Published in final edited form as:
J Neurolinguistics. 2012 September 1; 25(5): 408–422. doi:10.1016/j.jneuroling.2009.08.006.

A Neural Theory of Speech Acquisition and Production

Frank H. Guenther[1,2,3,4,†] and **Tony Vladusich**[1,4]

[1]Department of Cognitive and Neural Systems, Boston University, 677 Beacon Street, Boston, MA, 02215

[2]Division of Health Sciences and Technology, Harvard University - Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[3]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA 02129, USA

[4]Center of Excellence for Learning in Education, Science and Technology, Boston University, 677 Beacon Street, Boston, MA, 02215

Abstract

This article describes a computational model, called DIVA, that provides a quantitative framework for understanding the roles of various brain regions involved in speech acquisition and production. An overview of the DIVA model is first provided, along with descriptions of the computations performed in the different brain regions represented in the model. Particular focus is given to the model's *speech sound map*, which provides a link between the sensory representation of a speech sound and the motor program for that sound. Neurons in this map share with "mirror neurons" described in monkey ventral premotor cortex the key property of being active during both production and perception of specific motor actions. As the DIVA model is defined both computationally and anatomically, it is ideal for generating precise predictions concerning speech-related brain activation patterns observed during functional imaging experiments. The DIVA model thus provides a well-defined framework for guiding the interpretation of experimental results related to the putative human speech mirror system.

Keywords

Speech production; motor control; neural model; fMRI; mirror system

1. INTRODUCTION

The production of speech sounds requires the integration of auditory, somatosensory, and motor information represented in the temporal, parietal, and frontal lobes of the cerebral cortex, respectively. Together with sub-cortical structures—such as the cerebellum, basal ganglia and the brain stem—these cortical regions and their functional connections form a functional unit which we term the *speech motor control system*. The speech motor control system is engaged during even the simplest speech tasks, such as babbling, imitating or

†Corresponding author: Telephone: (617) 353-5765, Fax Number: (617) 353-7755, guenther@cns.bu.edu

reading single syllables and words (e.g., Fiez & Petersen, 1998; Guenther, 1995; Turkeltaub et al., 2002). The present article describes a computational and neural framework, called DIVA, that provides a quantitative account of the interactions between cortical motor, somatosensory, and auditory regions during speech output (see Guenther, Ghosh, & Tourville, 2006, for computational details), thereby providing a "neural theory of speech production", in keeping with this special issue's theme of a neural theory of language.

Our focus here is on the computations performed by the cerebral cortex, particularly the premotor region of frontal cortex (see Barlow, 1999; Duffy, 1995; Kent, 1997; Zemlin, 1998, for detailed discussions of other speech-related brain regions). We discuss how the properties of a specific class of neuron in the DIVA model, known as *speech sound map neurons* (e.g., Guenther, 1992, 1994; Guenther et al., 2006), resemble "mirror neurons" described in the F5 region of monkey premotor cortex (Gallese et al., 1996; Rizzolatti et al., 1996). As discussed in subsequent detail, we hypothesize that the speech sound map resides in left ventral premotor cortex and posterior Broca's area (BA 6, 44) of the human brain, and it is a crucial component of the network responsible for speech acquisition, as well as mature speech production. Some researchers (Rizzolatti & Arbib, 1998) have posited that monkey area F5, or the region immediately in front of area F5 (Petrides et al., 2005), is homologous to Broca's area in humans, and that human speech evolved on the basis of the functional properties of mirror neurons. Although this hypothesis appears plausible, research on the putative role of mirror neurons in human speech motor control would benefit greatly from computational and neural frameworks with which to synthesize, interpret and predict empirical results (e.g., Bonaiuto et al., 2007). Here we summarize data from several brain imaging and behavioral studies which have provided support for key functional predictions of the DIVA model, and we suggest how the DIVA model can be used to further examine the functional properties of mirror neurons in the speech motor control system.

2. OVERVIEW OF THE DIVA MODEL

Figure 1 schematizes the cortical components of the DIVA model. Each box in the diagram corresponds to a set of neurons, or map, and arrows correspond to synaptic projections that transform one type of neural representation into another. The outputs of the model control an articulatory synthesizer that produces an acoustic signal (Maeda, 1990). The articulator movements and acoustic signal produced by the model have been quantitatively compared to the speech sounds and movements produced by human speakers, as detailed elsewhere (e.g., Callan et al., 2000;Guenther, 1995;Guenther et al., 1998;Guenther et al., 1999;Nieto-Castanon et al., 2005;Perkell et al., 2004a,b).

The production of a speech sound in the DIVA model starts with activation of neurons associated with that sound in the model's *speech sound map*. A "speech sound" can be a phoneme, syllable, or even short syllable sequence, with the syllable being the most typical unit represented by a single "neuron" in the speech sound map, with each model neuron assumed to correspond to a small population of neurons in the cortex. Activation of the speech sound map neurons leads to motor commands that arrive in primary motor cortex via two control subsystems. A *feedforward control system* projects directly from the speech sound map to articulatory control units in the cerebellum and primary motor cortex. A *feedback control system*—which is itself composed of an *auditory feedback control subsystem*, and a *somatosensory feedback control subsystem*—involves slower, indirect projections passing through sensory brain areas. The functions of these various subsystems are described below.

The DIVA model provides a description of how a human infant learns to produce speech sounds through babbling and imitation processes. According to the model, a combination of

motor, auditory, and somatosensory information generated during early random and reduplicated babbling is used to tune the synaptic projections between the sensory error maps and motor cortex via a *feedback control map* in right ventral premotor cortex. Later in imitation learning, the error maps register the discrepancy between the intended and the actual sensory states. The sensory-motor transformations learned during babbling allow detected sensory errors to be mapped into corrective motor commands during the imitation stage.

The imitation stage describes how syllable-specific learning occurs when an infant is presented with a new speech sound to learn, corresponding to an infant learning a sound from his/her native language. Detection of a novel sound leads to activation of previously unused speech sound map neurons for that sound. The model first learns an *auditory target* for the new sound, represented as a time-varying acoustic signal. This auditory target is encoded in the synaptic projections from the speech sound map to the auditory error map in Figure 1. The target encodes the allowable variability of the acoustic signal throughout the duration of the syllable. The use of target *regions*, rather than *points*, is an important aspect of the DIVA model that provides a unified explanation for a wide range of speech production phenomena (see Guenther, 1995, for details). The speech sound map neurons that represent the new sound will also be used to produce the sound, as described next. Neurons in the speech sound map are therefore active both when perceiving a sound and when producing the same sound. This prediction (Guenther, 1992,1994) is supported by data from a recent functional MRI experiment on speech production and perception (Wilson et al., 2004).

In the next step of the imitation learning process, the model attempts to produce the sound by activating the speech sound map neurons corresponding to the sound. This leads to readout of a feedforward command (represented by the arrow from the speech sound map to the articulator velocity and position maps in Figure 1). On the first attempt to produce a novel sound, no tuned feedforward command for the sound will exist. Thus, the model predicts that readout of the feedforward command will result in auditory errors, and the system must employ the auditory feedback control subsystem to help shape the ongoing attempt to produce the sound by transforming auditory errors into corrective motor commands via the *feedback control map* in right ventral premotor cortex. Auditory error cell activity represents error in formant frequency space; this error must be transformed into a corrective command in a motoric, or articulatory, representation. The transformation from auditory errors to corrective motor commands (i.e., the mapping from Directions in sensory space Into Velocities of Articulators from which the model gets its name) is learned during the model's babbling cycle, and mathematically this transformation corresponds to the pseudoinverse of the Jacobian matrix relating the auditory and motor spaces (see Guenther et al., 2006 for details). Previous computer simulations have shown how such a mapping can be learned by a biologically plausible neural network (e.g., Guenther et al., 1998).

On each attempt to produce the sound, command signals in the feedforward control subsystem are updated to incorporate the refined commands generated by the auditory feedback control subsystem on that attempt. This results in a more accurate feedforward command for the next attempt. Eventually the feedforward command by itself is sufficient to produce the sound in normal circumstances. That is, the feedforward command is accurate enough that it generates little or no auditory error during production of the sound and thus does not invoke the auditory feedback control subsystem. At this point, the new sound can be produced fluently. As production of the speech sound is repeated, a *somatosensory target region* for the sound is learned, analogous to the auditory target region mentioned above. This target represents the expected tactile and proprioceptive sensations associated with the

sound and is used in the somatosensory feedback control subsystem to detect somatosensory errors.

3. THE MIRROR NEURON SYSTEM IN MONKEYS AND HUMANS

The hypothesized speech sound map neurons share the key properties of "mirror neurons" identified in the F5 region of monkey frontal premotor cortex. Mirror neurons exhibit the remarkable property of spiking during both the active production and passive observation of certain motor actions (di Pellegrino et al., 1992; Gallese et al., 1996; Ferrari et al., 2003; Kohler et al., 2002; Rizzolatti et al., 1996). Mirror neurons encode complex actions, such as grasping, rather than the individual movements that comprise an action. A given mirror neuron may fire spikes, for example, when a monkey grasps a piece of fruit with the hand or when the monkey observes a human grasping fruit in a similar fashion. Mirror neurons related to communicative mouth movements (Ferrari et al., 2003) have been found in the region of monkey premotor cortex immediately lateral to the region for grasping movements (di Pellegrino et al., 1992). It has been proposed that this area corresponds to BA 44 of Broca's area in the human brain (Rizzolatti & Arbib, 1998).

Some functional MRI studies in humans support the notion that Broca's area plays a central role in the mirror representation of hand and finger gestures (Iacoboni et al., 1999; Tai et al., 2004; Iacoboni & Depretto, 2006), in addition to its classical association with speech motor control. Based on these and related data, the putative mirror representation in Broca's area has been implicated in imitation learning and the production and perception of human speech (Arbib & Rizzolatti, 1998; Iacoboni & Dapretto, 2006; Rizzolatti et al., 1996). Other fMRI studies (Lingnau et al., 2009) and theoretical analyses (Hickok, 2009; Lotto et al., 2009) have, however, questioned the functional role of the putative human mirror system in relation to the claim that "we understand action because the motor representation of that action is activated in our brain" (Rizzolatti et al., 2001, p. 661). As Mahon and Caramazza (2008, p. 62) ask: "Do mirror neurons become activated only after perceptual analysis and recognition of the sensory stimulus, or is the activation of mirror neurons directly and causally implicated in that perceptual analysis?" We claim that answers to functional questions concerning the roles of putative speech mirror representations are best considered within the broader context of how humans acquire and produce speech utterances. Below we emphasize some key functional properties of the speech sound map (Section 4) and summarize empirical evidence supporting various associated components of the DIVA model of speech acquisition and production (Section 5), before returning to a discussion of the functional issues in (Sections 6, 7 & 8).

4. THE SPEECH SOUND MAP: DIVA'S MIRROR

According to the DIVA model, higher-level prefrontal cortex regions involved in phonological encoding of an intended utterance sequentially activate speech sound map neurons that correspond to the syllables to be produced. Activation of these neurons leads to the readout of feedforward motor commands to the primary motor cortex. It is this feedforward control of speech sounds via neurons in the speech sound map that we liken to the activity of mirror neurons during action production. In addition to driving the complex articulator movements required to produce speech sounds, neurons in the speech sound map learn the expected pattern of acoustic stimulation associated with a specific syllable, represented as trajectories of the formant frequencies (or ratios of formant frequencies; see Guenther et al., 1998) defining a target speech sound. As described above, this learning is rapid and takes place along the pathways between the speech sound map and the auditory error map during the perception of sample speech sounds (Figure 1).

In order to commit a new neuron in the speech sound map to a particular auditory target region, the speech sound map must first be activated by the new speech sound along pathways from the auditory regions of superior temporal cortex (pathways not shown in Figure 1). These pathways are generally omitted from current implementations of the DIVA model for simplicity, although they are necessary for a more complete description of speech acquisition and production. The inputs along these pathways may come from either the auditory state map or some higher-order categorical representation of speech sounds in the temporal cortex (see Guenther & Gjaja, 1996;Guenther, et al., 2004, for discussions of categorical speech sound representation in auditory cortical maps). Indeed, it is this driving input from auditory cortex which we liken to activation of mirror neurons during action observation. Due to its capacity to mediate between auditory and motor speech representations, the speech sound map plays a pivotal role in imitation learning in the DIVA model. As discussed below, the functional connectivity between the speech sound map and the auditory and motor cortices predicted by the DIVA model opens up new avenues for future fMRI studies of the speech-mirroring hypothesis. The next section describes previous studies that have successfully tested key functional predictions of the DIVA model.

5. TESTING THE MODEL WITH BEHAVIORAL AND BRAIN IMAGING EXPERIMENTS

An important feature of the DIVA model that differentiates it from other computational models of speech production is that all of the model's components have been associated with specific anatomical locations in the brain. These locations, specified in the Montreal Neurological Institute (MNI) coordinate frame, are based on the results of neurophysiological and neuroanatomical studies of speech production and articulation (see Guenther et al., 2006, for details). Since the model's components correspond to groups of neurons at specific anatomical locations, it is possible to generate simulated fMRI activations to compare against actual data. The relationship between the signal measured in blood oxygen level dependent (BOLD) fMRI and electrical activity of neurons has been studied intensively in recent years (e.g., Heeger et al., 2000; Rees et al., 2000; Logothetis, et al., 2001). It is well-known that the BOLD signal is relatively sluggish compared to electrical neural activity. That is, for a very brief burst of neural activity, the BOLD signal will begin to rise and continue rising well after the neural activity stops, peaking about 4-6 seconds after the neural activation burst before falling down to the starting level. Such a hemodynamic response function (HRF), is used to transform activities from the model neurons into simulated fMRI activity.

In our modeling work, each model neuron is meant to correspond to a small population of neurons that fire together. The output of a neuron corresponds to the average number of action potentials per second of the population of neurons. This output is sent to other neurons in the network, where it is multiplied by synaptic weights to form synaptic inputs to these neurons. The activity level of a neuron is calculated as the sum of all the synaptic inputs to the neuron (both excitatory and inhibitory), and if the net activity is above zero, the neuron's output is equal to this activity level. If the net activity is below zero, the neuron's output is zero. It has been shown that the magnitude of the BOLD signal typically scales proportionally with the average firing rate of the neurons in the region where the BOLD signal is measured (e.g., Heeger et al., 2000; Rees et al., 2000). It has been noted elsewhere, however, that the BOLD signal actually correlates more closely with local field potentials, which are thought to arise primarily from averaged postsynaptic potentials (corresponding to the inputs of neurons), than it does to the average firing rate of an area (Logothetis et al., 2001). In accord with this finding, the fMRI activations that we generate from our models are determined by convolving the total inputs to our modeled neurons, rather than the firing-rate outputs, with an idealized hemodynamic response function (see Guenther et al., 2006,

for details). Figure 2 shows fMRI activations measured in a syllable production fMRI experiment (top) and simulated activations from the DIVA model when producing the same speech sounds (bottom). Also shown are the hypothesized locations of each of the model's neuron types (middle). Comparison of the experimental and simulated activation patterns indicates that the model qualitatively accounts for most of the activation found during syllable activation. Below we discuss experimental findings that support the key functional components of the model: the auditory and somatosensory feedback control subsystems and the feedforward control subsystems.

5.1 AUDITORY FEEDBACK CONTROL

It is well established that auditory feedback plays an important role in tuning the speech motor control system. According to the DIVA model, axonal projections from speech sound map neurons in the left ventral premotor cortex and posterior inferior frontal gyrus to higher-order auditory cortical areas embody the auditory target region for the speech sound currently being produced. That is, they represent the auditory feedback that should arise when the speaker hears himself/herself producing the current sound. This target is compared to incoming auditory information from the auditory periphery, and if the current auditory feedback is outside the target region, neurons in the auditory error map of the posterior superior temporal gyrus and planum temporale become active. These error signals are then transformed into corrective motor commands through projections from the auditory error map to motor cortex. The auditory target projections from the speech sound map to the auditory cortical areas inhibit auditory error map neurons. If the incoming auditory signal is within the target region, this inhibition cancels the excitatory effects of the incoming auditory signal. If the incoming auditory signal is outside the target region, the inhibitory target region will not completely cancel the excitatory input from the auditory periphery, resulting in activation of auditory error neurons. The use of an inhibitory target region linked to projections from the speech sound map to auditory error neurons constitutes a unique functional prediction of the DIVA model in relation to mirror neurons (see Section 5).

Once the model has learned appropriate feedforward commands for a speech sound as described in the preceding section, it can correctly produce the sound using just those feedforward commands. That is, no auditory error will arise during production, and thus the auditory feedback control subsystem will not be activated. However, if an externally imposed perturbation occurs, such as a real-time "warping" of the subject's auditory feedback so that he hears himself/herself producing the wrong sound (c.f. Houde & Jordan, 1998), the auditory error neurons will become active and attempt to correct for the perturbation. Due to neural transmission delays and the delay between muscle activation and the resulting movement, these corrective commands will be delayed by approximately 75-150 ms relative to the onset of an unexpected perturbation.

These hypotheses were tested in an fMRI study involving real-time perturbation of the first formant frequency (F1) of the speaker's acoustic signal (Tourville et al., 2008). In this study, subjects produced one-syllable words (e.g., "bet", "head") in the scanner. On 1 in 4 trials (randomly dispersed), the subject's auditory feedback was perturbed by shifting F1 of his/her own speech upward or downward by 30% in real time (18 ms delay). This frequency shift is not typically noticeable to the subject. Subjects were scanned on a 3-Tesla Siemens Trio scanner using a sparse sampling, event-triggered fMRI protocol. Each trial was 12 seconds long. At the beginning of a trial, a word was projected on a video screen for two seconds, and the subject produced the word during this period. Two seconds after the word disappeared, two whole-brain scans were collected. These scans were timed to occur during the peak of the hemodynamic response due to speaking the word (noting that the hemodynamic response to a brief burst of neural activity takes approximately 4-6 seconds to peak). This protocol, schematized in Figure 3, allows the subject to speak in silence (other

than the sound of his/her own speech) and avoids artifacts that can arise if scanning occurs during movement of the speech articulators (e.g., Munhall, 2001).

According to the DIVA model, auditory error neurons should be active in the perturbed trials but not the unperturbed trials; thus one should see activation of the auditory error map in the *perturbed speech - unperturbed speech* contrast. Figure 4 shows the areas with significant activation (fixed effects analysis, statistics controlled for a false discovery rate of 0.05) in this contrast. As predicted by the model, auditory error activation is evident in the posterior superior temporal gyrus and planum temporale. The activation peak was located in the posterior end of the left planum temporale (crosshairs in Figure 4); this area has been implicated as an auditory-motor interface for speech (Buchsbaum et al., 2001;Hickok et al., 2003). Furthermore, the model predicts that auditory error map activity will lead to corrective motor commands in the motor cortical areas. The previous version of the DIVA model (Guenther et al., 2006) predicted that this motor cortical activity should be bilateral and focused in primary motor cortex. The experimental results, however, indicated right-lateralized premotor activity rather than bilateral primary motor activity. For this reason, the model as presented in Figure 1 contains a new component, termed the feedback control map, located in right ventral premotor cortex. Functionally, this map is hypothesized to contain neurons coding corrective motor commands for detected sensory errors. A second experiment involving a somatosensory perturbation (Section 4.2) provided further support for the existence of this map.

The DIVA model also produces sound output that can be quantitatively compared to the vocalizations of human subjects in the perturbation experiment. The speech of subjects in the fMRI study was, therefore, recorded and analyzed to identify whether they were compensating for the perturbation during the perturbed trials (as predicted by the model), and to estimate the delay of such compensation. The gray shaded areas in Figure 5 represent the 95% confidence interval for normalized F1 values during the vowel for upward perturbation trials (dark shading) and downward perturbation trials (light shading). Subjects showed clear compensation for the perturbations, starting approximately 100-130 ms after the start of the vowel. Simulation results from the DIVA model are indicated by the dashed line (upward perturbation) and solid line (downward perturbation). The model's productions fall within the 95% confidence interval of the subjects' productions, indicating that the model can quantitatively account for compensation seen in the fMRI subjects.

The results of this study supports several key aspects of the DIVA model's account of auditory feedback control in speech production: (a) the brain contains auditory error neurons that signal the difference between a speaker's auditory target and the incoming auditory signal; (b) these error neurons are located in the posterior superior temporal gyrus and supratemporal plane, particularly in the planum temporale of the left hemisphere; and (c) unexpected perturbation of a speaker's auditory feedback results in a compensatory articulatory response within approximately 75-150 ms of the perturbation onset. In addition, they suggested modification of the model to include a right-lateralized ventral premotor feedback control map.

5.2 SOMATOSENSORY FEEDBACK CONTROL

Like auditory information, somatosensory information has long been known to be important for speech production. The DIVA model posits a somatosensory feedback control subsystem operating alongside the auditory feedback control subsystem described above. The model's *somatosensory state map* corresponds to the representation of tactile and proprioceptive information from the speech articulators in primary and higher-order somatosensory cortical areas in the postcentral gyrus and supramarginal gyrus. The model's *somatosensory error map* is hypothesized to reside in the supramarginal gyrus, a region that has been implicated

in phonological processing for speech perception (e.g., Caplan et al., 1995; Celsis et al., 1999) and production (Damasio & Damasio, 1980; Geschwind, 1965). According to the model, neurons in this map become active during speech if the speaker's tactile and proprioceptive feedback from the vocal tract deviates from the somatosensory target region for the sound being produced. The output of the somatosensory error map then propagates to motor cortex through synapses that are tuned during babbling to encode the transformation from somatosensory errors into motor commands that correct those errors. Analogous to the transformation of auditory errors into motor commands described above, this transformation corresponds mathematically to the pseudoinverse of the Jacobian matrix relating the somatosensory and motor spaces (see Guenther et al., 2006 for details).

To test the model's prediction of a somatosensory error map in the supramarginal gyrus, we performed an fMRI study that involved unexpected blocking of the jaw during speech production (Tourville et al., 2008). This intervention should activate somatosensory error neurons since it creates a mismatch between the desired and actual somatosensory state. Subjects read two-syllable pseudo-words shown on a screen (e.g., "abi", "agi"). In 1 of 7 productions (randomly dispersed), a small, stiff balloon lying between the molars was rapidly inflated (within 100 ms) to a diameter of 1-1.5cm during the first vowel of the utterance. The balloon has the effect of blocking upward jaw movement for the start of the second syllable. A pilot articulometry study confirmed that subjects compensate for the balloon inflation by producing more tongue raising to overcome the effects of the immobilized jaw. The remainder of the experimental paradigm was similar to that described above for the auditory perturbation experiment.

Compared to unperturbed speech, perturbed speech caused significantly more activation in a wide area of the cerebral cortex, including portions of the frontal, temporal, and parietal lobes. The strongest activations were found in the supramarginal gyrus bilaterally (left half of Figure 6); this is consistent with the location of the hypothesized somatosensory error map in the DIVA model (see model simulation result in right half of Figure 6). Another activation peak was found in right hemisphere ventral motor/premotor cortex. This right-lateralized frontal activity was not predicted by the previous version of the DIVA model, and it provides further support for the right-lateralized feedback control map that has been added to the latest version of the model (Figure 1).

5.3 FEEDFORWARD CONTROL

According to the DIVA model, projections from the speech sound map in left ventral premotor areas to primary motor cortex, supplemented by cerebellar projections, constitute feedforward motor commands for syllable production (dark shaded portion of Figure 1). These projections might be interpreted as constituting a *gestural score* (see Browman & Goldstein, 1989) or *mental syllabary* (see Levelt, & Wheeldon, 1994). The primary motor and premotor cortices are well-known to be strongly interconnected (e.g., Krakauer & Ghez, 1999;Passingham, 1993). Furthermore, the cerebellum is known to receive input via the pontine nuclei from premotor cortical areas, as well as higher-order auditory and somatosensory areas that can provide state information important for choosing motor commands (e.g., Schmahmann & Pandya, 1997), and projects heavily to the primary motor cortex (e.g., Middleton & Strick, 1997). Damage to the superior paravermal region of the cerebellar cortex results in ataxic dysarthria, a motor speech disorder characterized by slurred, poorly coordinated speech (Ackermann et al., 1992). This finding is in accord with the view that this region is involved in providing the precisely-timed feedforward commands necessary for fluent speech.

Early in development, infants do not possess accurate feedforward commands for all speech sounds. Only after they practice producing the sounds of their language can feedforward

commands be tuned. In the DIVA model, feedforward commands for a syllable are tuned on each production attempt. The model predicts that, on the first attempt to produce a new sound, infants will rely very heavily on auditory feedback control to produce the sound. The corrective commands issued by the auditory feedback control subsystem during the current attempt to produce the sound become stored in the feedforward command pathway for use on the next attempt. We hypothesize that the superior paravermal region of the cerebellum is involved in this process (see Ghosh, 2004, for details). Each subsequent attempt to produce the sound results in a better feedforward command and less auditory error. This cycle continues until the feedforward command is capable of producing the sound without producing any auditory error, at which point the auditory feedback subsystem no longer contributes to speech motor output unless speech is perturbed in some way or the sizes and shapes of the articulators change. As the speech articulators grow, the auditory feedback control subsystem continues to provide corrective commands that are subsumed into the feedforward controller, thus allowing the feedforward controller to stay properly tuned despite changes in the sizes and shapes of the speech articulators over the course of a lifetime. Computer simulations of the DIVA model's adaptation to changes in vocal tract shape during infancy and childhood are provided in Callan et al. (2000).

The model's account of feedforward control leads to the following predictions. If a speaker's auditory feedback of his/her own speech is perturbed for an extended period (e.g., over many consecutive productions of a syllable), corrective commands issued by the auditory feedback control subsystem will eventually become incorporated into the feedforward commands. If the perturbation is then removed, the speaker will show "after-effects". The speaker's first few productions after normal feedback is restored will therefore show signs of the adaptation of the feedforward command that occurred when the feedback was perturbed. Effects of this type have been reported in speech sensorimotor adaptation experiments (e.g., Houde & Jordan, 1998).

We investigated these effects more closely in a sensorimotor adaptation experiment involving sustained perturbation of the first formant frequency during speech. In this study (Villacorta et al., 2007), subjects performed a speech production experiment that involved four phases: (a) a *baseline phase* in which the subject produced 15 repetitions of a short list of words with normal auditory feedback (each repetition of the list corresponding to one epoch), (b) a *ramp phase* during which a shift in F1 was gradually introduced to the subject's auditory feedback (epochs 16-20), (c) a *training phase* in which the full F1 perturbation (a 30% shift of F1) was applied on every trial (epochs 21-45), and (d) a *post-test phase* in which the subject received unaltered auditory feedback (epochs 46-65). The subjects' *adaptive response*—defined as the percent change in F1 compared to the baseline phase in the direction opposite the perturbation—is shown by the solid line with standard error bars in Figure 7. The shaded band in Figure 7 represents the 95% confidence interval for simulations of the DIVA model performing the same experiment (see Villacorta et al., 2007, for details). With the exception of only one epoch in the ramp phase (denoted by a filled circle in Figure 7), the model's productions did not differ significantly from the experimental results. Most critically, the model correctly predicts the time course of the after-effect that is observed in the fourth experimental phase.

6. THE FUNCTIONAL ROLES OF SPEECH MIRRORING IN DIVA

In the DIVA model, the speech sound map plays several key functional roles. First, it facilitates the acquisition of discrete speech sound units from the auditory state map, or a related categorical representation of speech sounds, in superior temporal cortex (Guenther & Gjaja, 1996; Guenther, et al., 2004). Second, feedback projections from the speech sound map to the auditory and somatosensory error maps in sensory cortices define the expected

sensory feedback signals associated with the production of specific target sounds. Third, feedforward motor command signals from the speech sound map to the articulator velocity and position maps in motor cortex and cerebellum directly elicit the motor programs for speech sounds. The DIVA model therefore clarifies how sensory and motor information can converge on single units exhibiting the mirroring property, without positing that mirror representations are responsible for "action understanding" (Arbib & Rizzolatti, 1998; Gallese et al., 1996; Rizzolatti et al., 1996; Rizzolatti & Fabbri-Destro, 2008). As described by Lotto et al. (2009), speech sound map units in DIVA may perform the additional role of priming units in the speech sound recognition system via top-down modulation of superior temporal cortex. Such top-down signals would help to explain how context-sensitive "co-articulation" effects known to occur in speech production—effects that have previously been simulated within the DIVA framework (Guenther, 1995)—are transferred to speech perception, without the need to postulate a unitary speech perception-and-production stage. The DIVA model therefore helps to reconcile the sensory and motor theories of speech perception, while providing an answer to the question of what role mirror-like representations play in speech production and perception (Mahon & Caramazza, 2008).

7. TESTING THE SPEECH MIRRORING HYPOTHESIS

The DIVA model makes some unique functional predictions that set it apart from other theories of the possible role(s) of mirror representations in speech production and perception (Arbib & Rizzolatti, 1998; see also Wilson et al., 2004; Hickok, 2009; Lotto et al., 2009) and motor control (Bonaiuto et al., 2007; Gallese et al., 1996; Rizzolatti et al., 1996; Rizzolatti & Fabbri-Destro, 2008). According to DIVA, when an infant, or an adult learning an unfamiliar-sounding language, listens to a speaker producing a new speech sound, previously unused speech sound map neurons become active. Projections from these neurons to the auditory cortex rapidly become tuned to the acoustic characteristics of that sound. These projections thus represent a target auditory trace for that sound. Additional projections from the speech sound map neurons to the primary motor cortex (both directly and via a cerebellar loop) represent (initially poorly tuned) feedforward commands for producing the newly learned sound. These feedforward command pathways become tuned over repeated attempts to produce the sound, with each attempt initiated by activating these same speech sound map neurons. This learning is driven by the initial mismatch between the newly acquired sound target and the infant's own production attempt as represented in the auditory state map. These auditory error signals are then transformed into a corrective motor command, and this corrective command is added to the feedforward command for the next attempt. As the feedforward commands improve, fewer error signals are generated and thus the contribution of the feedback control system gradually diminishes. The DIVA model thus predicts that mirror neurons emerge *as a consequence of imitation learning*, rather than driving the imitation-learning process themselves (e.g., see Iacoboni et al., 1999).

The DIVA model also sheds light on the issue of how perceptual and motor reference frames become "aligned" via the development of mirror neurons (Gallese et al., 1996; Iacoboni & Dapretto, 2006; Rizzolatti et al., 1996). As learning progresses in DIVA, speech sound map neurons gradually "acquire" the feedforward motor command programs corresponding to the rapidly-acquired auditory target sounds. The link between perception and action thus arises in the DIVA model because the motor reference frame is brought into register with the auditory reference frame (see Guenther et al., 1998, for a discussion of reference frames in speech production). The model therefore predicts a causal relationship between the speech sounds acquired in auditory coordinates and their associated motor programs: Individuals with more distinctive auditory speech representations—those people better able to discriminate between similar speech sounds—should produce more distinctive speech utterances than those with poorer auditory discrimination. Data from several studies of

speech production support this prediction (Perkell et al., 2004a,b; Villacorta et al., 2007). The DIVA model also makes the currently untested prediction that individuals with more distinctive auditory and motor speech representations will exhibit statistically more *separable* patterns of fMRI activation in both Broca's area and higher-order auditory cortex during speech perception and production experiments (Kriegeskorte et al., 2006).

Another prediction of the DIVA model is that projections from the speech sound map to the auditory and somatosensory errors maps have the effect of inhibiting expected auditory inputs from one's own speech. Evidence of inhibition in auditory areas of the superior temporal gyrus during one's own speech comes from several different sources, including recorded neural responses during open brain surgery (Creutzfeldt et al., 1989a,b), magnetoencephalography (MEG) studies (Houde et al., 2002; Numminen & Curio, 1999; Numminen et al., 1999), and positron emission tomography (PET) studies (Wise et al., 1999). Data regarding the prediction of an inhibitory effect of the speech sound map on the supramarginal gyrus during speech production is currently lacking, although this brain region has been implicated in phonological processing for speech perception (e.g., Caplan et al., 1995; Celsis et al., 1999) and speech production (Geschwind, 1965; Damasio & Damasio, 1980).

The DIVA model also predicts that the tuning of projections from the speech sound map to the auditory and somatosensory error maps will exhibit different time courses. Due to the necessity to rapidly acquire new auditory targets during imitation learning, the projections from the speech sound map to the auditory error map need to learn quickly and remain stable over long time periods. This argument does not, however, apply to projections to the somatosensory error map, which need to learn targets slowly over multiple production attempts. Since somatosensory target information cannot be gained entirely from listening or viewing a speaker producing a new sound, somatosensory targets must instead be learned by monitoring one's own correct self-productions after the speaker has learned adequate feedforward commands for producing the sound.

The final set of predictions we will discuss here concerns the modulation of efferent and afferent pathways centered on the speech sound map. In particular, the DIVA model predicts that pathways from the speech sound map to motor cortex are modulated by a GO signal, computed by the SMA and basal ganglia, which controls speaking rate (e.g., Guenther, 1995). This signal is represented by the arrow from the Initiation Map to the Articulator Velocity and Position Maps in Figure 1. Speaking rate varies proportionally with the magnitude of the GO signal (which itself varies between the normalized values of zero and one): a larger GO signal is associated with a higher speaking rate. When the GO signal is zero, outputs from the speech sound map to primary motor cortex are gated off. This gating mechanism plays a crucial role in preventing the obligatory imitation of perceived speech sounds. As far as we are aware, investigations of the functional properties of mirror neurons have not yet addressed such a gating function.

8. CONCLUDING REMARKS

This article has described a quantitative neural theory of speech acquisition and production that provides a unified account for a wide range of speech acoustic, kinematic, and neuroimaging data. The model posits three interacting subsystems for the neural control of speech production: an auditory feedback control subsystem, a somatosensory feedback control subsystem, and a feedforward control subsystem. The feedforward control subsystem is proposed to involve cortico-cortical projections from premotor to motor cortex, as well as contributions from the cerebellum. The auditory feedback control subsystem involves projections from premotor cortex to higher-order auditory cortex that encode

auditory targets for speech sounds, as well as projections from higher-order auditory cortex to motor cortex that transform auditory errors into corrective motor commands. The somatosensory feedback control subsystem involves projections from premotor cortex to higher-order somatosensory cortex that encode somatosensory targets for speech sounds, as well as projections from somatosensory error neurons to motor cortex that encode corrective motor commands. The speech sound map coordinates the activities of these various maps during normal speech acquisition and production, providing a conduit between the perceptual and motor aspects of speech control. We expect that the quantitative nature of the DIVA formulation—as it applies to fMRI studies of the human mirror system—will help facilitate rapid advances in understanding the speech sound map in Broca's area and its functional connectivity with related brain regions.

Acknowledgments

F.H.G. supported in part by the National Institute on Deafness and other Communication Disorders (R01 DC02852, F. Guenther PI). T.V. supported in part by the National Science Foundation (NSF SBE-0354378).

REFERENCES

Ackermann H, Vogel M, Petersen D, Poremba M. Speech deficits in ischaemic cerebellar lesions. Journal of Neurology. 1992; 239:223–227. [PubMed: 1597689]

Barlow, SM. Handbook of clinical speech physiology. Singular; San Diego: 1999.

Bonaiuto J, Rosta E, Arbib M. Extending the mirror neuron system model, I. Audible actions and invisible grasps. Biological Cybernetics. 2007; 96:9–38. [PubMed: 17028884]

Browman CP, Goldstein L. Articulatory gestures as phonological units. Phonology. 1989; 6:201–251.

Buchsbaum BR, Hickok G, Humphries C. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. Cognitive Science. 2001; 25:663–678.

Buckner RL, Raichle ME, Miezin FM, Petersen SE. Functional anatomic studies of memory retrieval for auditory words and visual pictures. Journal of Neuroscience. 1996; 16:6219–6235. [PubMed: 8815903]

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. Journal of Speech, Language, and Hearing Research. 2000; 43:721–736.

Caplan D, Gow D, Makris N. Analysis of Lesions by Mri in Stroke Patients with Acoustic-Phonetic Processing Deficits. Neurology. 1995; 45:293–298. [PubMed: 7854528]

Celsis P, Boulanouar K, Doyon B, Ranjeva JP, Berry I, Nespoulous JL, Chollet F. Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. NeuroImage. 1999; 9:135–144. [PubMed: 9918735]

Creutzfeldt O, Ojemann G, Lettich E. Neuronal-Activity in the Human Lateral Temporal-Lobe .1. Responses to Speech. Experimental Brain Research. 1989a; 77:451–475.

Creutzfeldt O, Ojemann G, Lettich E. Neuronal-Activity in the Human Lateral Temporal-Lobe .2. Responses to the Subjects Own Voice. Experimental Brain Research. 1989b; 77:476–489.

Damasio H, Damasio AR. The anatomical basis of conduction aphasia. Brain. 1980; 103:337–350. [PubMed: 7397481]

Dapretto M, Davies MS, Pfeifer JH, Scott AA, Sigman M, Bookheimer SY, et al. Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorders. Nature Neuroscience. 2006; 9:28–30.

DeLong, MR. The basal ganglia. In: Kandel, ER.; Schwartz, JH.; Jessell, TM., editors. Principles of Neural Science. 4th ed.. McGraw Hill; New York: 1999. p. 853-867.

DeLong MR, Wichman T. Basal ganglia-thalamocortical circuits in Parkinsonian signs. Clinical Neuroscience. 1993; 1:18–26.

Dronkers NF. A new brain region for coordinating speech articulation. Nature. 1996; 384:159–161. [PubMed: 8906789]

Duffy, JR. Motor speech disorders: Substrates, differential diagnosis, and management. Mosby; St. Louis: 1995.

Ferrari PF, Gallese V, Rizzolatti G, Fogassi L. Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. European Journal of Neuroscience. 2003; 17:1703–1714. [PubMed: 12752388]

Fiez JA, Petersen SE. Neuroimaging studies of word reading. Proceedings of the National Academy Sciences USA. 1998; 95:914–921.

Gallese V, Fadiga L, Fogassi L, Rizzolatti G. Action recognition in the premotor cortex. Brain. 1996; 119:593–609. [PubMed: 8800951]

Georgiou N, Bradshaw JL, Iansek R, Phillips JG, Mattingley JB, Bradshaw JA. Reduction in external cues and movement sequencing in Parkinson's disease. Journal of Neurology, Neurosurgery and Psychiatry. 1994; 57:368–370.

Geschwind N. Disconnexion syndromes in animals and man, II. Brain. 1965; 88:585–644. [PubMed: 5318824]

Guenther, FH. Boston University Ph.D. Dissertation. 1992. Neural Models of Adaptive Sensory-motor Control for Flexible Reaching and Speaking.

Ghosh, SS. Boston University Ph.D. Dissertation. Boston University; Boston, MA: 2004. Understanding cortical and cerebellar contributions to speech production through modeling and functional imaging.

Guenther FH. A neural network model of speech acquisition and motor equivalent speech production. Biological Cybernetics. 1994; 72:43–53. [PubMed: 7880914]

Guenther FH. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. Psychological Review. 1995; 102:594–621. [PubMed: 7624456]

Guenther FH, Gjaja MN. The perceptual magnet effect as an emergent property of neural map formation. Journal of the Acoustical Society of America. 1996; 100:1111–1121. [PubMed: 8759964]

Guenther FH, Hampson M, Johnson D. A theoretical investigation of reference frames for the planning of speech movements. Psychological Review. 1998; 105:611–633. [PubMed: 9830375]

Guenther FH, Espy-Wilson CY, Boyce SE, Matthies ML, Zandipour M, Perkell JS. Articulatory tradeoffs reduce acoustic variability during American English /r/ production. Journal of the Acoustical Society of America. 1999; 105:2854–2865. [PubMed: 10335635]

Guenther FH, Nieto-Castanon A, Ghosh SS, Tourville JA. Representation of sound categories in auditory cortical maps. J Speech Lang Hear Res. 2004; 47:46–57. [PubMed: 15072527]

Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. Brain and Language. 2006; 96:280–301. [PubMed: 16040108]

Heeger DJ, Huk AC, Geisler WS, Albrecht DG. Spikes versus BOLD: What does neuroimaging tell us about neuronal activity? Nature Neuroscience. 2000; 3:631–633.

Hickok G. Eight problems for the mirror neuron theory of action understanding in monkeys and humans. Journal of Cognitive Neuroscience. 2009; 21:1229–1243. [PubMed: 19199415]

Hickok G, Buchsbaum B, Humphries C, Muftuler T. Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. Journal of Cognitive Neuroscience. 2003; 15:673–682. [PubMed: 12965041]

Houde JF, Jordan MI. Sensorimotor adaptation in speech production. Science. 1998; 279:1213–1216. [PubMed: 9469813]

Houde JF, Nagarajan SS, Sekihara K, Merzenich MM. Modulation of the auditory cortex during speech: an MEG study. Journal of Cognitive Neuroscience. 2002; 14:1125–1138. [PubMed: 12495520]

Iacoboni M, Woods RP, Brass M, Bekkering H, Mazziotta JC, Rizzolatti G. Cortical mechanisms of human imitation. Science. 1999; 286:2526–2528. [PubMed: 10617472]

Iacoboni M, Dapretto M. The mirror neuron system and the consequences of its dysfunction. Nature Reviews Neuroscience. 2006; 7:942–951.

Kent, RD. The speech sciences. Singular; San Diego: 1997.

Krakauer, J.; Ghez, C. Voluntary movement. In: Kandel, ER.; Schwartz, JH.; Jessell, TM., editors. Principles of Neural Science. 4th ed.. McGraw Hill; New York: 1999. p. 756-781.

Kriegeskorte N, Goebel R, Bandettini P. Information-based functional brain mapping. Proceedings of the National Academy of Sciences USA. 2006; 103:3863–3868.

Levelt WJ, Wheeldon L. Do speakers have access to a mental syllabary? Cognition. 1994; 50:239–269. [PubMed: 8039363]

Lingnau A, Gesierich B, Caramazza A. Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans. Proceedings of the National Academy of Sciences USA. 2009; 106:9925–9930.

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]

Lotto AJ, Hickok GS, Holt LL. Reflections on mirror neurons and speech perception. Trends in Cognitive Sciences. 2009; 13:110–114. [PubMed: 19223222]

Maeda, S. (1990). Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. In W.J.Hardcastle & A.

Mahon BZ, Caramazza A. A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. Journal of Physioliology (Paris). 2008; 102:59–70.

Marchal, editor. Speech production and speech modeling. Kluwer Academic Publishers; Boston: p. 131-149.

Middleton FA, Strick PL. Cerebellar output channels. International Review of Neurobiology. 1997; 41:61–82. [PubMed: 9378611]

Munhall KG. Functional imaging during speech production. Acta Psychologica. 2001; 107:95–117. [PubMed: 11388145]

Nathaniel-James DA, Fletcher P, Frith CD. The functional anatomy of verbal initiation and suppression using the Hayling Test. Neuropsychologia. 1997; 35:559–566. [PubMed: 9106283]

Nieto-Castanon A, Guenther FH, Perkell JS, Curtin H. A modeling investigation of articulatory variability and acoustic stability during American English /r/ production. Journal of the Acoustical Society of America. 2005; 117:3196–3212. [PubMed: 15957787]

Numminen J, Curio G. Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. Neuroscience Letters. 1999; 272:29–32. [PubMed: 10507535]

Numminen J, Salmelin R, Hari R. Subject's own speech reduces reactivity of the human auditory cortex. Neuroscience Letters. 1999; 265:119–122. [PubMed: 10327183]

Passingham, RE. The frontal lobes and voluntary action. Oxford University Press; Oxford: 1993.

Paus T, Petrides M, Evans AC, Meyer E. Role of the human anterior cingulate cortex in the control of oculomotor, manual, and speech responses: a positron emission tomography study. Journal of Neurophysiology. 1993; 70:453–469. [PubMed: 8410148]

Perkell JS, Guenther FH, Lane H, Matthies ML, Stockmann E, Tiede M, Zandipour M. Cross-subject correlations between measures of vowel production and perception. Journal of the Acoustical Society of America. 2004a; 116:2338–2344. [PubMed: 15532664]

Perkell JS, Matthies ML, Tiede M, Lane H, Zandipour M, Marrone N, Stockmann E, Guenther FH. The distinctness of speakers' /s-sh/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. Journal of Speech, Language, and Hearing Research. 2004b; 47:1259–1269.

Petrides M, Cadoret G, Mackey S. Orofacial somatomotor responses in the macaque monkey homologue of broca's area. Nature. 2005; 435:1235–1238. [PubMed: 15988526]

Rees G, Friston K, Koch C. A direct quantitative relationship between the functional properties of human and macaque V5. Nature Neuroscience. 2000; 3(7):716–723.

Rizzolatti G, Arbib MA. Language within our grasp. Trends in Neurosciences. 1998; 21:188–194. [PubMed: 9610880]

Rizzolatti G, Fadiga L, Gallese V, Fogassi L. Premotor cortex and the recognition of motor actions. Cognitive Brain Research. 1996; 3:131–141. [PubMed: 8713554]

Rizzolatti G, Fogassi L, Gallese V. Neurophysiological mechanisms underlying the understanding and imitation of action. Nat Rev Neurosci. 2001; 2:661–670. [PubMed: 11533734]

Rizzolatti G, Fabbri-Destro M. The mirror system and its role in social cognition. Current Opinion in Neurobiology. 2008; 18:179–184. [PubMed: 18706501]

Rogers MA, Phillips JG, Bradshaw JL, Iansek R, Jones D. Provision of external cues and movement sequencing in Parkinson's disease. Motor Control. 1998; 2:125–132. [PubMed: 9644283]

Schmahmann JD, Pandya DN. The cerebrocerebellar system. International Review of Neurobiology. 1997; 41:31–60. [PubMed: 9378595]

Tourville JA, Reilly KJ, Guenther FH. Neural mechanisms underlying auditory feedback control of speech. NeuroImage. 2008; 39:1429–1443. [PubMed: 18035557]

Turkeltaub PE, Eden GF, Jones KM, Zeffiro TA. Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. NeuroImage. 2002; 16:765–780. [PubMed: 12169260]

Villacorta V, Perkell JS, Guenther FH. Sensorimotor adaptation to acoustic perturbations in vowel formants. Journal of the Acoustical Society of America. 2004; 115:2430. Program of the 147th Meeting of the Acoustical Society of America.

Villacorta, V. Massachusetts Institute of Technology PhD Dissertation. Cambridge, MA: 2005. Sensorimotor adaptation to perturbations of vowel acoustics and its relation to perception.

Villacorta VM, Perkell JS, Guenther FH. Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. Journal of the Acoustical Society of America. 2007; 122:2306–2319. [PubMed: 17902866]

Wilson SM, Saygin AP, Sereno MI, Iacoboni M. Listening to speech activates motor areas involved in speech production. Nature Neuroscience. 2004; 7:701–702.

Wise RJ, Greene J, Buchel C, Scott SK. Brain regions involved in articulation. Lancet. 1999; 353:1057–1061. [PubMed: 10199354]

Zemlin, WR. Speech and hearing science: Anatomy and physiology. 4th edition. Allyn and Bacon; Boston: 1998.

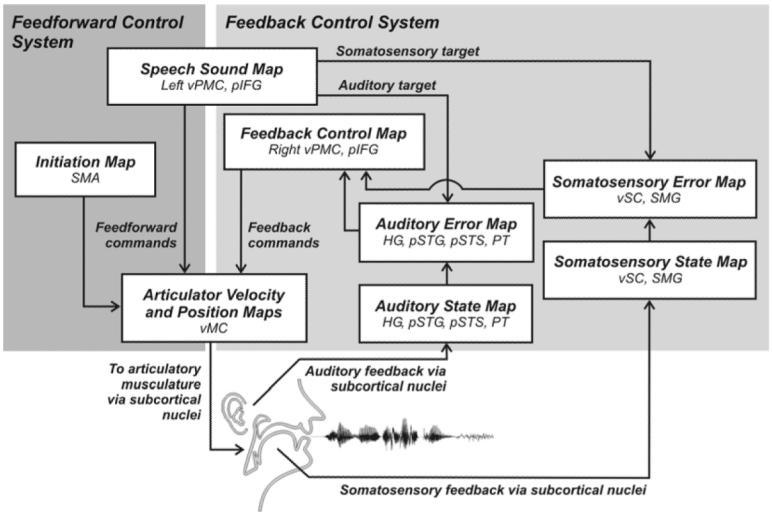

Figure 1.

Schematic of the cortical components of the DIVA model of speech acquisition and production. The mediating neural representation linking auditory and motor reference frames is the *speech sound map*, proposed to reside in the left posterior inferior frontal gyrus (Broca's area) and adjoining ventral premotor cortex. Additional details of the model are described in the text.

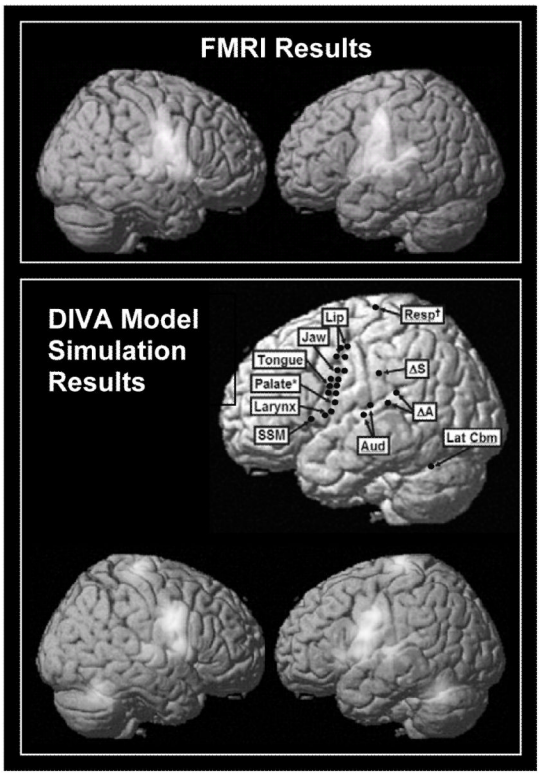

Figure 2.

Top. Lateral surfaces of the brain indicating locations of significant activations (random effects; statistics controlled at a false discovery rate of 0.05) measured in an fMRI experiment of single syllable production (*speech - baseline* contrast, where the baseline task consisted of silently viewing the letters YYY on the video screen). **Middle right.** Lateral surface of the brain indicating locations of the DIVA model components in the left hemisphere. Medial regions (superior paravermal cerebellum and deep cerebellar nuclei) are not visible. Unless otherwise noted, labels along the central sulcus correspond to the motor (anterior) and somatosensory (posterior) representation for each articulator. **Bottom.** Simulated fMRI activations from the DIVA model when performing the same speech task as the subjects in the fMRI experiment. [Abbreviations: Aud = auditory state neurons; ΔA = auditory error neurons; ΔS = somatosensory error neurons; Lat Cbm = superior lateral cerebellum; Resp = motor respiratory region; SSM = speech sound map. *Palate representation is somatosensory only. †Respiratory representation is motor only.]

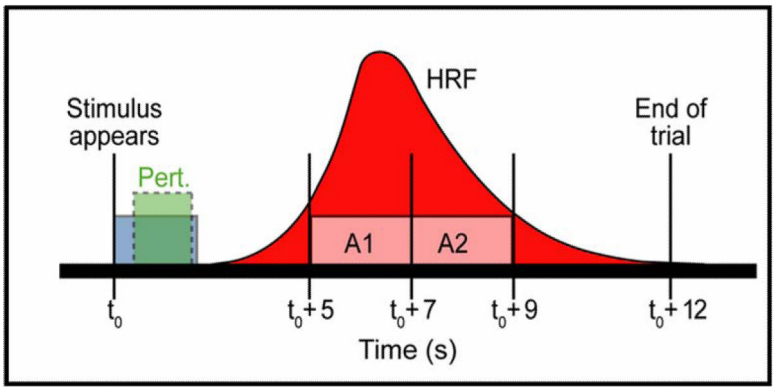

Figure 3.
Timeline for a single trial in the fMRI speech perturbation protocol. The subject reads the stimulus out loud during stimulus presentation, when the scanner is not collecting images and is thus quiet. Images are acquired approximately 2 seconds after articulation ceases. [Abbreviations: HR = estimated hemodynamic response; A1,A2 = acquisition periods of two full brain scans.]

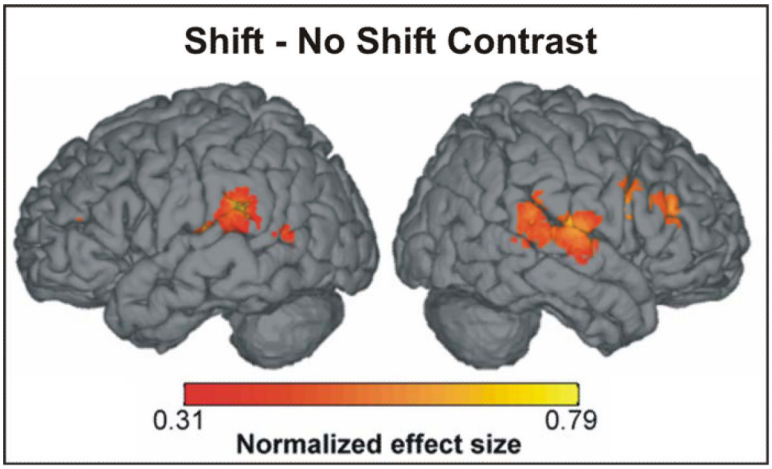

Figure 4.
Regions of significant activation in the *perturbed speech - unperturbed speech* contrast of an fMRI speech perturbation experiment investigating the effects of unexpected perturbation of auditory feedback (30% shift of the first formant frequency during single word reading. Peak activations were found in the superior temporal gyrus bilaterally and right hemisphere ventral premotor cortex/inferior frontal gyrus.

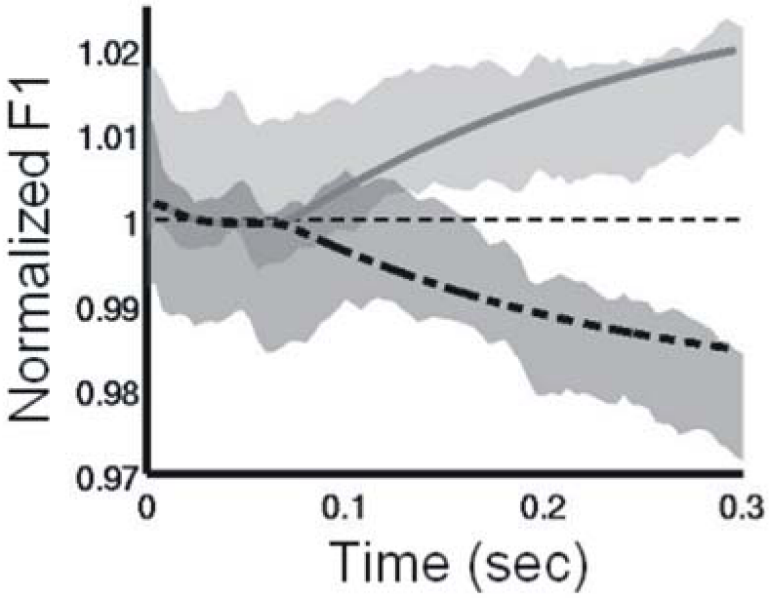

Figure 5.

Comparison of first formant frequency (F1) trajectories produced by the DIVA model (lines) and human subjects (shaded regions) when F1 is unexpectedly perturbed during production of a syllable. Utterances were perturbed by shifting F1 upward or downward by 30% throughout the syllable. Traces are shown for 300 ms starting from the onset of the perturbation at the beginning of vocalization. Shaded areas denote the 95% confidence interval for normalized F1 values during upward (dark) and downward (light) perturbations in the experimental study. Lines indicate values obtained from a DIVA model simulation of the auditory perturbation experiment. Both the model and the experimental subjects show compensation for the perturbation starting approximately 75-150 ms after perturbation onset.

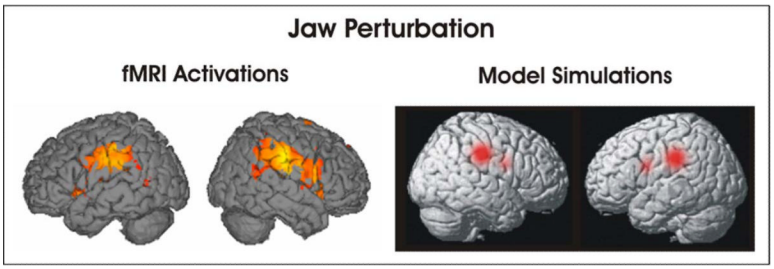

Figure 6.
The left half of the figure shows regions of significant activation in the *perturbed speech* - *unperturbed speech* contrast of an fMRI experiment investigating the effects of unexpected jaw perturbation during single word reading. The right half of the figure shows the results of simulations of the DIVA model during jaw-perturbed speech made prior to the experiment.

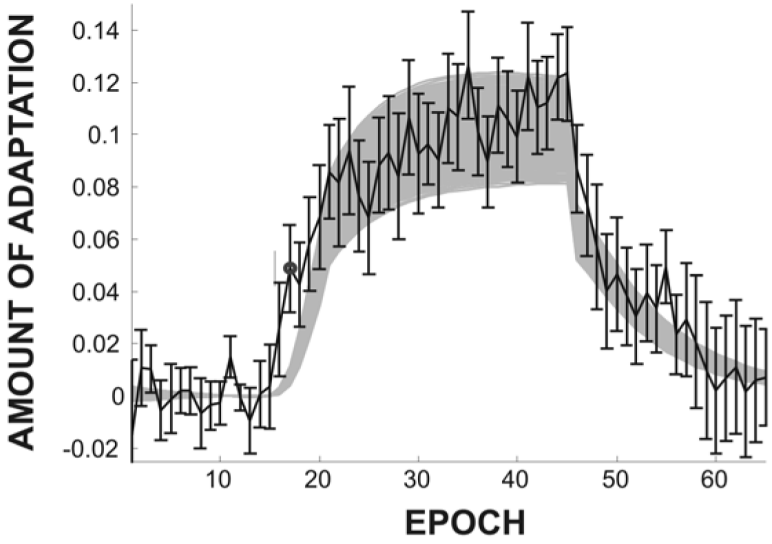

Figure 7.

Adaptive response to systematic perturbation of F1 during a sensorimotor adaptation experiment (solid lines) compared to DIVA model simulations of the same experiment (shaded area). The solid line with standard error bars represents experimental data collected from 20 subjects. The shaded region represents the 95% confidence interval derived from DIVA model simulations. The total duration of the experiment (65 epochs) was approximately 100 minutes. The horizontal dashed line indicates the baseline F1 value.