

Published in final edited form as:

Neuroimage. 2010 September ; 52(3): 862–874. doi:10.1016/j.neuroimage.2009.10.023.

The integration of large-scale neural network modeling and functional brain imaging in speech motor control

E. Golfinopoulos¹, J.A. Tourville¹, and F.H. Guenther^{1,2,3,4}

¹Department of Cognitive and Neural Systems, Boston University, Boston, MA, USA

²Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA USA

³Division of Health Sciences and Technology, Harvard University – Massachusetts Institute of Technology, Cambridge, MA USA

⁴Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA USA

Abstract

Speech production demands a number of integrated processing stages. The system must encode the speech motor programs that command movement trajectories of the articulators and monitor transient spatiotemporal variations in auditory and somatosensory feedback. Early models of this system proposed that independent neural regions perform specialized speech processes. As technology advanced, neuroimaging data revealed that the dynamic sensorimotor processes of speech require a distributed set of interacting neural regions. The DIVA (Directions into Velocities of Articulators) neurocomputational model elaborates on early theories, integrating existing data and contemporary ideologies, to provide a mechanistic account of acoustic, kinematic, and functional magnetic resonance imaging (fMRI) data on speech acquisition and production. This large-scale neural network model is composed of several interconnected components whose cell activities and synaptic weight strengths are governed by differential equations. Cells in the model are associated with neuroanatomical substrates and have been mapped to locations in Montreal Neurological Institute stereotactic space, providing a means to compare simulated and empirical fMRI data. The DIVA model also provides a computational and neurophysiological framework within which to interpret and organize research on speech acquisition and production in fluent and dysfluent child and adult speakers. The purpose of this review article is to demonstrate how the DIVA model is used to motivate and guide functional imaging studies. We describe how model predictions are evaluated using voxel-based, region-of-interest-based parametric analyses and inter-regional effective connectivity modeling of fMRI data.

Keywords

computational model; functional magnetic resonance imaging; structural equation modeling; speech production; DIVA model

© 2009 Elsevier Inc. All rights reserved

*Correspondence should be addressed to: Elisa Golfinopoulos, Department of Cognitive and Neural Systems, Boston University, 677 Beacon Street, Boston, MA, 02215, USA. Tel: +1 617 353 6181; Fax: +1 617 353 7755; egolfino@cns.bu.edu..

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

The theoretical foundations of speech comprehension and production modeling date back to the nineteenth century when it was observed that language disorders resulted from lesions to particular cortical regions (Broca, 1861; Wernicke, 1874). These observations led to the first modeling frameworks that linked independent brain regions with speech comprehension and production (Lichtheim, 1885). Since that time, pioneering human electrical stimulation studies mapping cortical representations of the speech articulators (Penfield and Rasmussen, 1950; Penfield and Roberts, 1959) and investigations on patients with stroke-related lesions (Damasio and Damasio, 1980; Goodglass, 1993) have elaborated initial theories, providing greater detail into the functional segregation of brain regions involved in speech production and perception. With the advent of functional brain imaging techniques and associated computational modeling methods (e.g., Friston et al., 1993; McIntosh et al., 1994; Friston et al., 1997; McKeown et al., 1998; Friston et al., 2003), new empirical research has stimulated a theoretical shift in the field of speech neuroscience from functional *segregation* to functional *integration*. Modern theoretical perspectives propose that speech perception and production might be more accurately characterized by a large network of interacting brain areas rather than local independent modules (e.g., Gracco, 1991; Baum and Pell, 1999; Hickok and Poeppel, 2000; Sidtis and Van Lancker Sidtis, 2003; Friederici and Alter, 2004; Riecker et al., 2005; Indefrey and Levelt, 2004; Pulvermüller, 2005).

Despite the growing consensus that multiple, interacting cortical and subcortical brain areas are associated with the coordination of speech, a complete account of how the regions interact to produce fluent speech is still lacking. The DIVA (Directions into Velocities of Articulators) neural network model of speech acquisition and production attempts to bridge this gap by characterizing different processing stages of speech motor control both computationally and neurophysiologically (Guenther, 1994, 1995; Guenther et al., 1998, 2006). The model, schematized in Figure 1, learns to control the movements of a computer-simulated vocal tract (a modified version of the synthesizer described by Maeda, 1990) in order to produce an acoustic signal. Since the model can generate a time-varying sequence of articulator positions and formant frequencies, the kinematics and acoustics of an extensive set of speech production data may be simulated and tested against that of both fluent (Guenther, 1994, 1995; Guenther et al., 1998, 2006; Nieto-Castanon et al., 2005) and dysfluent adult speakers (Max et al., 2004; Robin et al., 2008; Civier et al., submitted). The model can also account for aspects of child development that influence speech production (e.g., Guenther, 1995; Terband et al., in press) such as the dramatic restructuring of the size and shape of the vocal tract over the first few years of life (Callan et al., 2000).

The DIVA model is comprised of several interconnected components that are associated with cortical, subcortical, and cerebellar human brain regions. The model's neural substrates, projections, and computational characterizations are based, when possible, on the structural and functional descriptions of regions and projections in the human brain from human clinical, neuroimaging, and microstimulation studies and non-human primate single cell recording and anterograde and retrograde tracing studies (Guenther et al., 2006). Model cells are mapped to locations in Montreal Neurological Institute (MNI; Mazziotta et al., 2001) stereotactic space, a standard reference frame often used when analyzing neuroimaging data. Since the model's components represent clusters of neurons, and cell activities are both mathematically and stereotactically defined, it is possible to generate simulated functional magnetic resonance imaging (fMRI) activations at specific anatomical locations (Guenther et al., 2006). The resulting simulated fMRI activations can then be compared to the results of neuroimaging studies of speech production (Guenther et al., 2006; Tourville et al., 2008) and used both to interpret the neuroimaging data and further refine the model (when imaging results are not fully captured by the model). The DIVA model can also assist in motivating inter-regional

effective connectivity modeling of functional imaging data and, likewise, effective connectivity modeling can be used to test the DIVA model's predictions of modulated projections (Tourville et al., 2008). Therefore, the model is particularly well-suited for guiding and interpreting neuroimaging studies of speech acquisition and production.

In this review, we attempt to demonstrate that large-scale computational neural network modeling can act as a structuring framework for neuroimaging research. Together empirical research and neural network modeling can contribute more to our understanding of the underlying functional neural networks than either one separately.

Materials and Methods

DIVA model overview

A schematic representation of the DIVA neural network model is illustrated in Figure 1. The system is composed of thirteen modules encompassing two types of motor control schemes: feedforward and feedback control. Under feedforward control, speech production is based on well-learned speech motor programs that are acquired through previous attempts at producing the speech sound targets. The feedback control subsystem iteratively guides the articulators by monitoring mismatches between the sensory expectations of a speech motor program and the incoming sensory feedback associated with that speech motor program. In DIVA, the feedforward and feedback commands are integrated when generating speech motor commands. As a result, the model remains sensitive to errors during speech production, while maintaining the ability to produce fluent speech.

The combined feedforward and feedback signals constitute the total speech motor command that originates from cells within the *Articulator Velocity and Position Maps*. The term *map* is used here to refer to a set of neurons. The *Articulator Velocity and Position Maps* each consist of eight antagonistic pairs of cells corresponding to the eight degrees of freedom of the vocal tract model (Maeda, 1990). Activity in the *Position Map* represents a motor command describing jaw height, tongue shape, tongue body position, tongue tip position, lip protrusion, larynx height, upper lip height, and lower lip height. This command is sent to the articulatory synthesizer, resulting in movements of the vocal tract model. The neurons within the *Articulator Velocity and Position Maps* are hypothesized to lie in overlapping positions along the caudoventral portion of the precentral gyrus (See Figure 4 for a schematic of the DIVA model components to their hypothesized neuroanatomical locations). Montreal Neurological Institute (MNI) coordinates for the proposed location of these cells in the SPM2 canonical brain are provided in Table 1 along with a selection of the citations used to estimate these locations. A more detailed description of the mapping of the DIVA model cells onto specific neuroanatomical locations is provided in Guenther et al. (2006).

In order to generate spatio-temporally varying commands to the articulators, the DIVA model must first learn the relationship between movements of the speech articulators and the corresponding sensory feedback, a mapping that subserves both the feedforward and feedback control subsystems. After this initial “babbling” stage, the DIVA model learns the sensory expectations, or *targets*, for frequently encountered speech sounds and the speech motor programs that effectuate these sensory targets. The model takes as input an acoustic speech sound target representing a phoneme, syllable, word, or short phrase to be produced. Through a process of trial and error, the model learns to acquire each speech target over the course of several production attempts largely relying on auditory feedback control. Gradually, these commands are subsumed by the feedforward control subsystem and, under normal conditions, the model can rely primarily on feedforward control.

The feedforward control subsystem is initiated by the activation of a cell in the *Speech Sound Map*. The cells of the *Speech Sound Map* correspond to the “mental syllabary” described by Levelt and colleagues (e.g., Levelt and Wheeldon, 1994; Levelt et al., 1999). In particular, each cell in this map represents a phoneme or frequently encountered multi-phonemic speech sound, with syllables being the most typical sound type represented. The activation of one of these cells will result in the production of the corresponding speech sound. Cells in the *Speech Sound Map* are also hypothesized to be active during speech *perception* when the auditory expectations of the active speech sound target are tuned. These cells are hypothesized to lie in the left posterior inferior frontal gyrus and adjacent ventral premotor cortex.

The excitatory feedforward commands projecting from the *Speech Sound Map* to the *Articulator Velocity* and *Position Maps* can be thought of as a “motor program” or “gestural score” (e.g., Browman and Goldstein, 1989), i.e., a time series of articulatory gestures used to produce the corresponding speech sound. The feedforward control subsystem is also mediated by a trans-cerebellar pathway. Projections from the cerebellum are thought to contribute precisely-timed feedforward commands (Ghosh, 2005). The cells within the cerebellum are hypothesized to lie bilaterally in the anterior paravermal cerebellar cortex.

Commands from the *Articulator Velocity* and *Position Maps* are released to the articulators when the activity of the appropriate cell in the *Initiation Map* is non-zero. According to the model, each speech motor program in the *Speech Sound Map* is associated with a cell in the *Initiation Map*. Motor commands associated with the selected speech motor program are released when the corresponding *Initiation Map* cell becomes active. The *Initiation Map* is hypothesized to lie bilaterally within the supplementary motor area. The timing of the *Initiation Map* activation and the release of a speech motor program is hypothesized to be dependent on a reciprocal connection with the basal ganglia, including bilateral caudate, putamen, pallidum, and thalamus.

When confronted with a novel speech task or an unexpected change in environmental settings, the DIVA model is able to detect and correct the current speech motor program using feedback control. Studies have demonstrated that both auditory (Houde and Jordan, 1998; Jones and Munhall, 2005; Bauer et al., 2006; Purcell and Munhall, 2006) and proprioceptive and tactile afferent feedback (MacNeilage et al., 1967; Lindblom et al., 1979; Gay et al., 1981; Kelso and Tuller, 1983; Abbs and Gracco, 1984; Gracco and Abbs, 1985; Tremblay et al., 2003; Nasir and Ostry, 2006) are used to guide articulator movements during speech. Accordingly, the feedback control subsystem is further divided into two independent pathways responsible for auditory and somatosensory feedback control.

The auditory and somatosensory feedback control subsystems utilize a similar set of cortical computational mechanisms along each of their respective pathways, but the reference frames in which these two subsystems operate are dependent on the sensory modality. The feedback control subsystem computes the difference between the sensory feedback and sensory expectations of the speech target to detect errors in these reference frames and correct the current speech motor program when it is off-target. In the current implementation of the model, the acoustic reference frame corresponds to the first three formant frequencies of the acoustic signal while the somatosensory reference frame corresponds to proprioceptive and tactile feedback from the articulators.

The sensory response to self-generated speech is represented in the *Auditory* and *Somatosensory State Maps* which are hypothesized to lie along the supratemporal plane and the inferior postcentral gyrus, respectively. Expected sensory consequences of the sounds to be produced are encoded in the forward model projections from the *Speech Sound Map* to the *Somatosensory* and *Auditory Target Maps*. These expectations correspond to a convex target

region that encodes acceptable ranges (in acoustic and somatosensory reference frames) for the current speech target (Guenther, 1995). The cells of the *Auditory Target Map* are hypothesized to lie in bilateral planum temporale and superior temporal gyrus while the cells of the *Somatosensory Target Map* are hypothesized to lie in bilateral ventral somatosensory cortex and anterior supramarginal gyrus. An additional trans-cerebellar pathway is also hypothesized to contribute to the sensory predictions that form the auditory and somatosensory targets. The cerebellar cells are hypothesized to lie bilaterally in the superior lateral cerebellar cortex.

Activity within the *Auditory* and *Somatosensory Error Maps* represents the difference between the realized and expected sensory responses for the current speech sound. The cells of the *Auditory* and *Somatosensory Error Maps* are topographically superimposed on the cells of the *Auditory* and *Somatosensory Target Maps*. The *Error Maps* receive inhibitory inputs from the *Target Maps* and excitatory inputs from the *State Maps*. If the incoming sensory response falls outside of the expected target region, the net activity in the *Error Map* represents an error signal. This error is transformed into corrective motor commands via projections from the sensory *Error Maps* to the articulator velocity cells in the *Feedback Control Map*. These transformations constitute inverse models. In the Results section we demonstrate how the DIVA model was used to guide a functional MRI experiment and how the results of that experiment motivated the inclusion of the right-lateralized *Feedback Control Map* in ventral premotor cortex.

Generating DIVA model simulations

The neural network model is described by a system of ordinary differential equations that are provided in Table 2. A more detailed description of the mathematical formulation of the DIVA model is provided in Guenther et al. (2006). Model simulations may be generated using the source code developed in our laboratory and publicly available at the website <http://speechlab.bu.edu/DIVAcodes.php>. The current release of the model includes a graphical user interface to assist researchers in running simulations and in changing parameters. The articulatory synthesizer and a library of learned speech sounds are also included.

Generating simulated fMRI activations

Since the model's components correspond to neurons at specific anatomical locations and their activities are mathematically defined, it is possible to generate simulated fMRI activations during a DIVA model simulation of a speech production task. Simulating fMRI activations requires an understanding of the relationship between neural activity and the fMRI blood oxygen level dependent (BOLD) signal. This relationship remains an issue of some debate. It has been proposed that neurophysiological correlates of the behaviorally evoked blood flow changes are the spiking activity or the neuronal output (Heeger et al., 2000; Rees et al., 2000; Mukamel et al. 2005). However, this claim has recently been challenged by research using simultaneous electrophysiological and BOLD measurements in primary visual cortex of alert behaving monkeys (Goense and Logothetis, 2008). Goense and Logothetis (2008) demonstrated that there were periods when the time course of the average neuronal output (the multiple-unit activity) diverged from that of the local field potentials and did not remain elevated for the time course of the stimulus. Local field potentials are thought to reflect a weighted average of input signals on the dendrites and cell bodies of neurons (Raichle and Mintun, 2006). Since the transient response of the multiple-unit activity was not reflected in the BOLD response, the researchers suggest that the BOLD response was likely driven by a more sustained mechanism such as the local field potentials (Goense and Logothetis, 2008). Similar results have been reported by other studies (Mathiesen et al., 1998; Logothetis et al., 2001; Viswanathan and Freeman, 2007), providing growing evidence that the BOLD signal is

more representative of synaptic activity as reflected by local field potentials than spiking activity.

In accordance with this conclusion, the fMRI activations in the DIVA model are determined by using the total synaptic inputs to our modeled cells, rather than their activity. Each model cell is hypothesized to correspond to a population of neurons that fire together. The output of the model cell is analogous to a neuron's firing rate (the number of action potentials per second) of the population of neurons. This output is then multiplied by synaptic weights, forming postsynaptic inputs to other cells in the network. In order to generate simulated fMRI activations in the DIVA model, a cell's activity level is calculated as the sum of the cell's synaptic inputs (both excitatory and inhibitory). If the net activity of the cell's synaptic inputs is less than zero, the cell's activity level is considered to be zero.

The brain's vascular response to a transient change in neural activity is delayed, and this delay has been estimated at four to seven seconds depending on the brain region and task (Belin et al., 1999; Yang et al., 2000). To account for this hemodynamic delay in the DIVA model, fMRI activity is simulated using an idealized hemodynamic response function (HRF; Friston et al., 1995). The HRF function is convolved with the activity of each cell over the course of the entire speech task, in order to transform the cell activity into a temporally smoothed time series of simulated fMRI activity. A surface representation is then rendered with a corresponding hemodynamic response value at each cell location for each condition. Responses are then spatially smoothed with a spherical Gaussian point spread function. Activity in the modeled regions can then be compared across conditions, and the resulting simulated contrast image can be superimposed on a co-registered anatomical dataset. By default, simulated BOLD responses are superimposed on a cortical surface representation of the canonical anatomical volume included in the SPM software package. The voxel-based time series data can also be displayed as a movie of simulated fMRI activity over the time course of the speech task.

Results

Integrating modeling and fMRI

Voxel-based analysis—Localization of statistically significant behaviorally evoked changes in the brain using voxel-based analysis of fMRI data continues to produce a host of new findings regarding brain–behavior relationships in speech production and perception. In this approach, functional volumes are first spatially normalized to match a standard template. This step is performed so that data collected from multiple subjects are aligned in a common space, permitting intersubject comparisons. Spatial normalization also provides a means for reporting activation sites according to coordinates in a standard space, allowing for comparison of findings across studies. Task-related effects are then estimated for every voxel in the acquired functional volumes. For every subject, a set of contrast volumes is created that represents a specified comparison of task-related effects. Group-level significance tests are then based on comparisons of each voxel in each subject's contrast volume. The resulting parametric map is then thresholded to indicate which voxels differ significantly across the contrasted conditions.

Alone, the voxel-based analysis can provide an automated and standardized means of assessing the effects of variables on the activity of voxels at specific stereotactic locations across the entire brain. In combination with large-scale computational neural network modeling, the voxel-based method is valuable for comparing peak locations from experimental studies of individual conditions and groups of subjects with simulated fMRI activations. These comparisons can corroborate and strengthen initial model-based hypotheses or assumptions, generate new insights about functional neural networks, and result in the design of new experimental or model-based investigations. Combining model simulations with experimental

voxel-based results, therefore, can provide a synergistic approach to clarifying the neural substrates underlying a specific variable under study.

As an illustrative example, DIVA simulations were used to motivate a recent fMRI investigation to test model predictions regarding the brain regions involved in auditory feedback-based control of speech (Tourville et al., 2008). The model was trained to pronounce the monosyllabic word /bed/ with normal auditory feedback. Following successful training, the first formant of the model's auditory feedback of its own speech was shifted upward and downward by 30% in two additional simulations (See Tourville et al., 2008 for more details). According to the model, the mismatch between expected and actual auditory feedback will result in activation of cells in the *Auditory Error Map* which, in turn, drive compensatory motor commands issued from *Articulator Velocity and Position Maps*. The simulations indicated that the DIVA model compensated for the auditory perturbation by shifting the first formant of the model's acoustic output in the direction opposite the shift. The compensatory response was associated with increased activation in bilateral posterior peri-Sylvian cortex (the location of the *Auditory Error Map*), ventral Rolandic cortex (the location of the *Articulator Velocity and Position Maps*), and superior lateral cerebellum (the location of the Auditory Feedback Control Subsystem's Cerebellar Module).

These predictions were compared to data acquired in an fMRI study involving eleven neurologically normal, right-handed speakers of American English (Tourville et al., 2008). In the experiment, auditory feedback of self-produced speech was perturbed on randomly distributed trials while subjects read monosyllabic words from a screen. All words had the form consonant-vowel-consonant with /e/ (as in "bed") as the center vowel. The perturbed feedback trials were randomly interspersed with trials in which normal auditory feedback was provided so that subjects could not anticipate the perturbations. Speakers compensated for the altered feedback, as the model predicted, by shifting the first formant of their output in the direction opposite the induced perturbation. Moreover, the fMRI results of a fixed effects voxel-based analysis demonstrated that during the conditions in which there was a mismatch between the expected and actual auditory feedback, there was increased activity in bilateral higher-order auditory cortical areas, including the posterior superior temporal gyrus and planum temporale. Furthermore, activity was found in the precentral gyrus as predicted, though the relatively low anatomical precision of voxel-wise group analysis prevents distinguishing between primary motor cortical activity (predicted by the model) and premotor cortical activity. Compensating for the shift in auditory feedback was also associated with increased activity in right lateral prefrontal cortex and right inferior cerebellum, regions that were not predicted by the model.

A comparison of the simulated and empirical data is illustrated in Figure 2. The correspondence between the model and human subject voxel-based results, particularly within the peri-Sylvian areas, illustrates the ability of the DIVA model to qualitatively account for speech production-related neuroimaging data at a voxel-wise level. By specifying a computational neuroanatomy using MNI coordinates, the model-based framework can be tested more directly with voxel-based empirical data. Moreover the agreements in functional localization between the model simulations and the fMRI experiment strengthen and confirm the model-based hypotheses of the neural mechanisms proposed to underlie those areas that the model accurately predicts. Since the DIVA model correctly predicted the neural responses in higher-level auditory cortical areas, the theoretical basis upon which these neural predictions were made was reinforced. These results demonstrate how a large-scale computational neural network modeling framework in conjunction with empirical data can be used to substantiate hypotheses and assumptions and therefore offer a validated framework in which to interpret future voxel-based results.

While it is informative to identify which regions with increased activity were successfully predicted by the DIVA model, it is equally important to note the activity for which the model failed to account. One of the main unpredicted voxel-based results was that frontal cortex activity in response to the perturbation was right-lateralized and consisted primarily of premotor, rather than primary motor, activity. However voxel-based group analyses have inherent problems when one attempts to make fine anatomical distinctions (Nieto-Castanon et al., 2003) such as distinguishing between primary motor and premotor cortex in the precentral gyrus. For this reason we supplement voxel-wise analysis with region-of interest (ROI) based analyses as discussed next.

ROI-based analysis—Although voxel-based analysis is widely used, it suffers from well-known limitations. Chief among these is its inability to account for the substantial anatomical variability across subjects, despite non-linear warping of data prior to analysis. Nieto-Castanon and colleagues (2003) provided a measure of the extent of inter-subject variability for a number of easily-identified anatomical regions of interest (ROIs). For example, the mean overlap of voxels common to Heschl's gyrus, an ROI representing primary auditory cortex, across two subjects following normalization was 31% of the total voxels belonging to Heschl's gyrus in both subjects. The overlap dropped to 13% for three subjects and across nine subjects, a relatively small population for imaging studies, no overlap was found. In other words, there was not a single voxel in the standard stereotactic space that fell within Heschl's gyrus across all subjects. This variability was typical of the 12 temporal and parietal lobe regions of interest that were analyzed in the left and right hemispheres (see Nieto-Castanon et al., 2003). Therefore, the standard normalization procedure falls short of ensuring alignment of the structural, and presumably functional, regions across even a small subject cohort.

The standard means of accounting for inter-subject anatomical variability is to spatially smooth functional data after normalization with an isotropic smoothing kernel of 6–12 mm. While increasing the power of voxel-based analyses, this method has the effect of blurring regional boundaries. The blurring effect is particularly problematic across the banks of major sulci. For instance, two adjacent points along the dorsal and ventral banks of the Sylvian fissure that are separated by less than a millimeter in the 3-D volume space may lie several centimeters apart with respect to distance along the cortical sheet. Their cortical distance is a much better measure of their “functional distance:” the adjacent dorsal and ventral points could be the somatosensory larynx representation and the auditory association cortex, respectively. Isotropic smoothing in the 3-D volume ignores this distinction, blurring responses from the two regions, resulting in loss of statistical sensitivity and, perhaps, misleading findings.

The problems associated with voxel-based analysis can be overcome by comparing functional responses within like anatomical regions of interest across subjects. This region-of-interest (ROI)-based method eliminates the need for non-linear warping and smoothing of functional data. By eliminating these preprocessing steps, ROI-based analysis greatly improves the statistical sensitivity over standard voxel-based techniques and avoids potential erroneous anatomical localization of functional data due to imperfect spatial normalization procedures (Nieto-Castanon et al., 2003).

By itself, ROI analysis provides a means of assessing the effects of variables on the activity of like anatomical regions across subjects. In combination with large-scale computational neural network modeling, the ROI analysis permits the comparison of significant regions for particular conditions or contrasts from an experiment with like neuroanatomical regions hypothesized to be modulated in the model under the same conditions. When anatomical ROIs correspond to regions hypothesized to underlie the model, a direct comparison can be made between the regional BOLD responses and simulated activity. Like the voxel-based method, these comparisons should corroborate and strengthen initial model-based hypotheses and

assumptions, generate new insights about functional neural networks that guide improvements to the model, and result in the design of new experimental or model-based investigations. An advantage of ROI analysis over the voxel-based method, however, is that by testing hypotheses regarding the response of a specific set of ROIs, the number of statistical tests performed is greatly reduced compared to a typical voxel-based analysis, increasing statistical sensitivity.

We have developed a region-of-interest parcellation scheme that demarcates anatomical boundaries for the cerebral cortex and cerebellum based on an individual subject's neuroanatomical landmarks (See Figure 3; Tourville and Guenther, 2003). This parcellation scheme is a modified version of the scheme described by Caviness et al. (1996) and was designed to provide an automatic and standardized method to delineate ROIs that are particularly relevant for neuroimaging studies of speech. For example, the modified system distinguishes between superior temporal gyrus and superior temporal sulcus (a partition that is not included in the Caviness et al. scheme). The decision to partition these ROIs was based on results of imaging studies that suggest a phoneme processing center within the superior temporal sulcus (e.g., Binder et al., 2000; Scott et al., 2000). The parcellation scheme includes many of the regions hypothesized to underlie the DIVA neural network model, including the anterior supramarginal gyrus, pallidum, Heschl's gyrus, inferior frontal gyrus *pars opercularis*, frontal operculum, posterior superior temporal gyrus, planum temporale, putamen, superior lateral cerebellum, superior medial cerebellum, supplementary motor area, thalamus, ventral motor cortex, ventral premotor cortex, and ventral somatosensory cortex. Therefore, the modified parcellation scheme provides a means to compare functional responses in like anatomical regions across subjects and permits the testing of predictions generated from the DIVA model.

In the auditory feedback perturbation study, ROI analysis was used to determine which ROIs were significant for each contrast of interest, to test the laterality of responses in the auditory feedback control network (*shift – no shift* contrast), and to extract data for effective connectivity analysis (Tourville et al., 2008). Like the results from the voxel-based analysis, results from the ROI analysis for the auditory feedback control network supported the DIVA model's prediction for increased mediation of bilateral posterior superior temporal gyrus and planum temporale. Significant responses were also noted in right ventral motor cortex, right ventral premotor cortex, and right anterior medial cerebellum. Figure 5 illustrates the inter-hemispheric response differences of the regions tested that were found to significantly differ in at least one of the three contrasts of interest (*no shift – baseline*, *shift – baseline*, *shift – no shift*). Significant right lateralization for the auditory feedback control network was apparent only in the ventral premotor cortex ($t_{\text{two-tailed}} > 2.23$; $df = 10$; $p < 5\%$). This contrasts with significant left hemisphere lateralization of inferior frontal gyrus *pars opercularis*, ventral premotor cortex, and ventral motor cortex in the normal (unperturbed) speech production network (*no shift – baseline* contrast), indicating that auditory feedback control of speech is right-lateralized in the frontal cortex. This finding was not predicted by the DIVA model. In the next subsection we discuss effective connectivity analyses that were used to further characterize this auditory feedback control network and to motivate modifications to the DIVA model.

Effective connectivity analysis—Whereas voxel-based and ROI-based analyses of neuroimaging data focus on the functional specialization of brain regions, inter-regional effective connectivity analyses focus on the functional integration of brain regions. Structural equation modeling (SEM) provides a means to quantitatively assess the interactions between a number of preselected brain regions and is widely used to make effective connectivity inferences from neuroimaging data (Penny et al., 2004). SEM requires the specification of a causal, directed anatomical model. The relative strength of the directed connections is determined by path coefficients, which are estimated by minimizing the difference between the observed inter-regional covariance matrix and the covariance matrix implied by the model.

If the data are partitioned into two different task manipulations, a stacked model approach can be used to make inferences about changes in effective connectivity between the task manipulations (Della-Maggioire et al., 2000).

Structural equation modeling is a confirmatory tool; effective connectivity is assessed to test the feasibility of hypothesized path, or structural, models. SEM is therefore well-suited for testing the connectivity implied by a large-scale computational model. While it would be just as feasible to test the connectivity implied by the DIVA model using other connectivity analysis techniques including dynamic causal modeling (DCM; Friston et al., 2003), in the discussion to follow we show how SEM can be used to confirm or rebut hypothesized connections since this technique, unlike DCM, can be used with the sparse sampling fMRI paradigms typically used in our speech production studies.

In the auditory feedback perturbation study, SEM was used to assess projections between five sensory and motor cortical regions, including the left and right superior temporal gyrus, the right inferior frontal gyrus *pars triangularis*, the right ventral premotor cortex, and the right ventral motor cortex. Network connectivity, schematized in Figure 6, was constrained by interactions conceptualized within the DIVA model as well as consideration of the voxel-wise and ROI analysis results. For example, the DIVA model describes reciprocal connections between ventral frontal and posterior temporal cortex that enable a comparison of expected and realized acoustic consequences and drive compensatory movements in the event an error is detected. Effective connectivity was therefore assessed between those ventral frontal and posterior temporal regions that were found to be significantly more active in the *shift – no shift* contrast.

The implied covariance of the network provided a good fit to the empirical covariance across all conditions and met all goodness-of-fit criteria (see Tourville et al., 2008 for details). Estimates of the path coefficients for the *shift* (red) and *no shift* (green) conditions are provided in Figure 6. Pair-wise comparisons of the coefficients demonstrated that connection strengths from the left posterior superior temporal gyrus to the right posterior superior temporal gyrus, from the left posterior superior temporal gyrus to the right ventral premotor cortex, and from the right posterior superior temporal gyrus to the right inferior frontal gyrus *pars triangularis* were significantly greater in the *shift* condition, indicating that these pathways were engaged to a greater degree when auditory feedback control was invoked.

Based on the results of the effective connectivity analysis and recent findings in our laboratory on somatosensory feedback control (Golfinopoulos et al., in preparation), we determined that extensions should be made to the DIVA model to include projections from bilateral sensory cortical areas to right ventral premotor cortex. This constituted a change in the model since the earlier version included projections from the sensory error maps directly to primary motor cortex bilaterally, rather than to right-lateralized ventral premotor cortex. In the next subsection we describe the model modification process.

Computational Model Refinement—The utility of a large-scale computational neural network model lies in its testability. We have demonstrated how predictions of the DIVA model are well-suited for testing in neuroimaging experiments. In our illustrative example, several hypotheses generated from the model were supported. However, the empirical data made it clear that feedback-based articulator control relied on right-lateralized contributions of the ventral premotor and inferior frontal cortex. Based on this finding, we added the right-lateralized *Feedback Control Map*, highlighted in Figure 7, to the model. Projections from bilateral *Auditory* and *Somatosensory Error Maps* convey error information to the *Feedback Control Map*. The projections from the *Feedback Control Map* to the *Articulator Velocity and Position Maps* transform these error signals into feedback-based articulator velocity commands

that contribute to the outgoing motor program. Activation within the *Feedback Control Map* is hypothesized to maintain representations of higher order aspects of movement such as full articulatory gestures and/or their auditory correlates.

The *Feedback Control Map* is hypothesized to lie in the right ventral premotor cortex and its location in MNI space (MNI $xyz = [60, 14, 34]$) was based on the peak ventral premotor activation noted in the *shift – no shift* contrast of the auditory feedback control study. However, the location and function of this map were not based solely on the auditory feedback perturbation study. Recent work in our laboratory has demonstrated the recruitment of right-lateralized inferior frontal gyrus *pars triangularis* and ventral premotor cortex for feedback-based correction of speech motor commands to somatosensory perturbations (Golfinopoulos et al., in preparation). Right-lateralized contributions from ventral premotor cortex and/or adjacent inferior frontal cortex to feedback-based control of vocal fundamental frequency (Fu et al., 2006; Houde and Nagarajan, 2007; Toyomura et al., 2007) and visuomotor tracking (Grafton et al., 2008) has also been demonstrated. Ventral premotor activity was found to correlate with the magnitude of compensatory responses to fundamental frequency perturbations (Houde and Nagarajan, 2007) and feedback-based learning of arm movements (Grafton et al., 2008). The addition of a right-lateralized *Feedback Control Map* to the DIVA model, therefore, is consistent with a number of experimental findings.

The addition of the *Feedback Control Map* improves the model's functional neuroanatomical account of feedback control without reducing the model's other functional capabilities. This modification demonstrates an additional, powerful advantage of integrating large-scale neurocomputational modeling with empirical research. Modifications motivated by the most recent research must fit within an established theoretical framework and maintain the explanatory power of the model. With the addition of the *Feedback Control Map*, for instance, it was necessary to ensure that the model could capture the findings of both auditory and somatosensory feedback-based control of speech, since both sensory reference frames were originally accounted for by the model. Ultimately, a robust large-scale computational model is one that not only fits existing data, but may also be modified to fit new findings. It is necessary, however, that these modifications accommodate initial model capabilities so that the model can continue to account for a diverse and extensive data set.

Although we have limited our examples in this article to fitting fMRI data, it is also as feasible to test DIVA model predictions using empirical data from other techniques such as electroencephalography (EEG), magnetoencephalography (MEG), electrocorticography (ECOG), and transcranial magnetic stimulation (TMS). It is our hope that this review will inspire other researchers to use the DIVA model to motivate and guide their experiments, using any technique available to them. In the end, our goal is to successfully integrate experimental research and theory to develop a valid and comprehensive model that withstands the test of time.

Discussion

Classical theories of speech production and perception attributed particular speech and language disorders to dedicated, independent regions in the brain (Lichtheim, 1885). Modern theory posits a more distributed network that involves multiple interacting cortical and subcortical brain areas (e.g., Gracco, 1991; Baum and Pell, 1999; Hickok and Poeppel, 2000; Sidtis and Van Lancker Sidtis, 2003; Friederici and Alter, 2004; Riecker et al., 2005; Indefrey and Levelt, 2004; Pulvermüller, 2005). Although most models of speech production and perception rely on simplified frameworks and certain fallible constraints, many have been catalysts for innovative investigations and empirical findings that either confirmed or refuted initial predictions. To do so, the model must be specified in enough detail to generate testable

predictions and amenable to modifications in the (inevitable) event that some new empirical data do not support the model's predictions.

The DIVA neural network model proposes a control scheme using a dual-sensory (auditory and somatosensory) reference frame interacting with a feedforward control subsystem to account for speech acquisition and production. Through numerous computer simulations over the past two decades, the DIVA model architecture has succeeded in accounting for a very wide range of kinematic (Guenther, 1994; 1995; Guenther et al., 1998; Callan et al., 2000; Guenther et al., 2006), acoustic (Guenther et al., 1998; Tourville et al., 2008), and fMRI data (Guenther et al., 2006; Tourville et al., 2008), as well as aspects of communication disorders that can be simulated through damage to the model's components (e.g., Max et al., 2004; Civier et al., submitted; Terband et al., in press). The model thus constitutes a “unifying theory” that accounts for a far wider range of empirical findings from disparate fields (including neuroimaging, communication disorders, and motor control) than prior models of speech.

Nonetheless, it is important to note that the model in its current form still provides an incomplete account of the neural bases of speech. For example, there is evidence that the inferior cerebellum, anterior insula, and anterior cingulate gyrus may also be involved in the detection of sensory error in speech production, based on recent fMRI studies (Christoffels et al., 2007; Tourville et al., 2008; Golfinopoulos et al., in preparation). Clarifying the functional and computational contributions of these regions through carefully designed experiments may make the model more comprehensive and may also provide greater functionality to the current circuitry. In addition, while the DIVA model incorporates brain regions that mediate the acquisition and execution of sensorimotor programs of speech sounds, it does not account for those brain areas likely responsible for higher-level syllable sequence planning. We have recently developed a new neural network model of the mechanisms underlying the selection, sequencing, and initiation of speech movements, termed GODIVA (Bohland et al., in press), to account for areas such as the inferior frontal sulcus, the pre-supplementary motor area, and the caudate nucleus which are hypothesized to contribute to sequence planning and timing the release of planned speech movements to the vocal tract. Once these two models are fully integrated, the computational characterizations of regions that contribute to the *Initiation Map*, which are elaborated in the GODIVA model, will be more thoroughly described in the DIVA model.

An additional aspect of the neural control of speech not yet included in the DIVA model is the possible interaction between auditory and somatosensory reference frames. Currently, the DIVA model employs two independent sensory reference frames. It seems unlikely, however, that these two sensory reference frames influence feedback-based motor control independently (e.g., see Hickok and Poeppel, 2000, 2004). Evidence for connections between the superior temporal and inferior parietal regions is not substantial in the non-human primate literature, but Aboitiz and García (1997) have proposed that in human evolution these connections became increasingly developed, linking the auditory system with inferior parietal regions such as the supramarginal gyrus. Whether this linkage underlies a combined somato-auditory reference frame and whether such a combined reference frame is functionally critical for the production of speech sounds remains to be clarified.

We are also implementing improvements to the DIVA model software to incorporate simulated ROI-based analysis in addition to the simulated voxel-based analysis. The ROI analysis permits the omission of such pre-processing steps as non-linear normalization and smoothing and would provide an additional quantitative measure to assess significant regions and significant laterality effects in the simulated activations acquired from the DIVA model. Such a simulated ROI analysis could then be compared quantitatively with region of interest analyses of neuroimaging data acquired from speech production experiments. In the long run, we anticipate

adding new ROIs to the model as well as partitioning certain ROIs into smaller regions when motivated by experimental results.

Despite many outstanding issues concerning the neural bases of speech, we believe neuroanatomically and computationally driven approaches such as that embodied by the DIVA model are essential for moving research into the neural bases of complex perceptual, motor, and cognitive tasks beyond simplistic “box and arrow” characterizations. Such neurocomputational models not only theoretically motivate experiments to test model predictions, but they also offer empirical research a framework in which to understand the relationship between neural dynamics and behavior in both neurologically normal and disordered populations.

Acknowledgments

This research was supported by the National Institute on Deafness and other Communication Disorders (R01 DC02852, F. Guenther PI) and by CELEST, an NSF Science of Learning Center (NSF SBE-0354378). Imaging was made possible with support from the National Center for Research Resources grant P41RR14075 and the MIND Institute. The authors would like to thank Jon Brumberg for his patience and generous assistance and the entire Speechlab (past and present) for contributing ideas that continue to enhance the DIVA model.

References

- Abbs JH, Gracco VL. Control of complex motor gestures: Orofacial muscle responses to load perturbation of lip during speech. *J. Neurophysiol* 1984;51(4):705–723. [PubMed: 6716120]
- Aboitiz F, García V. The evolutionary origin of the language areas in the human brain. A neuroanatomical perspective. *Brain Res. Rev* 1997;25:381–396. [PubMed: 9495565]
- Ackermann H, Vogel M, Peterson D, Poremba M. Speech deficits in ischaemic cerebellar lesions. *Neurology* 1992;239:223–227.
- Baciu M, Abry C, Segebarth C. Equivalence motrice et dominance hémisphérique. Le cas de la voyelle u.. Étude IRMf. Actes XXIIIème Journées d'Etude sur la Parole 2000:213–216.
- Ballard KJ, Granier JP, Robin DA. Understanding the nature of apraxia of speech: Theory, analysis, and treatment. *Aphasiology* 2000;14(10):969–995.
- Baum S, Pell M. The neural bases of prosody: Insights from lesion studies and neuroimaging. *Aphasiology* 1999;13:581–608.
- Bauer JJ, Mittal J, Larson CR, Hain TC. Vocal responses to unanticipated perturbations in voice loudness feedback: an automatic mechanism for stabilizing voice amplitude. *J. Acoust. Soc. Am* 2006;119:2363–2371. [PubMed: 16642849]
- Belin P, Zatorre R, Hoge R, Evans A, Pike B. Event-Related fMRI of the auditory cortex. *NeuroImage* 1999;10:417–429. [PubMed: 10493900]
- Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* 2000;10:512–528. [PubMed: 10847601]
- Bohland JW, Guenther FH. An fMRI investigation of syllable sequence production. *NeuroImage* 2006;32:821–841. [PubMed: 16730195]
- Bohland JW, Guenther FH, Bullock D. Neural representations and mechanisms for the performance of simple speech sequences. *J. Cognitive Neurosci.* in press.
- Boling W, Reutens DC, Olivier A. Functional topography of the low postcentral area. *J. Neurosurg* 2002;97(2):388–395. [PubMed: 12186467]
- Broca P. Remarques sur la siège de la faculté de la parole articulée, suivies d'une observation d'aphémie perte de parole. *Bull. Soc. Anat* 1861;36:330–357.
- Browman CP, Goldstein L. Articulatory gestures as phonological units. *Phonology* 1989;6:201–251.
- Brown S, Ngan E, Liotti M. A larynx area in the human motor cortex. *Cereb. Cortex* 2008;18(4):837–845. [PubMed: 17652461]

- Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, Gallese V, Seitz RJ, Zilles K, Rizzolatti G, Freund H-J. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur. J. Neurosci* 2001;13:400–404. [PubMed: 11168545]
- Buchsbaum BR, Hickok G, Humphries C. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Sci* 2001;25(5):663–678.
- Callan DE, Kent RD, Guenther FH, Vorperian HK. An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system. *J. Speech Lang. Hear. R* 2000;43:721–736.
- Caviness VS, Meyer J, Makris N, Kennedy D. MRI-based topographic parcellation of human neocortex: an anatomically specified method with estimate of reliability. *J. Cogn. Neurosci* 1996;8:566–587.
- Christoffels IK, Formisano E, Schiller NO. Neural correlates of verbal feedback processing: an fMRI study employing overt speech. *Hum. Brain Mapp* 2007;28:868–879. [PubMed: 17266104]
- Civier O, Tasko S, Guenther FH. Overreliance on auditory feedback may lead to sound/syllable repetitions: simulations of stuttering and fluency-inducing conditions with a neural model of speech production. *J. Fluency Disord.* submitted.
- Damasio H, Damasio AR. The anatomical basis of conduction aphasia. *Brain* 1980;103(2):337–350. [PubMed: 7397481]
- Della-Maggiore V, Sekuler AB, Grady CL, Bennett PJ, Sekuler R, McIntosh AR. Corticolimbic interactions associated with performance on a short-term memory task are modified by age. *J. Neurosci* 2000;20(22):8410–8416. [PubMed: 11069948]
- Fesl G, Moriggl B, Schmid UD, Naidich TP, Herholz K, Yousry TA. Inferior central sulcus: variations of anatomy and function on the example of the motor tongue area. *NeuroImage* 2003;20:601–610. [PubMed: 14527621]
- Fink GR, Corfield DR, Murphy K, Kobayashi I, Dettmers C, Adams L, Frackowiak RS, Guz A. Human cerebral activity with increasing inspiratory force: a study using positron emission tomography. *J. Appl. Physiol* 1996;81(3):1295–1305. [PubMed: 8889766]
- Friederici AD, Alter K. Lateralization of auditory language functions: a dynamic dual pathway model. *Brain Lang* 2004;89:267–276. [PubMed: 15068909]
- Friston KJ, Büchel C, Fink GR, Morris J, Rolls E, Dolan RJ. Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage* 1997;6:218–229. [PubMed: 9344826]
- Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ. Functional connectivity: the principal-component analysis of large (PET) data sets. *J. Cereb. Blood. Flow. Metabol* 1993;13:5–14.
- Friston KJ, Harrison L, Penny W. Dynamic causal modeling. *NeuroImage* 2003;19:1273–1302. [PubMed: 12948688]
- Friston KJ, Holmes AP, Poline J-B, Grasby PJ, Williams SCR, Frackowiak RSJ, Turner R. Analysis of fMRI time series revisited. *NeuroImage* 1995;2:45–53. [PubMed: 9343589]
- Fu CH, Vythelingum GN, Brammer MJ, Williams SC, Amaro E Jr, Andrew CM, Yaguez L, van Haren NE, Matsumoto K, McGuire PK. An fMRI study of verbal self-monitoring: neural correlates of auditory verbal feedback. *Cereb. Cortex* 2006;16:969–977. [PubMed: 16195470]
- Gay T, Lindblom B, Lubker J. Production of bite-block vowels: Acoustic equivalence by selective compensation. *J. Acoust. Soc. Am* 1981;69(3):802–810. [PubMed: 7240561]
- Ghosh, SS. PhD thesis. Boston University; Boston, MA: 2005. Understanding cortical and cerebellar contributions to speech production through modeling and functional imaging.
- Ghosh SS, Tourville JA, Guenther FH. A neuroimaging study of premotor lateralization and cerebellar involvement in the production of phonemes and syllables. *J. Speech Lang. Hear. Res* 2008;51:1183–1202. [PubMed: 18664692]
- Goense JBM, Logothetis NK. Neurophysiology of the BOLD fMRI signal in awake monkeys. *Curr. Biol* 2008;18:631–640. [PubMed: 18439825]
- Golfinopoulos E, Tourville JA, Bohland JW, Ghosh SS, Guenther FH. Neural circuitry underlying somatosensory feedback control in speech production. in preparation.
- Goodglass, H. Understanding aphasia. Academic Press; New York: 1993.

- Gracco, VL. Sensorimotor mechanisms in speech motor control. In: Peters, H.; Hultsijn, W.; Starkweather, CW., editors. *Speech Motor Control and Stuttering*. Elsevier; North Holland: 1991. p. 53-78.
- Gracco VL, Abbs JH. Dynamic control of the perioral system during speech: Kinematic analyses of autogenic and nonautogenic sensorimotor processes. *J. Neurophysiol* 1985;54(2):418–432. [PubMed: 4031995]
- Grafton ST, Schmitt P, Van Horn J, Diedrichsen J. Neural substrates of visuomotor learning based on improved feedback and prediction. *NeuroImage* 2008;39:1383–1395. [PubMed: 18032069]
- Guenther FH. A neural network model of speech acquisition and motor equivalent speech production. *Biol. Cybern* 1994;72:43–53. [PubMed: 7880914]
- Guenther FH. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychol. Rev* 1995;102:594–621. [PubMed: 7624456]
- Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang* 2006;96:280–301. [PubMed: 16040108]
- Guenther FH, Hampson M, Johnson D. A theoretical investigation of reference frames for the planning of speech movements. *Psychol. Rev* 1998;105(4):611–633. [PubMed: 9830375]
- Guenther FH, Nieto-Castanon A, Ghosh SS, Tourville JA. Representation of sound categories in auditory cortical maps. *J. Speech Lang. Hear. R* 2004;47(1):46–57.
- Hashimoto Y, Sakai KL. Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: an fMRI study. *Hum. Brain Mapp* 2003;20:22–28. [PubMed: 12953303]
- Heeger DJ, Huk AC, Geisler WS, Albrecht DG. Spikes versus BOLD: what does neuroimaging tell us about neuronal activity? *Nat. Neurosci* 2000;3(7):631–633. [PubMed: 10862687]
- Hickok G, Poeppel D. Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci* 2000;4:131–138. [PubMed: 10740277]
- Hickok G, Poeppel D. Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition* 2004;92:67–99. [PubMed: 15037127]
- Houde JF, Jordan MI. Sensorimotor adaptation in speech production. *Science* 1998;279(5354):1213–1216. [PubMed: 9469813]
- Houde, F.; Nagarajan, SS. Program of the 154th Meeting of the Acoustical Society of America, *Journal of the Acoustical Society of America*. Vol. vol. 122. Louisiana; New Orleans: 2007. How is auditory feedback processed during speaking?; p. 3087
- Indefrey P, Levelt J. The spatial and temporal signatures of word production components. *Cognition* 2004;92(1–2):101–144. [PubMed: 15037128]
- Johnson MD, Ojemann GA. The role of the human thalamus in language and memory: evidence from electrophysiological studies. *Brain Cognition* 2000;42(2):218–230.
- Jones JA, Munhall KG. Remapping auditory-motor representations in voice production. *Curr. Biol* 2005;15(19):1768–1772. [PubMed: 16213825]
- Kelso JAS, Tuller B. 'Compensatory articulation' under conditions of reduced afferent information: a dynamic formulation. *J. Speech Lang. Hear. Res* 1983;26:217–224.
- Levelt WJ, Wheeldon L. Do speakers have access to a mental syllabary? *Cognition* 1994;50:239–269. [PubMed: 8039363]
- Levelt WJ, Roelofs A, Meyer AS. A theory of lexical access in speech production. *Behav. Brain Sci* 1999;22(1):1–38. [PubMed: 11301520]
- Lichtheim L. On aphasia. *Brain* 1885;7:433–484.
- Lindblom B, Lubker J, Gay T. Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *J. Phonetics* 1979;7:147–161.
- Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A. Neurophysiological investigation of the basis of the fMRI signal. *Nature* 2001;412:150–157. [PubMed: 11449264]
- Lotze M, Erb M, Flor H, Huelsmann E, Godde B, Grodd W. fMRI evaluation of somatotopic representation in human primary motor cortex. *NeuroImage* 2000a;11(5 Pt 1):473–581. [PubMed: 10806033]
- Lotze M, Seggewies G, Erb M, Grodd W, Birbaumer N. The representation of articulation in the primary sensorimotor cortex. *Neuroreport* 2000b;11(13):2985–2989. [PubMed: 11006980]

- MacNeilage PF, Rootes TP, Chase RA. Speech Production and perception in a patient with severe impairment of somesthetic perception and motor control. *J. Speech Lang. Hear. Res* 1967;10(3):440–467.
- Maeda, S. Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. In: Hardcastle, WJ.; Marchal, A., editors. *Speech Production and Speech Modeling*. Kluwer Academic Publishers; Boston: 1990. p. 131-149.
- Mathiesen C, Caesar K, Akgoren N, Lauritzen M. Modification of activity-dependent increases of cerebral blood flow in rat cerebellum. *J. Physiol* 1998;560:181–189.
- Max L, Guenther FH, Gracco VL, Ghosh SS, Wallace ME. Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: A theoretical model of stuttering. *CICSD* 2004;31:105–122.
- Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paus T, Simpson G, Pike B, Holmes C, Collins L, Thompson P, MacDonald D, Iacoboni M, Schormann T, Amunts K, Palomero-Gallagher N, Geyer S, Parsons L, Narr K, Kabani N, Le Goualher G, Feidler J, Smith K, Boomsma D, Hulshoff Pol H, Cannon T, Kawashima R, Mazoyer B. A four-dimensional probabilistic atlas of the human brain. *J. Am. Med. Inform. Assoc* 2001;8:401–430. [PubMed: 11522763]
- McCarthy G, Allison T, Spencer DD. Localization of the face area of human sensorimotor cortex by intracranial recording of somatosensory evoked potentials. *J. Neurosurg* 1993;79(6):874–884. [PubMed: 8246056]
- McKeown MJ, Makeig S, Brown GG, Jung TP, Kindermann SS, Bell AJ, Sejnowski TJ. Analysis of fMRI data by blind separation into independent spatial components. *Hum. Brain Mapp* 1998;6:160–188. [PubMed: 9673671]
- McIntosh AR, Gonzalez-Lima F. Structural equation modeling and its application to network analysis in functional imaging. *Hum. Brain Mapp* 1994;2:2–22.
- McNeil, MR.; Robin, DA.; Schmidt, RA. *Apraxia of speech: Definition and differential diagnosis*. 2nd ed.. Thieme; New York: 2007.
- Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, Zilles K. Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *NeuroImage* 2001;13(4):684–701. [PubMed: 11305897]
- Mukamel R, Gelbard H, Arieli A, Hasson , Fried I, Malach R. Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. *Science* 2005;309:951–954. [PubMed: 16081741]
- Nasir SM, Ostry DJ. Somatosensory precision in speech production. *Curr. Biol* 2006;16:1918–1923. [PubMed: 17027488]
- Nieto-Castanon A, Ghosh SS, Tourville JA, Guenther FH. Region of interest based analysis of functional imaging data. *NeuroImage* 2003;19:1303–1316. [PubMed: 12948689]
- Nieto-Castanon A, Guenther FH, Perkell JS, Curtin HD. A modeling investigation of articulatory variability and acoustic stability during American English /r/ production. *J. Acoust. Soc. Am* 2005;117(5):3196–3212. [PubMed: 15957787]
- Olthoff A, Baudewig J, Kruse , Dechent P. Cortical sensorimotor control in vocalization: a functional magnetic resonance imaging study. *Laryngoscope* 2008;118(11):2091–2096. [PubMed: 18758379]
- Penfield, W.; Rasmussen, T. *The cerebral cortex of man. A clinical study of localization of function*. Macmillan; New York: 1950.
- Penfield, W.; Roberts, L. *Speech and Brain Mechanisms*. Princeton University Press; Princeton: 1959.
- Penny WD, Stephan KE, Mechelli A, Friston KJ. Modelling functional integration: a comparison of structural equation and dynamic causal models. *NeuroImage* 2004;23:S264–S274. [PubMed: 15501096]
- Pulvermüller F. Brain mechanisms linking language and action. *Nat. Rev., Neurosci* 2005;6(7):576–582. [PubMed: 15959465]
- Purcell DW, Munhall KG. Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. *J. Acoust. Soc. Am* 2006;120:966–977. [PubMed: 16938984]
- Raichle ME, Mintun MA. Brain work and brain imaging. *Annu. Rev. Neurosci* 2006;3:716–723.
- Rees G, Friston K, Koch C. A direct quantitative relationship between the functional properties of human and macaque V5. *Nat. Neurosci* 2000;3(7):716–723. [PubMed: 10862705]

- Riecker A, Mathiak K, Wildgruber D, Erb M, Hertrich I, Grodd W, Ackermann H. fMRI reveals two distinct cerebral networks subserving speech motor control. *Neurology* 2005;64:700–706. [PubMed: 15728295]
- Rivier F, Clarke S. Cytochrome oxidase, acetylcholinesterase, and NADPHdiaphorase staining in human supratemporal and insular cortex: evidence for multiple auditory areas. *NeuroImage* 1997;6(4):288–304. [PubMed: 9417972]
- Rizzolatti G, Fadiga L, Gallese V, Fogassi L. Premotor cortex and the recognition of motor actions. *Cogn. Brain Res* 1996;3:131–141.
- Robin, DA.; Guenther, FH.; Narayana, S.; Jacks, A.; Tourville, JA.; Ramage, AE.; Lancaster, JL.; Franklin, C.; Ghosh, S.; Fox, PT. A transcranial magnetic stimulation virtual lesion study of speech. *Proceedings of the Conference on Motor Speech*; Monterey, California. 2008.
- Scott SK, Blank CC, Rosen S, Wise RJ. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 2000;123:2400–2406. [PubMed: 11099443]
- Sidtis JJ, Van Lancker-Sidtis D. A neurobehavioral approach to dysprosody. *Semin. Speech Lang* 2003;24:93–105. [PubMed: 12709883]
- Terband H, Maassen B, Guenther FH, Brumberg J. Computational neural modeling of speech motor control in childhood apraxia of speech. *J. Speech Lang. Hear. R.* in press.
- Tourville, JA.; Guenther, FH. Tech Rep CAS/CNS-03-022. Boston University; Boston, MA: 2003. A cortical and cerebellar parcellation system for speech studies.
- Tourville JA, Reilly KJ, Guenther FH. Neural mechanisms underlying auditory feedback control of speech. *NeuroImage* 2008;39:1429–1443. [PubMed: 18035557]
- Toyomura A, Koyama S, Miyamaoto T, Terao A, Omori T, Murohashi H, Kuriki S. Neural correlates of auditory feedback control in human. *Neuroscience* 2007;146:499–503. [PubMed: 17395381]
- Tremblay S, Shiller DM, Ostry DJ. Somatosensory basis of speech production. *Nature* 2003;423:866–869. [PubMed: 12815431]
- Van Buren JM. Confusion and disturbance of speech from stimulation in the vicinity of the head of the caudate nucleus. *J. Neurosurg* 1963;20:148–157. [PubMed: 14192084]
- Viswanathan A, Freeman RD. Neurometabolic coupling in cerebral cortex reflects synaptic more than spiking activity. *Nat. Neurosci* 2007;10:1308–1312. [PubMed: 17828254]
- Wernicke, C. *Der aphasische Symptomencomplex: Eine psychologische Studie auf anatomischer Basis.* Kohn und Weigert; Breslau: 1874.
- Yang Y, Engelen A, Engelen W, Xu S, Stern E, Silbersweig DA. A silent event-related functional MRI technique for brain activation studies without interference of scanner acoustic noise. *Magn. Reson. Med* 2000;43:185–190. [PubMed: 10680681]

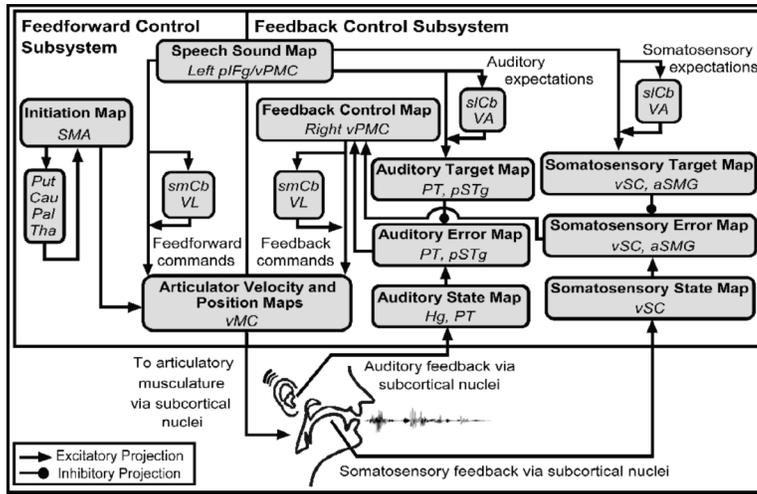
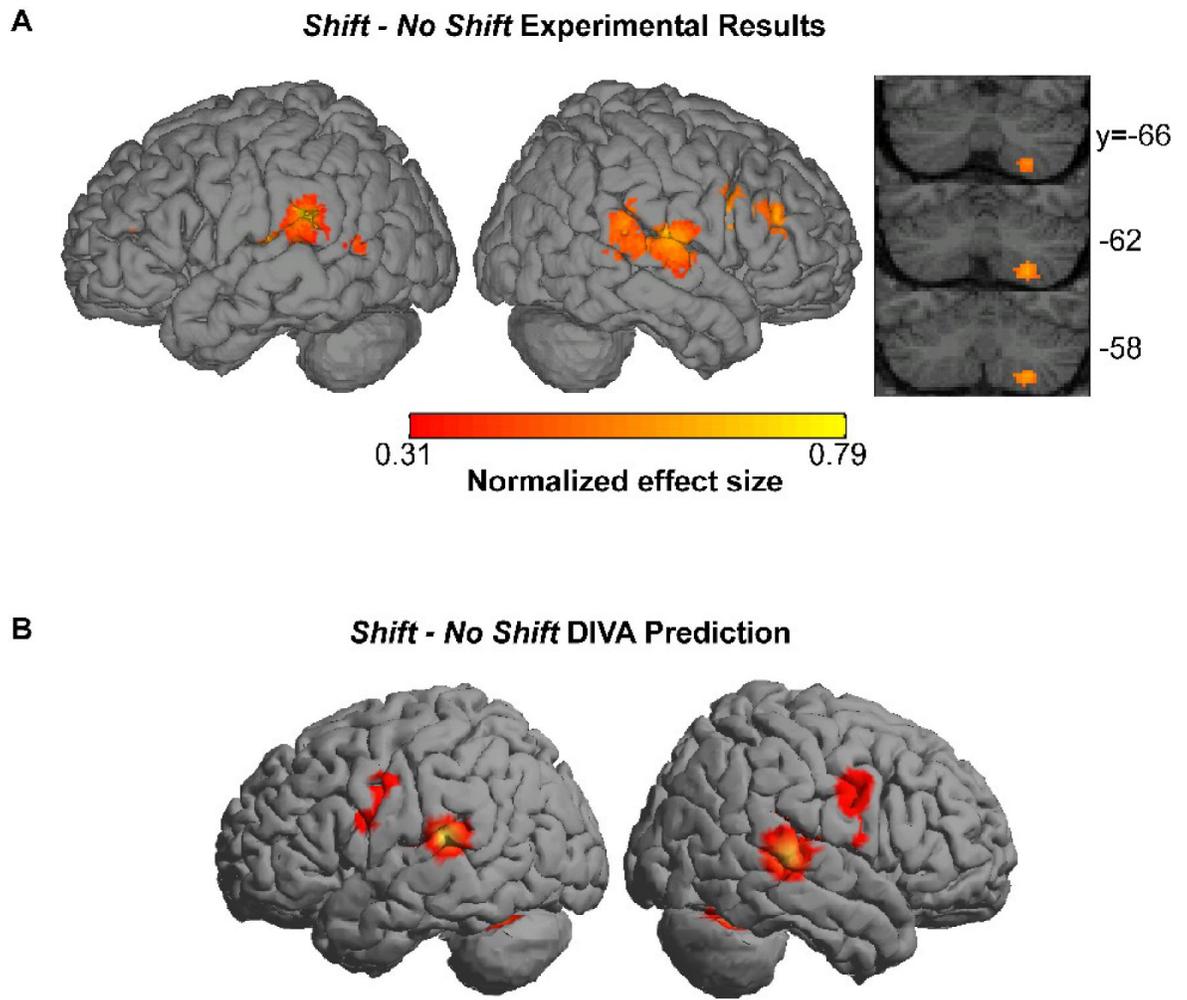


Figure 1. Schematic of the DIVA neural network model. Each box corresponds to a set of neurons (or map) and arrows between the boxes correspond to synaptic projections that transform one type of neural representation into another. The model is divided into two basic systems: the Feedforward Control Subsystem on the left and the Feedback Control Subsystem on the right. The neural substrates underlying this integrated control scheme include the premotor and primary motor cortices, somatosensory cortices, auditory cortices, the cerebellum, and the basal ganglia. Matlab source code for the DIVA model and a version of the Maeda (1990) vocal tract that may be used to produce simulated speech output is available at our website (<http://speechlab.bu.edu/software.php>). Abbreviations: aSMg = anterior supramarginal gyrus; Cau = caudate; Pal = pallidum; Hg = Heschl's gyrus; pIFg = posterior inferior frontal gyrus; pSTg= posterior superior temporal gyrus; PT = planum temporale; Put = Putamen; sICb = superior lateral cerebellum; smCb = superior medial cerebellum; SMA = supplementary motor area; Tha = thalamus; VA = ventral anterior nucleus of the cerebellum; VL = ventral lateral nucleus of the thalamus; vMC = ventral motor cortex; vPMC = ventral premotor cortex; vSC = ventral somatosensory cortex.



B *Shift - No Shift* DIVA Prediction

Figure 2. Empirical (A) and simulated (B) fMRI data demonstrating the comparable responses for the *shift - no shift* contrast. Figure is reprinted from Tourville et al. (2008) with permission from the authors.

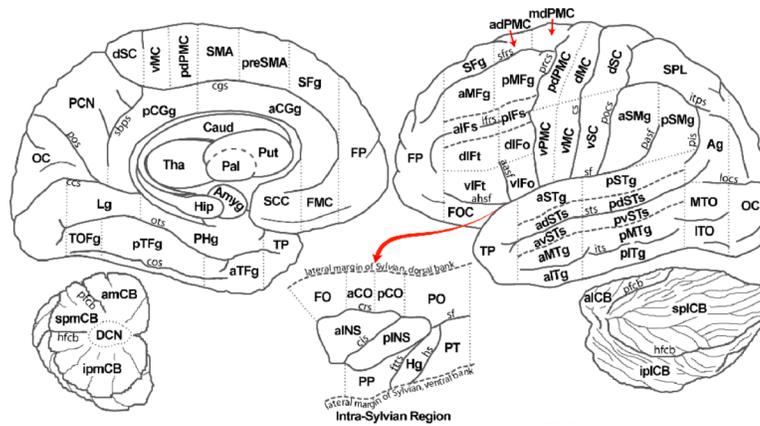


Figure 3.

A cortical and cerebellar parcellation scheme based on the Caviness et al. (1996) parcellation scheme. Dashed lines indicate boundaries between adjacent regions. The intra-Sylvian region is schematized as an exposed flattened surface as indicated by the red arrow. The detached labeled cerebellum is also shown in the lower left and lower right. ROI abbreviations: aCGg = anterior cingulate gyrus; aCO = anterior central operculum; adPMC = anterior dorsal premotor cortex; adSTs = anterior dorsal superior temporal sulcus; Ag = angular gyrus; aIFs = anterior inferior frontal sulcus; aINS = anterior insula; aITg = anterior inferior temporal gyrus; aICB = anterior lateral cerebellum; amCB = anterior medial cerebellum; aMFg = anterior middle frontal gyrus; aMTg = anterior middle temporal gyrus; Amyg = amygdala; aPH = anterior parahippocampal gyrus; aSMg = anterior supramarginal gyrus; aSTg = anterior superior temporal gyrus; aTFg = anterior temporal fusiform gyrus; avSTs = anterior ventral superior temporal sulcus; Caud = caudate; DCN = deep cerebellar nuclei; dIFo = dorsal inferior frontal gyrus, *pars opercularis*; dIFt = dorsal inferior frontal gyrus, *pars triangularis*; dMC = dorsal primary motor cortex; dSC = dorsal somatosensory cortex; FMC = frontal medial cortex; FO = frontal operculum; FOC = fronto-orbital cortex; FP = frontal pole; Hg = Heschl's gyrus; Hip = hippocampus; iplCB = inferior posterior lateral cerebellum; ipmCB = inferior posterior medial cerebellum; ITO = inferior temporal occipital gyrus; Lg = lingual gyrus; mdPMC = middle dorsal premotor cortex; MTO = middle temporal occipital gyrus; OC = occipital cortex; Pal = pallidum; pCGg = posterior cingulate gyrus; pCO = posterior central operculum; PCN = precuneus cortex; pdPMC = posterior dorsal premotor cortex; pdSTs = posterior dorsal superior temporal sulcus; PHg = parahippocampal gyrus; pIFs = posterior inferior frontal sulcus; pINS = posterior insula; pITg = posterior inferior temporal gyrus; pMFg = posterior middle frontal gyrus; pMTg = posterior middle temporal gyrus; PO = parietal operculum; PP = planum polare; pPH = posterior parahippocampal gyrus; preSMA = pre-supplementary motor area; pSMg = posterior supramarginal gyrus; pSTg = posterior superior temporal gyrus; PT = planum temporale; pTFg = posterior temporal fusiform gyrus; pvSTs = posterior ventral superior temporal sulcus; Put = putamen; SCC = subcallosal cortex; SFg = superior frontal gyrus; slCB = superior lateral cerebellum; SMA = supplementary motor area; smCB = superior medial cerebellum; SPL = superior parietal lobule; splCB = superior posterior lateral cerebellum; spmCB = superior posterior medial cerebellum; Tha = thalamus; TOF = temporal occipital fusiform gyrus; TP = temporal pole; vIFo = ventral inferior frontal gyrus, *pars opercularis*; vIFt = ventral inferior frontal gyrus, *pars triangularis*; vMC = ventral primary motor cortex; vPMC = ventral premotor cortex; vSC = ventral somatosensory cortex. Sulci abbreviations: aasf = anterior association ramus of the Sylvian fissure; ahsf = anterior horizontal ramus of the Sylvian fissure; ccs = calcarine sulcus; cgs = cingulate sulcus; cis = central insular sulcus; cs = central sulcus; cos = collateral sulcus; crs = circular sulcus; fts = first transverse temporal sulcus; hfcB = horizontal fissure of the cerebellum; hs = Heschl's

sulcus; ifrs = inferior frontal sulcus; itps = intraparietal sulcus; its = inferior temporal sulcus; locs = lateral occipital sulcus; ots = occipitotemporal sulcus; pasf = posterior ascending Sylvian fissure; pfcfb = primary fissure of the cerebellum; pis = primary intermediate sulcus; pocs = postcentral sulcus; pos = parieto-occipital sulcus; prcs = precentral sulcus; sbps = subparietal sulcus; sf = Sylvian fissure; sfrs = superior frontal sulcus; sts = superior temporal sulcus

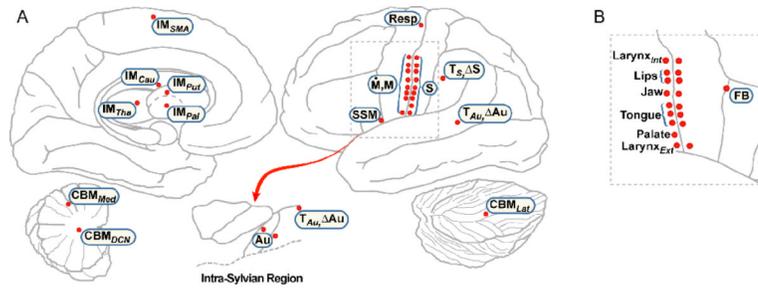


Figure 4.

Neuroanatomical mapping of the DIVA model. A. The location of DIVA model component sites (red dots) are plotted on a schematic of the left hemisphere. Medial regions are shown on the left, lateral regions on the right. B. A schematic of the right hemisphere lateral Rolandic and inferior frontal region. The corresponding contralateral region in the left hemisphere is outlined by the dashed box in A. The right hemisphere plot demonstrates the location of the Feedback Control Map and the location of motor and somatosensory representations of the articulators. *Abbreviations:* ΔAu = auditory error map; ΔS = somatosensory error map; Au = auditory state map; CBM_{DCN} = deep cerebellar nuclei; CBM_{Lat} = lateral cerebellum; CBM_{Med} = medial cerebellum; FB = feedback control map; IM_{Cau} = caudate initiation map; IM_{SMA} = supplementary motor area initiation map; IM_{Tha} = thalamus initiation map; IM_{Pal} = pallidum initiation map; IM_{Put} = putamen initiation map; $Larynx_{Int}$ = intrinsic larynx; $Larynx_{Ext}$ = extrinsic larynx; M = articulator position map; \dot{M} = articulator velocity map; $Resp$ = respiratory motor cells; S = somatosensory state map; SSM = speech sound map; T_{Au} = auditory target map; T_S = somatosensory target map.

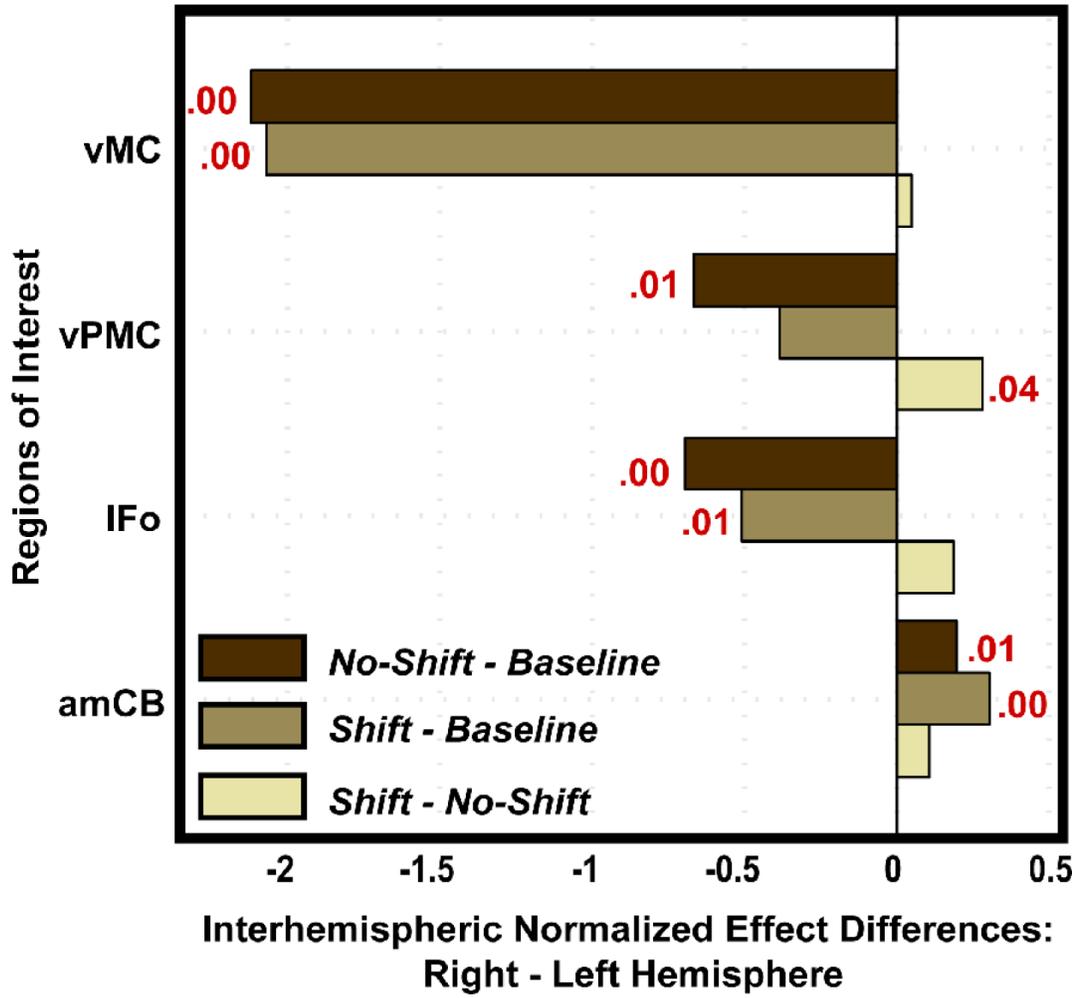


Figure 5. The difference between the normalized effects of the right and left hemispheres are shown for four ROIs that were found to differ significantly in at least one contrast (*no shift – baseline*, *shift – baseline*, *shift – no shift*). Positive difference values indicate a greater effect in the right hemisphere. The *p*-value is provided in red for those tests that resulted in a significant laterality effect. The laterality tests demonstrated left lateralized effects in the ventral frontal ROIs in the *no shift – baseline* contrast; this lateralized effect shifted to the right hemisphere in the *shift – no shift* contrast in ventral premotor cortex. Abbreviations: amCB = anterior medial cerebellum; IFo = inferior frontal gyrus, *pars opercularis*; vPMC = ventral premotor cortex; vMC = ventral motor cortex.

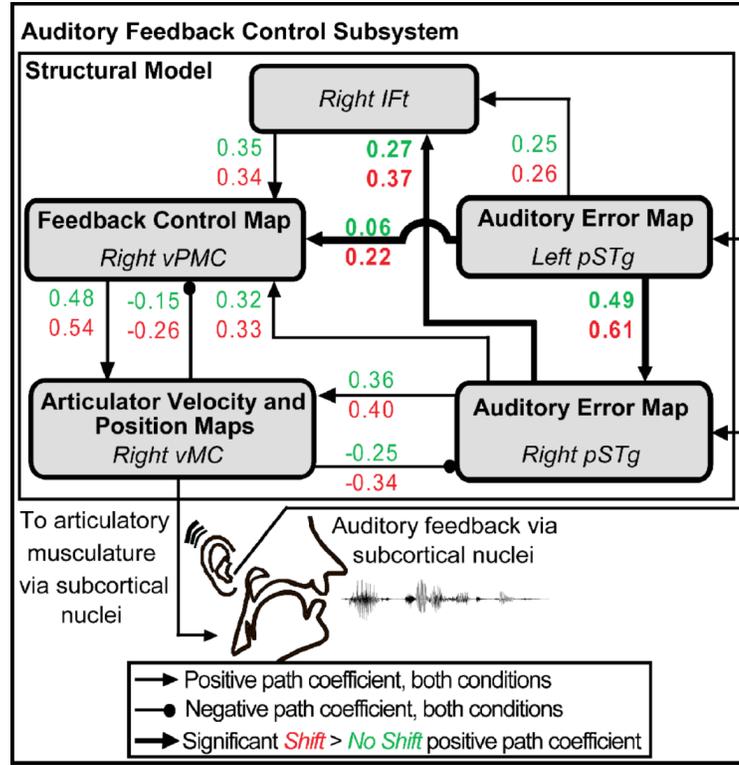


Figure 6. Schematic of the path diagram evaluated by structural equation modeling. Effective connectivity within the network of regions shown was significantly modulated by the auditory feedback perturbation. Path coefficients for all projections shown were significant in both conditions except that from right vMC to right vPMC (*no shift* condition $p = 0.07$). Pair-wise comparisons of path coefficients in the two conditions revealed significant interactions (highlighted in bold) due to the shift in auditory feedback in the projections between left pSTg to right pSTg, from left pSTg to right vPMC, and from right pSTg to right IFt. Abbreviations: IFt = inferior frontal gyrus, *pars triangularis*; pSTg = posterior superior temporal gyrus; vMC = ventral motor cortex; vPMC = ventral premotor cortex.

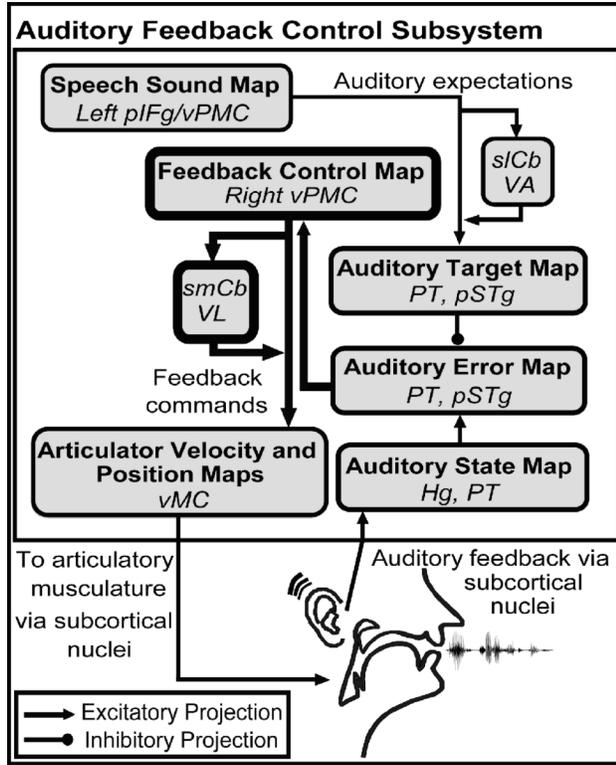


Figure 7. Schematic highlighting the right-lateralized Feedback Control Map and associated projections (indicated in bold). Abbreviations: Hg = Heschl's gyrus; pIFg = posterior inferior frontal gyrus; pSTg = posterior superior temporal gyrus; PT = planum temporale; slCB = superior lateral cerebellum; smCB = superior medial cerebellum; VA = ventral anterior nucleus of the thalamus; VL = ventral lateral nucleus of the thalamus; vMC = ventral motor cortex; vPMC = ventral premotor cortex.

Table 1

The locations of the DIVA neural network components in MNI space.

| Model Components | Hypothesized Neural Correlates | MNI Coordinates | | Select Citations Facilitating the Mapping Derivation |
|---------------------------|--|-----------------------|----------------------|--|
| | | Left Hemisphere | Right Hemisphere | |
| | | (x, y, z) | (x, y, z) | |
| Speech Sound Map | <i>Left Posterior Inferior Frontal Gyrus/Ventral Premotor Cortex</i> | (-56, 10, 2) | | Mapping hypothesized based on imaging studies (e.g., Ghosh et al., 2008; Tourville et al., 2008), apraxia of speech studies (e.g., Ballard et al., 2000; McNeil et al., 2007), and research on mirror neurons (e.g., Rizzolatti et al., 1996; Buccino et al., 2001). |
| Cerebellar Modules | <i>Superior Medial Cerebellum</i> | (-18, -59, -22) | (16, -59, -23) | Mapping hypothesized based on ataxic dysarthria studies (e.g., Ackermann et al., 1992) and Ghosh et al. (2008). |
| | <i>Superior Lateral Cerebellum</i> | (-36, -59, -27) | (40, -60, -28) | Mapping hypothesized based on a meta-analysis of pseudoword reading studies (Indefrey and Levelt, 2004) and imaging studies (e.g., Ghosh et al., 2008; Tourville et al., 2008). |
| | <i>Deep Cerebellar Nuclei</i> | (-10.3, -52.9, -28.5) | (14.4, -52.9, -29.3) | Mapping hypothesized to relay cerebellar output to the cerebral cortex during speech production. |
| Auditory State Map | <i>Heschl's Gyrus</i> | (-37.4, -22.5, 11.8) | (39.1, -20.9, 11.8) | Mapping hypothesized based on imaging studies (Guenther et al., 2004; Ghosh et al., 2008) and supported by Rivier and Clarke (1997) and Morosan et al. (2001). |
| | <i>Planum Temporale</i> | (-57.2, -18.4, 6.9) | (59.6, -15.1, 6.9) | |

| Model Components | Hypothesized Neural Correlates | MNI Coordinates | | Select Citations Facilitating the Mapping Derivation |
|--|--|----------------------|---------------------|---|
| | | Left Hemisphere | Right Hemisphere | |
| | | (x, y, z) | (x, y, z) | |
| Auditory Target/Error Maps | <i>Planum Temporale</i> | (-39.1, -33.2, 14.3) | (44, -30.7, 15.1) | Mapping hypothesized based on Buchsbaum et al. (2001) and Hickok and Poeppel (2004) and supported by Tourville et al. (2008). |
| | <i>Posterior Superior Temporal Gyrus</i> | (-64.6, -33.2, 13.5) | (69.5, -30.7, 5.2) | |
| Somatosensory State Map | <i>Ventral Somatosensory Cortex</i> | | | Mapping hypothesized based on imaging studies (e.g., Ghosh et al., 2008; Tourville et al., 2008). |
| Tongue 1 | | (-60.2, -2.8, 27) | (62.9, -1.5, 28.9) | Mapping hypothesized based on Boling et al. (2002) and supported by somatosensory evoked potential study of humans (McCarthy et al., 1993). |
| Tongue 2 | | (-60.2, -0.5, 23.3) | (66.7, -1.9, 24.9) | |
| Tongue 3 | | (-60.2, 0.6, 20.8) | (64.2, 0.1, 21.7) | |
| Upper Lip | | (-53.9, -7.7, 47.2) | (59.6, -10.2, 40.6) | Mapping hypothesized based on McCarthy et al. (1993). |
| Lower Lip | | (-56.4, -5.3, 42.1) | (59.6, -6.9, 38.2) | |
| Jaw | | (-59.6, -5.3, 33.4) | (62.1, -1.5, 34) | Mapping hypothesized based on motor jaw representation. |
| Larynx (Intrinsic) | | (-53, -8, 42) | (53, -14, 38) | Mapping hypothesized based on motor larynx representation. |
| Larynx (Extrinsic) | | (-61.8, 1, 7.5) | (65.4, 1.2, 12) | |
| Palate | | (-58, -0.7, 14.3) | (65.4, -0.4, 21.6) | Mapping hypothesized based on McCarthy et al. (1993). |
| Somatosensory Target/Error Maps | <i>Anterior Supramarginal Gyrus</i> | (-62.1, -28.4, 32.6) | (66.1, -24.4, 35.2) | Mapping hypothesized based on Baciu et al. (2000) and Hashimoto and Sakai (2003) and supported by Golfinopoulos et al. (in preparation). |
| Feedback Control Map | <i>Ventral Premotor Cortex</i> | | (60, 14, 34) | Mapping hypothesized based on |

| Model Components | Hypothesized Neural Correlates | MNI Coordinates | | Select Citations Facilitating the Mapping Derivation |
|---|-------------------------------------|-----------------|------------------|--|
| | | Left Hemisphere | Right Hemisphere | |
| | | (x, y, z) | (x, y, z) | |
| | | | | Tourville et al. (2008) and Golfinopoulos et al. (in preparation) |
| Initiation Map | <i>Supplementary Motor Area</i> | (0, 0, 68) | (2, 4, 62) | Mapping hypothesized based on imaging studies (e.g., Bohland and Guenther, 2006; Ghosh et al., 2008). |
| | <i>Caudate</i> | (-12, -2, 14) | (14, -2, 14) | Mapping hypothesized based on Van Buren (1963). |
| | <i>Putamen</i> | (-26, -2, 4) | (30, -14, 4) | Mapping hypothesized based on imaging studies (e.g., Bohland and Guenther, 2006; Ghosh et al., 2008; Tourville et al., 2008). |
| | <i>Globus Pallidus</i> | (-24, -2, -4) | (24, 2, -2) | Mapping hypothesized based on imaging studies (e.g., Bohland and Guenther, 2006; Ghosh et al., 2008; Tourville et al., 2008) |
| | <i>Thalamus</i> | (-10, -14, 8) | (10, -14, 8) | Mapping hypothesized based on electrical stimulation study Johnson and Ojemann (2000) and imaging studies (e.g., Bohland and Guenther, 2006; Ghosh et al., 2008; Tourville et al., 2008) |
| Articulator Position and Velocity Maps | <i>Ventral Primary Motor Cortex</i> | | | Mapping hypothesized based on imaging studies (e.g., Bohland and Guenther, 2006; Ghosh et al., 2008) and meta-analysis of pseudoword reading studies (Indefrey and Levelt, 2004). |

| Model Components | Hypothesized Neural Correlates | MNI Coordinates | | Select Citations Facilitating the Mapping Derivation |
|---------------------------|--------------------------------|----------------------|---------------------|---|
| | | Left Hemisphere | Right Hemisphere | |
| | | (x, y, z) | (x, y, z) | |
| Tongue 1 | | (-60.2, 2.1, 27.5) | (62.9, 2.5, 28.9) | Mapping hypothesized based on Motor Tongue Area of Fesl et al. (2003). |
| Tongue 2 | | (-60.2, 3, 23.3) | (66.7, 2.5, 24.9) | |
| Tongue 3 | | (-60.2, 4.4, 19.4) | (64.2, 3, 22) | |
| Upper Lip | | (-53.9, -3.6, 47.2) | (59.6, -7.2, 42.5) | Mapping hypothesized based on Penfield and Roberts (1959) and Lotze et al. (2000a, 2000b). |
| Lower Lip | | (-56.4, 0.5, 42.3) | (59.6, -3.6, 40.6) | |
| Jaw | | (-59.6, -1.3, 33.2) | (62.1, 3.9, 34) | Mapping hypothesized based on Penfield and Roberts (1959). |
| Larynx (Intrinsic) | | (-53, 0, 42) | (53, 4, 42) | Mapping hypothesized based on Brown et al. (2008) and supported by Penfield Rasmussen (1950) and Penfield and Roberts (1959). |
| Larynx (Extrinsic) | | (-58.1, 6, 6.4) | (65.4, 5.2, 10.4) | |
| Respiration | | (-17.4, -26.9, 73.4) | (23.8, -28.5, 70.1) | Mapping hypothesized based on Fink et al. (1996) and Olthoff et al. (2009). |

Table 2

The equations of the DIVA neural network components.

| Model Components | Cell Equations | Description of Variables |
|--------------------------|--|---|
| Speech Sound Map | $P_i(t) = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ sound is produced or perceived} \\ 0 & \text{otherwise} \end{cases}$ | Cells are hypothesized to be active during speech production, driving the movement trajectories of the articulators, and during speech perception, when auditory expectations are tuned. |
| Articulator Velocity Map | $\dot{M}_{Feedforward}(t) = P(t)z_{PM}(t) - M(t)$ | $Z_{PM}(t)$ encodes the feedforward motor commands for the speech sound currently being produced. $M(t)$ specifies the current articulator position. $\dot{M}_{Feedforward}$ encodes feedforward articulator velocity commands. |
| Auditory State Map | $Au(t) = f_{AcAu}(Acoust(t - \tau_{AcAu}))$ | $Acoust(t)$ is the acoustic signal resulting from the current articulator movement. τ_{AcAu} is the delay associated with transmitting the auditory signal from the periphery to the cerebral cortex. f_{AcAu} transforms the acoustic signal into a corresponding auditory cortical map representation, $Au(t)$. |
| Auditory Target Map | $T_{Au}(t) = P(t - \tau_{PAu})z_{PAu}(t)$ | τ_{PAu} is the transmission delay for the signals from the Speech Sound Map to the Auditory Target Map. $z_{PAu}(t)$ are synaptic weights that encode the auditory expectations, $T_{Au}(t)$, for the current sound being produced. |
| Auditory Error Map | $\Delta Au(t) = Au(t) - T_{Au}(t)$ | $\Delta Au(t)$ is the difference between the actual auditory responses and the auditory expectations for the current speech sound. |
| Somatosensory State Map | $S(t) = f_{ArS}(Artic(t - \tau_{ArS}))$ | $Artic(t)$ is the state of the articulators resulting from the current articulator movement. τ_{ArS} is the delay associated with transmitting the articulator state to the cerebral cortex. f_{ArS} transforms the articulatory state into the corresponding somatosensory cortical map representation, $S(t)$. |
| Somatosensory Target Map | $T_S(t) = P(t - \tau_{PS})z_{PS}(t)$ | τ_{PS} is the transmission delay for the signals from the Speech Sound Map to the Somatosensory Target Map. $z_{PS}(t)$ are the |

| Model Components | Cell Equations | Description of Variables |
|---------------------------------|--|--|
| | | synaptic weights that encode the somatosensory expectations, $T_S(t)$, for the current sound being produced. |
| Somatosensory Error Map | $\Delta S(t) = S(t) - T_S(t)$ | $\Delta S(t)$ is the difference between the actual somatosensory responses and the somatosensory expectations for the current speech sound. |
| Feedback Control Map | $\dot{M}_{Feedback}(t) = \Delta A u(t - \tau_{AuM}) z_{AuM} + \Delta S(t - \tau_{SM}) z_{SM}$ | Z_{AuM} and Z_{SM} are synaptic weights that transform directional sensory error signals into corrective motor velocities. τ_{AuM} and τ_{SM} are cortico-cortical transmission delays. $\dot{M}_{Feedback}$ encodes articulator feedback velocity commands. |
| Initiation Map | $I_i(t) = \begin{cases} 1 & \text{if the } i^{th} \text{ sound is produced} \\ 0 & \text{otherwise} \end{cases}$ | Commands from the <i>Articulator Velocity</i> and <i>Position Maps</i> are released to the articulatory synthesizer when the i^{th} cell in the <i>Initiation Map</i> , $I_i(t)$, is non-zero. |
| Articulator Position Map | $M(t) = M(0) + a_{ff} \int_0^t \dot{M}_{Feedforward}(t) I_i(t) dt + a_{fb} \int_0^t \dot{M}_{Feedback}(t) I_i(t) dt$ | $M(0)$ is the initial configuration of the vocal tract. a_{fb} and a_{ff} are weights relating the contributions of feedback and feedforward control, respectively. |