To appear in: *Proceedings of the 4th International Nijmegen Speech Motor Conference*, Nijmegen, The Netherlands, June 13-16, 2001.

Neural Modeling of Speech Production

Frank H. GUENTHER[1,2]
[1]*Department of Cognitive and Neural Systems, Boston University, 677 Beacon Street, Boston, MA, USA*
[2]*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA*
guenther@cns.bu.edu

Abstract. This paper describes a neural model of speech production and perception-production interactions. This model has been developed to account for a wide variety of experimental data, ranging from kinematic analyses of articulator movements to functional imaging studies of the human brain. We have also tested predictions based on the model with these and other experimental techniques. Hypothesized neural correlates of the model's components have been identified to facilitate testing of model predictions with techniques such as fMRI. The model also serves as a framework for interpreting and organizing the accumulating mass of data from functional imaging studies of the human brain.

1. Introduction: The DIVA Model of Speech Production

Our laboratory has developed a neural network model of speech motor skill acquisition and speech production, called the DIVA model, that explains a wide range of data on contextual variability, motor equivalence, coarticulation, and speaking rate effects (Guenther, 1994, 1995a,b; Guenther, Hampson, and Johnson, 1998; Guenther and Micci Barreca, 1997). This model is schematized in Figure 1. Each block in the model corresponds to a set of neurons that constitute a neural representation. Model parameters, corresponding to synaptic weights, are tuned during a babbling phase in which random movements of the speech articulators provide tactile, proprioceptive, and auditory feedback signals that are used to train three neural mappings indicated by filled semicircles in the figure. These mappings are later used for phoneme production.

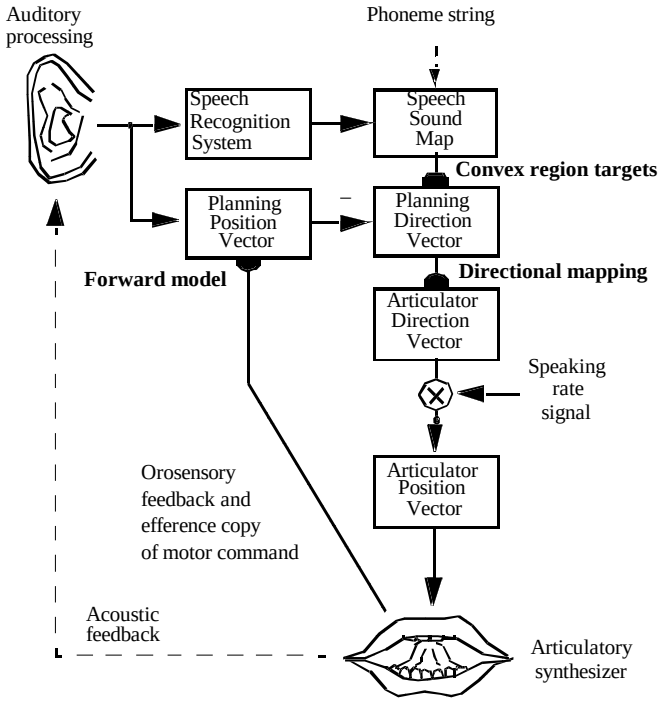

Figure 1. Overview of the DIVA model. Filled semicircles represent learned neural mappings.

The synaptic weights of the first mapping, labeled "convex region targets" in the figure, encode auditory and orosensory targets for each phoneme the model learned during babbling. To explain how infants learn phoneme-specific and language-specific limits on acceptable articulatory and acoustic variability, the learned speech sound targets take the form of multidimensional regions, rather than points, in auditory and orosensory spaces. The notion of phonemic targets as multidimensional regions provides a simple and unified explanation for many

long-studied speech phenomena, including aspects of anticipatory and carryover coarticulation, contextual variability, motor equivalence, velocity/distance relationships, and speaking rate effects (Guenther, 1995a).

The second neural mapping, labeled "directional mapping" in the figure, transforms desired movement directions in auditory and orosensory spaces into movement directions in an articulator space closely related to the vocal tract musculature. This mapping is related to the Moore-Penrose (MP) pseudoinverse of the Jacobian matrix relating the auditory, somatosensory, and articulatory spaces; in effect, the model learns an approximation of the MP pseudoinverse during babbling. The use of this mapping to control the model's articulator movements is thus closely related to pseudoinverse-style control techniques in robotics (e.g., Ligeois, 1977), and the resulting controller is capable of automatically compensating for constraints and/or perturbations applied to the articulators (Guenther, 1994, 1995a; Guenther and Micci Barreca, 1997), thus accounting for the motor equivalent capabilities observed in humans when speaking with a bite block or lip perturbation.

The third mapping, labeled "forward model" in the figure, transforms orosensory feedback from the vocal tract and an efference copy of the motor outflow commands into a neural representation of the auditory signal that corresponds to the current vocal tract shape. This forward model allows the system to control speech movements without relying on auditory feedback, which may be absent or too slow for use in controlling ongoing articulator movements.

2. Hypothesized Neural Correlates of the DIVA Model

One advantage of the neural network approach is that it allows one to analyze the brain regions involved in speech in terms of a well-defined theoretical framework, thus allowing a deeper understanding of the brain mechanisms underlying speech. Figure 2 illustrates hypothesized neural correlates for several central components of the DIVA model. These hypotheses are based on a number of neuroanatomical and neurophysiological studies, including lesion/aphasia studies, MEG, PET, and fMRI imaging studies, and single-cell recordings from cortical and subcortical areas in animals.

The pathway labeled 'a' in the figure corresponds to projections from premotor cortex to primary cortex, hypothesized to underlie feedforward control of the speech articulators. Pathway b represents hypothesized projections from premotor cortex (lateral BA 6) to higher-order auditory cortical areas in the superior temporal gyrus (BA 22) and orosensory areas in the supramarginal gyrus (BA 40). These "efference copy" projections are hypothesized to carry target sensations associated with motor plans in premotor cortex. For example, premotor cortex cells representing the syllable /bli/ project to higher-order auditory cortex cells; these projections represent an expected sound pattern (i.e., the auditory representation of the speaker's own voice while producing /bli/). Similarly, projections from premotor cortex to orosensory areas in the supramarginal gyrus represent the expected pattern of somatosensory stimulation during /bli/ production. Pathway b is hypothesized to encode the convex region targets for speech sounds in the DIVA model, corresponding to the pathway between the Speech Sound Map and Planning Direction Vector in Figure 1.

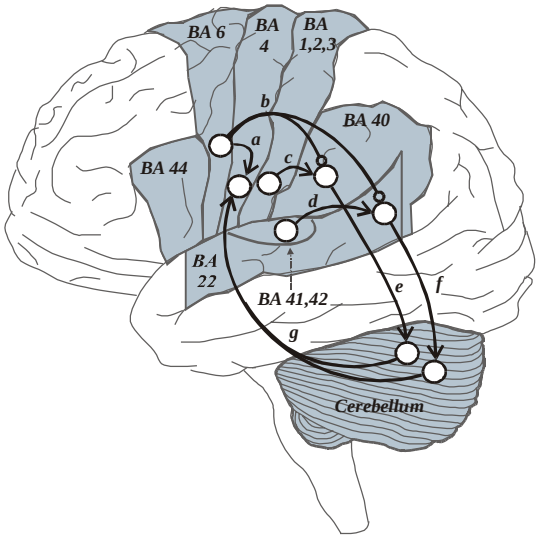

Figure 2. Hypothesized neural correlates of several central components of the DIVA model. BA = Brodmann's Area. See text for details.

One interesting aspect of the model in Figure 2 is the role of auditory cortical areas in speech production as well as speech perception. According to the model, auditory "targets" project from premotor cortical areas to the posterior superior temporal gyrus (pathway b), where they are compared to incoming auditory information from primary auditory cortex (pathway d). The difference between the target and the actual auditory signal represents an "error" signal that is mapped through the cerebellum (pathway f), which transforms the auditory error into a motor velocity signal that can act to zero this error (pathway g). This projection through the cerebellum to motor cortex forms a component of the Directions Into Velocities of Articulators mapping that gives the DIVA model its name. Evidence that auditory cortical areas in the superior temporal gyrus and temporal plane are involved in speech production comes from a number of neuroimaging studies. For example, Hickok et al. (2000) report activation in left posterior STG areas (planum temporale, superior temporal sulcus) during a PET visual object naming task in which the subject's auditory feedback of his/her own productions was masked with noise. Bookheimer et al. (1995) report activations near primary auditory cortex in a similar task. Paus et al. (1996) also reported activation in the area of the left planum temporale during a PET object naming task. These authors attributed this activation to "motor-to-sensory discharges", compatible with pathway b in Figure 2. This interpretation receives support from a magnetoencephalography (MEG) study by Levelt et al. (1998), who showed that the auditory cortical activations during speech production slightly preceded the initiation of articulatory processes. All of these results provide support for the notion of auditory perceptual targets for speech production, in keeping with a central aspect of the DIVA model (e.g., Guenther, 1995b; Guenther et al., 1998; see also Perkell et al., 1995; Bailly et al., 1993).

The model also proposes a novel role for the supramarginal gyrus (BA 40) in speech production. This brain region has been implicated in phonological processing for speech perception (e.g., Caplan, Gow, and Makris, 1995; Celsis et al., 1999), as well speech production (Geshwind, 1965; Damasio and Damasio, 1980). The current model proposes that, among other things, the supramarginal gyrus represents the difference between target oral sensations (projecting from premotor cortex via pathway b in Figure 2) and the current state of the vocal tract (projecting from somatosensory cortex via pathway c). This difference represents the desired movement direction in orosensory coordinates and is hypothesized to map through the cerebellum to motor cortex, thus constituting a second component of the Direction Into Velocities of Articulators mapping.

Not shown in Figure 2, for the sake of clarity, is the insular cortex (BA 43), buried within the sylvian fissure. The anterior insula has been shown to play an important role in speech articulation (e.g., Dronkers, 1996). This region is contiguous with the frontal operculum, which includes portions of the premotor and motor cortices related to oral movements. We adopt the view that the anterior insula has similar functional properties to the premotor and motor cortices. This view receives support from fMRI studies showing activation of anterior insula during non-speech tongue movements (Corfield et al., 1999), PET results showing concurrent primary motor cortex and anterior insula activations during articulation (Fox et al., 2001), and PET results showing concurrent lateral premotor cortex and anterior insula activations during articulation (Wise et al., 1999).

An important purpose of the model outlined in Figure 2 is to generate predictions that serve as the basis for focused functional imaging studies of brain function during speech. For example, the model of Figure 2 predicts that perturbation of a speech articulator such as the lip during speech should cause an increase in activation in the supramarginal gyrus, since the perturbation will cause a larger mismatch between orosensory expectations and the actual orosensory feedback signal. The model further predicts that extra activation will be seen in the cerebellum and motor cortex under the perturbed condition, since pathway e in Figure 2 would transmit the extra supramarginal gyrus activation to the cerebellum and on to motor cortex (pathways e, g). We are currently testing these and other predictions of the model using fMRI and MEG.

References

Bailly, G., Laboissière, R., and Schwartz, J.L. (1991). Formant trajectories as audible gestures: An alternative for speech synthesis. *Journal of Phonetics*, **19**, pp. 9-23.

Bookheimer, S.Y., Zeffiro, T.A., Blaxton, T., Gaillard, W., and Theodore, W. (1995). Regional cerebral blood flow during object naming and word reading. *Human Brain Mapping*, **3**, pp. 93-106.

Caplan, D., Gow, D., and Makris, N. (1995). Analysis of lesions by MRI in stroke patients with acoustic-phonetic processing deficits. *Neurology*, **45**, pp. 293-298.

Celsis, P., Boulanouar, K., Ranjeva, J.P., Berry, I., Nespoulous, J.L., and 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**, pp. 135-144.

Corfield, D.R., Murphy, K., Josephs, O., Fink, G.R., Frackowiak, R.S.J., Guz, A., Adams, L., and Turner, R. (1999). Cortical and subcortical control of tongue movement in humans: A functional neuroimaging study using fMRI. *Journal of Applied Physiology*, **85**, pp. 1468-1477.

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

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

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

Fox, P.T., Huang, A., Parsons, L.M., Xiong, J., Zamarippa, F., Rainey, L., and Lancaster, J.L. (2001). Location-probability profiles for the mouth region of human primary motor-sensory cortex: Model and validation. *NeuroImage*, **13**, pp. 196-209.

Geschwind, N. (1965). Disconnexion syndromes in animals and man. I. *Brain*, **88**, pp. 237-294.

Guenther, F.H. (1995b). A modeling framework for speech motor development and kinematic articulator control. *Proceedings of the XIIIth International Conference of Phonetic Sciences (*vol. 2, pp. 92-99). Stockholm, Sweden: KTH and Stockholm University.

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

Guenther, F.H., and Gjaja, M.N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America*, **100**, pp. 1111-1121.

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

Guenther, F.H., Husain, F.T., Cohen, M.A., and Shinn-Cunningham, B.G. (1999b). Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America*, **106**, pp. 2900-2912.

Guenther, F.H., and Micci Barreca, D. (1997). Neural models for flexible control of redundant systems. In: P. Morasso and V. Sanguineti (eds.), *Self-organization, Computational Maps, and Motor Control* (pp. 383-421). Amsterdam: Elsevier-North Holland.

Hickok, G., Erhard, P., Kassubek, J., Helms-Tillery, A.K., Naeve-Velguth, S., Strupp, J.P., Strick, P.L., and Ugurbil, K. (2000). A functional magnetic resonance imaging study of the role of left posterior superior temporal gyrus in speech production: Implications for the explanation of conduction aphasia. *Neuroscience Letters*, **287**, pp. 156-160.

Levelt, W.J.M., Praamstra, P., Meyer, A.S., Helenius, P., and Salmelin, R. (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience*, **10**, pp. 553-567.

Ligeois, A. (1977). Automatic supervisory control of the configuration and behavior of multibody mechanisms. *IEEE Transactions on Systems, Man, and Cybernetics, 7*(12), 869-871.

Paus, T., Perry, D.W., Zatorre, R.J., Worsley, K.J., and Evans, A.C. (1996). Modulation of cerebral blood flow in the human auditory cortex during speech: Role of motor-to-sensory discharges. *European Journal of Neuroscience*, **8**, pp. 2236-2246.

Perkell, J. S., Matthies, M. L., Svirsky, M. A., & Jordan, M. I. (1995). Goal-based speech motor control: A theoretical framework and some preliminary data. *Journal of Phonetics*, **23**, pp. 23-35.

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

Acknowledgement
This research was supported by the National Institute on Deafness and other Communication Disorders.