of neurons located primarily in the left ventral premotor cortex. This area includes the rostral portion of the ventral precentral gyrus and nearby regions in the posterior inferior frontal gyrus and anterior insula. When the node becomes activated in order to produce a speech sound, motor commands are sent to the motor cortex via both a feedforward control system and a feedback control system.

The feedforward control system generates previously learned motor programs for speech sounds in two steps. First, a cortico-basal ganglia loop is responsible for launching the motor program at the correct moment in time, which involves activation of an *initiation map* in the supplementary motor area located on the medial wall of the frontal cortex. Second, the motor programs themselves are responsible for generating feedforward commands for producing learned speech sounds. These commands are encoded by projections from the speech sound map to an *articulator map* in the ventral primary motor cortex of the precentral gyrus bilaterally. Further, these projections are supplemented by a cerebellar loop that passes through the pons, cerebellar cortex lobule VI, and the ventrolateral nucleus of the thalamus.

The auditory feedback control subsystem involves axonal projections from the speech sound map to the *auditory target map* in the higher-order auditory cortical areas in the posterior auditory cortex. These projections encode the intended auditory signal for the speech sound being produced and thus can be compared to incoming auditory information from the auditory periphery via the medial geniculate nucleus of the

thalamus that is represented in the model's *auditory state map*. The targets are time-varying regions that allow a degree of variability in the acoustic signal during a syllable, rather than an exact, singular point (Guenther, 1995). If the current auditory feedback is outside of this target region, the *auditory error map* in the posterior auditory cortex is activated, and this activity transforms into corrective motor commands through projections from the auditory error nodes to the *feedback control map* in the right ventral premotor cortex, which in turn projects to the articulator map in the ventral motor cortex.

The somatosensory feedback control subsystem works in parallel with the auditory subsystem. We hypothesise that the main components are located in the ventral somatosensory cortex, including the ventral postcentral gyrus and adjoining supramarginal gyrus. Projections from the speech sound map to the *somatosensory target map* encode the intended somatosensory feedback to be compared to the *somatosensory state map*, which represents current proprioceptive information from the speech articulators. The somatosensory feedback arrives from cranial nerve nuclei in the brain stem via the ventral posterior medial nucleus of the thalamus. Nodes in the *somatosensory error map* are activated if there is a mismatch between the intended and current somatosensory states and, as for the auditory subsystem, this activation transforms into corrective motor commands via the feedback control map in right ventral premotor cortex.

## 4. Quantitative Fits to Behavioural and Neural Data

The DIVA model provides a unified explanation of a number of speech production phenomena and as such can be used as a theoretical framework to investigate both normal and disordered speech production. Predictions from the model have guided a series of empirical studies and, in turn, the findings have been used to further refine the model. These studies include, but are not limited to, investigations of sensorimotor adaptation (Villacorta, Perkell, & Guenther, 2007), speech sequence learning (Segawa, Tourville, Beal, & Guenther, 2015), somatosensory feedback control (Golfinopoulos et al., 2011), and auditory feedback control (Niziolek & Guenther, 2013; Tourville, Reilly, & Guenther, 2008). Here, we will focus on studies of auditory feedback control to illustrate quantitative fits of the DIVA model to behavioural and neural data.

To recap, the DIVA model posits that axonal projections from the speech sound map in the left ventral premotor cortex to higher-order auditory cortical areas encode the intended auditory signal for the speech sound currently being produced. This auditory target is compared to incoming auditory information from the periphery, and if the auditory feedback is outside the target region, neurons in the auditory error map in the posterior auditory cortex become active. This activation is then transformed into corrective motor commands through projections from the auditory error map to the motor cortex via a feedback control map in left inferior frontal cortex. Once the model has learned feedforward commands for a speech sound, it can correctly produce the sound without depending on auditory feedback. However, if an unexpected perturbation occurs, such as real-time manipulations imposed on auditory feedback so that the subject perceives themselves as producing the incorrect sound, the auditory error map will

22

become active and try to correct for the perturbation. Such a paradigm allows the testing of the DIVA model's account of auditory feedback control during speech production.

To test these model predictions regarding auditory feedback control, we performed two studies involving auditory feedback perturbations during speech in an MRI scanner to measure subject responses to unexpected perturbations both behaviourally and neurally (Niziolek & Guenther, 2013; Tourville et al., 2008). In both studies, speakers produced monosyllabic utterances; on 25% of trials, the first and/or second formant frequency was unexpectedly perturbed via a digital signal processing algorithm in near real-time (Cai, Boucek, Ghosh, Guenther, & Perkell, 2008; Villacorta et al., 2007). The formant shifts have an effect of moving the perceived vowel towards another vowel in the vowel space. In the DIVA model, this shift would result in auditory error signals and subsequent compensatory movements of the articulators. Analysis of the acoustic signal indicated that in response to unexpected shifts in formants, participants had rapid compensatory responses within the same syllable as the shift. Figure 5 shows that productions from the DIVA model in response to the perturbations fall within the distribution of productions of the speakers, supporting the model's account of auditory feedback control of speech.

Neuroimaging results in both studies highlighted the neural circuitry involved in compensation. During the normal feedback condition, neural activity was left-lateralised in the ventral premotor cortex (specifically, the posterior inferior frontal gyrus pars opercularis and the ventral precentral gyrus), consistent with the DIVA simulation

(Figure 6). During the perturbed condition, activity increased in both hemispheres in the posterior superior temporal cortex, supporting the DIVA model prediction of auditory error maps in these areas (Figure 7). Perturbed speech was also associated with an increase in ventral premotor cortex activity in the right hemisphere; in the model, this activity is associated with the feedback control map, which translates auditory error signals into corrective motor commands. Furthermore, structural equation modelling was used to examine effective connectivity within the network of regions contributing to auditory feedback control, and revealed an increase in effective connectivity from the left posterior temporal cortex to the right posterior temporal and ventral premotor cortices (Tourville et al., 2008), consistent with the model's prediction of a right-lateralised feedback control map involved in transforming auditory errors into corrective motor commands.

The DIVA model makes further predictions regarding how the auditory feedback controller interacts with the feedforward controller if perturbations are applied for an extended period of time (i.e., over many consecutive productions). Specifically, the corrective motor commands from the auditory feedback control subsystem will eventually update the feedforward commands so that, if the perturbation is removed, the speaker will show residual after-effects; i.e., the speaker's first few utterances after normal auditory feedback has been restored will show effects of the adaptation of the feedforward command in the form of residual "compensation" to the now-removed perturbation.

To test these predictions, we conducted a sensorimotor adaptation experiment using sustained auditory perturbation of F1 during speech (Villacorta et al., 2007). Subjects performed a speech production task with four phases during which they repeated a short list of words (one list repetition = one epoch): (1) a *baseline phase* where they produced 15 epochs with normal feedback; (2) a *ramp phase* of 5 epochs, over which a perturbation was gradually increased to a 30% shift from the baseline F1; (3) a *training phase* of 25 epochs where the perturbation was applied to every trial; and (4) a *posttest phase* where auditory feedback was returned to normal for the final 20 epochs. A measure of adaptive response, calculated as the percent change in F1 in the direction opposite the perturbation, is shown by the solid line connecting data points in Figure 8, along with the associated standard error bars. The data show evidence of adaptation during the hold phase, as well as the predicted after-effects in the posttest phase.

Simulations of the DIVA model were performed on the same adaptation paradigm, with one version of the model tuned to incorporate the auditory acuity of each subject. The dashed line shows the DIVA simulation results when modelling the subject with the lowest auditory acuity, and a bold solid line shows the simulation results for the subject with the best auditory acuity. Grey shaded regions represent the 95% confidence intervals derived from the model simulations across all subjects. Notably, with the exception of four epochs in the baseline phase (during which the subjects were more variable than the model), the model's productions were not statistically significantly different from the experimental results.

## 5.  Utility of Neurocomputational Models

An accurate neurocomputational model can provide us with mechanistic insights into speech disorders of neurological origin, which in turn can be used to better understand and, in the longer run, treat these communication disorders. For example, various "damaged" versions of the model can be created and simulated to see which one best corresponds to the behaviour and brain activity seen in a particular communication disorder. This knowledge provides insight into exactly what functionality is impaired and what is spared in the disorder, which in turn can guide the development of optimised therapeutic treatments for overcoming the impairment.

Figure 9 shows the components of the DIVA model associated with various speech disorders. Regardless of the aetiology of the disorder, the severity and nature of the speech impairment will depend on whether the neural damage affects feedforward control mechanisms, feedback control mechanisms, or a combination of the two. In a developing speech system, the feedback control system is central to tuning feedforward motor commands. Once developed, the feedforward commands can generate speech with little input from the feedback system. Damage to the feedback control system in mature speakers, therefore, will have limited effects on speech output (as evidenced by the largely preserved speech of individuals who become deaf in adulthood), whereas substantial damage to the feedforward control system will typically cause significant motor impairment. To date, the DIVA model has been considered with respect to a number of motor speech disorders, including dysarthria and apraxia of speech, as briefly summarised in the following paragraphs.

*Dysarthria* is an umbrella term for a range of disorders of motor execution characterised by weakness, abnormal muscle tone, and impaired articulation (Duffy, 2013). Dysarthria type varies by lesion site as well as by perceptual speech characteristics. For example, *ataxic dysarthria* (AD in Figure 9) is associated with damage to the cerebellum and results in uncoordinated and poorly timed articulations, often characterised by equal stress on syllables and words, irregular articulatory breakdowns, vowel distortions, and excess loudness variations (Darley, Aronson, & Brown, 1969). In the DIVA model, the cerebellum has a number of important roles in speech motor control, which can account for the speech characteristics of the disorder. First, the cerebellum plays an essential role in learning and generating finely timed, smoothly coarticulated feedforward commands to the speech articulators. Damage to this functionality is likely the main cause of motor disturbances in ataxic dysarthria. Second, the cerebellum is hypothesised to contribute to feedback control as it is involved in generating precisely timed auditory and somatosensory expectations (targets) for speech sounds, and it is also likely involved in generating corrective commands in response to sensory errors via projections between the right premotor and bilateral primary motor cortices. Damage to the cerebellum, therefore, is expected to affect both the feedforward and feedback control systems according to the DIVA model (but see Parrell, Agnew, Nagarajan, Houde, & Ivry, 2017).

Two types of dysarthria are associated with impaired function of the basal ganglia – hypokinetic and hyperkinetic dysarthria (HoD and HrD, respectively, in Figure 9). HoD commonly occurs in individuals with Parkinson's disease, a neurodegenerative disease

that involves depletion of striatal dopamine, and results in monopitch and loudness, imprecise consonants, reduced stress, and short rushes of speech (Darley et al., 1969). The effect of dopamine depletion is twofold in that it weakens the direct pathway involved in facilitating motor output and strengthens the indirect pathway involved in inhibiting motor output. The sum effect is a reduction in articulatory movements, decreased pitch and loudness range, and delays in initiation and ending of movements. The DIVA model accounts for these changes in the initiation circuit; here, underactivation results in initiation difficulties and a reduced GO signal that controls movement speed. HrD occurs in individuals with Huntington's disease and is perceptually recognised by a harsh voice quality, imprecise consonants, distorted vowels, and irregular articulatory breakdowns (Darley et al., 1969). In contrast to Parkinson's disease, HrD appears to involve a shift in balance away from the indirect pathway and toward the direct pathway, resulting in abnormal involuntary movements of the speech articulators, which corresponds to an overactive initiation circuit in the DIVA model.

*Apraxia of speech* (AOS in Figure 9) is a disorder of speech motor planning and programming that is distinct from both dysarthria (in that it does not involve muscle weakness) and aphasia (in that it does not involve language impairment; Duffy, 2013). It can occur developmentally, known as *childhood apraxia of speech*, or as a result of stroke, traumatic brain injury, or neurodegenerative disease, such as primary progressive apraxia of speech. It is most often associated with damage to the left inferior frontal gyrus, anterior insula, and/or ventral precentral gyrus. According to the DIVA model, damage to these areas affects the speech sound map and thus the representations of

frequently produced sound sequences. The speech sound map is a core component of the motor programs for these sequences, so damage to it will strongly affect the feedforward commands for articulating them, in keeping with the characterization of apraxia of speech as an impairment of speech motor programming. It's also plausible, according to the model, that such damage may affect the readout of sensory expectations for these sound sequences to higher-order auditory and somatosensory cortical areas, leading to impaired feedback control mechanisms that compare the expected and realised sensory information, though this proposition has not been thoroughly tested experimentally. Recently, Ballard and colleagues (2018) published the first investigation of adaptive (feedforward) and compensatory (feedback) responses in patients with apraxia of speech. Their results indicated an adaptive response to sustained perturbations in the first formant for the patient group but, surprisingly, not for age-matched controls. In addition, compensatory responses to pitch perturbations were considered normal for both groups. While these results are in contrast to the DIVA model predictions, further studies are needed to understand the relationship between the extent of damage to speech sound map areas and the control of speech. Methodological differences between this study and previous adaptation studies also warrant further investigation to elucidate the role of feedforward and feedback control in this population.

We end this section with a brief treatment of a striking example of how neurocomputational models can help guide the development of therapeutic technologies, in this case developing a speech neural prosthesis for individuals with *locked-in syndrome*, which is characterised by a total loss of voluntary movement but intact

cognition and sensation. Insights from the DIVA model were used to guide the development of a brain-computer interface (BCI) that translated cortical signals generated during attempted speech in order to drive a speech synthesiser that produced real-time audio feedback (Guenther et al., 2009). The BCI utilised an intracortical electrode (Bartels et al., 2008; Kennedy, 1989) permanently implanted in the speech motor cortex of a volunteer with locked-in syndrome. A schematic of the system, interpreted within the DIVA model framework, is provided in Figure 10. The input to the BCI was derived from motor/premotor cortical neurons that are normally responsible for generating speech movements, in essence replacing the motor periphery that was no longer functional due to a brain stem stroke. The BCI produced audio output that was transduced by the participant's intact auditory system; this auditory feedback could be compared to the desired auditory signal (auditory target) for the sounds being produced since the neural circuitry for generating the auditory targets was also intact (auditory target pathway in Figure 10). The BCI was limited to producing vowel-like sounds by controlling the values of the first three formant frequencies of a formant synthesiser.

The system capitalised on two key insights derived from the DIVA model. The first insight was that it should be possible to decode the intended formant frequencies for vowels that the participant was attempting to produce. This is because the implanted area, at the border between premotor and primary motor cortex in the left hemisphere, is believed to be involved in the generation of feedforward motor commands that are intended to reach the auditory target for the vowel. Statistical analysis of the neural firing patterns during an attempted production of a vowel sequence verified this prediction. The

detected formant frequencies were sent to a formant synthesiser that produced corresponding acoustic output within approximately 50 ms of neural firing, which is similar to the delay between neural firing and sound output in a healthy talker.

The second insight was that the participant should be able to use real-time auditory feedback of his attempted productions to improve his performance with practice. This is because the participant's auditory feedback control system was fully intact (see Figure 10), allowing his brain to iteratively improve its (initially poor) feedforward motor programs for producing vowels with the BCI by detecting (through audition) and correcting (through the BCI) production errors as detailed in Section 4. This prediction was also verified; the BCI user was able to significantly improve his success rate in reaching vowel targets as well as his endpoint error and movement time from the first 25% of trials to the last 25% of trials in a session (see Guenther et al., 2009, for details).

Although the speech output produced by the participant with locked-in syndrome using the speech BCI was rudimentary – consisting only of vowel-to-vowel movements that were substantially slower and more variable than normal speech – it is noteworthy that this performance was obtained using only 2 electrode recording channels. Future speech BCIs can take advantage of state-of-the-art systems with 100 or more electrode channels, which should allow far better control of a speech synthesiser than the 2-channel system used by Guenther et al. (2009), providing the promise for an eventual system that can restore conversational speech capabilities to those suffering from locked-in syndrome.

## 6. Concluding Remarks

This article has mapped a brief history of research into speech motor control before and after the publication of Levelt's *Speaking.* At the time of publication, a number of distinct theories of speech motor control had been proposed (and their limitations debated). Levelt laid out a broad theoretical framework that would guide speech and language research for the next 30 years, leading to ever more sophisticated quantitative models of linguistic processes. In parallel, the advent of new technologies – particularly cognitive neuroimaging – accelerated our ability to non-invasively study the areas of the brain involved in both normal and disordered speech motor control. These technological advances have supported the development and experimental testing of neurocomputational models of speech production, most notably the DIVA model, which has been used to provide a unified account of a wide range of neural and behavioural findings regarding speech motor control. This in turn is leading to a better understanding of motor speech disorders, setting the stage for the creation of novel, targeted treatments for these disorders.

# References

Abbs, J. H. (1986). Invariance and variability in speech production: A distinction between linguistic intent and its neuromotor implementation. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 202-219). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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.

Baldo, J. V., Wilkins, D. P., Ogar, J., Willock, S., & Dronkers, N. F. (2011). Role of the precentral gyrus of the insula in complex articulation. *Cortex, 47*(7), 800-807.

Ballard, K. J., Halaki, M., Sowman, P. F., Kha, A., Daliri, A., Robin, D., . . . Guenther, F. (2018). An investigation of compensation and adaptation to auditory perturbations in individuals with acquired apraxia of speech. *Frontiers in Human Neuroscience, 12*(510), 1-14. doi:10.3389/fnhum.2018.00510

Bartels, J., Andreasen, D., Ehirim, P., Mao, H., Seibert, S., Wright, E. J., & Kennedy, P. (2008). Neurotrophic electrode: method of assembly and implantation into human motor speech cortex. *Journal of Neuroscience Methods, 174*(2), 168-176. doi:10.1016/j.jneumeth.2008.06.030

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronkers, N. F. (2003). Voxel-based lesion–symptom mapping. *Nature Neuroscience, 6*(5), 448. doi:10.1038/nn1050

Broca, P. (1861). Remarks on the seat of the faculty of articulated language, following an observation of aphemia (loss of speech). *Bulletin de la Société Anatomique, 6*, 330-357.

Broca, P. (1865). Sur le siège de la faculté du langage articulé (15 juin). *Bulletins de la Société Anthropologque de Paris, 6*, 377-393.

Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2008). A system for online dynamic perturbation of formant trajectories and results from perturbations of the Mandarin triphthong/iau. *Proceedings of the 8th ISSP*, 65-68.

Caplan, D. (2001). Functional neuroimaging studies of syntactic processing. *Journal of Psycholinguistic Research, 30*(3), 297-320.

Daniloff, R., & Moll, K. (1968). Coarticulation of lip rounding. *Journal of Speech, Language, and Hearing Research, 11*(4), 707-721. doi:10.1044/jshr.1104.707

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech, Language, and Hearing Research, 12*(3), 462-496. doi:10.1044/jshr.1203.462

Duffy, J. R. (2013). *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management* (3rd ed.). St Louis, MO: Mosby.

Düsterhöft, F., Häusler, U., & Jürgens, U. (2004). Neuronal activity in the periaqueductal gray and bordering structures during vocal communication in the squirrel monkey. *Neuroscience, 123*(1), 53-60. doi:10.1016/j.neuroscience.2003.07.007

Fairbanks, G. (1954). Systematic research in experimental phonetics: 1. A theory of the

    speech mechanism as a servosystem. *Journal of Speech & Hearing Disorders, 19*,

    133-139. doi:10.1044/jshd.1902.133

Fel'dman, A. G. (1966a). Functional tuning of the nervous system with control of

    movement or maintenance of a steady posture-II. Controllable parameters of the

    muscles. *Biophysics, 11*, 565-578.

Fel'dman, A. G. (1966b). Functional tuning of the nervous system with control of

    movement or maintenance of a steady posture, III, Mechanographic Analysis of

    Execution by Man of the Simplest Motor Tasks. *Biophysics, 11*, 766-775.

Folkins, J. W., & Abbs, J. H. (1975). Lip and jaw motor control during speech:

    Responses to resistive loading of the jaw. *Journal of Speech, Language, and

    Hearing Research, 18*(1), 207-220. doi:10.1044/jshr.1801.207

Ghosh, S. S., Tourville, J. A., & Guenther, F. H. (2008). A neuroimaging study of

    premotor lateralization and cerebellar involvement in the production of phonemes

    and syllables. *Journal of Speech, Language, and Hearing Research, 51*(5), 1183-

    1202. doi:10.1044/1092-4388(2008/07-0119

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.

    doi:10.1016/j.neuroimage.2010.12.065

Gracco, V. L., & Abbs, J. H. (1985). Dynamic control of the perioral system during

    speech: kinematic analyses of autogenic and nonautogenic sensorimotor

    processes. *Journal of Neurophysiology, 54*(2), 418-432.

Guenther, F. H. (1994). A neural network model of speech acquisition and motor

    equivalent speech production. *Biological Cybernetics, 72*(1), 43-53.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a

    neural network model of speech production. *Psychological Review, 102*(3), 594-

    621. doi:10.1037/0033-295X.102.3.594

Guenther, F. H. (2016). *Neural control of speech*. Cambridge, MA: MIT Press.

Guenther, F. H., Brumberg, J. S., Wright, E. J., Nieto-Castanon, A., Tourville, J. A.,

    Panko, M., . . . Andreasen, D. S. (2009). A wireless brain-machine interface for

    real-time speech synthesis. *PLOS ONE, 4*(12), e8218.

    doi:10.1371/journal.pone.0008218

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of

    the cortical interactions underlying syllable production. *Brain & Language, 96*(3),

    280-301.

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of

    reference frames for the planning of speech movements. *Psychological Review,*

    *105*(4), 611-633. doi:10.1037/0033-295X.105.4.611-633

Itabashi, R., Nishio, Y., Kataoka, Y., Yazawa, Y., Furui, E., Matsuda, M., & Mori, E.

    (2016). Damage to the left precentral gyrus is associated with apraxia of speech in

    acute stroke. *Stroke, 47*(1), 31-36. doi:10.1161/strokeaha.115.010402

Jürgens, U. (2009). The neural control of vocalization in mammals: A review. *Journal of Voice, 23*(1), 1-10. doi:10.1016/j.jvoice.2007.07.005

Jürgens, U., & Ploog, D. (1970). Cerebral representation of vocalization in the squirrel monkey. *Experimental Brain Research, 10*(5), 532-554. doi:10.1007/BF00234269

Jürgens, U., & Richter, K. (1986). Glutamate-induced vocalization in the squirrel monkey. *Brain Research, 373*(1-2), 349-358. doi:10.1016/0006-8993(86)90349-5

Kennedy, P. R. (1989). The cone electrode: a long-term electrode that records from neurites grown onto its recording surface. *Journal of Neuroscience Methods, 29*(3), 181-193. doi:10.1016/0165-0270(89)90142-8

Kent, R. (1977). Coarticulation in recent speech production models. *Jornal of Phonetics, 5*(1), 15-133.

Larson, C. R. (1991). On the relation of PAG neurons to laryngeal and respiratory muscles during vocalization in the monkey. *Brain Research, 552*(1), 77-86. doi:10.1016/0006-8993(91)90662-F

Levelt, W. J. (1989). *Speaking: From intention to articulation* (Vol. 1). Cambridge, MA: MIT press.

Lindblom, B., Lubker, J., & Gay, T. (1979). Formant frequencies of some fixed-mandible vowel and a model of speech motor programming by predictive simulation. *Journal of Phonetics, 7*, 147-162. doi:10.1121/1.2016039

MacNeilage, P. F. (1970). Motor control of serial ordering of speech. *Psychological Review, 77*(3), 182. doi:10.1037/h0029070

Müller-Preuss, P., & Jürgens, U. (1976). Projections from the 'cingular'vocalization area in the squirrel monkey. *Brain Research, 103*(1), 29-43. doi:10.1016/0006-8993(76)90684-3

Niziolek, C. A., & Guenther, F. H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *Journal of Neuroscience, 33*(29), 12090-12098. doi:10.1523/JNEUROSCI.1008-13.2013

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. *Journal of Neuroscience, 37*(38), 9249-9258. doi:10.1523/jneurosci.3363-16.2017

Penfield, W., & Rasmussen, T. (1950). *The cerebral cortex of man; a clinical study of localization of function*. Oxford, England: Macmillan.

Penfield, W., & Roberts, L. (1959). Speech and brain mechanisms: Princeton, NJ. Princeton University Press.

Perkell, J. S. (1981). On the use of feedback in speech production. *Advances in Psychology, 7*, 45-52. doi:10.1016/S0166-4115(08)60177-6

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature, 331*(6157), 585. doi:10.1038/331585a0

Recasens, D. (1989). Long range coarticulation effects for tongue dorsum contact in VCVCV sequences. *Speech Communication, 8*(4), 293-307. doi:10.1016/0167-6393(89)90012-5

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*(4), 333-382. doi:10.1207/s15326969eco0104_2

Schoch, B., Dimitrova, A., Gizewski, E., & Timmann, D. (2006). Functional localization in the human cerebellum based on voxelwise statistical analysis: a study of 90 patients. *Neuroimage, 30*(1), 36-51. doi:10.1016/j.neuroimage.2005.09.018

Segawa, J. A., Tourville, J. A., Beal, D. S., & Guenther, F. H. (2015). The neural correlates of speech motor sequence learning. *Journal of Cognitive Neuroscience, 27*(4), 819-831. doi:10.1162/jocn_a_00737

Sörös, P., Sokoloff, L. G., Bose, A., McIntosh, A. R., Graham, S. J., & Stuss, D. T. (2006). Clustered functional MRI of overt speech production. *Neuroimage, 32*(1), 376-387. doi:10.1016/j.neuroimage.2006.02.046

Theys, C., De Nil, L., Thijs, V., Van Wieringen, A., & Sunaert, S. (2013). A crucial role for the cortico-striato-cortical loop in the pathogenesis of stroke-related neurogenic stuttering. *Human Brain Mapping, 34*(9), 2103-2112. doi:10.1002/hbm.22052

Thoms, G., & Jürgens, U. (1987). Common input of the cranial motor nuclei involved in phonation in squirrel monkey. *Experimental Neurology, 95*(1), 85-99. doi:10.1016/0014-4886(87)90009-4

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage, 39*(3), 1429-1443. doi:10.1016/j.neuroimage.2007.09.054

Turkeltaub, P. E., Eden, G. F., Jones, K. M., & Zeffiro, T. A. (2002). Meta-analysis of the

    functional neuroanatomy of single-word reading: method and validation.

    *Neuroimage, 16*(3), 765-780. doi:10.1006/nimg.2002.1131

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to

    feedback perturbations of vowel acoustics and its relation to perception. *Journal*

    *of the Acoustical Society of America, 122*(4), 2306-2319. doi:10.1121/1.2773966

Wernicke, C. (1874). *Der aphasische Symptomencomplex: eine psychologische Studie*

    *auf anatomischer Basis*. Breslau, Germany: Cohn & Weigert.
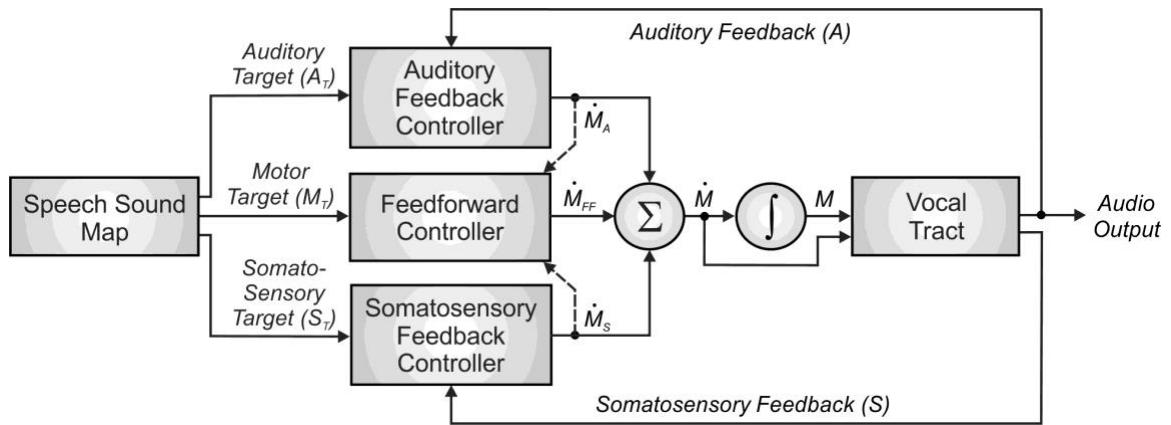
Figure 1. Control scheme utilised by the DIVA model for speech sound production. See text for details.
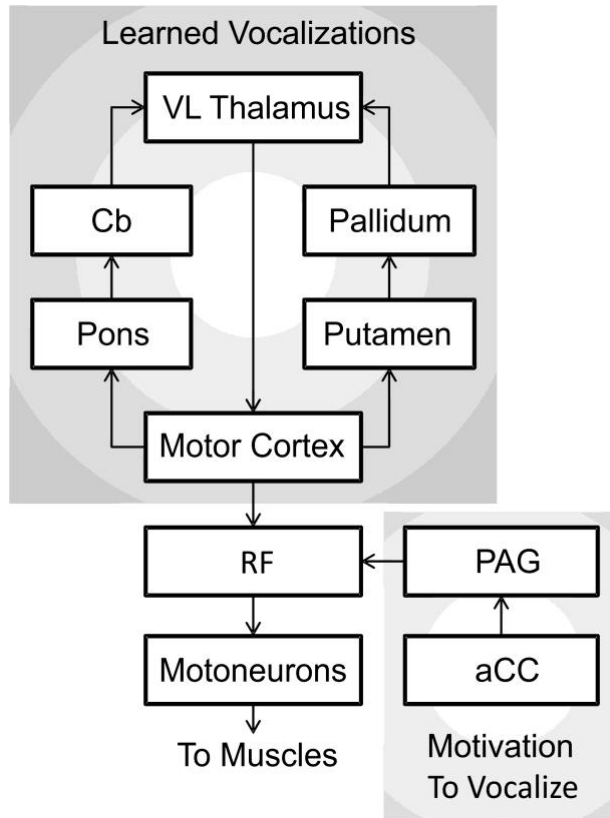
Figure 2. Schematic of the primate vocalization system proposed by Jürgens (2009). aCC = anterior cingulate cortex; Cb = cerebellum; PAG = periaqueductal grey matter; RF = reticular formation; VL = ventral lateral nucleus of the thalamus.
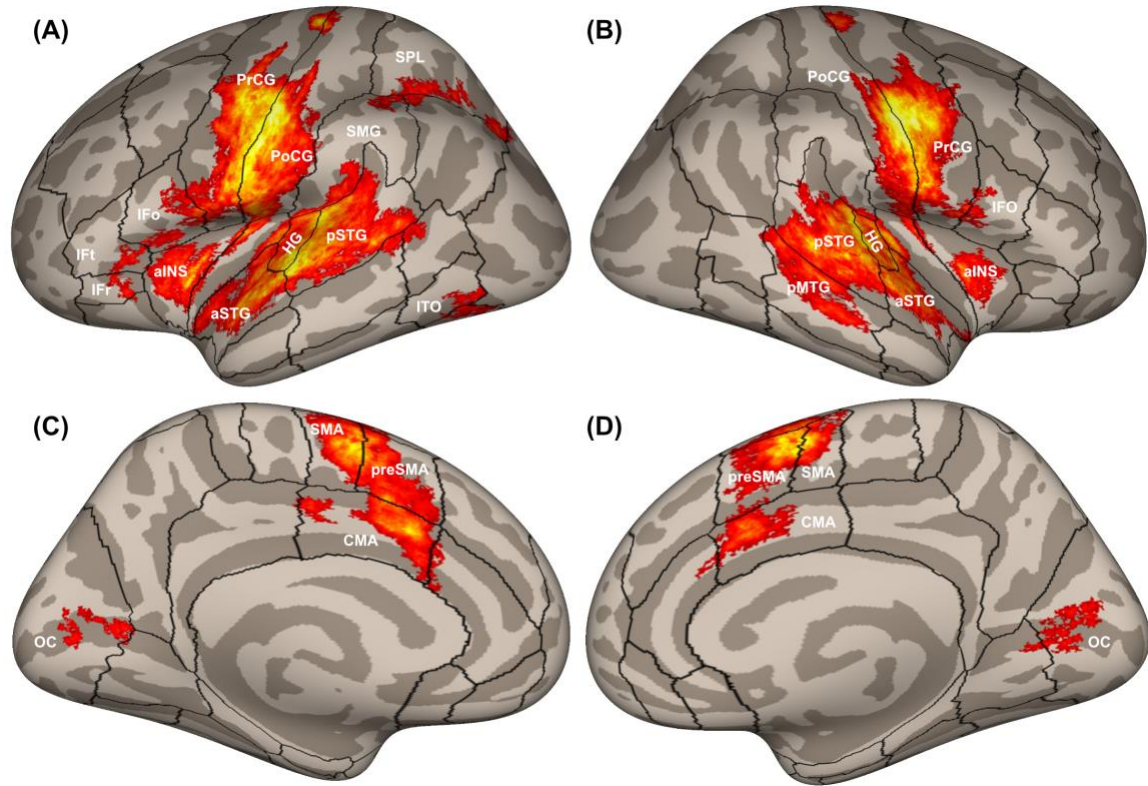
Figure 3. Cortical activity measured with fMRI in 116 participants while reading aloud simple utterances, plotted on inflated cortical surfaces. Boundaries between cortical regions are represented by black outlines. The panels show (A) left and (B) right hemisphere views of the lateral cortical surface; and (C) left and (D) right hemisphere views of the medial cortical surface. aINS, anterior insula; aSTG, anterior superior temporal gyrus; CMA, cingulate motor area; HG, Heschl's gyrus; IFo, inferior frontal gyrus pars opercularis; IFr, inferior frontal gyrus pars orbitalis; IFt, inferior frontal gyrus pars triangularis; ITO, inferior temporo-occipital junction; OC, occipital cortex; pMTG, posterior middle temporal gyrus; PoCG, postcentral gyrus; PrCG, precentral gyrus;

preSMA, pre-supplementary motor area; pSTG, posterior superior temporal gyrus; SMA, supplementary motor area; SMG, supramarginal gyrus; SPL, superior parietal lobule.
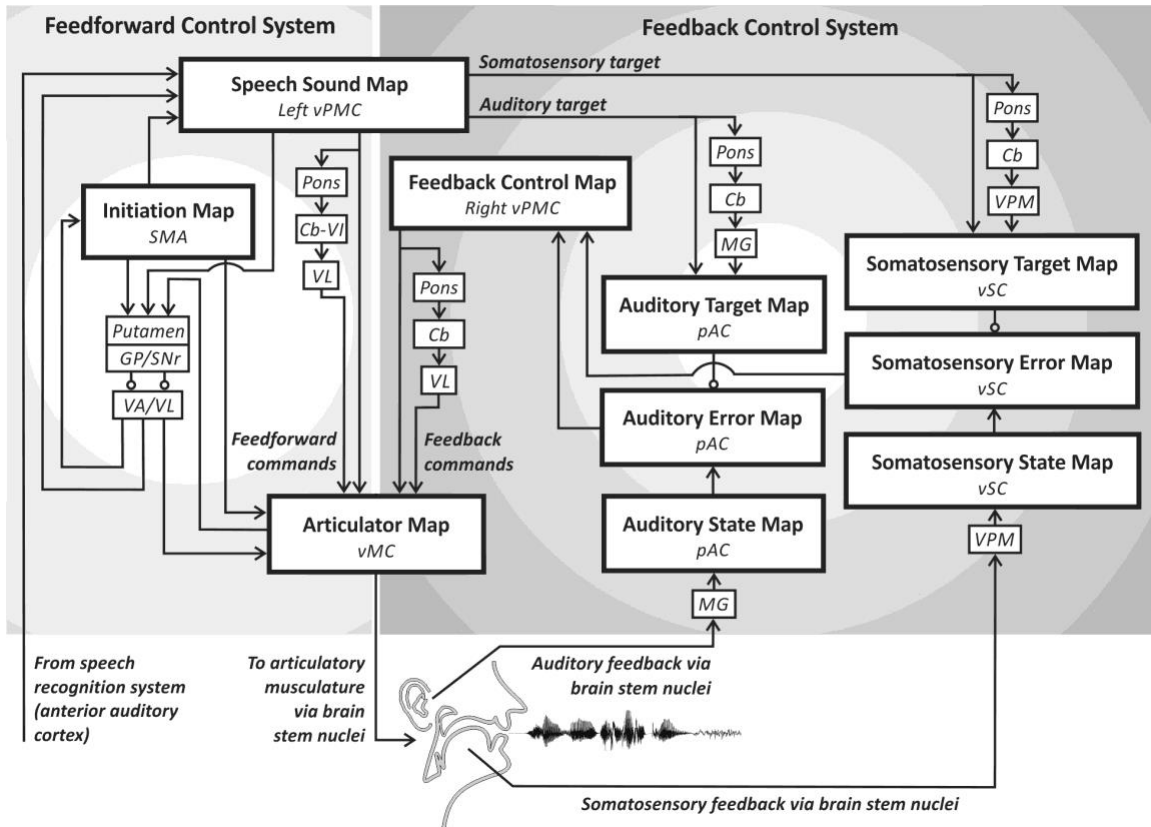
Figure 4. Neural correlates of the DIVA model. Each box indicates a set of model nodes that is associated with a specific type of information and hypothesised to reside in the brain regions shown in italics. See text for details. Cb, cerebellum; Cb-VI, cerebellum lobule VI; GP, globus pallidus; MG, medial geniculate nucleus of the thalamus; pAC, posterior auditory cortex; SMA, supplementary motor area; SNr, substantia nigra pars reticula; VA, ventral anterior nucleus of the thalamus; VL, ventral lateral nucleus of the thalamus; vMC, ventral motor cortex; VPM, ventral posterior medial nucleus of the thalamus; vPMC, ventral premotor cortex; vSC, ventral somatosensory cortex.
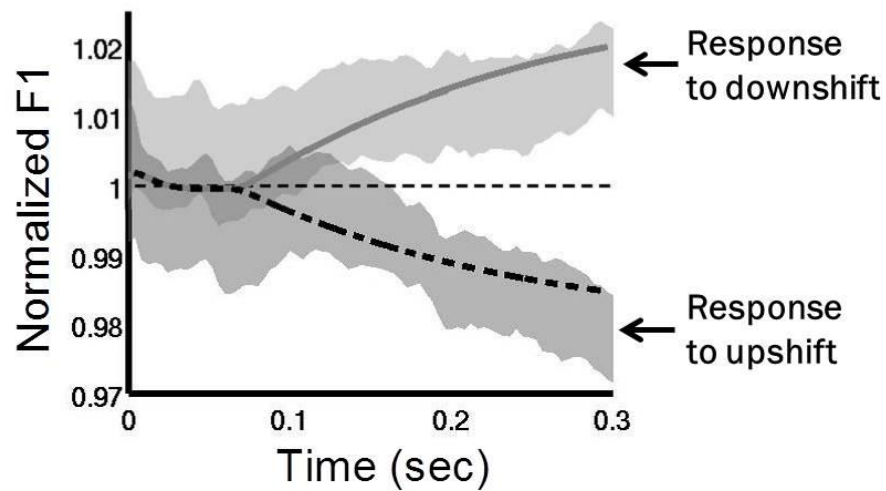
Figure 5. Normalised first formant response to perturbations in F1 in the DIVA model and in experimental subjects (adapted from Tourville et al., 2008). DIVA model productions in response to an upward perturbation are shown by the dashed line and to a downward perturbation by the solid line. The grey shaded areas show 95% confidence intervals for speakers responding to the same perturbations (Tourville et al., 2008). The DIVA model productions fall within the distribution of the productions of the speakers.
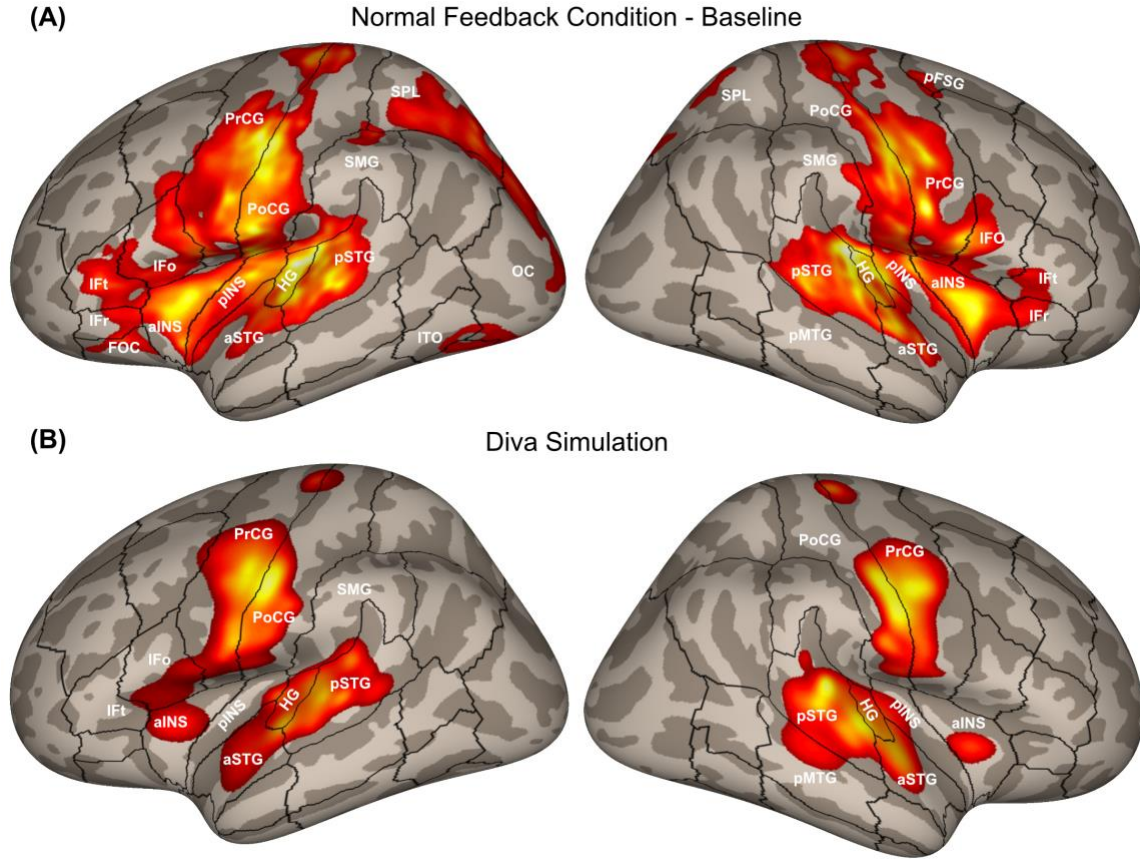
**(A)** Normal Feedback Condition - Baseline

**(B)** Diva Simulation

Figure 6. (A) Cortical activity during the normal feedback speech condition from a pooled analysis of formant perturbation studies by Tourville et al. (2008) and Niziolek and Guenther (2013). (B) Cortical activity generated by a DIVA model simulation of the normal feedback condition. aINS, anterior insula; aSTG, anterior superior temporal gyrus; HG, Heschl's gyrus; IFo, inferior frontal gyrus pars opercularis; IFr, inferior frontal gyrus pars orbitalis; IFt, inferior frontal gyrus pars triangularis; OC, occipital cortex; pINS, posterior insula; pFSG, posterior superior frontal gyrus; pMTG, posterior middle temporal gyrus; PoCG, postcentral gyrus; PrCG, precentral gyrus; pSTG,

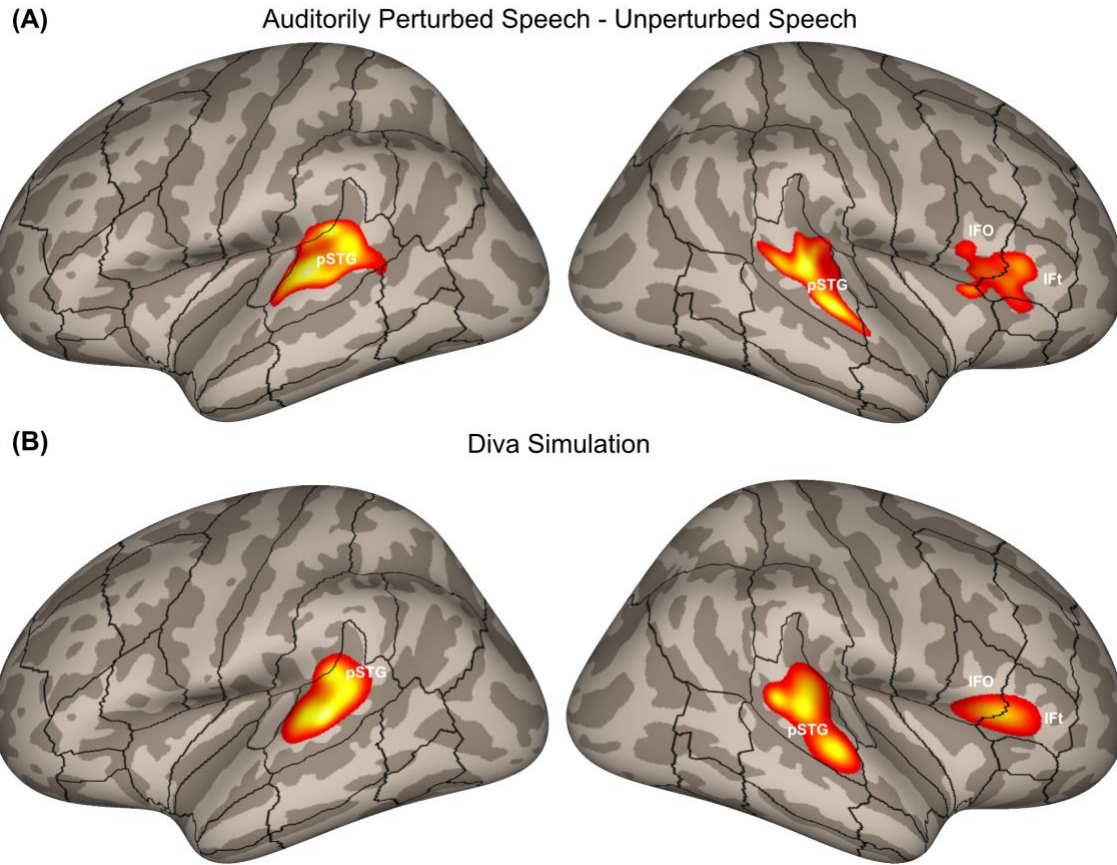posterior superior temporal gyrus; SMG, supramarginal gyrus; SPL, superior parietal lobule.

Figure 7. (A) Areas of increased cortical activity in response to auditory perturbations from a pooled analysis of formant perturbation studies by Tourville et al. (2008) and Niziolek and Guenther (2013) plotted on inflated cortical surfaces. (B) Cortical activity generated by a DIVA model simulation of the auditory perturbation experiment. IFo, inferior frontal gyrus pars opercularis; IFt, inferior frontal gyrus pars triangularis; pSTG, posterior superior temporal gyrus.
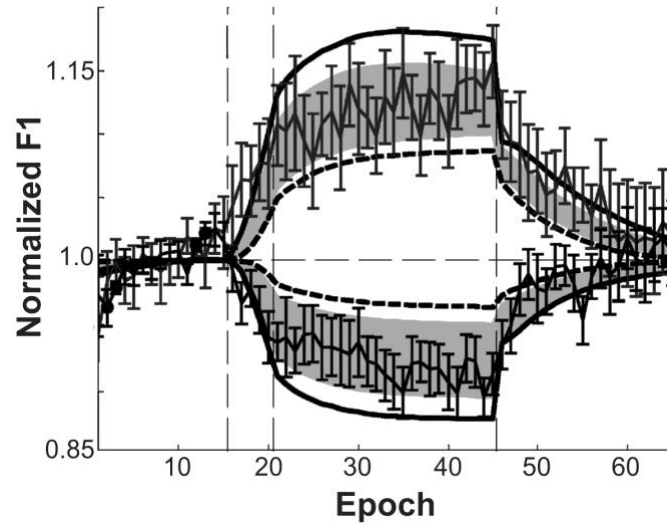
Figure 8. Comparison of normalised adaptive first formant response to perturbation of F1 during a sensorimotor adaptation experiment to simulations of the DIVA model (adapted from Villacorta et al., 2007). Vertical dashed lines indicate progression from baseline to ramp, training, and posttest phases over the course of the experiment. Thin solid line with standard error bars indicates data from 20 participants. Shaded region shows 95% confidence intervals from the DIVA model simulations, with model simulation results for the subjects with the lowest and highest auditory acuity represented by the bold dashed line and bold solid line, respectively.
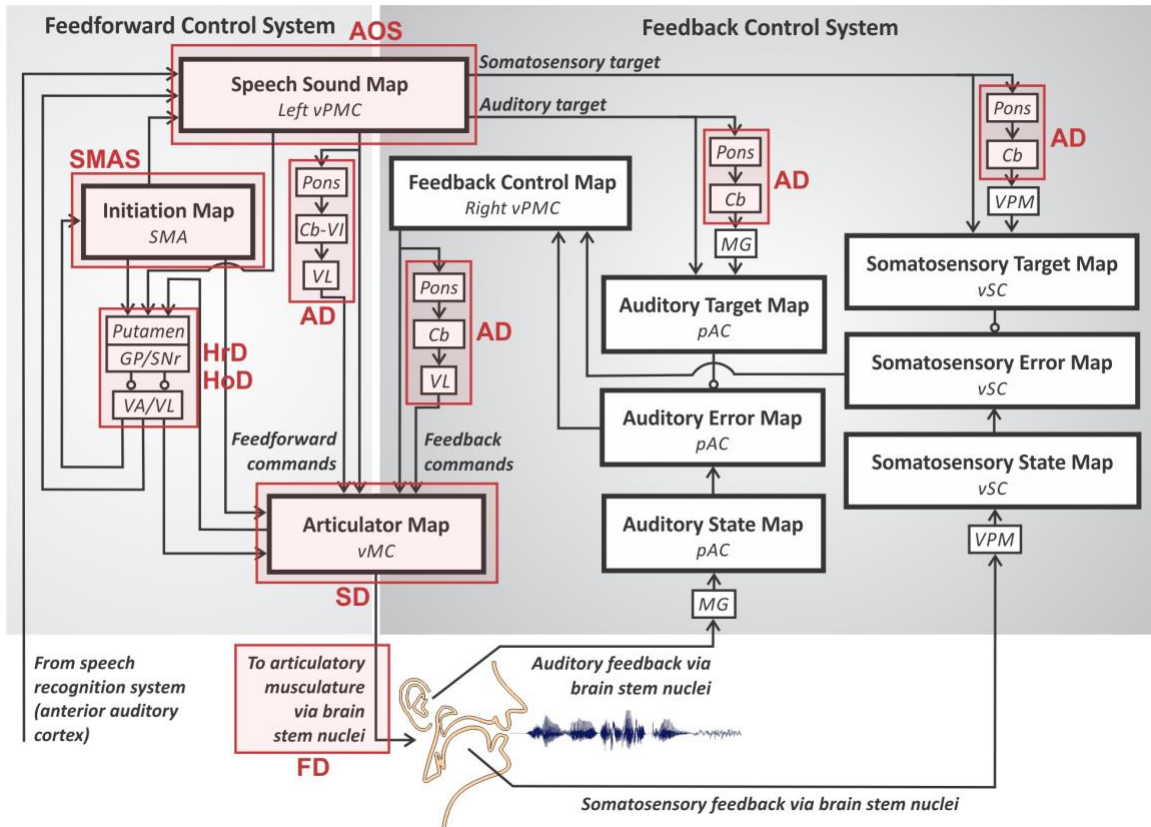
Figure 9. Locus of neural damage for common speech motor disorders within the DIVA model (adapted from Guenther, 2016). AD, ataxic dysarthria; AOS, apraxia of speech; FD, flaccid dysarthria; HoD, hypokinetic dysarthria; HrD, hyperkinetic dysarthria; SD, spastic dysarthria; SMAS, supplementary motor area syndrome. See caption of Figure 4 for anatomical abbreviations.
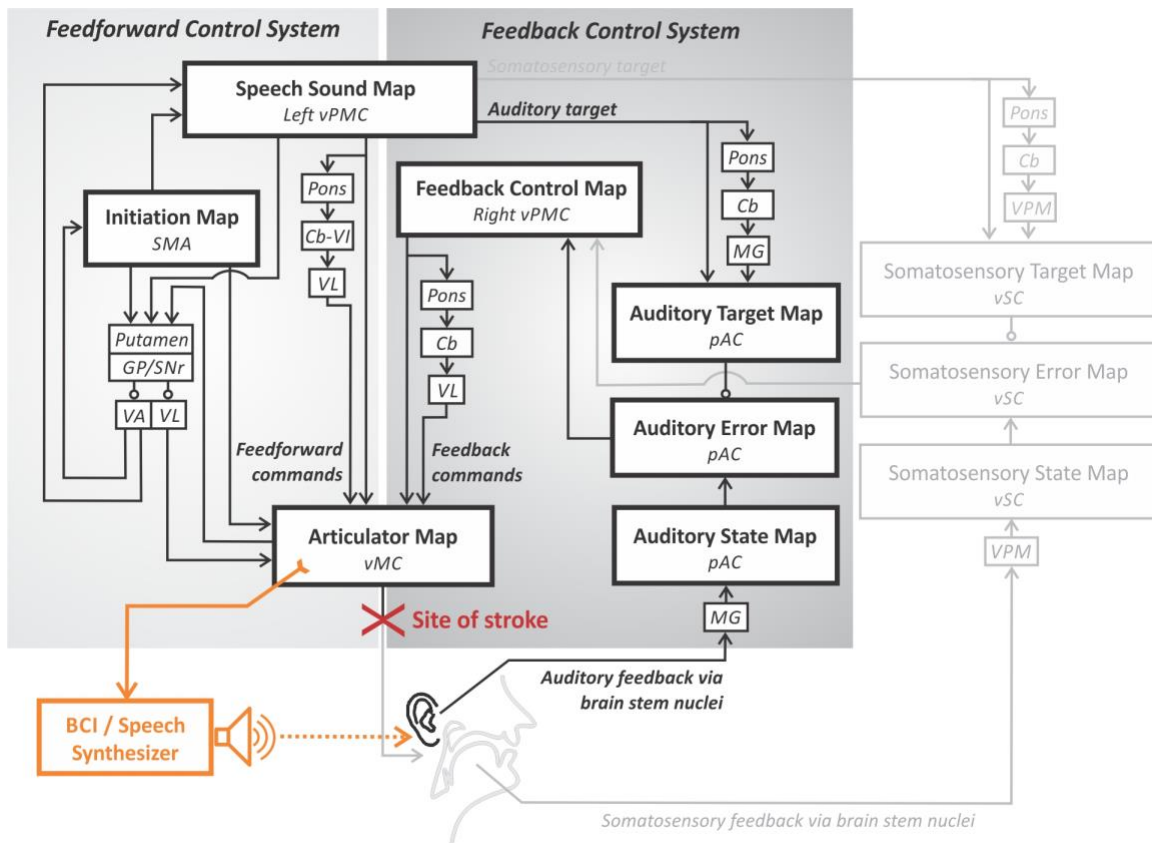
Figure 10. Schematic of a brain-computer interface (BCI) for restoring speech capabilities to a locked-in volunteer (Guenther et al., 2009) within the DIVA model framework. See caption of Figure 4 for anatomical abbreviations.