A Model of Cortical and Cerebellar Function in Speech

Frank H. Guenther[†‡] and **Satrajit S. Ghosh**[†]

† Department of Cognitive and Neural Systems, Boston University, Boston, MA 02215

‡ Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139

E-mail: guenther@cns.bu.edu

ABSTRACT

This paper describes a neural model of speech acquisition and production that accounts for a wide range of experimental data. The model's components correspond to regions of the cerebral cortex and cerebellum that become active during speech production tasks. The model is defined mathematically, and computer simulations verify its ability to account for a wide variety of experimental results concerning speech movements. The model also generates quantitative predictions that can be tested with brain imaging techniques, and it provides a basis for interpreting the functional effects of neurological damage.

1. INTRODUCTION

Over the past decade, our research group has developed, tested, and refined a neural network model of the control of speech movements called the DIVA model [1-5]. The model is described mathematically and implemented in computer simulations that control movements of an articulatory synthesizer. This paper describes the current state of the model with reference to the brain regions thought to correspond to the model's components and concludes with a treatment of how the model relates to speech disorders due to neurological damage.

In the model, production of a phoneme or syllable starts with the activation of speech sound map cells corresponding to the sound to be produced. For simplicity, each speech sound map cell in the computer simulations corresponds to a single phoneme or syllable. For example, there is one speech sound map cell for each vowel, and one for each commonly produced syllable. These cells are hypothesized to correspond to "mirror neurons" that have been found in numerous studies of premotor cortex [6-9], including studies of speech [10, 11]. They can also be interpreted as a "mental syllabary" [12]. The activities of the speech sound map cells in the model's premotor cortex are as follows:

$$(1) \qquad \begin{aligned} P_i(t) &= 1 \quad \textit{if sound } i \textit{ is being produced} \\ P_i(t) &= 0 \quad \textit{otherwise} \end{aligned}$$

When a premotor cortex speech sound map cell is activated in order to start production of the corresponding speech sound, it sends signals to cells in the model's somatosensory, auditory, and primary motor cortical areas. These signals lead to production of the syllable through two control subsystems acting in parallel: a feedback-based control subsystem and a feedforward control subsystem. In the feedback control subsystem, projections from premotor and motor cortical areas to auditory and somatosensory cortical areas form "forward models" that encode sensory expectations for the sound being produced (see Section 2). These expectations are compared to the current sensory state, and an error signal arises if there is a mismatch (see Section 3). This mismatch is then mapped into corrective motor commands by projections from the sensory error cells to the motor cortex (see Section 4). The feedforward control subsystem consists of projections from premotor cortex to motor cortex. These feedforward commands are tuned by monitoring the movements of the feedback controller (see Section 5).

2. FORWARD MODELS THAT PREDICT SENSORY EVENTS

The model posits that signals from the premotor cortex travel to the auditory and somatosensory cortical areas through synaptic weights that encode sensory expectations for the sound being produced. These weights are schematized in Figure 1 by the open circles at the ends of the pathways projecting from the motor and premotor cortices (P and M) to cells in the auditory and somatosensory cortical areas (ΔS and ΔA). These "forward models" are hypothesized to include both cortical and cerebellar components, with the cerebellar contribution being particularly important for fine temporal details.

The first set of synaptic weights, z_{PAu}, corresponds to the pathways projecting from the premotor cortex (labeled P) to cells in the higher-order auditory cortical areas (ΔA) in Figure 1. These weights encode an expected auditory trace for each speech sound. They can be tuned while listening to syllables from the native language and/or listening to correct self-productions. They are hypothesized to encode a spatiotemporal "target region" for the sound in auditory coordinates [see 2, 3]. During production of the sound, this target region is compared to the current auditory state in the auditory cortical areas, and any discrepancy between the target and the current auditory state will lead to a command signal to motor cortex that acts to correct this discrepancy (see Section 4).

ISBN 1-876346-48-5

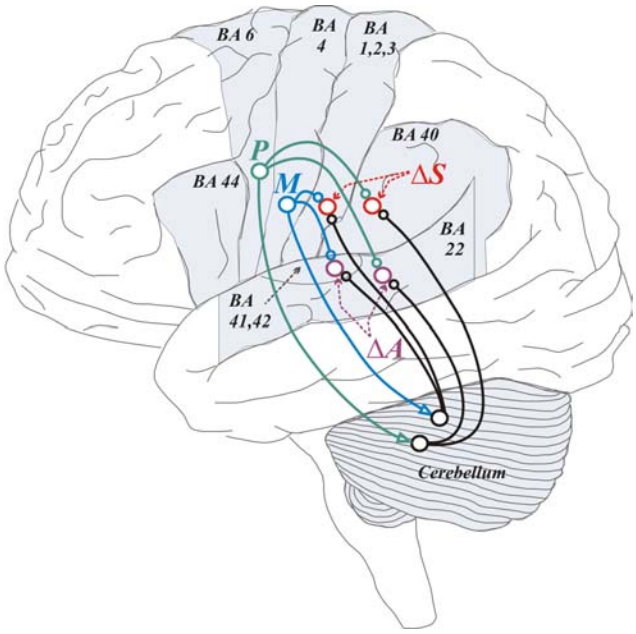

Figure 1: Projections from cells in the motor and premotor cortical areas (*M* and *P*) constitute "forward models" that encode sensory expectations for the current motor actions. These forward models are hypothesized to include both cortico-cortical and cerebellar components.

A second set of synaptic weights, z_{PS}, corresponds to the pathways projecting from the premotor cortex (*P*) to cells in the higher-order somatosensory cortical areas (ΔS) in Figure 1. These weights encode the expected somatic sensation corresponding to the active syllable. A spatiotemporal somatosensory target region can be estimated by monitoring the somatosensory consequences of producing the syllable and averaging these somatosensory consequences over many production attempts.

In addition to the forward model projections from the relatively abstract speech sound representation in the premotor cortex, the model includes synaptic weights z_{MS} and z_{MA} from primary motor cortex to lower-level somatosensory and auditory cortical areas (schematized by the projections from *M* to ΔS and ΔA in Figure 1). These projections represent the expected somatosensory and auditory consequences of the current motor commands, and deviations from these expectations drive corrective movements in the manner described in Section 4.

3. AUDITORY AND SOMATOSENSORY REPRESENTATIONS

The model posits ***auditory state cells*** that correspond to the representation of speech-like sounds in auditory cortical areas (BA 41, 42, 22). The activity of these cells is represented by the following equation:

$$(2) \qquad Au(t) = f_{AcAu}(Acoust(t - \tau_{AcAu}))$$

where f_{AcAu} is the function that transforms an acoustic signal into the corresponding auditory map representation and τ_{AcAu} is the time it takes an acoustic signal transduced by the cochlea to make its way to the auditory cortical areas. The model also has ***auditory error cells*** in these same cortical regions that encode the difference between auditory target regions for the sound being produced and the current auditory state as represented by $Au(t)$:

$$(3) \qquad \begin{aligned} \Delta Au(t) = Au(t) &- P(t - \tau_{PAu}) z_{PAu}(t) \\ &- M(t - \tau_{MAu}) z_{MAu}(t) \end{aligned}$$

where τ_{PAu} and τ_{MAu} are the propagation delays for the signals from premotor and motor cortex to auditory cortex, and z_{MAu} and $z_{PAu}(t)$ encode the auditory expectations for the sound as described in Section 2. The auditory error cells become active during production if the speaker's auditory feedback deviates from the auditory target region for the speech sound being produced.

The model also includes ***somatosensory state cells*** that correspond to the representation of speech articulators in somatosensory cortical areas (BA 1,2,3,40,43):

$$(4) \qquad S(t) = f_{ArS}(Artic(t - \tau_{ArS}))$$

where f_{ArS} is a function that transforms the current state of the articulators into the corresponding somatosensory map state (e.g., positions and velocities of articulators).

Somatosensory error cells code the difference between the somatosensory target region for a sound and the current somatosensory state:

$$(5) \qquad \begin{aligned} \Delta S(t) = S(t) &- P(t - \tau_{PS}) z_{PS}(t) \\ &- M(t - \tau_{MS}) z_{MS}(t) \end{aligned}$$

where τ_{PS} and τ_{MS} are the propagation delays from premotor and motor cortex to somatosensory cortex, and $z_{MS}(t)$ and $z_{PS}(t)$ encode the somatosensory expectations for the sound as described Section 2. The somatosensory error cells become active during production if the speaker's somatosensory feedback from the vocal tract deviates from the somatosensory target region for the speech sound being produced.

4. FEEDBACK CONTROL SIGNALS IN MOTOR CORTEX

According to the model, feedforward and feedback-based control signals are combined in motor cortex. The model's ***motor cortex velocity cells*** correspond to "phasic" cells found in motor cortex single-cell recording studies [e.g., 13]. The model includes two sets of motor velocity cells: one that encodes a feedforward control signal and one that encodes a feedback control signal.

Feedback control signals project from sensory error cells (see Section 3) to the motor cortex, both directly and via the cerebellum. These "inverse model" projections are illustrated in Figure 2 and are governed by the following equation:

$$\dot{M}_{Feedback}(t) = \Delta Au(t-\tau_{AuM})z_{AuM} + \Delta S(t-\tau_{SM})z_{SM} \quad (6)$$

where z_{AuM} and z_{SM} are synaptic weights that transform directional sensory error signals into motor velocities that correct for these errors. The model's name, DIVA, derives from this mapping from sensory ***d***irections ***i***nto ***v***elocities of ***a***rticulators. Mathematically speaking, the weights z_{AuM} and z_{SM} approximate the pseudoinverse of the Jacobian of the function relating articulator positions (M) to the corresponding sensory state (Au, S). These weights can be tuned during babbling by monitoring the relationship between movement commands and their sensory consequences.

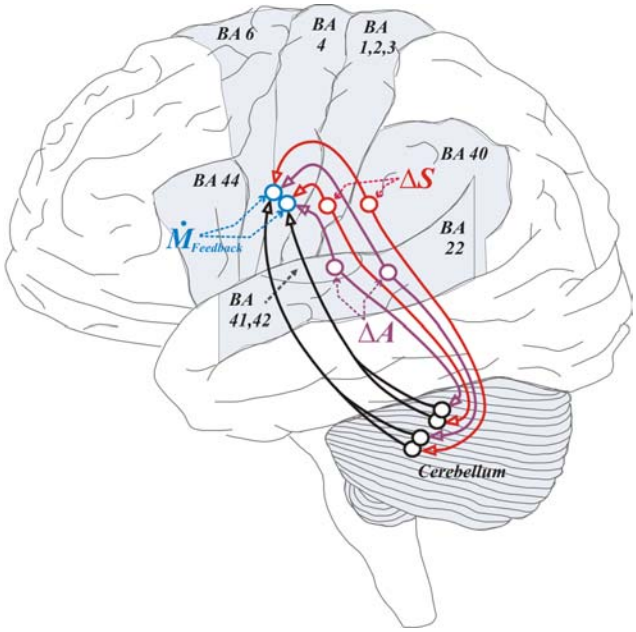

Figure 2: Projections from cells in the auditory and somatosensory cortical areas (ΔA, ΔS) constitute "inverse models" that transform sensory error signals into corrective motor actions. These inverse models, which are hypothesized to include cortico-cortical as well as cerebellar components, are responsible for generating feedback-based control signals in motor cortex.

5. FEEDFORWARD CONTROL SIGNALS IN MOTOR CORTEX

The feedforward motor command, hypothesized to project from ventrolateral premotor cortex to primary motor cortex both directly and via the cerebellum (see Figure 3), is represented by the following equation in the model:

$$\dot{M}_{Feedforward}(t) = P(t-\tau_{PM})z_{PM}(t) - M(t)\,. \quad (7)$$

The weights $z_{PM}(t)$ encode the feedforward motor command for the speech sound being produced. This command can be learned over time by averaging the motor commands from previous attempts to produce the sound (see Section 6).

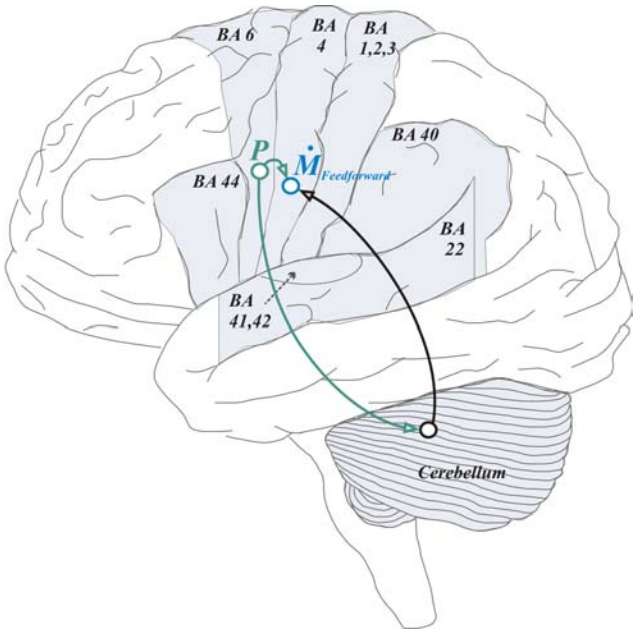

Figure 3: Feedforward control signals are hypothesized to project from the premotor cortex to primary motor cortex, both directly and via the cerebellum.

6. COMBINING FEEDFORWARD AND FEEDBACK-BASED CONTROL SIGNALS IN MOTOR CORTEX

The model's ***motor cortex position cells*** correspond to "tonic" cells found in motor cortex single-cell recording studies [e.g., 13]. They represent the length of a muscle or muscle synergy, and they act as a command to the motor periphery. Their activity is governed by the following equation:

$$M(t) = M(0) + \alpha_{ff}\int_0^t \dot{M}_{Feedforward}(t)g(t)dt + \alpha_{fb}\int_0^t \dot{M}_{Feedback}(t)g(t)dt \quad (8)$$

where α_{fb} and α_{ff} are parameters that determine how much the model is weighted toward feedback control and

ISBN 1-876346-48-5

feedforward control, respectively, and $g(t)$ is a speaking rate signal (or "Go" signal; [14]) that is 0 when not speaking and 1 when speaking at a maximum rate. This speaking rate signal is believed to correspond to the effects of basal ganglia modulation of motor cortical commands [14].

Before an infant has any practice producing a speech sound, the contribution of the feedforward control signal to the overall motor command should be small since it will not yet be tuned. Therefore, during the first few productions, the primary mode of control will be feedback-based control. During these early productions, the feedforward control system is "tuning itself up" by monitoring the motor commands generated by the feedback control system (see also [15]). The feedforward system gets better and better over time, all but eliminating the need for feedback-based control except when external constraints are applied to the articulators (e.g., a bite block, as in [16]) or auditory feedback is artificially perturbed (as in [17]). As the speech articulators get larger with growth, the feedback-based control system provides corrective commands that are eventually subsumed into the feedforward controller. This allows 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 [see 18].

7. ARTICULATORY AND ACOUSTIC STATES

The model also contains variables corresponding to the current articulatory and acoustic state. These values do not correspond to any brain cell activities; they correspond instead to the physical positions of the articulators and the resulting acoustic signal. The articulatory state describes the positions of the seven articulators in the Maeda articulatory synthesizer [19], and is governed by the following equation in the model:

$$Artic(t) = f_{MAr}(M(t - \tau_{MAr})) + Pert(t) \quad (9)$$

where f_{MAr} is the function relating the motor cortex position command to the Maeda parameter values, τ_{MAr} is the time it takes for a motor command to have its effect on the articulatory mechanism, and *Pert* is the effect of external perturbations on the articulators. The acoustic state is determined from the articulatory state as follows:

$$Acoust(t) = f_{ArAc}(Artic(t)) \quad (10)$$

where f_{ArAc} is the transformation performed by Maeda's articulatory synthesis software.

8. SIMULATIONS OF THE MODEL

Comparisons between experimental data and computer simulations of the model controlling movements of a simulated vocal tract have been reported elsewhere [1-3]. These simulations show that the model is capable of accounting for a wide range of speech production phenomena, including motor equivalence, speaking rate effects, coarticulation, and contextual variability in speech movements.

9. CORRESPONDENCE WITH fMRI DATA

Figure 4 shows brain activity while ten subjects produced CVCV nonsense utterances in a functional magnetic resonance imaging experiment performed using a 3 Tesla Siemens scanner at the Massachusetts General Hospital NMR Center. Although the model as described above provides an account for most of the cortical and cerebellar activity seen in this figure, activation in the supplementary motor area (SMA), visible at the top medial portion of the brain in Figure 4, is not accounted for by the model. Based on the existing literature on SMA function, we believe this area is involved in the selection and sequencing of strings of speech sounds, along with the basal ganglia. We are currently working on expanding our model to include a computational description of SMA and basal ganglia function.

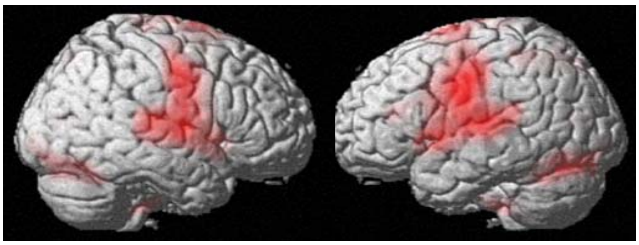

Figure 4: Brain activity while producing CVCV utterances, as measured with fMRI.

10. ACCOUNTING FOR SPEECH DISORDERS WITH THE MODEL

A major long-term goal of our modeling work is to provide a mechanistic account for speech disorders due to brain damage. Because the components of the model correspond to brain regions, speech disorders arising from damage to these brain regions can be simulated by damage to the corresponding components of the model. For example, according to the model an expected effect of Wernicke's aphasia (damage to higher-order auditory and perhaps somatosensory cortical areas) is pure feedforward control of speech (see Figure 3 as compared to Figure 2), including a loss of ability to compensate for speech errors. The expected effect of damage to Broca's area, particularly if the lesion extends into premotor cortex, is an inability to instigate the production of speech "chunks" (e.g., phrases, syllables, or phonemes). The size of the speech chunks affected would depend on the antero-posterior extent of the lesion, with more posterior lesion sites affecting smaller chunks of sound than more anterior lesion sites. Damage to the model's cerebellum, as in ataxic dysarthria, eliminates the fine temporal details of forward and inverse models and feedforward control signals, which leads to

ISBN 1-876346-48-5

poorly timed movements. A potential cause of stuttering in the model is an inappropriately low value of α_{ff} and/or an inappropriately high value of α_{fb} in Equation 8 when normal auditory feedback is available (see also [20]). This biases the model toward feedback control, which is unstable except for very slow movements and can lead to stuttering-like behavior for faster speech. This weak feedforward control signal may arise due to damage in the white matter pathways projecting to primary motor cortical areas (see [21]).

REFERENCES

[1] F. H. Guenther. "A neural network model of speech acquisition and motor equivalent speech production", *Biol Cybern,* **72**, pp. 43-53, 1994.

[2] F. H. Guenther. "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production", *Psychol Rev,* **102**, pp. 594-621, 1995.

[3] F. H. Guenther, M. Hampson, D. Johnson. "A theoretical investigation of reference frames for the planning of speech movements", *Psychol Rev,* **105**, pp. 611-633, 1998.

[4] J. S. Perkell, F. H. Guenther, H. Lane, M. L. Matthies, P. Perrier, J. Vick, R. Wilhelms-Tricarico, M. Zandipour. "A theory of speech motor control and supporting data from speakers with normal hearing and profound hearing loss", *Journal of Phonetics,* **28**, pp. 233-272, 2000.

[5] F. H. Guenther. "Neural models of adaptive sensory-motor control for flexible reaching and speaking". Boston University Ph.D. Dissertation, Boston, MA, 1992.

[6] E. Kohler, C. Keysers, M. A. Umilta, L. Fogassi, V. Gallese, G. Rizzolatti. "Hearing sounds, understanding actions: action representation in mirror neurons", *Science,* **297**, pp. 846-848, 2002.

[7] G. Rizzolatti, L. Fadiga, L. Fogassi, V. Gallese. "Resonance behaviors and mirror neurons", *Arch Ital Biol,* **137**, pp. 85-100, 1999.

[8] G. Rizzolatti, L. Fadiga, V. Gallese, L. Fogassi. "Premotor cortex and the recognition of motor actions", *Brain Res Cogn Brain Res,* **3**, pp. 131-141, 1996.

[9] V. Gallese, L. Fadiga, L. Fogassi, G. Rizzolatti. "Action recognition in the premotor cortex", *Brain,* **119**, pp. 593-609, 1996.

[10] L. Fadiga, L. Craighero, G. Buccino, G. Rizzolatti. "Speech listening specifically modulates the excitability of tongue muscles: a TMS study", *Eur J Neurosci,* **15**, pp. 399-402, 2002.

[11] M. Sundara, A. K. Namasivayam, R. Chen. "Observation-execution matching system for speech: a magnetic stimulation study", *Neuroreport,* **12**, pp. 1341-1344, 2001.

[12] W. J. Levelt and L. Wheeldon. "Do speakers have access to a mental syllabary?", *Cognition,* **50**, pp. 239-269, 1994.

[13] J. F. Kalaska, D. A. Cohen, M. L. Hyde, M. Prud'homme. "A comparison of movement direction-related versus load direction- related activity in primate motor cortex, using a two-dimensional reaching task", *J Neurosci,* **9**, pp. 2080-2102, 1989.

[14] D. Bullock and S. Grossberg. "Neural dynamics of planned arm movements: emergent invariants and speed- accuracy properties during trajectory formation", *Psychol Rev,* **95**, pp. 49-90, 1988.

[15] M. Kawato and H. Gomi. "A computational model of four regions of the cerebellum based on feedback-error learning", *Biol Cybern,* **68**, pp. 95-103, 1992.

[16] B. Lindblom, J. Lubker, T. Gay. "Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation", *Journal of Phonetics,* **7**, pp. 147-161, 1979.

[17] J. Houde and M. I. Jordan. "Adaptation in speech production", *Science,* **279**, pp. 1213-1216, 1998.

[18] D. E. Callan, R. D. Kent, F. H. Guenther, H. K. Vorperian. "An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system", *J Speech Lang Hear Res,* **43**, pp. 721-736, 2000.

[19] S. Maeda, *in* "Speech production and speech modelling" (W. J. Hardcastle and A. Marchal, eds.), p. 131-149. Kluwer Academic Publishers, Boston, 1990.

[20] L. Max, V. L. Gracco, F. H. Guenther, I. Vincent, M. Wallace. "A sensorimotor model of stuttering: Insights from the neuroscience of motor control", *in* "4th World Conference on Fluency Disorders", Montreal, Quebec, in press.

[21] M. Sommer, M. A. Koch, W. Paulus, C. Weiller, C. Buchel. "Disconnection of speech-relevant brain areas in persistent developmental stuttering", *Lancet,* **360**, pp. 380-383, 2002.

Acknowledgements. This research was supported by NIH/NIDCD grants R01 DC02852 (F. Guenther, PI) and R01 DC01925 (J. Perkell, PI).

 ISBN 1-876346-48-5

ISBN 1-876346-48-5