A neural model of speech production and supporting experiments. Frank H. Guenther\textsuperscript{1,2,3} & Joseph S. Perkell\textsuperscript{2,1}, 1 Dept. of Cognitive & Neural Systems, Boston University, Boston, MA, USA, 2 Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA, 3 Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA. [Full Paper Available on CD]

This paper describes the DIVA model of speech production and presents results of experiments designed to test and refine the model. According to the model, production of a phoneme or syllable starts with activation of a speech sound map cell (in left ventral premotor cortex) corresponding to the sound to be produced. This leads to production of the sound through two motor subsystems: a feedback control subsystem and a feedforward control subsystem. In the feedback control subsystem, signals from the premotor cortex travel to the auditory and somatosensory cortical areas through tuned synapses that encode sensory expectations for the sound being produced. These expectations take the form of time-varying auditory and somatosensory target regions. The target regions are compared to the current auditory and somatosensory state, and any discrepancy between the target and the current state leads to a corrective command signal to motor cortex. In the feedforward control subsystem, signals project from premotor cortex to primary motor cortex, both directly and via the cerebellum. These signals are tuned with practice by monitoring the commands from previous attempts to produce the sound, initially under feedback control. Feedforward and feedback-based control signals are combined in the model’s motor cortex to form the overall motor command. We present experimental results that support two theoretical characteristics of the model: its use of auditory target regions (including hypothesized effects of perceptual acuity on production target size), and its ability to achieve stable acoustic results using motor equivalent tradeoffs between articulatory gestures.