Cortical Interactions Underlying the Production of Speech Sounds

Frank H. Guenther[1,2,3]

[1]Department of Cognitive and Neural Systems
Boston University
677 Beacon Street
Boston, MA, 02215
Telephone: (617) 353-5765
Fax Number: (617) 353-7755
Email: guenther@cns.bu.edu

[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

***Journal of Communication Disorders*, in press**

ABSTRACT:

Speech production involves the integration of auditory, somatosensory, and motor information in the brain. This article describes a model of speech motor control in which a feedforward control system, involving premotor and primary motor cortex and the cerebellum, works in concert with auditory and somatosensory feedback control systems that involve both sensory and motor cortical areas. New speech sounds are learned by first storing an auditory target for the sound, then using the auditory feedback control system to control production of the sound in early repetitions. Repeated production of the sound leads to tuning of feedforward commands which eventually supplant the feedback-based control signals. Although parts of the model remain speculative, it accounts for a wide range of kinematic, acoustic, and neuroimaging data collected during speech production and provides a framework for investigating communication disorders that involve malfunction of the cerebral cortex and interconnected subcortical structures.

LEARNING OUTCOMES:

(1) Readers will be able to describe several types of learning that occur in the sensory-motor system during babbling and early speech. (2) Readers will be able to identify three neural control subsystems involved in speech production. (3) Readers will be able to identify regions of the brain involved in monitoring auditory and somatosensory feedback during speech production. (4) Readers will be able to identify regions of the brain involved in feedforward control of speech.

KEYWORDS:

Speech production, motor control, neural model, fMRI

ACKNOWLEDGEMENTS

Supported by the National Institute on Deafness and other Communication Disorders (R01 DC02852, F. Guenther PI). The author thanks Jason Tourville for his assistance in preparing this article.

INTRODUCTION

The production of speech sounds requires the integration of diverse information sources in order to generate the intricate patterning of muscle activations required for fluency. Accordingly, a large portion of the cerebral cortex is involved in even the simplest speech tasks, such as reading a single word or syllable (e.g., Fiez & Petersen, 1998; Turkeltaub, Eden, Jones, & Zeffiro, 2002). Broadly speaking, there are three main types of information involved in the production of speech sounds: auditory, somatosensory, and motor, represented in the temporal, parietal, and frontal lobes of the cerebral cortex, respectively. These regions and their interconnections, along with subcortical structures such as the cerebellum, basal ganglia, and brain stem, constitute the neural control system responsible for speech production.

This article describes a neural model of speech production that provides a quantitative account for the interactions between motor, somatosensory, and auditory cortical areas during speech. Although the model is computationally defined, the discussion herein will focus on the hypothesized functions of the modeled brain regions without consideration of associated equations. The interested reader is referred to Guenther, Ghosh, & Tourville (2005) for details concerning the mathematical and computer implementation of the model. We focus here on the computations performed by the cerebral cortex and cerebellum; detailed treatments of the involvement of additional brain regions in speech production can be found in Barlow (1999), Duffy (1995), Kent (1997), and Zemlin (1998).

OVERVIEW OF THE DIVA MODEL

Figure 1 schematizes the main components of the DIVA model. Each box in the diagram corresponds to a set of neurons (or *map*) in the model, and arrows correspond to synaptic projections that transform one type of neural representation into another. The model is implemented in computer simulations that control an articulatory synthesizer (Maeda, 1990) in order to produce an acoustic signal. The articulator movements and acoustic signal produced by the model can be compared to the productions of human speakers; the results of many such comparisons are described elsewhere (e.g., Callan, Kent, Guenther, & Vorperian, 2000; Guenther, 1995; Guenther, Hampson, & Johnson, 1998; Guenther et al., 1999; Nieto-Castanon, Guenther, Perkell, & Curtin, 2005; Perkell et al., 2004a,b).

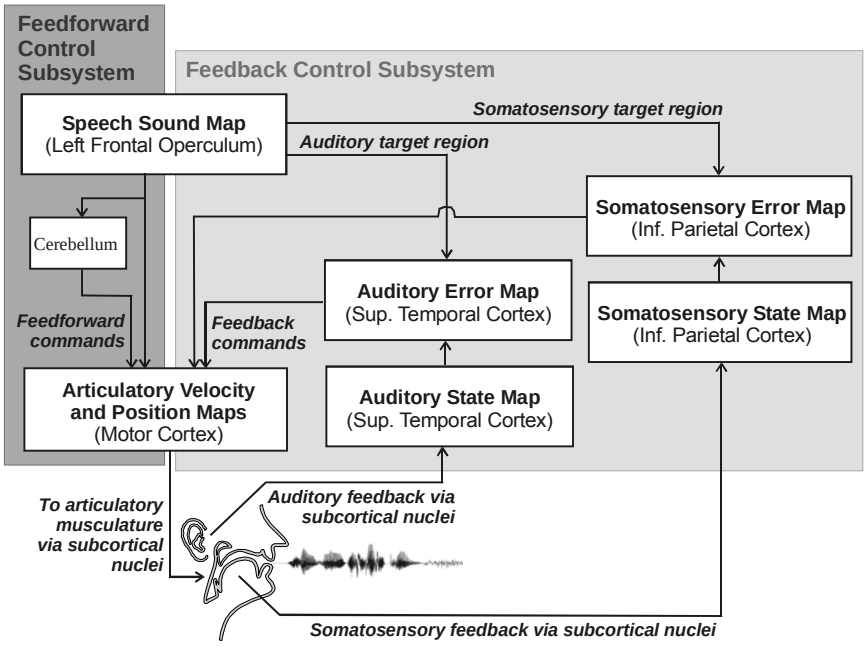

Figure 1. Schematic of the DIVA model of speech acquisition and production. Projections to and from the cerebellum are simplified for clarity.

The production of a speech sound in the DIVA model starts with activation of a *speech sound map* cell, hypothesized to lie in the inferior and posterior portion of Broca's area, a region sometimes

referred to as the frontal operculum. (A “speech sound” as defined herein can be a phoneme, syllable, or word. For the sake of readability, we will typically use the term “syllable” to refer to a single speech sound unit, represented by its own speech sound map cell in the model.) Activation of the speech sound map cell leads to motor commands that arrive in motor cortex via two control subsystems: a feedforward control subsystem, and a feedback control subsystem. The feedback control subsystem can be further broken into two components: an auditory feedback control subsystem, and a somatosensory feedback control subsystem.

Before it can produce speech sounds, the model must undergo a training process analogous to infant babbling and early word imitation. The synaptic projections labeled *Feedback commands* in Figure 1 are tuned during a babbling phase in which semi-random articulator movements are used to produce auditory and somatosensory feedback; this combination of articulatory, auditory, and somatosensory information is used to tune the synaptic projections between the sensory error maps and the model’s motor cortex. The learning in this stage is not phoneme- or syllable-specific; the learned sensory-motor transformations will be used for all speech sounds that will be learned later.

In the next learning stage, the model is presented with sample speech sounds to learn, much like an infant is exposed to the sounds of his/her native language. These sounds take the form of time-varying acoustic signals corresponding to a phoneme, syllable, or word spoken by a human speaker. Based on these samples, the model learns an auditory target for each sound, encoded in the synaptic projections from the speech sound map to the higher-order auditory cortical areas (*Auditory target region* in Figure 1). The targets consist of time-varying regions (or ranges) that encode the allowable variability of the acoustic signal throughout the syllable. The use of target *regions* (rather than point targets) is an important aspect of the DIVA model that provides a unified explanation for a wide range of speech production phenomena, including motor equivalence, contextual variability, anticipatory coarticulation, carryover coarticulation, and speaking rate effects (see Guenther, 1995 for details).

Learning of a sound’s auditory target involves activation of a speech sound map cell (which will represent the sound for production) when the model “hears” the sound sample. This in turn leads to tuning of the synapses projecting from that cell to the auditory cortical areas. Later the same speech sound map cell can be activated to generate the motor commands necessary for producing the sound. Thus the speech sound map cells are activated both when perceiving a sound and when producing the same sound. Neurons with this property, called *mirror neurons,* have been identified in monkey premotor area F5, which is believed to be analogous to Broca’s area in humans (Rizzolatti & Arbib, 1998; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996), where the speech sound map cells are hypothesized to reside.

After an auditory target for a sound has been learned, the model can attempt to produce the sound. On the first attempt, the model will not have a tuned feedforward command for the sound; thus it must rely heavily on the auditory feedback control subsystem to produce the sound. On each attempt to produce the sound, the feedforward command is updated to incorporate the 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 very few auditory errors during production of the sound and thus does not invoke the auditory feedback control subsystem. At this point the model can fluently produce the speech sound.

As the model repeatedly produces a speech sound, it learns a *somatosensory target region* for the sound, 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.

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., 2005 for details). Since the model’s components correspond to groups of neurons at specific anatomical locations, it is possible to generate simulated fMRI activations from the model’s cell activities during a computer simulation.

The relationship between the signal measured in blood oxygen level dependent (BOLD) fMRI and electrical activity of neurons has been studied by numerous investigators in recent years. 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. We use such a hemodynamic response function (HRF), which is part of the SPM software package for fMRI data analysis (http://www.fil.ion.ucl.ac.uk/spm/), to transform neural activities in our model cells into simulated fMRI activity.

In our modeling work, each model cell is meant to correspond to a small population of neurons that fire together. The output of a cell corresponds to neural firing rate (i.e., the number of action potentials per second of the population of neurons). This output is sent to other cells in the network, where it is multiplied by synaptic weights to form synaptic inputs to these cells. The activity level of a cell is calculated as the sum of all the synaptic inputs to the cell (both excitatory and inhibitory), and if the net activity is above zero, the cell's output is equal to this activity level. If the net activity is below zero, the cell'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, Huk, Geisler, & Albrecht, 2000; Rees, Fristion, & Koch, 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, Pauls, Augath, Trinath, & Oeltermann, 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 (i.e., the activity level as defined above), rather than the outputs (firing rates), with an idealized hemodynamic response function (see Guenther et al., 2005 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 cell 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, with the notable exception of the supplementary motor area in the medial frontal lobe (not visible), which is not currently included in the model (see *Concluding Remarks*).

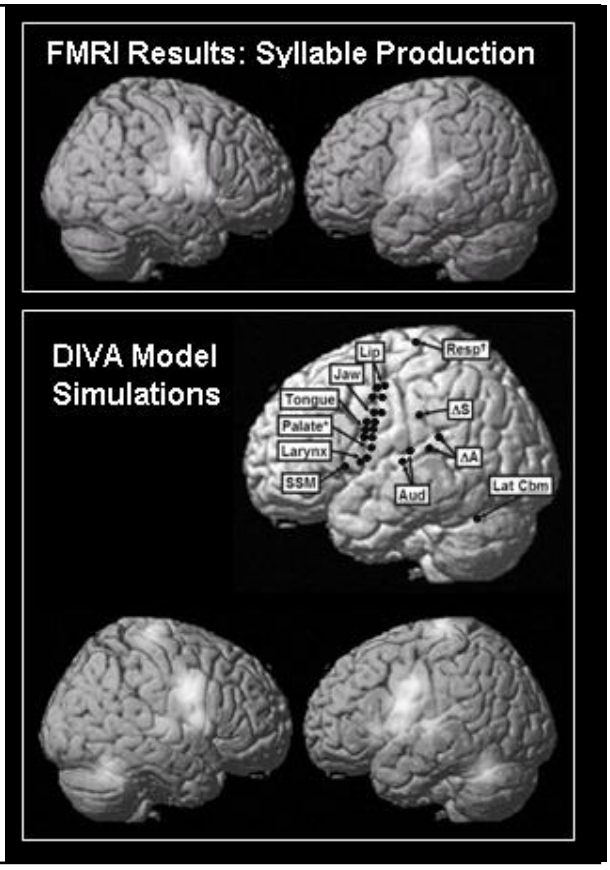

Figure 2. Top. Lateral surfaces of the brain indicating locations of significant activations (random effects; statistics controlled at a false discovery rate of 0.05) measured in an fMRI experiment of single syllable production (*speech – baseline* contrast, where the baseline task consisted of silently viewing the letters YYY on the video screen). **Middle right.** Lateral surfaces of the brain indicating locations of the DIVA model components. 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. Abbreviation key: Aud = auditory state cells; ΔA = auditory error cells; ΔS = somatosensory error cells; Lat Cbm = superior lateral cerebellum; Resp = motor respiratory region; SSM = speech sound map. *Palate representation is somato-sensory only. †Respiratory representation is motor only. **Bottom.** Simulated fMRI activations from the DIVA model when performing the same speech task as the subjects in the fMRI experiment.

The following sections detail the hypothesized neural subsystems for auditory feedback control, somatosensory feedback control, and feedforward control.

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 cells in the left frontal operculum to the 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, auditory error cells in the posterior superior temporal gyrus and planum temporale become active (*auditory error map* in Figure 1). These error cell activities are then transformed into corrective motor commands through projections from the auditory error cells to motor cortex.

The auditory target projections from the model's speech sound map to the auditory cortical areas inhibit auditory error map cells. 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 cells. Evidence of inhibition in auditory cortical areas in the superior temporal gyrus during one's own speech comes from several different sources, including recorded neural responses during open brain surgery (Creutzfeldt, Ojemann, & Lettich, 1989a,b), MEG measurements (Houde, Nagarajan, Sekihara, & Merzenich, 2002; Numminen & Curio, 1999; Numminen, Salmelin, & Hari, 1999), and PET measurements (Wise, Greene, Buchel, & Scott, 1999).

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 producing the wrong sound (cf. Houde & Jordan, 1998), the auditory error cells 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., 2005). 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, which is not noticeable to the subject). 11 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).

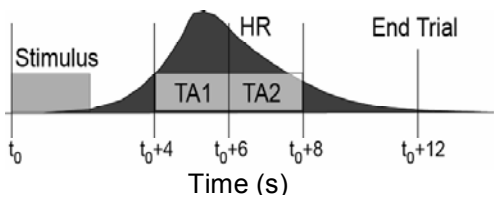

Figure 3. Timeline for a single trial in our fMRI protocol. The subject reads the stimulus out loud during stimulus presentation, when the scanner is not collecting images and is thus quiet. Images are acquired approximately 2 seconds after articulation ceases. HR = estimated hemodynamic response; TA = time of acquisition of brain volume scans.

According to the DIVA model, auditory error cells should be active in the perturbed trials but not the unperturbed trials; thus one should see auditory error cell activation 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 the *perturbed speech – unperturbed speech* contrast. As predicted by the model, auditory error cell 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, Hickok, & Humphries, 2001; Hickok, Buchsbaum, Humphries, & Muftuler, 2003).

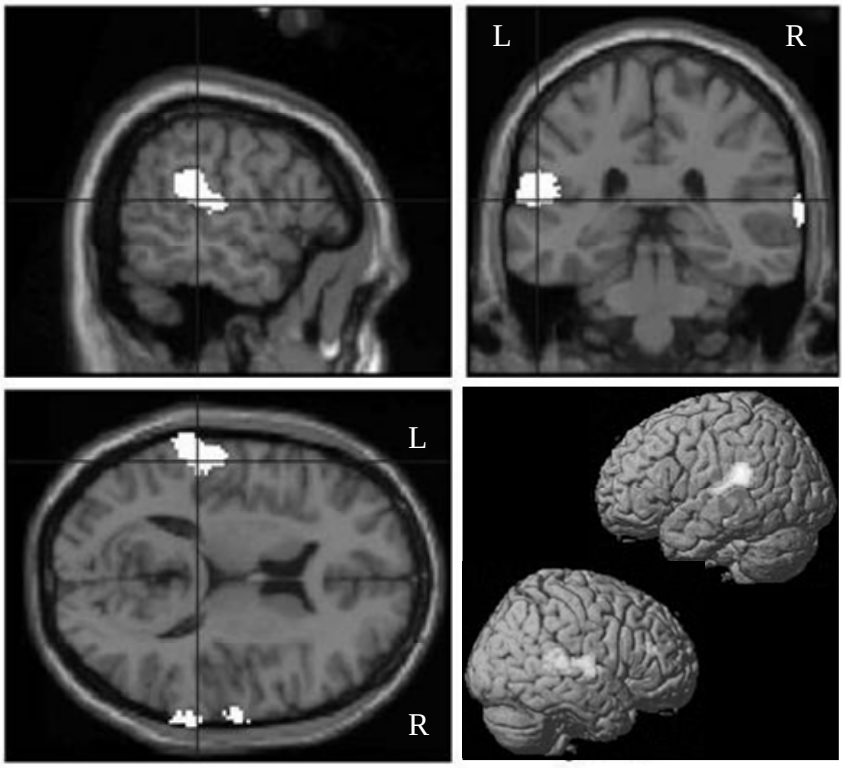

Figure 4. Regions of significant activation in the *perturbed speech – unperturbed speech* contrast of an fMRI experiment investigating the effects of unexpected perturbation of auditory feedback (30% shift of the first formant frequency during single word reading. The bottom right panel illustrates activations on the left lateral surface of the brain. The remaining panels illustrate slices passing through the left planum temporale, highlighting activations in this area as well as right hemisphere auditory cortical regions.

The speech of subjects in the fMRI study was recorded and analyzed to identify whether subjects were compensating for the perturbation within the perturbed trials, 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: (i) the brain contains auditory error cells that signal the difference between a speaker's auditory target and the incoming auditory signal; (ii) these error cells are located in the posterior superior temporal gyrus and supratemporal plane, particularly in the planum temporale of the left hemisphere; and (iii) unexpected perturbation of a speaker's auditory

feedback results in a compensatory articulatory response within approximately 75-150 ms of the perturbation onset.

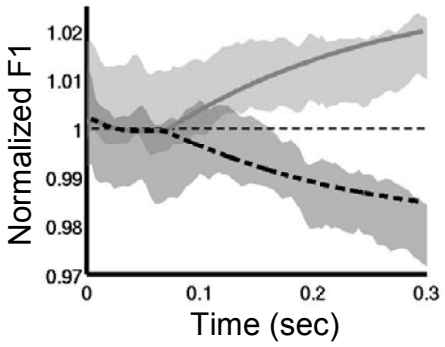

Figure 5. Comparison of first formant frequency (F1) trajectories produced by the DIVA model (lines) and human subjects (shaded regions) when F1 is unexpectedly perturbed during production of a syllable. Utterances were perturbed by shifting F1 upward or downward by 30% throughout the syllable. Traces 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.

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, Gow, & Makris, 1995; Celsis et al., 1999), and speech production (Damasio & Damasio, 1980; Geschwind, 1965). According to the model, cells 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.

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., 2005), an intervention that should activate somatosensory error cells since it creates a mismatch between the target and actual somatosensory state. 13 subjects were asked to read 2-syllable pseudo-words from the 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; this 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 as 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 and in the left frontal operculum, as shown in Figure 6, which utilizes a strict threshold to highlight the most active regions. The results of this study are in keeping with the DIVA model's hypothesis that somatosensory error cells in the supramarginal gyrus will be activated by unexpected jaw perturbation during speech.

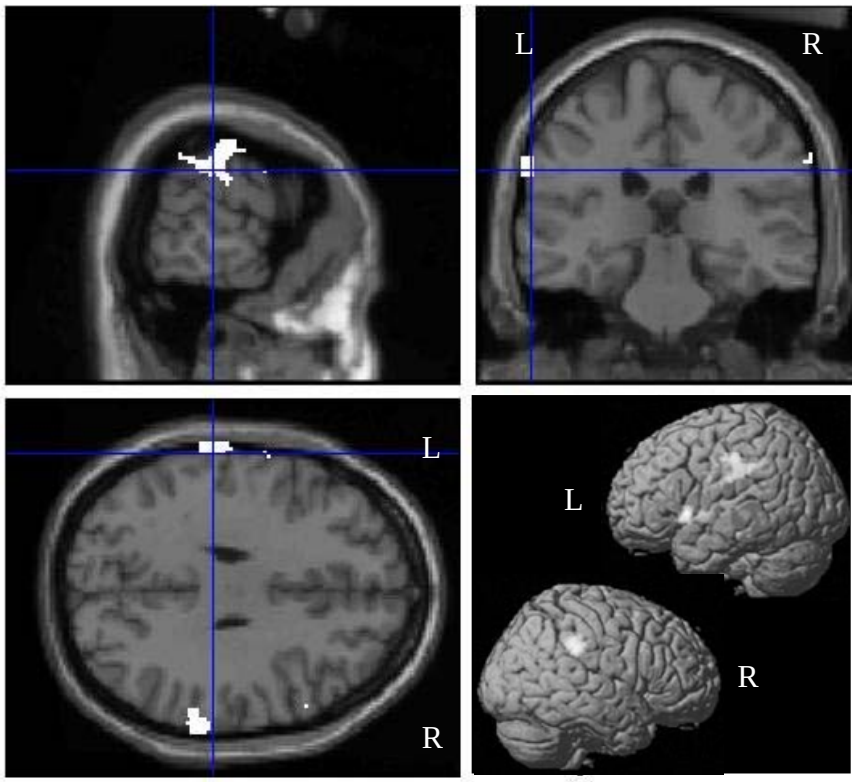

Figure 6. Regions of significant activation in the *perturbed speech – unperturbed speech* contrast of an fMRI experiment investigating the effects of unexpected jaw perturbation during single word reading. The bottom right panel illustrates activations on the left and right lateral surfaces of the brain. The remaining panels illustrate slices passing through the left supramarginal gyrus, highlighting activations in this area as well as right hemisphere supramarginal gyrus.

FEEDFORWARD CONTROL

According to the DIVA model, projections from premotor cortex (specifically, the left frontal operculum) 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, Vogel, Petersen, & Poremba, 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. On the first attempt to produce a new sound, the model relies very heavily on auditory feedback control to produce the sound since its feedforward commands are inaccurate, thus resulting in auditory errors that activate the feedback control subsystem. The corrective commands issued by the auditory feedback control subsystem during the current attempt to produce the sound become stored in the feedforward command 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, 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 production unless speech is perturbed in some way or the sizes or shapes of the articulators change. As the speech articulators get larger with growth, 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 dramatic changes in the sizes and shapes of the speech articulators over the course of a lifetime.

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"; i.e., the speaker's first few productions after normal feedback is restored will still show signs of the adaptation of the feedforward command that occurred when the feedback was perturbed. Effects of this type were reported in the speech sensorimotor adaptation experiment of 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, Perkell, & Guenther, 2004; Villacorta, 2005), 20 subjects performed a psychophysical experiment that involved four phases: 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), a *ramp phase* during which a shift in F1 was gradually introduced to the subject's auditory feedback (epochs 16-20), a *training phase* in which the full F1 perturbation (a 30% shift of F1) was applied on every trial (epochs 21-45), and a *post-test phase* in which the subject received unaltered auditory feedback (epochs 46-65). The subjects' *adaptive response* (i.e., 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, 2005 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. Notably, subjects showed an after-effect as predicted by the model, and the model provides an accurate quantitative fit to the time course of this after-effect.

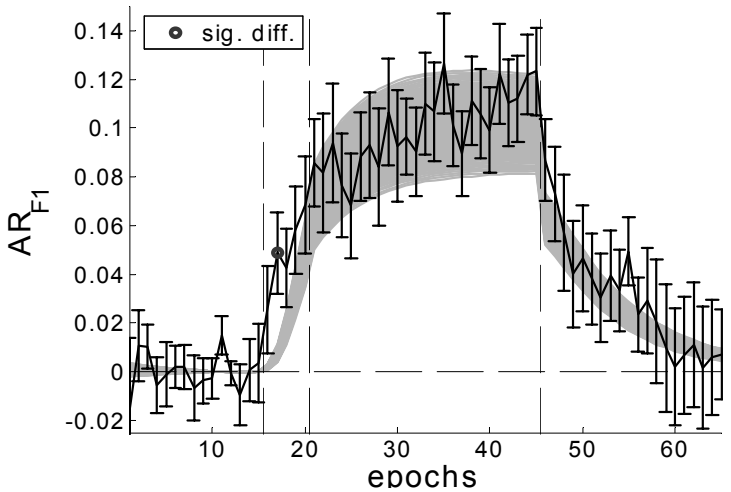

Figure 7. Adaptive response (AR) to systematic perturbation of F1 during a sensorimotor adaptation experiment (solid lines) compared to DIVA model simulations of the same experiment (shaded area). The solid line with standard error bars represents experimental data collected from 20 subjects. The shaded region represents the 95% confidence interval derived from DIVA model simulations. The vertical dashed lines indicate the transitions between the baseline, ramp, training, and post-test phases over the course of the experiment (approximately 100 minutes total duration). The horizontal dashed line indicates the baseline F1 value. [Adapted with permission from Villacorta, 2005.]

CONCLUDING REMARKS

This article has described a neural model of speech motor control 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 thought to involve cortico-cortical projections from premotor to motor cortex, as well as contributions from the cerebellum. The auditory feedback control subsystem is thought to involve 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 is thought to involve projections from premotor cortex to higher-order somatosensory cortex that encode somatosensory targets for speech sounds, as well as projections from somatosensory error cells to motor cortex that encode corrective motor commands.

Although the model described herein accounts for most of the activity seen in fMRI studies of single word or syllable production, it does not provide a complete account of the cortical and cerebellar mechanisms involved in speech. In particular, as currently defined, the DIVA model is given a string of sounds by the modeler, and the model produces this string in the specified order. Brain structures involved in the selection, initiation, and sequencing of speech movements are not treated in the preceding discussion; these include the anterior cingulate area, supplementary motor area (SMA), basal ganglia, and anterior insula (cf. Buckner, Raichle, Miezin, & Petersen, 1996; DeLong, 1999; DeLong & Wichman, 1993; Dronkers, 1996; Georgiou et al., 1994; Nathaniel-James et al., 1997; Paus, Petrides, Evans, & Meyer, 1993; Rogers, Phillips, Bradshaw, Iansek, & Jones, 1998). These areas are being incorporated into an expanded version of the model in ongoing research.

References

Ackermann, H., Vogel, M., Petersen, D., & Poremba, M. (1992). Speech deficits in ischaemic cerebellar lesions. *Journal of Neurology, 239,* 223-227.

Barlow, S.M. (1999). *Handbook of clinical speech physiology*. San Diego: Singular.

Browman, C.P. & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6,* 201-251.

Buchsbaum, B.R., Hickok, G., & Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science, 25,* 663-678.

Buckner, R.L., Raichle M.E., Miezin, F.M., & Petersen, S.E. (1996). Functional anatomic studies of memory retrieval for auditory words and visual pictures. *J Neurosci*, *16*, 6219-6235.

Callan, D.E., Kent, R.D., Guenther, F.H., & Vorperian, H.K. (2000). 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, 43,* 721-736.

Caplan, D., Gow, D., & Makris, N. (1995). Analysis of Lesions by Mri in Stroke Patients with Acoustic-Phonetic Processing Deficits. *Neurology, 45,* 293-298.

Celsis, P., Boulanouar, K., Doyon, B., Ranjeva, J.P., Berry, I., Nespoulous, J.L., & Chollet, F. (1999). Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. *NeuroImage, 9,* 135-144.

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

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

Damasio, H. & Damasio, A.R. (1980). The anatomical basis of conduction aphasia. *Brain, 103,* 337-350.

DeLong, M. R. (1999). The basal ganglia. In E.R.Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), *Principles of Neural Science* (4th ed., pp. 853-867). New York: McGraw Hill.

DeLong, M.R. & Wichman, T. (1993). Basal ganglia-thalamocortical circuits in Parkinsonian signs. *Clinical Neuroscience, 1,* 18-26.

Dronkers, N.F. (1996). A new brain region for coordinating speech articulation. *Nature, 384,* 159-161.

Duffy, J.R. (1995). Motor speech disorders: Substrates, differential diagnosis, and management, St. Louis: Mosby.

Fiez, J.A., & Petersen, S.E. (1998). Neuroimaging studies of word reading. *Proc Natl Acad Sci USA*, *95*, 914-921.

Georgiou, N., Bradshaw, J.L., Iansek, R., Phillips, J.G., Mattingley, J.B., & Bradshaw, J.A. (1994). Reduction in external cues and movement sequencing in Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry, 57,* 368-370.

Geschwind, N. (1965). Disconnexion syndromes in animals and man. II. *Brain, 88,* 585-644.

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

Guenther, F.H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review, 102,* 594-621.

Guenther, F.H., Espy-Wilson, C.Y., Boyce, S.E., Matthies, M.L., Zandipour, M., & Perkell, J.S. (1999). Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America, 105,* 2854-2865.

Guenther, F.H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review, 105,* 611-633.

Guenther, F.H., Ghosh, S.S., and Tourville, J.A. (2005). Neural modeling and imaging of the cortical interactions underlying syllable production *Brain and Language*, E-print ahead of publication.

Heeger, D.J., Huk, A.C., Geisler, W.S., & Albrecht, D.G. (2000). Spikes versus BOLD: What does neuroimaging tell us about neuronal activity? *Nature Neuroscience, 3(7)*, 631-633.

Hickok, G., Buchsbaum, B., Humphries, C. and Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience, 15*, pp. 673-682.

Houde, J.F. & Jordan, M.I. (1998). Sensorimotor adaptation in speech production. *Science, 279,* 1213-1216.

Houde, J.F., Nagarajan, S.S., Sekihara, K., & Merzenich, M.M. (2002). Modulation of the auditory cortex during speech: an MEG study. *Journal of Cognitive Neuroscience, 14,* 1125-1138.

Kent, R.D. (1997). *The speech sciences*. San Diego: Singular.

Krakauer, J. & Ghez, C. (1999). Voluntary movement. In E.R.Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), *Principles of Neural Science* (4th ed., pp. 756-781). New York: McGraw Hill.

Levelt, W.J. & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition, 50,* 239-269.

Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature, 412*, 150-157.

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. Marchal (Eds.), *Speech production and speech modeling* (pp. 131-149). Boston: Kluwer Academic Publishers.

Middleton, F.A. & Strick, P.L. (1997). Cerebellar output channels. *International Review of Neurobiology, 41,* 61-82.

Munhall, K.G. (2001). Functional imaging during speech production. *Acta Psychologica*. *107*, pp. 95-117.

Nathaniel-James, D.A., Fletcher, P., & Frith, C.D. (1997). The functional anatomy of verbal initiation and suppression using the Hayling Test. *Neuropsychologia, 35,* 559-566.

Nieto-Castanon, A., Guenther, F.H., Perkell, J.S., and Curtin, H. (2005). A modeling investigation of articulatory variability and acoustic stability during American English /r/ production. *Journal of the Acoustical Society of America, 117, 3196-3212*.

Numminen, J. & Curio, G. (1999). Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neuroscience Letters, 272,* 29-32.

Numminen, J., Salmelin, R., & Hari, R. (1999). Subject's own speech reduces reactivity of the human auditory cortex. *Neuroscience Letters, 265,* 119-122.

Passingham, R. E. (1993). *The frontal lobes and voluntary action*. Oxford: Oxford University Press.

Paus, T., Petrides, M., Evans, A. C., & Meyer, E. (1993). Role of the human anterior cingulate cortex in the control of oculomotor, manual, and speech responses: a positron emission tomography study. *Journal of Neurophysiology, 70,* 453-469.

Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Stockmann, E., Tiede, M., & Zandipour, M. (2004a). Cross-subject correlations between measures of vowel production and perception. *Journal of the Acoustical Society of America*, *116(4)* Pt. 1, 2338-2344.

Perkell, J.S., Matthies, M.L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E., & Guenther, F.H. (2004b). 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*, *47*, 1259-1269.

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

Rizzolatti, G. & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences, 21,* 188-194.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Research.Cognitive Brain Research, 3,* 131-141.

Rogers, M.A., Phillips, J.G., Bradshaw, J.L., Iansek, R., & Jones, D. (1998). Provision of external cues and movement sequencing in Parkinson's disease. *Motor Control, 2,* 125-132.

Schmahmann, J.D. & Pandya, D.N. (1997). The cerebrocerebellar system. *International Review of Neurobiology, 41,* 31-60.

Tourville, J.A., Guenther, F.H., Ghosh, S.S., Reilly, K.J., Bohland, J.W., & Nieto-Castanon, A. (2005). Effects of acoustic and articulatory perturbation on cortical activity during speech production. *NeuroImage* (*Proceedings of the 11th Annual Meeting of the Organization for Human Brain Mapping, Toronto*), 26(S1), p. S49.

Turkeltaub, P.E., Eden, G.F., Jones, K.M., & Zeffiro, T.A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *NeuroImage*, *16*, 765-780.

Villacorta, V., Perkell, J.S., and Guenther, F.H., (2004). Sensorimotor adaptation to acoustic perturbations in vowel formants. Program of the 147th Meeting of the Acoustical Society of America, *Journal of the Acoustical Society of America*, *115*, 2430.

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

Wise, R.J., Greene, J., Buchel, C., & Scott, S.K. (1999). Brain regions involved in articulation. *Lancet, 353,* 1057-1061.

Zemlin, W.R. (1998). *Speech and hearing science: Anatomy and physiology* (4th edition), Boston: Allyn and Bacon.

CONTINUING EDUCATION:

1. Which of these brain regions includes cells that register the difference between the desired and actual acoustic signal of a speaker?
(a) middle temporal gyrus
(b) planum temporale
(c) supplementary motor area
(d) cerebellum
(e) globus pallidus

2. Which of these brain regions appears to include cells that register the difference between the desired and actual somatosensory state of a speaker?
(a) middle temporal gyrus
(b) supplementary motor area
(c) temporal pole
(d) cerebellum
(e) supramarginal gyrus

3. Damage to the superior paravermal region of the cerebellar cortex usually results in:
(a) ataxic dysarthria, characterized by slurred, poorly articulated speech
(b) stuttering
(c) apraxia of speech, characterized by groping, deletion, and substitution errors
(d) no noticeable effect on speech
(e) mutism

4. Feedforward motor commands contribute substantially to speech except:
(a) when auditory feedback is not available
(b) early in development, before an infant has practice producing new words and syllables
(c) when auditory feedback is perturbed, e.g. by shifting the formant frequencies
(d) when the sensory cortical areas are damaged
(e) all of the above

5. Babbling and early speech are thought to involve:
(a) learning of the relationships between motor, somatosensory, and auditory information
(b) tuning of feedforward commands for new speech sounds
(c) learning of somatosensory targets for speech sounds
(d) learning of auditory targets for speech sounds
(e) all of the above