BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCE

Dissertation

NEURAL MECHANISMS UNDERLYING AUDITORY FEEDBACK CONTROL OF SPEECH

by

JASON A. TOURVILLE

B.A., Oberlin College, 1996

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2008

© Copyright by
JASON A. TOURVILLE
2008

Approved by

First Reader

Frank H. Guenther, Ph.D.
Associate Professor of Cognitive and Neural Systems

Second Reader

Daniel H. Bullock, Ph.D.
Professor of Cognitive and Neural Systems and Psychology

Third Reader

Joseph S. Perkell, Ph.D.
Adjunct Professor of Cognitive and Neural Systems

ACKNOWLEDGMENTS

The research described below was supported by R01 DC02852 from the National Institute on Deafness and other Communication Disorders (F. Guenther, PI). Neuroimaging was performed at the Athinoula A. Martinos Center for Biomedical Imaging, which is funded by grants from the National Center for Research Resources (P41RR14075) and the MIND institute.

Pursuing a doctoral degree from the Department of Cognitive and Neural Systems was challenge in terms of both the intellectual rigor and the personal commitment it required. The experience was rewarding and humbling. It may be that no doctoral student has ever required, and received, more support, patience, and encouragement than this one. All were generously provided by my advisor, Frank Guenther who deserves credit for opening the CNS door for me and then suffering through the consequences. I am forever indebted to Frank for the opportunity he provided me some 8+ years ago and for those he continues to provide today. Frank also deserves credit for providing a unique and wonderful work environment. I have worked with a number of remarkable students and post-docs while in the speechlab, and count many as cherished friends. I thank Kevin Reilly, Jon Brumberg, Maya Peeva, Elissa Golfinopoulus, Oren Civier, Carrie Niziolek, and Julie Yoo for their patient help. Satra Ghosh, Alfonso Nieto-Castanon, and Jay Bohland deserve their own ‘thank you’ because they have been patient and helpful for so long. I also thank the staff at the CNS department and my defense committee, particularly Joseph Perkell and Daniel Bullock, for their valued contributions to the preparation of this dissertation.

Those who attended my dissertation defense know the tremendous support I have from a very extended family, support that has remained unwavering for many, many, years. My surviving those years is due, in no small part, to them. My parents, Allen and Molly, deserve special thanks for giving me every opportunity a son of Ellenburg, New York could want and then some.

Finally, I happily acknowledge my wife, Christy. Her unconditional support and encouragement, her humor, love, and sacrifice made the pursuit that culminates with this document possible. Thank you.

NEURAL MECHANISMS UNDERLYING AUDITORY FEEDBACK CONTROL OF SPEECH

(Order No.)

JASON A. TOURVILLE

Boston University Graduate School of Arts and Science, 2008

Major Professor: Frank H. Guenther, Associate Professor of Cognitive and Neural Systems

ABSTRACT

Speech production relies on both feedforward and feedback control mechanisms. The DIVA model provides a quantitatively explicit description of the interaction between these controllers during speech production. According to the model, speech is initiated with the generation of feedforward motor commands. If these commands do not result in the intended speech sound and unexpected auditory feedback is detected, an auditory "error signal" is generated which is corrected by the feedback controller. The aim of this dissertation is to enhance our understanding of these processes using functional magnetic resonance imaging (fMRI).

The first part of this dissertation describes efforts to enhance and constrain the interpretation of fMRI experiments designed to test predictions of the DIVA model. A mapping of the model's components to specific brain regions follows a detailed review of the brain regions involved in speech production. This mapping permits anatomically specific *a priori* hypotheses regarding task-related modulation of brain activation. This work also informed the development of a whole-brain parcellation system tailored for

imaging studies of speech processing that, when used in conjunction with region-of-interest-based functional analysis methods, offers greater statistical power than traditional voxel-based methods.

An fMRI study investigating the neural mechanisms underlying auditory feedback control of speech is described in the second part of the dissertation. Subjects spoke monosyllabic words under normal and perturbed auditory feedback conditions. Acoustic measurements demonstrated compensation for the perturbation within approximately 136 ms of onset. Neural responses during perturbed feedback were greater in posterior superior temporal cortex bilaterally, consistent with the DIVA model's hypothesis that neurons coding mismatches between expected and actual auditory signals lie in these regions. Increased frontal activation, however, was restricted to the right hemisphere. Laterality tests indicated that (i) under normal feedback conditions, articulator control is predominantly left-lateralized and (ii) feedback-based articulator control relies on greater involvement of right hemisphere ventral frontal regions, especially premotor and inferior frontal cortex. Structural equation modeling revealed increased influence of bilateral auditory cortical areas on right frontal areas during shifted speech. These results led to the addition of a right-lateralized premotor feedback control map to the DIVA model.

TABLE OF CONTENTS

1 Introduction..... 1

1.1 Combining neural modeling and functional neuroimaging 3

1.2 The influence of auditory feedback on speech production 5

2 The functional neuroanatomy of speech production..... 8

2.1 Overview of the brain regions involved in speech production 8

2.2 Overview of the DIVA model 22

2.3 The neuroanatomy of the DIVA model.....27

2.3.1 Premotor cortex *Speech Sound Map*.....29

2.3.2 Motor velocity and position maps 33

2.3.3 Feedback control subsystem.....39

2.3.4 Feedforward motor control subsystem 50

2.4 Predicting BOLD responses from DIVA simulations 54

3 A parcellation system for functional imaging studies of speech..... 57

3.1 Cortical Regions of Interest.....61

3.2 Subcortical regions of interest.....68

3.3 Conclusions.....71

4 An investigation of the neural substrates of auditory feedback control of speech 74

4.1 Auditory feedback in the DIVA model 74

4.2 Identifying the neural substrates of auditory feedback control with fMRI.....75

4.3 Real-time auditory feedback perturbation 79

4.3.1 Imaging data analysis 81

4.3.2 Assessing effective connectivity by structural equation modeling.....85

4.4 Evidence of compensation.....90

4.5 The neural substrates of auditory feedback control.....93

4.5.1 Voxel-based analysis 93

4.5.2 Region of interest analysis.....98

4.6 Inter-regional interactions in the auditory feedback control network.....102

4.7 Discussion 104

4.7.1 Formant shift compensation 104

4.7.2 The auditory feedback control network.....107

4.7.3 Implications for the study of normal and disordered speech production 110
4.7.4 Implications for the DIVA model 114
4.7.5 Additional regions to be incorporated 116

5 Conclusions 120

Bibliography 126

Curriculum Vitae 153

LIST OF TABLES

Table 2-1. The location of DIVA cell components in MNI space. 28

Table 3-1. Cortical regions of interest and contributions to speech processes. 64

Table 3-2. Cortical ROIs outside core speech network..... 66

Table 3-3. Cerebellum regions of interest and their anatomical boundaries..... 69

Table 4-1. Peak voxel responses for baseline contrasts. 95

Table 4-2. Peak voxel and regional responses for the *shift* - *no shift* contrast. 97

Table 4-3. Effective connectivity determined by structural equation modeling. 104

LIST OF FIGURES

Figure 2-1. Brain regions involved in speech production. 11

Figure 2-2. Lateral cortical regions involved in speech production..... 14

Figure 2-3. Subcortical regions involved in speech production. 18

Figure 2-4. Regions of the cerebellum contributing to speech production. 20

Figure 2-5. The DIVA model of speech acquisition and production. 24

Figure 2-6. Neuroanatomical locations of the DIVA model components. 30

Figure 2-7. The DIVA model with neuroanatomical labels..... 32

Figure 2-8. Predicted and experimental BOLD responses..... 56

Figure 3-1. Cortical and cerebellar regions of interest for speech studies. 62

Figure 3-2. Cerebellum regions of interest. 70

Figure 4-1. Timeline of a single trial in the event-triggered sparse sampling protocol. 79

Figure 4-2. Structural equation model example. 85

Figure 4-3. First formant response to an induced shift in F1. 91

Figure 4-4. Responses for normal and shifted speech compared to baseline..... 94

Figure 4-5. BOLD responses in the *shift – no shift* contrast. 96

Figure 4-6. Selected regional responses from the *shift – no shift* contrast..... 99

Figure 4-7. Spatial profiles of peri-Sylvian ROI responses for the *shift – no shift* contrast. 100

Figure 4-8. Interhemispheric response differences for selected ROIs across conditions. 102

Figure 4-9. Schematic of the path diagram evaluated by structural equation modeling. 103

Figure 4-10. Updated DIVA model. 115

Figure 4-11. Neuranatomical locations of updated DIVA model. 117

LIST OF ABBREVIATIONS

aCGg	anterior cingulate gyrus
adPMC	anterior dorsal premotor cortex
adSTs	anterior dorsal superior temporal sulcus
Ag	angular gyrus
AGFI	adjusted goodness-of-fit index
aINS	anterior insula
alCB	anterior lateral cerebellum
amCB	anterior medial cerebellum
aMFg	anterior middle frontal gyrus
aMTg	anterior middle temporal gyrus
AOS	apraxia of speech
aSMg	anterior supramarginal gyrus
aSTg	anterior superior temporal gyrus
aTFg	anterior temporal fusiform gyrus
avSTs	anterior ventral superior temporal sulcus
BA	Brodmann's area
BOLD	blood oxygenation-level-dependent
CV	consonant vowel
CVC	consonant vowel consonant
DCN	deep cerebellar nuclei
DIVA	directions into velocities of articulators
dMC	dorsal primary motor cortex
DSP	digital signal processor
EEG	electroencephalography
EPI	echo planar imaging
F0	fundamental frequency
F1	first formant
F2	second formant
FDR	false discovery rate

FIR	finite impulse response
FMC	fronto-medial cortex
fMRI	functional magnetic resonance imaging
FO	frontal operculum
FOC	fronto-orbital cortex
FOV	field of view
FP	frontal pole
FWHM	full-width half maximum
GFI	goodness-of-fit index
Hg	heschl's gyrus
HRF	hemodynamic response function
ICBM	international consortium for brain mapping
IFo	inferior frontal gyrus, *pars opercularis*
IFt	inferior frontal gyrus, *pars triangularis*
iplCB	inferior posterior lateral cerebellum
ipmCB	inferior posterior medial cerebellum
Lg	lingual gyrus
LPC	linear predictive coding
mdPMC	middle dorsal premotor cortex
MEG	magnetoencephalography
MGN	medial geniculate nucleus
mMC	medial motor cortex
MNI	Montreal Neurological Institute
mPMC	medial premotor cortex
MRI	magnetic resonance imaging
MTO	middle temporal occipital gyrus
OC	occipital cortex
pCGg	posterior cingulate gyrus
pdPMC	posterior dorsal premotor cortex
pdSTs	posterior dorsal superior temporal sulcus
PET	positron emission tomography

PHg	parahippocampal gyrus
pINS	posterior insula
pMFg	posterior middle frontal gyrus
pMTg	posterior middle temporal gyrus
PO	parietal operculum
PP	planum polare
preSMA	pre-supplementary motor area
pSMg	posterior supramarginal gyrus
pSTg	posterior superior temporal gyrus
PT	planum temporale
pTFg	posterior temporal fusiform gyrus
pvSTs	posterior ventral superior temporal sulcus
PWS	persons who stutter
RMR	root mean square residual
RMSEA	root mean square error of approximation
ROI	region of interest
SCC	subcallosal cortex
SD	standard deviation
SE	standard error
SEM	structural equation modeling
SFg	superior frontral gyrus
slCB	superior lateral cerebellum
SMA	supplementary motor area
smCB	superior medial cerebellum
SPL	superior parietal lobule
splCB	superior posterior lateral cerebellum
SPM	statistical parametric mapping
spmCB	superior posterior medial cerebellum
SSM	speech sound map
TE	echo time
TOF	temporal occipital fusiform gyrus

TP	temporal pole
TR	repetition time
vMC	ventral primary motor cortex
vPMC	ventral premotor cortex
vSC	ventral somatosensory cortex

CHAPTER 1

INTRODUCTION

Speech production is marked by rapid, coordinated movements of multiple articulators. This is an impressive feat of motor control given the large number of muscles involved. Movement of the lips alone involves the coordination of at least thirteen different sets of muscles; tongue movements involve eight more (Zemlin, 1998). Muscles contributing to mandibular, velar, pharyngeal, and laryngeal control add many more to the total number of muscles involved in producing even the simplest monosyllable. Yet, fluent speakers meet these demands with relative ease, producing as many as 4-7 syllables per second (Tsao and Weismer, 1997). By two years of age, children learning American English have typically mastered the fine articulatory distinctions that differentiate the consonants /b/ and /d/, the fine timing control that differentiate /b/ and /p/. By grade school they've mastered nearly the full inventory of phonemes, which they can combine in any syllable structure to produce long, complex, intelligible utterances (McLeod and Bleile, 2003). The brain of a developing speaker quickly learns to orchestrate the interplay among the many muscles associated with speech production.

A vital role in achieving verbal fluency is played by auditory feedback. While many motor acts are aimed at achieving goals in three-dimensional space (e.g., reaching, grasping, throwing, walking, and handwriting), the primary goal of speech is a time-varying acoustic signal that transmits a linguistic message via the listener's auditory system. For spatial tasks, visual feedback of task performance plays an important role in

monitoring performance and improving skill level (Redding and Wallace, 2006; Huang and Shadmehr, 2007). Analogously, auditory information plays an important role in monitoring vocal output and achieving verbal fluency (Lane and Tranel, 1971, Cowie and Douglas-Cowie, 1983). Auditory feedback is crucial for on-line correction of speech production (Lane and Tranel, 1971, Xu et al., 2004, Purcell and Munhall, 2006b) and for the development and maintenance of stored motor plans (Cowie and Douglas-Cowie, 1983, Purcell and Munhall, 2006a, Villacorta et al., 2007).

The principal aim of the research described in this dissertation was to improve our understanding of the neural mechanisms underlying these processes. A functional magnetic resonance imaging (fMRI) experiment was designed to characterize the network of regions involved in auditory feedback control of speech production. The experiment tested *a priori* hypotheses derived from the DIVA (**D**irections **I**nto **V**elocities of **A**rticulators) model of speech acquisition and production (Guenther, 1995, Guenther et al., 1998, Guenther et al., 2006). To take advantage of the theoretical and computational framework provided by the DIVA model, the model's components were assigned to likely neuroanatomical substrates. This work permitted the generation of anatomically explicit simulated hemodynamic responses based on the model's cell activities, providing a powerful means for constraining the design and interpretation of functional imaging studies of speech. The investigation of the neural mechanisms underlying auditory feedback control of speech were further aided by combining functional imaging techniques optimized for studying speech production with neural and structural equation modeling.

1.1 Combining neural modeling and functional neuroimaging

Blood oxygenation level-dependent (BOLD) functional magnetic resonance imaging (fMRI; Belliveau et al., 1992, Kwong et al., 1992, Ogawa et al., 1993) has provided a powerful tool for the non-invasive study of human brain function. The development of BOLD fMRI has proven particularly beneficial to the study of speech given its uniquely human nature. The past decade and a half of imaging research has provided a tremendous amount of functional data regarding the brain regions involved in both speech perception and production (a recent *PubMed* search for "speech" and "fMRI" returned nearly 2000 entries since 1992). Over that time, some consensus regarding the brain regions involved in speech production has been built (e.g. Turkeltaub et al., 2002, Indefrey and Levelt, 2004). A clear picture of the specific functional contributions of each region or how those regions interact remains elusive, however.

Because DIVA is a neurobiologically plausible computational model, activity of the model's various components can be compared to brain activations in a straightforward manner. Thus, hypotheses derived from the model are well-suited for testing with neuroimaging methodologies. By associating the model's components with specific neuroanatomical substrates and simulating various speaking conditions, it is possible to generate "simulated neural responses" for those conditions that can then be directly compared to functional imaging data. A neuroanatomically explicit DIVA model, therefore, provides a means for designing and interpreting functional imaging studies from a unified theoretical framework, a powerful means of studying speech neural

processes. The assignment of the model's components to likely neural substrates is described in Chapter 2, *The functional neuroanatomy of speech production.* The chapter begins with a review of the brain regions involved in speech production and an overview of the DIVA model. This is followed by a detailed description of the model's components and the neural substrates hypothesized to support each. The chapter concludes with the description of a method for generating simulated fMRI data from model simulation results and an example of its application to a study of the brain regions involved in consonant-vowel (CV) syllable production.

The review of brain regions involved in speech production provided in Chapter 2 also served as a basis for the development of a parcellation system that defines anatomical regions of interest (ROIs) that cover the entire brain. The parcellation system is described in Chapter 3, *A parcellation system for functional imaging studies of speech*. A set of cortical, subcortical, and cerebellar regions of interest (ROIs) are described that reflect important anatomical distinctions for the study of speech. The characterization of functional responses within anatomically-defined ROIs (e.g., Nieto-Castanon et al., 2003) provides a more statistically sensitive means of assessing neural responses in functional imaging studies of speech. Together, the efforts described in Chapters 2 and 3 allow researchers to *i*) make explicit, testable predictions regarding the location of BOLD responses to experimental manipulation, greatly constraining the interpretation of functional imaging results, and *ii*) supplement traditional voxel-based functional analyses with more statistically powerful ROI-based techniques.

1.2 The influence of auditory feedback on speech production

The control of movement is often characterized as involving one or both of two broad classes of control. Under *feedback control,* task performance is monitored during execution, and deviations from the desired performance are corrected according to sensory information. Under *feedforward control,* task performance is executed from previously learned commands, without reliance on incoming task-related sensory information. Speech production involves both feedforward and feedback control, and auditory feedback has been shown to impact both control processes (Houde and Jordan, 1998, Jones and Munhall, 2005, Bauer et al., 2006, Purcell and Munhall, 2006a).

Early evidence of the influence of auditory feedback on speech came from studies showing that speakers modify the intensity of their speech in noisy environments (Lombard, 1911). Artificial disruption of normal auditory feedback in the form of temporally delayed feedback induces disfluent speech (Yates, 1963, Stuart et al., 2002). Recent studies have used transient, unexpected auditory feedback perturbations to demonstrate auditory feedback control of speech. Despite being unable to anticipate the perturbation, speakers respond to pitch (Larson et al., 2000, Donath et al., 2002, Jones and Munhall, 2002, Natke et al., 2003, Xu et al., 2004) and formant shifts (Purcell and Munhall, 2006b) by altering their vocal output in the direction opposite the shift. These compensatory responses act to steer vocal output closer to the intended auditory target.

The ease with which fluent speakers are able to coordinate the rapid movements of multiple articulators suggests that speech is also guided by a feedforward controller

(Neilson and Neilson, 1987). Our ability to speak effectively when noise completely masks auditory feedback (Lane and Tranel, 1971, Pittman and Wiley, 2001) and the long-term maintenance of intelligibility of post-lingually deafened individuals (Cowie and Douglas-Cowie, 1983, Lane and Webster, 1991) provide further evidence of feedforward control mechanisms. The existence of stored feedforward motor commands that are tuned over time by auditory feedback is confirmed by studies of sensorimotor adaptation (Houde and Jordan, 2002, Jones and Munhall, 2002, 2005, Purcell and Munhall, 2006a, Villacorta et al., 2007). Speakers presented with auditory feedback containing a persistent shift of the formant frequencies (which constitute important cues for speech perception) of their own speech, will adapt to the perturbation by changing the formants of their speech in the direction opposite to the shift. Following adaptation, utterances made immediately after removal or masking of the perturbation also typically contain formants that differ from baseline formants in the direction opposite to the induced perturbation (e.g., Purcell and Munhall, 2006a). These temporarily maintained compensations that follow adaptation are indicative of a transient reorganization of the sensory-motor neural mappings that underlie feedforward control in speech (e.g., Purcell and Munhall, 2006a). The same studies also illustrate that the feedforward speech controller continuously monitors auditory feedback and is modified when that feedback does not meet expectations.

The DIVA model provides a mechanistic account of these sensorimotor adaptation results. The model predicts that auditory or somatosensory errors will be corrected via its feedback-based control mechanism, and that these corrections will

eventually become coded into the feedforward controller if the errors are consistently encountered. In Chapter 4, *An investigation of the neural substrates of auditory feedback control of speech*, an fMRI experiment is described in which these predictions are explored. The study was designed to reveal the network of brain regions involved in auditory feedback control of speech and the interactions within that network.

CHAPTER 2

THE FUNCTIONAL NEUROANATOMY OF SPEECH PRODUCTION

2.1 Overview of the brain regions involved in speech production

With the proliferation of imaging studies of speech production, a consensus is building regarding the brain areas underlying speech motor control. Bohland and Guenther (2006) have described a "minimal network" of brain regions involved in speech production. The network was based on the conjunction of BOLD activity while speakers read aloud syllable sequences of varying complexity compared to a baseline letter viewing task (e.g., 'ta-ta-ta', 'ka-ru-ti', 'stri-stri-stri', 'kla-stri-splu'). Their analysis demonstrated the recruitment of a large bilateral network of cortical and subcortical regions that included portions of the medial and lateral frontal lobe, parietal lobe, temporal lobe, the thalamus, basal ganglia, and the cerebellum during speech production. It is perhaps unsurprising that the active regions commonly associated with speech production include those known to contribute to the planning and execution of movements (primary sensorimotor and premotor cortex, the supplementary motor area, the cerebellum, thalamus, and basal ganglia) and those associated with acoustic and phonological processing of speech sounds (the superior temporal gyrus).

The "minimal speech network" identified by Bohland & Guenther (2006), and described by several others (e.g., Fiez and Petersen, 1998, Riecker et al., 2000b, Turkeltaub et al., 2002, Ghosh et al., 2003, Indefrey and Levelt, 2004, Soros et al., 2006)

is illustrated in the brain images shown in Figures 2-1 through 2-4. Here and throughout this report, the term "activation" is used to describe a significant finding for some comparison of BOLD response across cognitive tasks. The comparison of BOLD responses typically takes the form of a simple subtraction of the estimated response associated with one test condition (e.g., a baseline condition such as lying quietly) in a particular brain region from the effect size associated with another test condition (e.g., speech production) in the same region. Such a comparison is referred to as a *cognitive contrast*. An *active* brain region, therefore, is one that demonstrates a BOLD response that is significantly greater than zero for a given cognitive contrast. The BOLD responses are typically compared on a voxel-by-voxel basis, in which case the term "region" refers to the location of an active voxel or cluster of voxels[1]. Identifying the location of active regions is made easy by co-registering a parametric map of the contrast statistics (e.g., a volume of *t* statistics) and an anatomical volume such that those voxels that survive a significance threshold are clearly indicated by some means, typically by a color map indicating the magnitude of the test statistic[2]. A general term for this method of analyzing and interpreting functional imaging data is "statistical parametric mapping." This

---

[1] Less commonly, responses from voxels within some prescribed area of the brain are pooled and then compared across test conditions. In that case the term "region" is used in the more traditional sense to describe an area of the brain defined by generally well-accepted anatomical landmarks, e.g., the 'supramarginal gyrus,' 'motor cortex,' 'basal forebrain,' etc. The advantages of the latter approach are discussed in Chapter 3.

[2] Plotting the supra-threshold test statistic demonstrates those voxels that are significant for a given contrast and the relative level of the statistical significance of the contrast effect size. Replacing the test statistic with the effect size for supra-threshold voxels, the method used for some plots shown in Chapter 4, provides a more direct illustration the magnitude of activation differences across the conditions of a contrast.

approach, which is discussed in more detail below, was used to reveal the brain areas involved in speech motor control (Figure 2-1).

The activations shown in Figure 2-1 (indicated by the red – yellow color gradient) and illustrated in greater anatomical detail in Figures 2-2 through 2-4 were derived from the pooled analysis of BOLD responses during three simple speech production tasks contrasted with a silent letter viewing baseline task (*speech – baseline*). The three speech tasks included in the analysis were production of monosyllable nonsense words (11 subjects), bi-syllable nonsense words (13 subjects), and monosyllable American English words (10 subjects) for a total of 34 subjects. The *speech – baseline* contrast was first determined from each subject's functional data; for every voxel the estimated response during the baseline condition was subtracted from the response during the speech production task, providing a *contrast volume* for each subject. Each voxel in the contrast volume represents the *effect size* associated with the *speech – baseline* contrast at that location in the functional data. The statistical significance of these differences was determined by performing a two-tailed *t*-test at each voxel ($>10^6$ tests) on the pooled contrast volumes (n = 34 in each test), i.e., the contrast volume for each subject was treated as a random effect. Because the subject data were pooled across several studies, the contrast effects for each subject were mean-normalized prior to statistical tests. The probabilities associated with the resulting volume of *t*-statistics were corrected to ensure a false discovery rate (FDR) below 5%. This thresholded statistical parametric map was co-registered with the canonical brain distributed with the SPM2 imaging analysis software

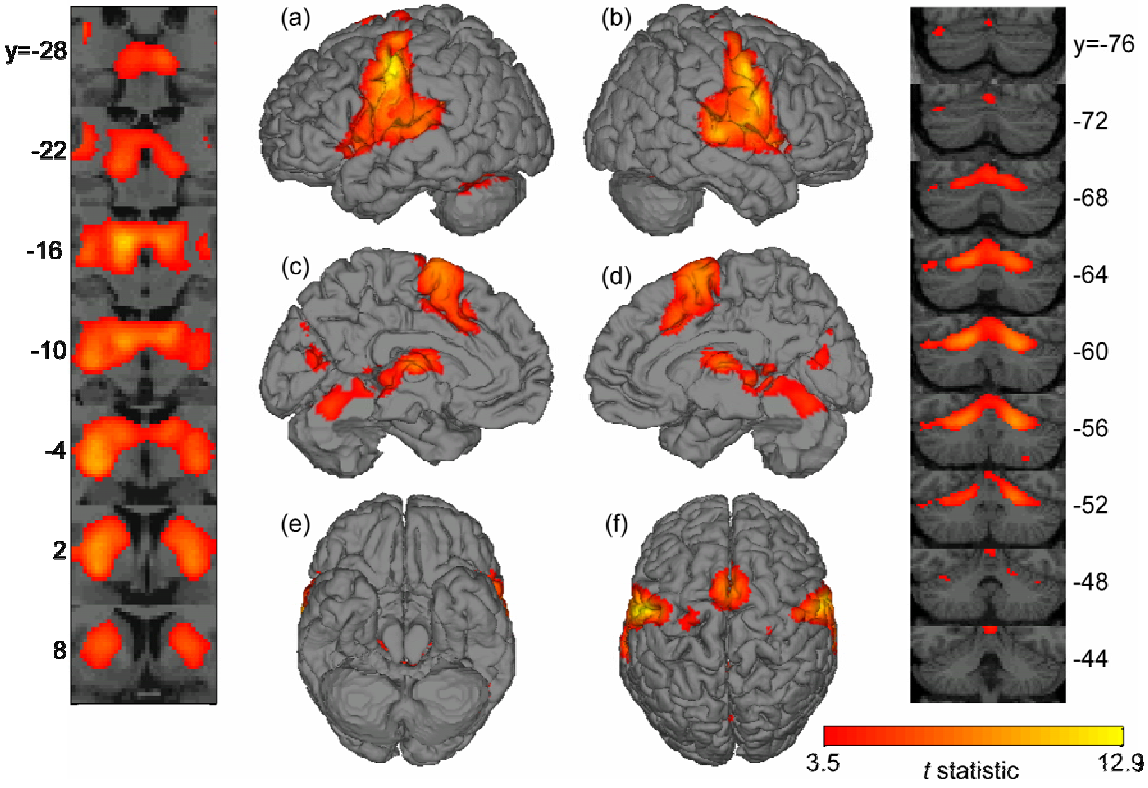


Figure 2-1. Brain regions involved in speech production.
Thresholded statistical parametric maps (FDR < 5%, $t > 3.48$) were created by contrasting simple speech production tasks with a silent baseline then superimposed on a representative brain to reveal the brain regions involved in speech production. The statistical map was derived from a voxel-wise mixed effects analysis of three speech production tasks vs. silence. Contrasts were calculated first for each individual subject. *t*-statistics for each voxel were then determined for each voxel by treating each subject contrast as a random effect. Those areas shown in color exhibited significantly greater BOLD responses during the speech task. *Far left*: A series of coronal slices through the central region of the brain reveals activation in the thalamus and basal ganglia. Numbers to the left of the images denote the anterior-posterior level (*y* coordinate) of the image in Montreal Neurological Institute (MNI; Mazziotta et al., 2001) stereotactic space (higher *y* values are more anterior). Figure 2-3 provides a more anatomically detailed illustration of the activity in these regions. *Center*: Activity is superimposed on renderings of the (a) left hemisphere lateral, (b) right lateral, (c) left medial, (d) right medial, (e) ventral, and (f) dorsal surfaces of the brain. The cortical regions associated with speech production include lateral sensorimotor cortex, medial and ventrolateral prefrontal cortex, peri-Sylvian cortex, and superior temporal cortex. A more detailed illustration of the location of these regions on the cerebral surface is provided in Figure 2-2. *Far right*: A series of coronal slices through the cerebellum reveals widespread bilateral activation in superior cerebellar cortex and a small cluster of activation in right inferior cerebellar cortex. Figure 2-3 provides a more anatomically detailed illustration of the activity in these regions.

package provided by the Wellcome Department of Imaging Neuroscience, University College London (Friston et al., 1995b; http://www.fil.ion.ucl.ac.uk/spm/). The co-registered supra-threshold *speech – baseline* statistical data is shown in Figure 2-1 superimposed on the SPM2 canonical brain.

Figure 2-1 demonstrates the cortical, subcortical, and cerebellar regions that comprise the "minimal speech network" as described by Bohland & Guenther (2006). These regions and their ostensible functional contributions to speech production are discussed briefly below. A more detailed treatment of the putative role each plays in speech motor control is provided in Section 2.3 *The neuroanatomy of the DIVA model*. The three figures that follow Figure 2-1 provide a more anatomically detailed illustration of the brain areas active during speech production. Those figures serve as anchors for the following discussion.

Lateral surface of the cerebral cortex. On the lateral surface of activation can be found along the lateral pre- and postcentral gyri, or Rolandic cortex, that straddle the central sulcus (Figure 2-2). This area contains the sensorimotor representations for orofacial muscles (Lotze et al., 2000a, Boling et al., 2002, Fesl et al., 2003). The primary motor representation, or Brodmann's area (BA) 4, lies anterior to the precentral sulcus; cells of the primary motor cortex project to the brainstem cranial nuclei that, in turn, project to motoneurons of the speech articulators. Electrical stimulation of these cells reliably disrupts speech production (Ojemann, 1991). More anteriorly on the lateral precentral gyrus is the premotor motor cortex (BA 6) representation of the speech articulators. BA 6 continues anteriorly to the precentral sulcus. Posterior to the central sulcus, the primary

somatosensory representations of the speech articulators are located along the lateral postcentral gyrus (BA 3, 1, and 2). Tactile and proprioceptive information from the articulators is relayed to somatosensory cortex via brainstem cranial nerve nuclei.

Anterior to the precentral gyrus, activation spreads into the ventral, posterior-most regions of the inferior frontal gyrus including BA 44 (the posterior portion of Broca's area in the left hemisphere), the frontal operculum, and adjacent anterior insula. Damage to this region in the left hemisphere has long been associated with the disruption of fluent speech (see, e.g., Duffy, 1995, Dronkers, 1996, Kent and Tjaden, 1997, Hillis et al., 2004) and has been described by Indefrey and Levelt (2004) as an interface between phonological encoding and articulation. Strong interconnectivity between ventral inferior frontal cortex and anatomically adjacent premotor and motor cortices has been demonstrated in monkeys (Wang et al., 2002, Fang et al., 2005, Wang et al., 2005). Functional connections between and within the same areas has been shown in humans (Greenlee et al., 2004, Greenlee et al., 2007).

The lateral surface (Figure 2-2) also shows activation along the superior temporal gyrus, the location of primary auditory and auditory association cortex (Rivier and Clarke, 1997, Scott and Johnsrude, 2003, Morosan et al., 2005). Auditory association cortex has been implicated in phonological processing. Activation of more anterior and ventral portions of the superior temporal gyrus has been correlated with increasing intelligibility of speech sounds while more posterior areas of the superior temporal gyrus are associated with phonetic processing (Scott et al., 2000, Giraud and Price, 2001).

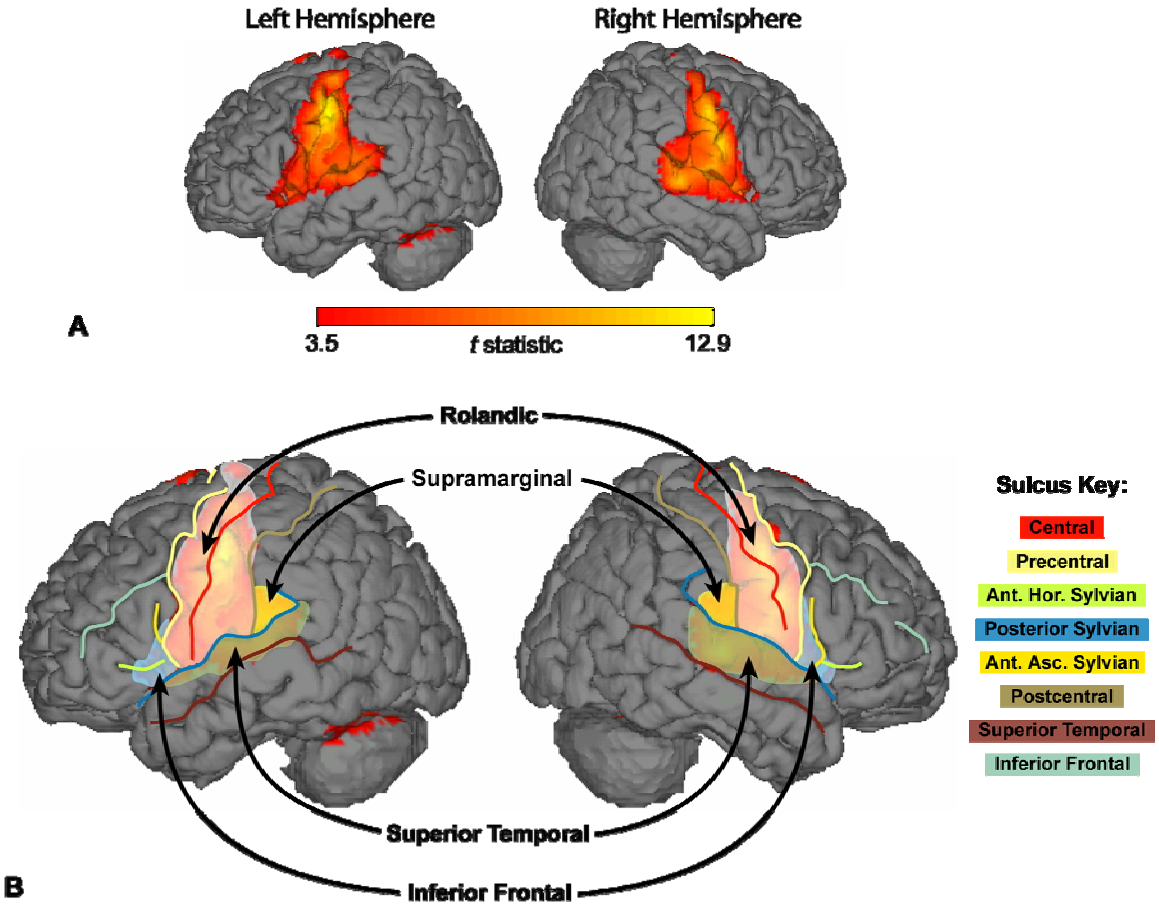


Figure 2-2. Lateral cortical regions involved in speech production.
(A). Activity on the lateral cortical surface resulting from the pooled analysis of simple speech production tasks is shown (cf., surface images (*a*) and (*b*) in Figure 2-1). Details of the analysis and plotting procedure are provided in the caption of Figure 2-1. (B). The locations of core lateral cortical regions involved in speech production and discussed throughout this report are illustrated on a blowup of the image shown in (A). Semi-transparent color overlays demark activation in Rolandic cortex (white), inferior frontal cortex (blue), anterior supramarginal cortex (yellow), and superior temporal cortex (green). Solid lines denote the major sulci of the lateral surface and are identified in the color key at the far right. The Rolandic cortex is divided by the central sulcus: the precentral gyrus is located anterior to the central sulcus and is bounded anteriorly by the precentral sulcus; the postcentral gyrus is located posterior to the central sulcus and is bounded posteriorly by the postcentral sulcus. The motor (BA 4) and premotor (BA 6) articulator representations are located on the lateral portion of the precentral gyrus and the somatosensory (BA 1, 2, 3) representations are located on the adjacent postcentral gyrus. The inferior frontal region, is divided by the anterior ascending (Ant. Asc.) and anterior horizontal (Ant. Hor.) branches of the Sylvian fissure. The area between the precentral sulcus and anterior ascending Sylvian corresponds to BA 44. BA 45 is located anterior to BA 44, between the two anterior branches of the Sylvian. Both are bounded medially by the inferior frontal sulcus. Simple speech production tasks result in activation in the ventral portions of BA 44 and 45 in both hemispheres, though activation is stronger in the left (the region commonly referred to as Broca's area). Posterior to the postcentral gyrus, the higher-level somatosensory representations of the articulators are found on the anterior supramarginal gyrus (BA 40). Primary and secondary auditory cortex is located on the superior temporal gyrus. The gyrus is bounded laterally by the superior temporal sulcus and medially by the posterior branch of the Sylvian fissure. The unlabeled activation in the left hemisphere is located in cerebellar cortex; activity in this region is illustrated in greater detail in Figure 2-4.

Although much of the activity in auditory cortical areas during speech can be attributed to hearing one's own voice while speaking, it has been demonstrated with fMRI that even covert (silent) speech activates auditory cortical areas (Hickok et al., 2000, Okada et al., 2003, Okada and Hickok, 2006), a finding that corroborates the MEG study of Numminen and Curio (1999) demonstrating modulation of auditory cortical activation during covert as well as overt speech. A region within the ventral bank of the posterior Sylvian fissure, the planum temporale (BA 42, medial portion of BA 22), is active during both speech perception and speech production tasks and has been proposed as a sensorimotor interface for speech (Hickok and Poeppel, 2007).

The activation in the posterior peri-Sylvian area extends to the inferior parietal cortex including the parietal operculum and anterior supramarginal gyrus, regions implicated in a number of processes related to speech production including verbal working memory (e.g., Jonides et al., 1998, Becker et al., 1999), phonetic discrimination (Caplan et al., 1995) and an interface between orthographical, phonological and lexical/semantic decision-making (Pugh et al., 2001). Damage to this area results in a number of speech production and perception deficits including Wernicke's aphasia and conduction aphasia (Damasio and Damasio, 1980, Goodglass, 1993). Reciprocal functional connections between inferior parietal cortex and inferior frontal cortex have been indicated in humans using electrical stimulation (Jurgens, 1984, Luppino et al., 1993, Lehericy et al., 2004, Matsumoto et al., 2004) providing evidence for a link between the lateral frontal and posterior regions associated with speech production. Such a pathway has been established in monkeys; the superior longitudinal fasciculus

(subdivision III) carries reciprocal projections between the monkey homologue of BA 45 (anterior Broca's area in humans) and surrounding cortex of the lateral frontal lobe and the inferior parietal and posterior superior temporal cortex (Schmahmann and Pandya, 2006).

Medial prefrontal cortex. Speech production also results in activation of the medial surface of the frontal lobe (Figure 2-1, middle row) including the supplementary motor area (SMA), pre-supplementary motor area (preSMA), and cingulate motor area. Studies in monkeys and humans have shown the SMA to be strongly interconnected with lateral motor and premotor cortex and portions of the basal ganglia (Jurgens, 1984, Luppino et al., 1993, Lehericy et al., 2004, Matsumoto et al., 2004, Matsumoto et al., 2007). Microstimulation of the SMA yields vocalization, word or syllable repetitions, and/or speech arrest (Penfield and Welch, 1951). Bilateral damage to these areas results in speech production deficits including transcortical motor aphasia (Jonas, 1981, Ziegler et al., 1997) and akinetic mutism (Nemeth et al., 1988, Adams, 1989, Mochizuki and Saito, 1990). The available lesion and imaging data both from studies of speech and other modalities have led a number of investigators to conclude that the SMA plays a critical role in the initiation of speech sounds during speech production (e.g., Jonas, 1987, Ziegler et al., 1997, Alario et al., 2006, Bohland and Guenther, 2006).

The thalamus and basal ganglia. The series of coronal slices shown in Figure 2-3 demonstrates activation during speech production in the basal ganglia (which include the putamen, globus pallidus, caudate, substantia nigra, and subthalamic nucleus) and the thalamus. These regions are heavily interconnected with the frontal cortex via multiple

cortico-basal ganglia-thalamo-cortical loops (Alexander et al., 1986, Alexander and Crutcher, 1990, Middleton and Strick, 2000). The architecture of the basal ganglia make them suitable for selectively enabling one output from a set of competing alternatives (Mink and Thach, 1993, Mink, 1996, Kropotov and Etlinger, 1999), a property evident in several computational models of basal ganglia function (e.g., Redgrave et al., 1999, Brown et al., 2004, Prescott et al., 2006). Damage to the basal ganglia results in disorders (e.g., Parkinson's disease, Tourette's syndrome) characterized by disruption of voluntary movement control (Albin et al., 1995). Basal ganglia lesions have been associated with several disturbances in speech including various forms of aphasia (Crosson, 1992) and articulatory sequencing/syllable switching deficits (Pickett et al., 1998).

Electrical stimulation studies have also suggested the involvement of the thalamus in the motor control of speech. A series of studies conduction by Ojemann and colleagues have assessed the effect of thalamic stimulation while subjects named pictured objects using the phrase "this is a __" (for a review, see Crosson, 1992, Johnson and Ojemann, 2000). Stimulation of the ventrolateral thalamus was found to disrupt respiratory patterns and resulted in slowed, slurred speech. Perservarative and omissions errors were noted during stimulation of the ventrolateral thalamus and the putamen as were naming errors (Johnson and Ojemann, 2000). These disruptive effects were observed more often, if not exclusively, with stimulation within the dominant hemisphere for language (e.g., the left hemisphere in right-handed persons). Stimulation of the ventral anterior nuclei, on the other hand, can elicit speech that subjects are unable to inhibit (Schaltenbrand, 1965, 1975). The observed compulsory speech included recitations of whole words and phrases

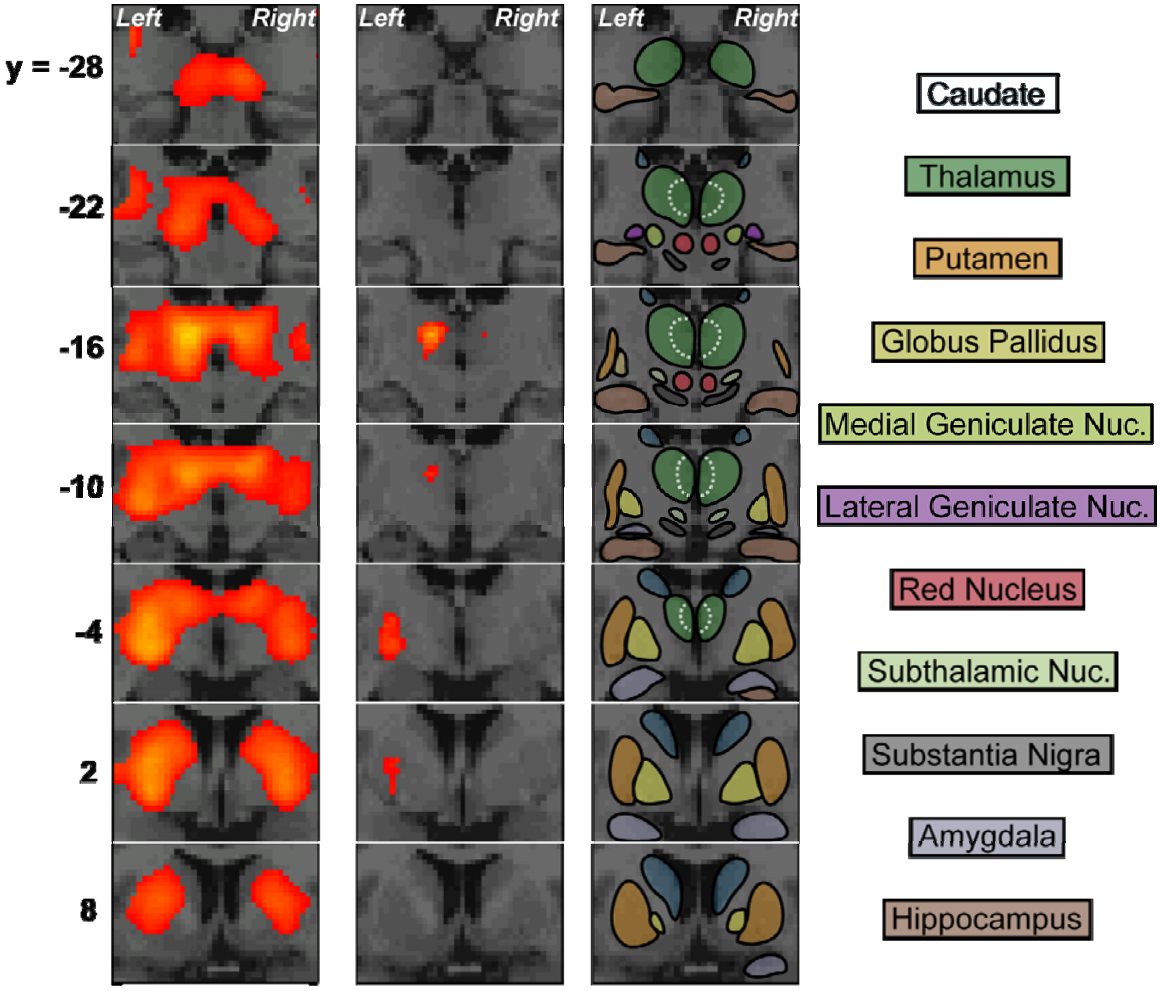


Figure 2-3. Subcortical regions involved in speech production.
Coronal slices revealing the thalamus, basal ganglia, and mesencephalon are shown. Numbers indicate the MNI *y* coordinate associated with the adjacent coronal image (a higher *y* coordinate indicates a more anterior slice). In each image, the left hemisphere is shown on the left, the right hemisphere on the right as indicated by the labels *Left* and *Right* found in the top image of each column. *Left*: Subcortical activation following the same threshold used in Figure 2-1. *Middle*: The location of peak subcortical responses is made more easily visible by implementing a more restrictive significance threshold. *Right:* Subcortical regions are outlined and color-coded according to the key at the far right. The dotted white lines within the thalamus approximate the boundary between the medial and lateral thalamic nuclei.

and again, was evoked more often from thalamic stimulation in the dominant hemisphere. Schaltenbrand (1965) noted that such elicitation did not result from cortical stimulation, a claim that was more recently reviewed and supported by Crosson (1992). However, an additional subcortical region, the head of the dominant caudate nucleus, has been shown

to evoke word production (Van Buren, 1963). Crosson (1992) notes similarities between the results of stimulation in the caudate nucleus and the anterior thalamic nuclei, both components of a higher-level cortico-basal ganglia-thalamo-cortical loop with the prefrontal cortex. The findings suggest that these areas serve similar functions, and may be involved in the release of a speech / language plan (cf. Schaltenbrand, 1975).

Although current neuroimaging studies have thus far paid relatively little attention to basal ganglia activity during speech, a few noteworthy observations have been made. Riecker et al. (2006) note a decrease in striatal activity with increased speaking rate, a pattern opposite that of the cortex and cerebellum. Bohland and Guenther (2006) note that overt production (compared to preparation alone) led to an increase in activity in the putamen bilaterally. This coincided with additional motor cortical activation and is likely part of the motor executive loop described above. Furthermore, additional sequence complexity led to increased activation in the anterior thalamus, caudate nucleus, and portions of prefrontal cortex, likely due to the increased planning load for these sequences.

The cerebellum. Figure 2-4 shows bilateral activation along the superior portion of the cerebellar cortex. The activation straddles the primary cerebellar fissure, falling in paravermal lobules V and VI, spreads laterally to lateral lobules VI and VII and antero-medially to vermal III and IV of the anterior lobe. Like the basal ganglia, the cerebellum is heavily interconnected with the cerebral cortex via the thalamus, in this case via a cortico-pontine-cerebello-thalamo-cortical loop. Multiple closed-loop circuits involving

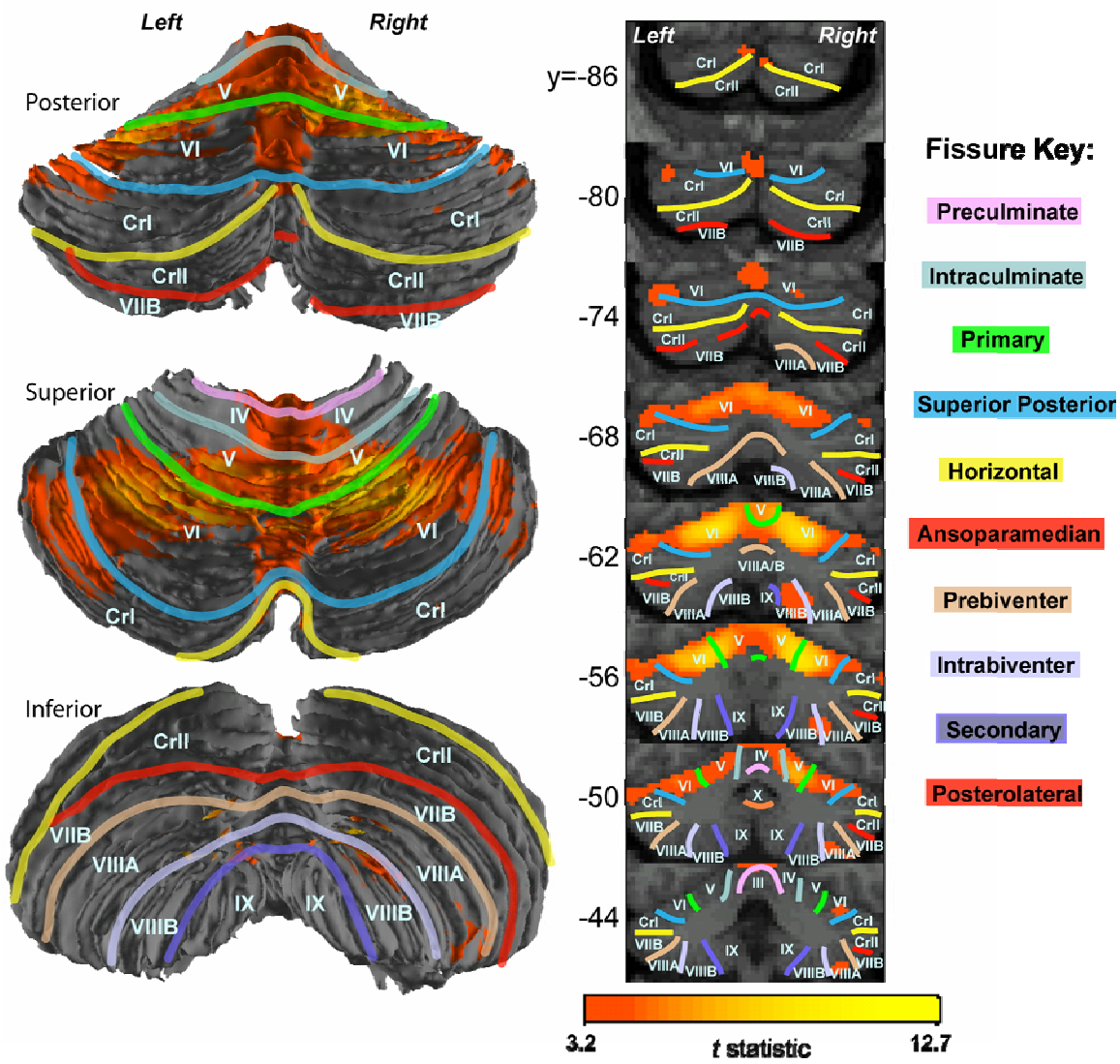


Figure 2-4. Regions of the cerebellum contributing to speech production.
Left: Activation during simple speech production tasks are shown on renderings of the posterior (top), superior (middle), and inferior (bottom) surfaces of the cerebellum (refer to the text or Figure 2-1 for a description of the analyses). The visible cerebellar lobules are labeled and the major fissures that form the lobule boundaries are outlined. *Middle*: A series of coronal slices through the cerebellum. Slices toward the top of the figure are more posterior. Activation during speech production is located predominantly in the superior cerebellar cortex (dorsal to the horizontal fissure) bilaterally. In each image, the left hemisphere is shown on the left, the right hemisphere on the right as indicated by the labels *Left* and *Right.* Activation peaks are located in paravermal lobules V and VI but extends laterally along lobule VI to Crus I and antero-medially to vermal IV, V, and VI. Activation in inferior cerebellar cortex is also noted within paravermal lobules VIIIA and VIIIB. *Right*: Color key for the major cerebellar fissures traced in the brain images.

the cerebellum and primary motor, premotor, and prefrontal (BA 46) cortex have been identified in non-human primates (Middleton and Strick, 1997, Schmahmann and Pandya, 1997, Middleton and Strick, 2001, Kelly and Strick, 2003, Voogd, 2003). Substantial cerebellar inputs to parietal cortex have also been identified in monkeys (Clower et al., 2001, Clower et al., 2005)

Speech deficits resulting from cerebellar stroke usually occur with damage to the superior cerebellar artery (Ackermann et al., 1992). This type of infarct can lead to ataxic dysarthria, a motor disorder that results in inaccurate articulation, prosodic excess, and phonatory-prosodic insufficiency (Darley et al., 1975). Lesion studies implicate both the anterior vermal region (Urban et al., 2003) as well as the more lateral paravermal region (Ackermann et al., 1992) in ataxic dysarthria. Cerebellar damage additionally results in increased duration of sentences, words, syllables, and phonemes (Ackermann and Hertrich, 1994, Kent and Tjaden, 1997) and deficits in short-term verbal rehearsal and planning for speech production (Silveri et al., 1998).

The activation locus in the inferior cerebellar cortex (in/near lobule VIII, particularly in the right hemisphere) of the right hemisphere has been consistently noted for studies of speech production. However, Bohland and Guenther (2006) reported increased activity in this region in response to increased complexity of syllable sequences and attributed it to working memory processes, a view supported by the working memory studies of Desmond and colleagues (Desmond et al., 1997, Chen et al., 2005).

The network of regions involved in speech processing described above is representative of the majority of reports from imaging studies that have compared overt

syllable production tasks to a suitable baseline (e.g., Lotze et al., 2000b, Riecker et al., 2000b, Bohland and Guenther, 2006, Soros et al., 2006). A tremendous number of imaging studies have investigated neural responses during word production tasks of one form or another (e.g., word repetition, picture naming, sentence reading, etc.). Meta-analyses of these findings again reveal a network of active regions very similar to that illustrated in Figures 2-1, 2-2, 2-3, and 2-4 (Fiez and Petersen, 1998, Turkeltaub et al., 2002, Indefrey and Levelt, 2004). With the addition of lexical and linguistic load to the speech production tasks, the "core" network remains but is supplemented by additional regions (discussed further below). Yet, essentially the same network responds during production of even a single vowel (e.g., Ghosh et al., 2003, Soros et al., 2006). In the section that follows, the components of the DIVA model are described in detail and assigned to various regions of the "minimal speech network" described above.

2.2 Overview of the DIVA model

The DIVA model is an adaptive neural network (Guenther et al., 1998, Guenther et al., 2006) that provides a quantitatively explicit description of the sensorimotor interactions involved in vocal articulator control. The model, schematized in Figure 2-5, consists of integrated feedforward and feedback control subsystems that learn to control a simulated vocal tract (a modified version of the synthesizer described by Maeda, 1990). Once trained, the model takes a speech sound as input, and generates a time varying sequence of articulator positions that command movements of the simulated vocal tract that produce the desired sound.

Each block in Figure 2-5 corresponds to a set of neurons that constitute a *neural representation*. When describing the model, the term *map* is used to refer to such a set of cells. The term *mapping* is used to refer to a transformation from one neural representation to another. These transformations are represented by arrows in Figure 2-5 and are assumed to be carried out by filtering cell activations in one map through synapses projecting to another map. The synaptic weights are tuned during a babbling phase meant to coarsely represent that typically experienced by a normally developing infant (e.g., Oller and Eilers, 1988). Random movements of the speech articulators provide tactile, proprioceptive, and auditory feedback signals that are used to learn the mappings between the different neural representations. After babbling[3], the model can quickly learn to produce new sounds from audio samples provided to it and is able to produce arbitrary combinations of the sounds it has learned.

[3] Though the modeled babbling stage is meant to functionally mimic that experienced during early childhood development, the time scales associated with the model and natural processes differ. The various stages of infant babbling, which begin approximately 2 months after birth, persist through much of the first year of life (Oller, 1980, Stark, 1980). The DIVA model babbling stage is complete after several iterative cycles through the model production steps. The DIVA model also does not have deal with changes in the shape of vocal tract, unlike a growing infant, however, Callan et al. (2000) showed that the model is capable of keeping up with such growth.

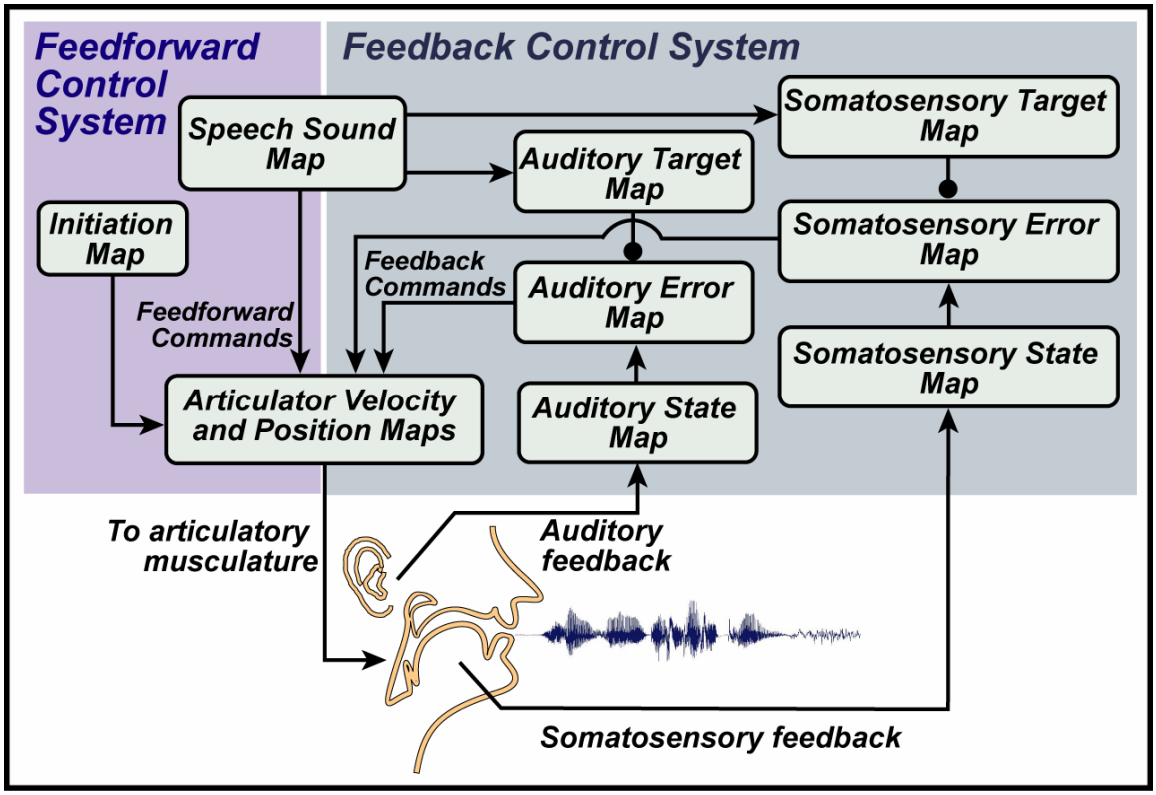


Figure 2-5. The DIVA model of speech acquisition and production.

Production of a phoneme or syllable begins in the model with the activation of a *Speech Sound Map* cell (see Figure 2-5) corresponding to the sound to be produced. Projections from the *Speech Sound Map* to the model's *Articulator Velocity Map* form a feedforward motor command that drives the simulated vocal tract. Additional projections from *Speech Sound Map* cells to *Auditory* and *Somatosensory Target Maps* encode the sensory expectations associated with the speech sound to be produced. Projections from the *Speech Sound Map* to the *Auditory Target Map* can be tuned while listening to phonemes and syllables from the native language or by listening to correct self-productions. After learning, these synapses encode a spatiotemporal target region for the sound in auditory coordinates. Likewise, projections from the *Speech Sound Map* to the

Somatosensory Target Map encode the expected somatic sensations corresponding to the active *Speech Sound Map* cell. This spatiotemporal somatosensory target region is estimated by monitoring the somatosensory consequences associated with production of a speech sound over many successful production attempts. During production of a sound, the sensory target regions are compared to the current sensory states; any discrepancy between the target and the current state, or sensory error, will result in the generation of a corrective motor command via projections from the sensory error maps to the articulator velocity map. These projections, which form the feedback motor command, are tuned during babbling by monitoring the relationship between sensory signals and the motor commands that generated them.

Feedforward and feedback motor commands are combined in the model's *Motor Position Map*; the combined motor command then drives the simulated vocal tract, resulting in production of the desired speech sound. The passing of motor commands to the vocal tract is gated by an *Initiation Map* that initiates production of each sound at the proper instant. The relative weight of feedforward and feedback commands in the overall motor command is dependent upon the size of the error signal associated with the activation of the *Speech Sound Map* cell. Initially, feedforward projections are poorly tuned, resulting in large error signals and greater reliance on the feedback command. The feedforward command is learned over time by averaging the overall motor commands from previous attempts to produce the sound. Eventually, the feedforward projections are able to drive production of the desired speech sound with minimal sensory error during normal speaking conditions and, thus, little reliance on feedback motor commands.

The DIVA model provides a unified explanation of a number of speech production phenomena including motor equivalence, contextual variability, anticipatory and carryover coarticulation, velocity/distance relationships, speaking rate effects, and speaking skill acquisition and retention throughout development (e.g., Guenther, 1994, Guenther, 1995, Guenther et al., 1998, Callan et al., 2000, Nieto-Castanon et al., 2005, Guenther et al., 2006). Because it can account for such a wide array of data, the DIVA model has provided the theoretical framework for a number of investigations of normal and disordered speech production. Predictions from the model have guided studies of the role of auditory feedback in normally hearing persons, deaf persons, and persons who have recently regained some hearing through the use of cochlear implants (Perkell et al., 2000, Perkell et al., 2004a, Perkell et al., 2004b, Lane et al., 2007, Perkell et al., 2007). The model has also been employed in investigations of the etiology of stuttering (Max et al., 2004) and apraxia of speech (Robin et al., 2008, Terband et al., 2008).

The following section describes work undertaken to extend the unified framework provided by the DIVA model to studies of the neural mechanism underlying speech processes. The model's components were assigned to likely neuroanatomical regions, thereby providing a means of generating "simulated neural responses" that can be used to guide the design and interpretation of functional imaging experiments. The discussion of this work is accompanied by a more detailed description of the DIVA model's components and the findings that guided their assignments to neural substrates.

2.3 The neuroanatomy of the DIVA model

This section describes the components of the DIVA model in detail and their hypothesized neuroanatomical locations. Model components were mapped to locations in the Montreal Neurological Institute (MNI; Mazziotta et al., 2001) standard reference frame based on relevant neuroanatomical and neurophysiological studies. While the majority of the studies described herein specifically addressed speech processes, investigations of other modalities (e.g., non-orofacial motor control) also proved informative. The MNI coordinates for each of the model's components are given in Table 2-1. The proposed locations were chosen to coincide with landmarks mapped to the representative canonical brain provided with the SPM image analysis software package (Friston et al., 1995b; http://www.fil.ion.ucl.ac.uk/spm/). Plotting functional activation on the SPM canonical brain is a popular means for disseminating of neuroimaging data. Mapping the model's components onto this brain (see Figure 2-6), then, provides a convenient way to compare simulated neural responses to the results of neuroimaging experiments. Unless otherwise noted, each cell type is represented in both hemispheres. There are currently no functional differences between the left and right hemisphere versions of a particular cell type in the model. However, future versions of the model will incorporate hemispheric differences in cortical processing as indicated by experimental findings (e.g., Zatorre et al., 1992, Tallal et al., 1993, Zatorre et al., 2002, Poeppel, 2003). Based on the hypothesized neural substrates for the DIVA model components, the schematic shown in Figure 2-5 has been updated and is shown below in Figure 2-7.

Model Components	Left Hemisphere			Right Hemisphere		
	x	y	z	x	y	z
Speech Sound Map						
Left vPMC/pIFg	-56	10	2			
Initiation Map						
SMA	0	0	68	2	4	62
Caudate	-12	-2	14	14	-2	14
Putamen	-26	-2	4	30	-14	4
Globus Pallidus	-24	-2	-4	24	2	-2
Thalamus (VL)	-10	-14	8	10	-14	8
Articulator Velocity and *Position Maps*						
Tongue						
1	-60.2	2.1	27.5	62.9	2.5	28.9
2	-60.2	3.0	23.3	66.7	2.5	24.9
3	-60.2	4.4	19.4	64.2	3	22
Lip						
Upper	-53.9	-3.6	47.2	59.6	-7.2	42.5
Lower	-56.4	0.5	42.3	59.6	-3.6	40.6
Jaw	-59.6	-1.3	33.2	62.1	3.9	34.0
Larynx (Intrinsic)	-53	0	42	53	4	42
Larynx (Extrinsic)	-58.1	6.0	6.4	65.4	5.2	10.4
Respiration	-17.4	-26.9	73.4	23.8	-28.5	70.1
Cerebellum						
smCB	-18	-59	-22	16	-59	-23
slCB	-36	-59	-27	40	-60	-28
Deep Cerebellar Nuclei	-10.3	-52.9	-28.5	14.4	-52.9	-29.3
Somatosensory State Map						
Tongue						
1	-60.2	-2.8	27.0	62.9	-1.5	28.9
2	-60.2	-0.5	23.3	66.7	-1.9	24.9
3	-60.2	0.6	20.8	64.2	0.1	21.7
Lip						
Upper	-53.9	-7.7	47.2	59.6	-10.2	40.6
Lower	-56.4	-5.3	42.1	59.6	-6.9	38.2
Jaw	-59.6	-5.3	33.4	62.1	-1.5	34.0
Larynx (Intrinsic)	-53	-8	42	53	-14	38
Larynx (Extrinsic)	-61.8	1	7.5	65.4	1.2	12
Palate	-58	-0.7	14.3	65.4	-0.4	21.6
Somatosensory Error Map						
Supramarginal Gyrus	-62.1	-28.4	32.6	66.1	-24.4	35.2
Auditory State Map						
Heschl's gyrus	-37.4	-22.5	11.8	39.1	-20.9	11.8
Planum temporale	-57.2	-18.4	6.9	59.6	-15.1	6.9
Auditory Error Map						
SPT	-39.1	-33.2	14.3	44	-30.7	15.1
pSTg	-64.6	-33.2	13.5	69.5	-30.7	5.2

Table 2-1. The location of DIVA cell components in MNI space.

2.3.1 Premotor cortex *Speech Sound Map*

The model's *Speech Sound Map*, as shown in Figure 2-6, is hypothesized to lie in the left ventral premotor cortex and posterior Broca's area[4]. Each cell in the *Speech Sound Map* corresponds to a different frequently encountered speech sound (i.e., phoneme, syllable, word, or short phrase) in one's native language[5]. In the model, every phoneme and frequent multi-phonemic speech sound (e.g., common syllable) is represented by a unique *Speech Sound Map* cell. In this way, the cells of the *Speech Sound Map* are analogous to the "mental syllabary" described by Levelt and colleagues (e.g., Levelt and Wheeldon, 1994, Levelt et al., 1999). Activation of a *Speech Sound Map* cell results in the production of the corresponding speech sound. In contrast, infrequent sounds are produced by activating the *Speech Sound Map* cells associated with the subunits that form the syllable, word, etc. For example, an infrequent syllable would be produced by producing the individual phonemes in the appropriate order, rather than utilizing a stored motor program for the whole syllable. The sequential activation of *Speech Sound Map* cells is hypothesized to arise from inputs from higher-level brain regions involved in phonological encoding of an intended utterance (e.g., anterior Broca's area and/or middle

---

[4] The term Broca's area here refers to the inferior frontal gyrus *pars opercularis* (BA 44; posterior Broca's area) and *pars triangularis* (BA 45; anterior Broca's area). Due to the large amount of inter-subject variability in the location of the ventral precentral sulcus as measured in stereotactic coordinates, it is difficult to differentiate the ventral premotor cortex and posterior Broca's area in fMRI or PET studies that involve averaging across subjects using standard normalization techniques (see Tomaiuolo et al., 1999 for related discussion, Nieto-Castanon et al., 2003).

[5] Although each sound is represented by a single *Speech Sound Map* cell in the model, it is expected that speech sound representations in actual premotor cortex are distributed across many neurons. These distributed representations would be more robust to cell death and would allow greater generalizability of learned motor programs to new sounds. These topics are beyond the scope of the current report, however.

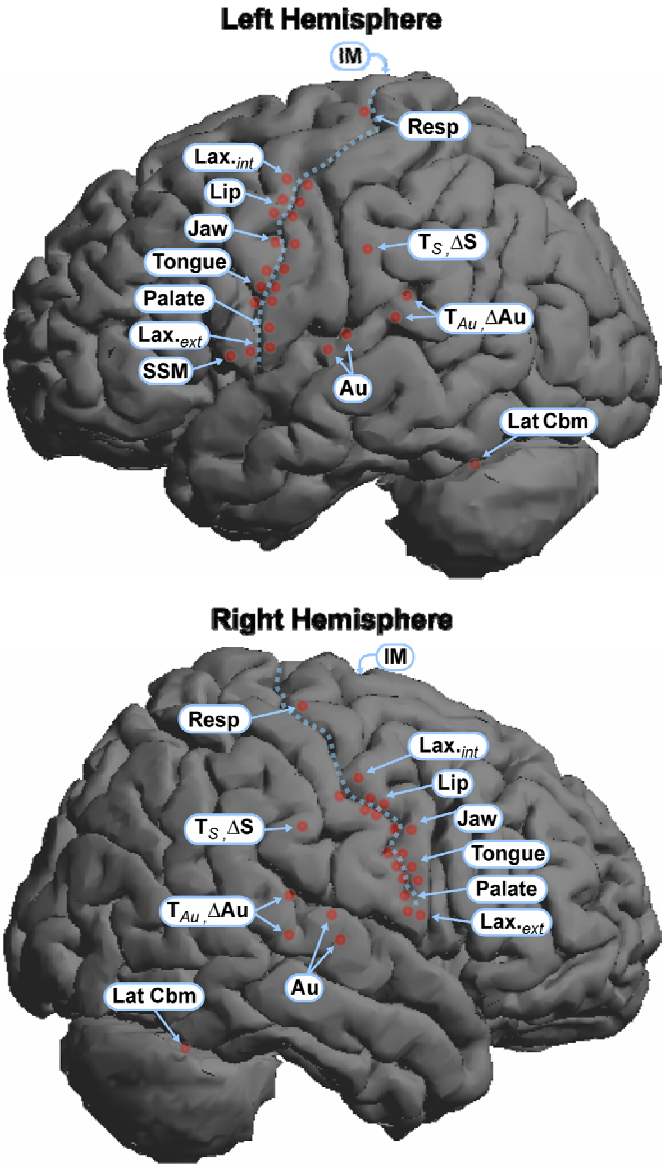


Figure 2-6. Neuroanatomical locations of the DIVA model components.
The location of DIVA model component sites (red dots) are plotted on renderings of the left (top) and right (bottom) lateral surfaces of the SPM2 canonical brain. Sites immediately anterior to the central sulcus (dotted line) represent cells of the model's *Articulator Velocity* and *Position Maps*. Sites located immediately posterior to the central sulcus represent cells of the *Somatosensory State Map.* Subcortical sites (basal ganglia, thalamus, paravermal cerebellum, deep cerebellar nuclei), are not shown. *Abbreviations*: Au = *Auditory State Map*; ΔAu = *Auditory Error Maps;* IM = *Initiation Map*; $Lax._{int}$, $Lax._{ext}$ = intrinsic and extrinsic larynx, Lat Cbm = lateral cerebellum; Resp: respiratory motor cells; ΔS = *Somatosensory Error Maps;* SSM = *Speech Sound Map*; T_{Au} =*Auditory Target Map;* T_S = *Somatosensory Target Map.*

frontal gyrus). The activation of these cells leads to the readout of feedforward motor commands to the primary motor cortex (see *Feedforward Control Subsystem* below), as well as a feedback control command if there is any error during production (see *Feedback Control Subsystem*). The feedforward command emanating from a *Speech Sound Map* cell can be thought of as a "motor program" or "gestural score", i.e., a time sequence of motor gestures used to produce the corresponding speech sound (c.f. Browman and Goldstein, 1989).

Speech Sound Map cells are hypothesized to lie in ventral premotor cortex and adjacent inferior frontal cortex because of their functional correspondence with "mirror neurons." Mirror neurons are so termed because they respond both during an action and while viewing (or hearing) that action performed by another animal or person (Rizzolatti et al., 1996, Kohler et al., 2002). These cells have been shown to code for complex actions such as grasping rather than the individual movements that comprise an action (Rizzolatti et al., 1988). Neurons within the *Speech Sound Map* are hypothesized to embody similar properties: activation during speech production drives complex articulator movement, and activation during speech perception tunes connections between the *Speech Sound Map* and sensory cortex.

Demonstrations of mirror neurons in humans have implicated left precentral gyrus for grasping actions (Tai et al., 2004), and left opercular inferior frontal gyrus for finger movements (Iacoboni et al., 1999). Mirror neurons related to communicative mouth movements have been found in monkey area F5 (Ferrari et al., 2003) immediately lateral to their location for grasping movements (di Pellegrino et al., 1992). This area has been

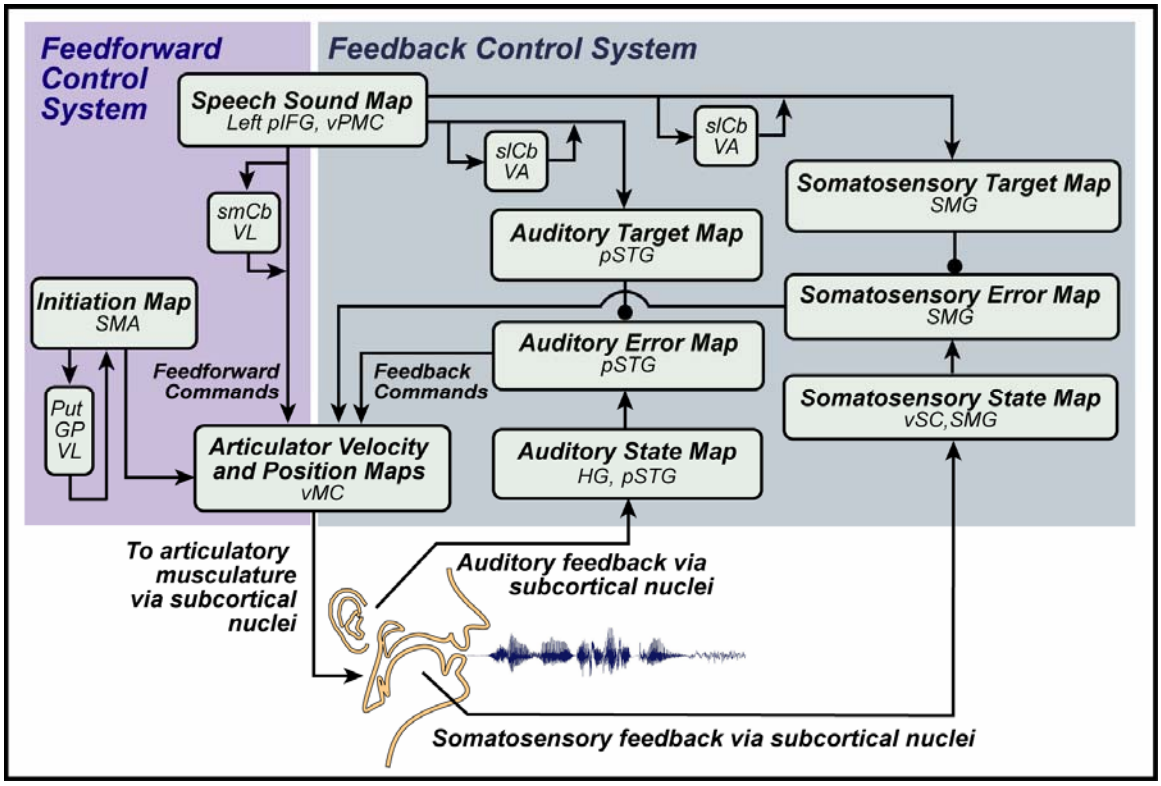


Figure 2-7. The DIVA model with neuroanatomical labels.

proposed to correspond to the caudal portion of ventral inferior frontal gyrus (Brodmann's area 44) in the human (see Rizzolatti and Arbib, 1998, Binkofski and Buccino, 2004)[6]. *Speech Sound Map* cells are therefore proposed to lie in ventral lateral premotor areas of the left hemisphere, including posterior portions of the inferior frontal gyrus. This proposal is consistent with the pooled fMRI results shown in Figure 2-2;

[6] The rare bifurcation of the left ventral precentral sulcus (posterior segment intersects the central sulcus, anterior segment intersects the anterior ascending branch of the Sylvian fissure) on the SPM standard brain makes it difficult to localize ventral BA 44. No clear sulcal landmark distinguishes BA 44 from BA 6. We have placed the *Speech Sound Map* region immediately behind the inferior end of the anterior ascending branch of the Sylvian fissure under the assumption that this area corresponds to ventral BA 44. The MNI coordinates chosen for the *Speech Sound Map* are consistent with the inferior frontal gyrus pars opercularis region.

activation is demonstrated in left ventral inferior frontal gyrus including BA 44 and the frontal operculum.

The equation governing *Speech Sound Map* cell activation in the model is:

$$(1)\quad \begin{array}{l} P_i(t) = 1 \; if \; the \; i^{th} \; sound \; is \; being \; produced \; or \; perceived \\ P_i(t) = 0 \; otherwise. \end{array}$$

Each time a new speech sound is presented to the model (as an acoustic sample) for learning, a new cell is recruited into the *Speech Sound Map* to represent that sound. The projections from the premotor *Speech Sound Map* cells to the auditory cortex represent a target auditory trace for that sound; this auditory target is subsequently used in the production of the sound (see *Feedback Control Subsystem* below for details), along with feedforward commands projecting from the *Speech Sound Map* cell to the motor cortex (detailed in *Feedforward Control Subsystem* below). According to the model, when an infant listens to a speaker producing a new speech sound, a previously unused *Speech Sound Map* cell becomes active, and projections from this cell to auditory cortical areas are tuned to represent the auditory signal corresponding to that sound. After the sound has been learned, activation of the *Speech Sound Map* cell leads to production of the corresponding sound via the model's feedforward and feedback subsystems.

2.3.2 Motor velocity and position maps

According to the model, feedforward and feedback-based control signals are combined in motor cortex. Three distinct subpopulations (maps) of motor cortical cells are thought to be involved in this process: one population representing positional commands to the speech articulators, one representing velocity commands originating from the

feedforward control subsystem, and one representing velocity commands originating from the feedback control subsystem. The model's *Motor Position Map* consists of eight antagonistic pairs of cells, each pair representing a position command for one of eight vocal articulators. Their activities at time *t* are represented by the 16-dimensional vector *M(t)* which is governed by the following equation:

$$(2) \quad M(t) = M(0) + \alpha_{ff} \int_0^t \dot{M}_{Feedforward}(t) g(t) dt + \alpha_{fb} \int_0^t \dot{M}_{Feedback}(t) g(t) dt$$

where $M(0)$ is the initial configuration of the vocal tract when starting an utterance, α_{fb} and α_{ff} are parameters that determine how much the model is weighted toward feedback control and feedforward control[7], respectively, and $g(t)$ is a speaking rate signal (arising from the *Initiation Map* in SMA) that is 0 when not speaking and 1 when speaking at a maximum rate. The 16-dimensional vectors $\dot{M}_{Feedforward}(t)$ and $\dot{M}_{Feedback}(t)$ represent activity in the model's *Articulator Velocity Maps.* $\dot{M}_{Feedforward}(t)$ encodes a feedforward articulator velocity signal that is determined by inputs from premotor cortex and the cerebellum; $\dot{M}_{Feedback}(t)$ encodes a feedback articulator velocity signal that is determined by inputs from sensory cortical areas; the sources of these command signals are discussed further in later sections (*Feedback Control Subsystem* and *Feedforward Control Subsystem*). The model's motor velocity and position cells are

---

[7] Under normal circumstances, both α_{fb} and α_{ff} are assumed to be 1. However, certain motor disorders may be associated with an inappropriate balance between feedforward and feedback control. For example, stuttering can be induced in the model by using an inappropriately low value of α_{ff} (Guenther and Ghosh, 2003).

hypothesized to correspond to "phasic" and "tonic" cells identified in motor cortex electrophysiological studies in monkeys (e.g., Kalaska et al., 1989), respectively.

The motor position command, *M(t)* is sent from ventral motor cortex to the articulatory musculature via brainstem nuclei (see Figure 2-6). Movements of the vocal tract articulators are produced according to the following equation:

$$(3) \quad Artic(t) = f_{MAr}(M(t - \tau_{MAr})) + Pert(t)$$

where f_{MAr} is a function that transforms each antagonistic pair of motor position cell activities into a single articulator position that serves as input to the Maeda articulatory synthesizer (Maeda, 1990), 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 if such perturbations are applied. The eight-dimensional vector *Artic* does not correspond to any cell activities in the model; it corresponds instead to the physical positions of the eight articulators[8] in the Maeda articulatory synthesizer (Maeda, 1990). The resulting vocal tract area function is converted into a digital filter that is used to synthesize an acoustic signal that forms the output of the model (e.g., Maeda, 1990).

Cells coding for the position and velocity of the tongue parameters in the model are hypothesized to correspond with the motor tongue area (MTA) as described by Fesl et al. (2003). The localization of this area is in agreement with imaging (Urasaki et al.,

---

[8] The eight articulators in the modified version of the Maeda synthesizer used herein correspond approximately to jaw height, tongue shape, tongue body position, tongue tip position, lip protrusion, larynx height, upper lip height and lower lip height. These articulators were based on a modified principal components analysis of midsagittal vocal tract outlines, and each articulator can be varied from -3.5 to +3.5 standard deviations from a neutral configuration.

1994, Corfield et al., 1999, Fox et al., 2001, Fesl et al., 2003) and physiological (Penfield and Rasmussen, 1950) studies of the primary motor region for tongue/mouth movements. Separate motor (and somatosensory) locations are designated for each tongue degree of freedom in the model. This expanded representation is consistent with the large tongue sensorimotor representation reported in imaging studies (e.g. Fesl et al., 2003).

A region superior and medial to the tongue region along the posterior bank of the precentral gyrus has been shown to produce lip movements in humans when stimulated electrically (Penfield and Roberts, 1959). Comparing production of syllables involving tongue movements to those involving lip movements, Lotze et al. (2000b) found the lip area to be approximately 1-2 cm from the tongue area in the directions described by Penfield. In another mapping study of motor cortex using fMRI, Lotze et al. (2000a) showed the lip region inferior and lateral with respect to the hand motor area, consistent with the Penfield electrical stimulation results. This area is hypothesized to code for the motor position and velocity of the model lip parameters. Upper and lower lip regions have been designated along the precentral gyrus superior and medial to the tongue representation. Data indicating the relative locations of upper and lower lip motor representations in humans is scarce. Currently, the upper lip motor representation is placed dorsomedial to the lower lip representation, mirroring the somatosensory organization (see *Somatosensory State Map* below). Penfield and Roberts also identified a primary motor region corresponding to jaw movements that lies between the lip and tongue representations along the posterior bank of the precentral sulcus.

The locus of the primary motor representation of the laryngeal muscles was not

made clear by Penfield's pioneering electrophysiological work (Penfield and Rasmussen, 1950). Though Penfield did mention a "vocalization" area adjacent to the lip area (Penfield and Roberts, 1959 p. 200), assumptions regarding the site of the human motor larynx area have relied heavily upon findings from non-human primates (e.g. Simonyan and Jurgens, 2003). Consequently, the motor larynx area is typically associated with the ventrolateral extreme of the precentral gyrus (e.g., Duffy, 2005, Ludlow, 2005). A recent imaging investigation sought to clarify the location of the motor larynx representation by comparing BOLD responses during vocal (vowel production) and non-vocal laryngeal tasks (forced glottal closure). A common area of activation across laryngeal tasks was noted bilaterally in dorsal orofacial motor cortex, very near the Penfield and Roberts "vocalization" area referenced above. This larynx-specific area was located immediately rostro-dorsal to the region activated during lip movements (motor lip area). A more ventral region within the Rolandic/frontal operculum, consistent with the monkey literature, was also active across laryngeal tasks. The investigators conclude that the dorsal region represents the intrinsic laryngeal muscles that control the size of the glottal opening. The opercular representation, it was speculated, likely represents the extrinsic laryngeal muscles that affect vocal tract resonances by controlling larynx height. Based on these findings, two sets cells representing laryngeal parameters of the Maeda articulator model (Maeda, 1990) associated with laryngeal functions have been assigned to MNI locations: cells in ventrolateral precentral gyrus (labeled *Motor Larynx, Extrinsic* in Table 2-1 and $Lax._{ext}$ in Figure 2-6) represent larynx height, whereas cells in the dorsomedial orofacial region of precentral gyrus (labeled *Motor Larynx, Intrinsic* in

Table 2-1 and *Lax.*$_{int}$ in Figure 2-6) represents a weighted sum of parameters representing glottal opening (Ag0) and glottal pressure (AgP).

Fink et al. (1996) demonstrated dorsolateral precentral gyrus activation during voluntary breathing using PET. The bilateral region noted in that study lay along the superior portion of primary motor cortex, well above the ventral motor representations of the articulators. In an fMRI study, Evans, Shea, and Saykin (1999) found a similar activation associated with volitional breathing along superior precentral gyrus medial to the Fink et al. findings and only in the left hemisphere. In the pooled production analysis, activity was noted in approximately the same regions as that described by Fink et al.: bilateral activation dorsomedial to and distinct from ventral motor activation (see dorsal surface image in Figure 2-1. This activity is hypothesized to be associated with the control of breathing (e.g., maintenance of appropriate subglottal pressure) required for speech production. Cells corresponding to voicing control parameters, specifically the parameter governing glottal pressure (AgP), were therefore assigned to this area of motor cortex in both hemispheres.

While the studies mentioned above indicate bilateral primary motor involvement during articulator movements, they do not explicitly show bilateral involvement of these areas during speech production (though Penfield & Roberts report a bilateral precentral gyrus region that causes "vocalization"). In their review of neuroimaging studies of speech production, however, Indefrey and Levelt (2004) noted consistent bilateral activation of lateral Rolandic cortex during overt speech when compared to silence. Activation along both banks of the lateral central sulcus was also demonstrated in the

pooled analysis of simple speech production tasks discussed above (Figure 2-2). While the response in Rolandic cortex was noted bilaterally, it was approximately 20% greater in the left hemisphere than in the right, a finding consistent with other neuroimaging studies of overt speech (Riecker et al., 2000a, Bohland and Guenther, 2006, Sidtis et al., 2006a). Based on the inter-hemispheric differences noted in the pooled analysis, the model's motor position and velocity cell populations are assumed to contain 20% more cells in the left hemisphere than the right hemisphere.

2.3.3 Feedback control subsystem

The feedback control subsystem in the DIVA model (Figure 2-7) carries out the following functions when producing a learned sound. First, activation of the *Speech Sound Map* cell corresponding to the sound in the model's premotor cortex leads to readout of learned auditory and somatosensory targets for that sound. These targets take the form of temporally varying regions in auditory and somatosensory spaces. The current auditory and somatosensory states, available through sensory feedback, are compared to these targets in the higher-order auditory and somatosensory cortices. If the current sensory state falls outside of the target region, an error signal arises in the higher-order sensory cortex. These error signals are then mapped into appropriate corrective motor commands via learned projections from the sensory error cells to the motor cortex. The feedback system components supporting these functions are discussed below.

Auditory State Map. The current acoustic state is determined from the articulatory state as follows:

$$(4) \quad Acoust(t) = f_{ArAc}(Artic(t))$$

where f_{ArAc} is the transformation performed by Maeda's articulatory synthesizer. The vector *Acoust(t)* does not correspond to brain cell activities; instead it corresponds to the physical acoustic signal resulting from the current articulator configuration. The representation of this acoustic signal in auditory cortical areas is given by the activation of cells in the model's *Auditory State Map*:

$$(5) \quad Au(t) = f_{AcAu}(Acoust(t - \tau_{AcAu}))$$

where *Au(t)* is a vector of *Auditory State Map* cell activities, f_{AcAu} is a function that transforms an acoustic signal into the corresponding auditory cortical map representation, and τ_{AcAu} is the time it takes for an acoustic signal to travel from the cochlea to the auditory cortex. In the current implementation of the DIVA model, the *Auditory State Map* is a 3-dimensional vector representing the 1st, 2nd and 3rd formants of the acoustic signal.

The *Auditory State Map* cells are hypothesized to lie within primary auditory cortex and the surrounding auditory association cortex (BA 41, 42). Therefore, the *Auditory State Map* is localized to regions along the medial portion of Heschl's gyrus and the anterior planum temporale (Rivier and Clarke, 1997, Morosan et al., 2001). Activation in these areas is consistently found in fMRI studies of speech perceptual processing (e.g., Scott and Johnsrude, 2003, Guenther et al., 2004) regardless of the intelligibility of the utterance (see Scott and Johnsrude, 2003 for discussion). Projections from this region to more posterior regions of the superior temporal gyrus, including the

planum temporale and posterior superior temporal sulcus, the hypothesized locations of the auditory target and error maps (see below), have been demonstrated in monkeys (e.g., Kaas and Hackett, 2000). Evidence for similar connectivity in humans has also been reported (Tardif and Clarke, 2001).

Somatosensory State Map. The model's *Somatosensory State Map* corresponds to the representation of speech articulators in somatosensory cortical areas (BA 3, 1, 2) and is governed as follows:

$$(6) \quad S(t) = f_{ArS}(Artic(t - \tau_{ArS}))$$

where *S(t)* is a 22-dimensional vector of *Somatosensory State Map* cell activities, f_{ArS} is a function that transforms the current state of the articulators into the corresponding somatosensory cortical map representation, and τ_{ArS} is the time required for somatosensory feedback from the periphery to reach higher-order somatosensory cortical areas. The function f_{ArS} transforms the articulatory state into a 22-dimensional somatosensory map representation *S(t)*. The first 16 dimensions of *S(t)* correspond to proprioceptive feedback representing the current positions of the 8 Maeda articulators, each represented by an antagonistic pair of cells as in the motor representation. The remaining 6 dimensions represent tactile feedback.

Tactile and proprioceptive representations of the articulators are hypothesized to lie along the inferior postcentral gyrus, roughly adjacent to their motor counterparts across the central sulcus. An anatomical marker for the tongue somatosensory region has been demonstrated using PET imaging (Boling et al., 2002) and electrical stimulation

(Picard and Olivier, 1983). The location of the tongue region is described as lying below the anterior apex of the triangular region of the inferolateral postcentral gyrus approximately 2 cm above the Sylvian fissure. This region of the postcentral gyrus was found to represent the tongue in a somatosensory evoked potential study of humans (McCarthy et al., 1993), a finding further supported by a similar procedure in the macaque (McCarthy and Allison, 1995). By generating potentials on either side of the central sulcus, both studies by McCarthy and colleagues demonstrate adjacent motor-somatosensory organization of the tongue representation.

McCarthy et al. (1995) also mapped the primary sensory representations of the lip and palate. The lip representation was located superior and medial to the tongue representation along the anterior bank of the postcentral gyrus at the apex of the inferior postcentral triangle and below the hand representation. Nakamura et al. (1998) localized the lip and tongue sensory representations to nearly identical regions of the postcentral gyrus using MEG. The palatal representation was located inferolateral to the tongue region roughly 1 cm above the Sylvian fissure. The relative locations of the lip, tongue, and palate were confirmed in the macaque (McCarthy and Allison, 1995). Consistent with early electrophysiological work (Penfield and Rasmussen, 1950) and a recent MEG study (Nakamura et al., 1998), the upper lip representation is placed dorsomedial to the lower lip representation.

Graziano, Taylor, Moore, and Cooke (2002) report early electrical stimulation work (Foerstner, 1936, Fulton, 1938) which depicts a sensory representation of the larynx at the inferior extent of the postcentral gyrus, near the Sylvian fissure. This location

mirrors the ventral motor larynx representation that lies on the inferior precentral gyrus.

Using the same reasoning as outlined above for the primary motor representation of articulators, bilateral somatosensory representations for each of the articulators are hypothesized to have a 20% leftward bias. As was the case for precentral activation, the pooled speech production results demonstrated greater left hemisphere activation along the postcentral gyrus (Figure 2-2).

Sensory Target and Error Maps. Projections from *Speech Sound Map* cells in the ventral posterior frontal and premotor cortex (lateral BA 6 and 44) to higher-order sensory cortical areas[9] provide auditory and somatosensory cortex with the sensory expectations associated with the active *Speech Sound Map* cell. In other words, the synaptic weights between the *Speech Sound Map* and sensory target maps encode the auditory and somatosensory *targets* associated with the speech sound being produced. Projections such as these, which predict the sensorimotor state resulting from a movement, are typically described as representing a *forward model* of the movement (e.g. Miall and Wolpert, 1996, Kawato, 1999, Desmurget and Grafton, 2000, Davidson and Wolpert, 2005). According to the model, the current auditory target is represented by the activation in the *Auditory Target Map* that lies along the posterior superior temporal gyrus (BA 22).

---

[9] Although currently treated as a single set of synaptic weights in the model, it is possible that this mapping may include a trans-cerebellar contribution (motor cortex -> pons -> cerebellum -> thalamus -> higher-order auditory cortex) in addition to a cortico-cortical contribution. Current data do not definitively resolve this issue. The weight matrix z_{PAu} (as well as z_{PS}, z_{SM}, and z_{AuM} defined below) can thus be considered as combining cortico-cortical and trans-cerebellar synaptic projections. Evidence for a trans-cerebellar contribution to the weight matrix z_{PM}, which encodes a feedforward command between the premotor and motor cortices, is considered in the next section.

Activation in the map is governed by long-range excitatory inputs from the *Speech Sound Map* according to the following equation:

$$(7) \quad T_{Au}(t) = P(t - \tau_{PAu}) z_{PAu}(t)$$

where τ_{PAu} is the propagation delay for the signals from premotor cortex to auditory cortex, and $z_{PAu}(t)$ are synaptic weights that encode the auditory expectations for the sound being produced. The weights $z_{PAu}(t)$ are tuned based on examples from other speakers producing the sound, and during one's own correct productions. Inhibitory connections from the *Auditory Target Map* to the *Auditory Error Map*[10] inhibit the portion of the *Auditory Error Map* that responds to the target region for the current speech sound. Additional excitatory inputs to the *Auditory Error Map* from the *Auditory State Map* provide an estimate of the current auditory state. Activity within the *Auditory Error Map* is therefore given by:

$$(8) \quad \Delta Au(t) = Au(t) - T_{Au}(t).$$

This activity represents the difference between the expected and realized auditory state for the current speech sound. If the current sensory state falls outside the expected target region, error signals arise that are mapped into corrective motor commands via projections to the motor cortex *Motor Velocity Map*.

According to the model, projections from the *Speech Sound Map* cell to higher order auditory (and somatosensory) cortex have the effect of inhibiting the regions of

[10] The sensory target and sensory error maps are hypothesized to lie in overlapping cortical regions and therefore share the same hypothesized location.

sensory cortex that encode the expected sensory feedback. By effectively "cancelling" the self-produced portion of the sensory feedback response, these signals function similarly to the motor-to-sensory projections original described by von Holst and Mittelstaedt (1950) and Sperry (1950). von Holst and Mittelstaedt (1950) proposed the 'principle of reafference' in which a copy of the expected sensory consequences of a motor command, termed an *efference copy*, was subtracted from the realized sensory consequences. A wide body of evidence suggests such a mechanism plays an important role in the motor control of eye and hand movements, as well as speech (e.g. Reppas et al., 2002, Cullen, 2004, Roy and Cullen, 2004, Heinks-Maldonado and Houde, 2005, Bays et al., 2006, Voss et al., 2006).

Hickok and colleagues have demonstrated an area within the posterior Sylvian fissure at the junction of the temporal and parietal lobes (the Sylvian-parietal-temporal, SPT, area) of the left hemisphere and another in the lateral posterior superior temporal sulcus bilaterally that respond during speech perception and speech production (Buchsbaum et al., 2001, Hickok and Poeppel, 2004). Such regions could support the comparison of efferent motor commands with auditory input as in the DIVA model's *Auditory Error Map*. In a review of imaging studies of speech processing, Wise and colleagues noted that several studies implicated area SPT as being "engaged in the motor act of speech" (Wise et al., 2001).

The hypothesized inhibitory influence on auditory cortex embodied by the model's SSM-to-*Auditory Target Map* projections is supported by recent anatomical and physiological data. Reciprocal connections between posterior inferior frontal gyrus and

both the supramarginal gyrus and posterior superior temporal gyrus in the human have been demonstrated by Matsumoto et al. (2004). The connectivity was revealed using a cortico-cortical evoked potential technique involving direct cortical stimulation in epilepsy patients. Inhibition of posterior superior temporal gyrus via this pathway during speech production has been demonstrated in a number of studies. Wise and colleagues, using positron emission tomography (PET) to indirectly assess neural activity, noted reduced superior temporal gyrus activation during speech production compared to a listening task (Wise et al., 1999). Likewise, comparisons of auditory responses during self-produced speech and while listening to recordings of one's own speech have demonstrated attenuation of auditory cortex responses during speech production (Numminen et al., 1999, Curio et al., 2000, Heinks-Maldonado et al., 2005, Heinks-Maldonado et al., 2006). Further evidence of auditory response suppression during self-initiated vocalizations is provided by single unit recordings from non-human primates; attenuation of auditory cortical responses prior to self-initiated vocalizations has been demonstrated in the marmoset (Eliades and Wang, 2003, 2005). Collectively, these findings support the inhibitory projections from ventral inferior frontal cortex to higher-order auditory cortex, either directly or indirectly (e.g., via the supramarginal gyrus) hypothesized in the DIVA model.

The Buchsbaum and Hickok studies (Buchsbaum et al., 2001, Hickok and Poeppel, 2004) have indicated that the posterior regions identified as responsive during both speech perception and production may be lateralized to the left hemisphere. Evidence for a bilateral contribution to auditory error representation from these regions,

however, is provided by the fMRI study of Hashimoto and Sakai (2003). Bilateral activation of the posterior superior temporal gyrus and the inferior supramarginal gyrus was noted when speakers listened to temporally delayed feedback of their own voices. The activity within the posterior superior temporal gyrus and superior temporal sulcus correlated with the size of the disfluency effect caused by the delayed auditory feedback. Based on this result, the auditory target and error cells were assigned bilaterally. Finally, to date, insufficient data are available to functionally differentiate the intra-Sylvian SPT and more lateral posterior superior temporal sulcus sites identified by Hickok and colleagues. Pending further investigation, the auditory target and error maps have therefore been assigned to both SPT and the more superior temporal sulcus locations (T_{Au}, ΔA in Figure 2-6).

The model's expected *somatosensory target* and current *state* representations are analogous to those described for the auditory system. Tactile and proprioceptive expectations are represented by the *Somatosensory Target Map* proposed to lie along the anterior supramarginal gyrus and surrounding cortex (BA 40, 1, 2, 3). Activation of the *Somatosensory Target Map* is given by:

$$(9) \quad T_S(t) = P(t - \tau_{PS}) z_{PS}(t)$$

where τ_{PS} is the propagation delay from premotor cortex to somatosensory cortex, and the weights $z_{PS}(t)$ encode somatosensory expectations for the sound being produced. It is hypothesized that these weights become tuned during correct self-productions of the corresponding speech sound. In other words, this learning follows the learning of an auditory target, which can be learned simply by monitoring examples from other

speakers. Somatosensory error is represented by activation within the *Somatosensory Error Map*:

$$(10)\quad \Delta S(t) = S(t) - T_S(t)\,.$$

The somatosensory target and error maps are hypothesized to lie within the inferior parietal cortex along the anterior supramarginal gyrus, posterior to the primary somatosensory representations of the speech articulators (T_S, ΔS in Figure 2-6). Hickok and colleagues (e.g. Hickok and Poeppel, 2004) have argued that speech motor commands and sensory feedback are integrated in the ventral parietal lobe, analogous to the visual-motor integration of the dorsal parietal lobe (Andersen, 1997, Rizzolatti et al., 1997). Reciprocal connections between area F5, a region thought to be a homologue of Broca's area (see Rizzolatti and Arbib, 1998, Binkofski and Buccino, 2004), and inferior parietal cortex have been demonstrated in the monkey (Luppino et al., 1999). These connections are believed to contribute to the sensorimotor transformations required to guide movements (see Rizzolatti and Luppino, 2001) such as grasping. Analogous connections are hypothesized to play a role in monitoring and guiding speech articulator movements. Neuroimaging studies of motor learning have noted cerebellar activity that is associated with the size or frequency of sensory error (e.g., Flament et al., 1996, Schreurs et al., 1997, Blakemore et al., 1999, Imamizu et al., 2000, Tesche and Karhu, 2000, Blakemore et al., 2001, Diedrichsen et al., 2005, Miall and Jenkinson, 2005, Imamizu et al., 2007, Grafton et al., 2008). It has been argued by many that the cerebellum uses sensory error to build forward models that generate sensory *predictions* (Imamizu et al., 2000, Blakemore et al., 2001, Kawato et al., 2003, O'Reilly et al., 2008), the role of

projections from the *Speech Sound Map* to sensory target maps in DIVA model. It is likely, therefore, that the cerebellum contributes to the attenuation of sensory target representation in sensory cortex (cf. Blakemore et al., 2001). For this reason, cerebellar side loops are hypothesized in the *Speech Sound Map*-to-sensory target map projection.

In the model, production errors represented by activations in the Auditory and/or *Somatosensory Error Maps* are mapped into corrective motor commands through learned pathways projecting from the sensory cortical areas to the motor cortex. These projections form a feedback control signal that is governed by the following equation:

$$(11)\quad \dot{M}_{Feedback}(t) = \Delta Au(t-\tau_{AuM})z_{AuM} + \Delta S(t-\tau_{SM})z_{SM}$$

where z_{AuM} and z_{SM} are synaptic weights that transform directional sensory error signals into motor velocities that correct for these errors, and τ_{AuM} and τ_{SM} are cortico-cortical transmission delays. The model's name, DIVA, derives from this mapping from sensory directions into velocities of articulators. The weights z_{AuM} and z_{SM} approximate a pseudoinverse of the Jacobian of the function relating articulator positions (*M*) to the corresponding sensory state (see Guenther et al., 1998 for a more detailed description). As such, z_{AuM} and z_{SM} represent what is often referred to as an *inverse model* (e.g., Wolpert and Kawato, 1998, Kawato, 1999) because they implement the transformation between the desired sensory outcome and the appropriate motor actions, i.e., an inverse kinematic transformation. These weights are hypothesized to be tuned during an early babbling stage by monitoring the relationship between movement commands and their sensory consequences. The feedback motor command, $\dot{M}_{Feeback}(t)$, is represented by the

activation of cells in the model's *Articulator Velocity Map*. The location of these cells was described in the *Motor velocity and position maps* section above.

The cerebellum is a likely contributor to the feedback motor command. The section describing the sensory target maps above referenced several studies that have demonstrated a correlation between the size and/or frequency of sensory error and cerebellar activation. It has been speculated that a representation of sensory errors in the cerebellum drives corrective motor commands (Penhune and Doyon, 2005, Grafton et al., 2008) and contributes to feedback-based motor learning (Wolpert et al., 1998, Ito, 2000, Tseng et al., 2007). For instance, the cerebellum has been hypothesized to support the learning of inverse kinematics (e.g., Wolpert and Kawato, 1998, Kawato, 1999), a role for which it is anatomically well-suited: the cerebellum receives inputs from higher-order auditory and somatosensory areas (e.g., Schmahmann and Pandya, 1997), and projects heavily to the motor cortex (Middleton and Strick, 1997). Based on the cerebellum's putative role in feedback-based motor learning, it is hypothesized to contribute to the mapping between sensory states and motor cortex, i.e., the projections that encode the feedback motor command (Figure 2-5).

2.3.4 Feedforward motor control subsystem

According to the model, projections from ventral premotor and posterior inferior frontal cortex to primary motor cortex (see Figure 2-7), constitute feedforward motor commands for the production of speech sounds. The primary motor and premotor cortices are well-known to be strongly interconnected (e.g., Passingham, 1993, Krakauer and Ghez, 1999). The feedforward motor command for production of a sound is represented in the model

by the following equation:

$$(12)\quad \dot{M}_{Feedforward}(t) = P(t)z_{PM}(t) - M(t).$$

The weights $z_{PM}(t)$ encode the feedforward motor command for the speech sound being produced. This command is learned over time by incorporating the corrective motor commands from the feedback control subsystem on the previous attempt into the new feedforward command. Before the model has any practice producing a speech sound, the contribution of the feedforward control signal to the overall motor command will be small since it will not yet be tuned. Therefore, during the first few productions, the primary mode of control will be feedback control. During these early productions, the feedforward control system "tunes itself" by monitoring the motor commands generated by the feedback control system (e.g., Kawato and Gomi, 1992). The feedforward system improves over time, all but eliminating the need for feedback-based control except when unexpected sensory feedback is encountered (e.g., changing vocal tract dynamics, a bite block, artificial auditory feedback perturbation). 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. As mentioned above, once an appropriate feedforward command sequence has been learned for a speech sound, this sequence will successfully produce the sound with very little, if any, contribution from the feedback subsystem; the feedback system will automatically become disengaged since no sensory errors will arise during production unless unexpected constraints are

placed on the articulators or the auditory signal is perturbed.

The feedforward motor command, $\dot{M}_{Feedforward}(t)$, is represented by the activation of cells in the model's *Articulator Velocity Map*. The location of these cells was previously described in the *Motor velocity and position maps* section above. The model's feedforward motor command is hypothesized to involve contributions from the cerebellum. 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 and Pandya, 1997), and projects heavily to the motor cortex (Middleton and Strick, 1997). In addition to the hypothesized contributions to motor learning described in previous sections, it is widely held that the cerebellum is also involved with learning and maintenance of feedforward motor commands (Kawato, 1999, Ohyama et al., 2003, though see Grafton et al., 2008). In the DIVA model, the cerebellum is therefore included as a side loop in the projection from the *Speech Sound Map* to the *Motor Velocity Map* (Figure 2-5). Lesions to anterior vermal and paravermal cerebellum have been associated with disruption of speech production (Ackermann et al., 1992, Urban et al., 2003). This region was active in the pooled speech production analysis, but the activation spread to adjacent paravermal and lateral cortex bilaterally (lobules III-VI; see Figure 2-4). Similar patterns of activation within bilateral superior cerebellar cortex during speech production have been reported (Riecker et al., 2000a, Wildgruber et al., 2001, Riecker et al., 2002). Model cells have therefore been placed bilaterally in two cerebellar cortical regions: anterior paravermal cortex (not visible in Figure 2-6) and

superior lateral cortex (*Lat. Cbm*) in Figure 2-6). At present, there is no explicit cerebellar component in the DIVA model and activity in both sets of cerebellar cells is determined by activity in the model's *Motor Position Map*. Cells have also been placed bilaterally in the thalamus at the peaks visible in Figure 2-3. These thalamic cells (not visible in Figure 2-6) lie within the lateral, ventral thalamus, the area containing nuclei that relay cerebellar output to the cerebral cortex (but see *Initiation Map* below).

Initiation Map. The final component of the feedforward subsystem is an *Initiation Map* hypothesized to lie within supplementary motor area (SMA, see Figure 2-6), a region thought to play a critical role in the initiation of speech motor programs (e.g., Jonas, 1987, Ziegler et al., 1997, Alario et al., 2006, Bohland and Guenther, 2006). Commands from the motor cortex ($M(t)$ in Equation 2) are released to the vocal tract when the activity of the appropriate cell in the SMA *Initiation Map* is non-zero. The activity of the *Initiation Map* corresponding to the i^{th} speech motor program is given by:

$$(13)\quad \begin{array}{l} I_i(t) = 1 \; if \; the \; i^{th} \; sound \; is \; being \; produced \; or \; perceived \\ I_i(t) = 0 \; otherwise. \end{array}$$

In other words, the *Initiation Map* acts as a simple multiplicative gate on the release of motor programs. In the current version of the model, this signal is equated to the "go" signal $g(t)$ described above (2).

The SMA is reciprocally interconnected with the precentral gyrus and the basal ganglia, a region that, based on clinical and physiological data, is also thought to contribute to motor gating (e.g., Van Buren, 1963, Albin et al., 1995, Pickett et al., 1998). Notably, in addition to motor and prefrontal afferents, the basal ganglia receive inputs

from virtually the entire cortex, including associative and limbic cortex (Parent and Hazrati, 1995). Thus, the basal ganglia are well-suited for integrating contextual cues for the purpose of gating motor commands (see Bullock et al., 1998 for a neural network model of motor control that includes a more sophisticated model of basal ganglia-based motor gating). Cells representing *Initiation Map* activity were placed bilaterally in the SMA according to the peak MNI responses from this region in the pooled speech production imaging results (Figure 2-2, Table 2-1). Cells were also placed at the location of peak activation in the caudate, putamen, globus pallidus, and thalamus (Figure 2-3, Table 2-1). The activations in the SMA, basal ganglia, and thalamus are a weighted sum of the *Motor Velocity* and *Motor Position* map activities.

2.4 Predicting BOLD responses from DIVA simulations

One of the primary goals of the DIVA model component mapping described in Section 2.3 was to provide a means for generating predictions that help guide future neuroimaging studies. By identifying the likely locations of the model's components within a standard reference frame, it is possible to run "simulated fMRI experiments" in which the model produces speech sounds under different speaking conditions and the model cell activities are then used to generate a simulated hemodynamic response pattern based on these cell activations. The simulated hemodynamic response patterns can then be used to constrain the interpretation of fMRI and/or positron emission tomography (PET) experiments in which human subjects produce the same (or similar) speech sounds in the same speaking conditions. An example of a simulated BOLD response from a

DIVA model simulation of a simple CV syllable production is shown in Figure 2-8. DIVA cell activations were plotted on the SPM2 canonical brain surface according to the component locations listed in Table 2-1 and shown in Figure 2-6 Responses derived from the DIVA simulation are shown on the left; experimental results from an fMRI study that compared CV syllable productions to a silent letter viewing baseline in 10 subjects is shown on the right.

The simulated BOLD responses on the left in Figure 2-8 were generated by comparing the model cells activations during the CV production to a baseline condition in which all cell activities were set to 0. To produce the simulated hemodynamic response for each condition, model cell activities were first normalized by the maximum possible activity of the cell. This was done to correct for differences in the dynamic ranges of the different cell types in the model. The resultant activity was then convolved with an idealized hemodynamic response function, generated using default settings of the function 'spm_hrf' from the SPM2 toolbox. This function was designed by the creators of SPM to approximate the transformation from cell activity to hemodynamic response in the brain. A brain volume was then constructed with the appropriate hemodynamic response values at each position. Responses were smoothed with a Gaussian kernel (FWHM = 12mm) then mapped to the SPM2 canonical brain. These procedures approximated the standard voxel-based analyses for experimental data including those used that provided the data shown at the right of Figure 2-8.

While it is informative to see how much of the fMRI activity in human subjects producing simple syllables can be accounted for by the model, the generation of novel predictions from the model that can be tested in future neuroimaging studies is perhaps a more useful application for simulated fMRI results. This capability is leveraged for the fMRI study of auditory feedback control of speech described in Chapter 4. Subjects' auditory feedback was unexpectedly perturbed during speech production to investigate the hypothesized *Auditory Error Map* location. A simulation of the experimental condition was performed prior to collecting fMRI data and the results mapped to a standard brain as described above (see Figure 4-5*B* below). The simulated results served as a guide for the interpretation of the experimental findings.

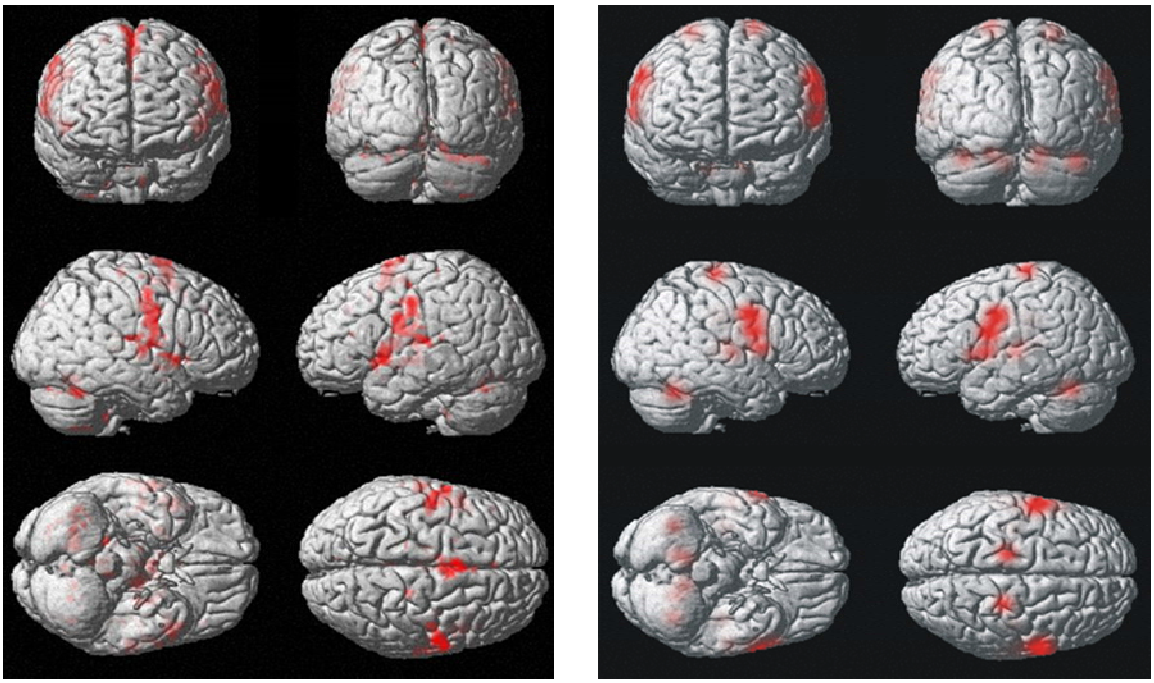


Figure 2-8. Predicted and experimental BOLD responses
Left = fMRI experimental data for CV *production – silent baseline* task. *Right*: Predicted BOLD response based on DIVA simulation of CV production. The simulated neural responses closely approximate the empirical data.

CHAPTER 3

A PARCELLATION SYSTEM FOR FUNCTIONAL IMAGING STUDIES OF SPEECH

In most functional imaging experiments, data from multiple subjects are averaged to obtain sufficient statistical power. Data are typically compared on a voxel-by-voxel basis following a full-brain normalization procedure meant to align the data from each subject in a standardized stereotactic space (Friston et al., 1995a, Mazziotta et al., 2001). This method of averaging subject data is confounded by the substantial anatomical variability that exists between subjects. Nieto-Castanon and colleagues (2003) provided a measure of the extent of inter-subject variability for a number of easily-identified anatomical regions of interest (ROIs). For example, the mean overlap of voxels common to Heschl's gyrus, an ROI representing primary auditory cortex, across two subjects following normalization was 31% of the total voxels belonging to Heschl's gyrus in both subjects. The overlap dropped to 13% for three subjects and across nine subjects, a relatively small population for imaging studies, no overlap was found. In other words, there was not a single voxel in the standard stereotactic space that fell within Heschl's gyrus across all subjects. This variability was typical of the 12 temporal and parietal lobe regions of interest that were analyzed in the left and right hemispheres (see Nieto-Castanon et al., 2003). Thus, the standard normalization procedure falls far short of ensuring alignment of the structural, and presumably functional, regions across even a small subject cohort.

The standard means of accounting for inter-subject anatomical variability is to

spatially smooth functional data after normalization with an isotropic smoothing kernel of 6-12 mm. While increasing the power of voxel-based analyses, this method has the effect of blurring regional boundaries. The blurring effect is particularly problematic across the banks of major sulci. For instance, two adjacent points along the dorsal and ventral banks of the Sylvian fissure that are separated by less than a millimeter in the 3-D volume space may lie several centimeters apart with respect to distance along the cortical sheet. Their cortical distance is a much better measure of their "functional distance:" the adjacent dorsal and ventral points could be the somatosensory larynx representation and the auditory association cortex, respectively. Isotropic smoothing in the 3-D volume ignores this distinction, blurring responses from the two regions, resulting in loss of statistical sensitivity and, perhaps, misleading findings.

The ROI-based functional imaging analysis software described by Nieto-Castanon et al. (2003) was designed to address this confound. The software compares functional responses across like anatomical regions of interest based on individual anatomical landmarks. This method, by accounting for inter-subject anatomical variability, greatly improves the statistical sensitivity over standard voxel-based technique and provides a more direct link between neuroanatomical structure and function. The remainder of this section provides a method for defining anatomical regions of interest that are relevant for imaging studies of speech.

To obtain a fine-grained functional map of the cortical interactions underlying speech, it is first necessary to parcellate the speech-related areas of cortex into smaller functional units. Traditionally defined speech-related cortical areas, such as "Wernicke's

area", "Broca's area", and "auditory cortex", involve large expanses of cortex and are often applied inconsistently in the literature. For example, portions of the supramarginal gyrus, angular gyrus, and/or middle temporal gyrus are sometimes included in the definition of Wernicke's area (Penfield and Roberts, 1959), while other researchers limit Wernicke's area to the posterior superior temporal gyrus and planum temporale (Keuhn et al., 1989, Martin, 1996). Similarly, Broca's area is sometimes limited to Brodmann's Area (BA) 44 (Martin, 1996), while other definitions also include BA 45 (Goodglass, 1993, Duvernoy, 1999). Even more confusing, the term auditory cortex is sometimes used to refer only to primary auditory cortex (BA 41) and other times to primary and higher-order auditory cortical areas (BA 42, 22, and 52), prompting the neuroanatomist Duvernoy (1999, p. 46) to note that "the precise localization of the auditory cortex seems difficult to define."

A finer-grained parcellation scheme based on anatomical landmarks has been created for the purpose of analyzing the volumes of different regions of cortex (Caviness et al., 1996). This system, developed and used extensively at the Center for Morphometric Analysis (CMA) at Massachusetts General Hospital, has allowed researchers to compare brains of neurologically normal subject populations to brains of individuals with psychiatric disorders such as schizophrenia in an attempt to identify the brain regions involved in these disorders. Many of the anatomical landmarks defining borders between different parcellation units, or regions of interest (ROIs), align approximately with cytoarchitectonic maps of cortex (e.g., the well-known Brodmann areas). It is commonly assumed that cytoarchitecture and normal function of a brain

region are closely related, as evidenced by the use of functional names for many of Brodmann's areas; e.g. BA 4 is commonly called primary motor cortex and BA 41 is commonly called primary auditory cortex in the neuroscience literature. The CMA parcellation scheme can therefore be thought of as a means to identify functional brain regions using anatomical landmarks that are clearly visible on structural MRI images (unlike cytoarchitectonic details, which are impossible to identify in standard structural MRI scans).

The Caviness et al. (1996) parcellation scheme was not specifically designed for the study of speech and speech disorders and, as such, it is not ideally suited for speech neuroimaging studies. Several of the ROIs in the CMA system are not defined at a fine-enough grain to distinguish functional contributions in regions that previous work suggests support multiple functional roles (e.g., those ROIs that delineate precentral and superior temporal cortex). In other cortical regions (e.g., the junction of the posterior superior temporal, supramarginal, and angular gyri), the CMA definitions introduce spurious regional distinctions for the study of speech processes. A modified version of the Caviness et al. parcellation scheme was therefore created that is specifically geared to speech studies.

Following a review of relevant physiological and imaging studies of speech processing, a set of speech-related cortical ROIs was defined. To assess the functional role of cortical regions not typically associated with speech processing, ROIs representing the remainder of the cerebral cortex were also defined and largely follow the conventions of the CMA system. In addition to modifying the CMA cerebral cortex

parcellation system, a set of ROIs within the cerebellum, a region known to play an important role in speech production, was defined based upon the anatomical atlas of Schmahmann and colleagues (2000).

3.1 Cortical Regions of Interest

The cortical ROI definitions are illustrated in Figure 3-1; the core regions involved in speech are highlighted in gray. A list of these regions and the hypothesized functional contribution(s) of each to speech processes is provided in Table 3-1. The table also lists, when applicable, the Brodmann areas that correspond to the cortical ROIs. Modifications of the Caviness et al. (1996) cortical parcellation system were based on the results of a number of recent neuroimaging studies of speech (see Table 3-1). Significant changes to the Caviness et al. system are described below.

Superior Temporal Sulcus: The dorsal and ventral banks of the superior temporal sulcus are defined separately from the surrounding temporal lobe gyri (superior and middle, respectively). This modification of the Caviness et al. definitions, which do not differentiate the cortex of the superior temporal sulcus from the adjacent superior and middle temporal gyri, was made to reflect findings that suggest a phoneme processing center within the superior temporal sulcus that is functionally distinct from surrounding gyral cortex (Wise et al., 1999, Belin et al., 2000, Binder et al., 2000, Scott et al., 2000). Cutting planes orthogonal to the cortical surface are made at the lateral margins of the dorsal and ventral surfaces of the superior temporal sulcus to delineate cortex lying

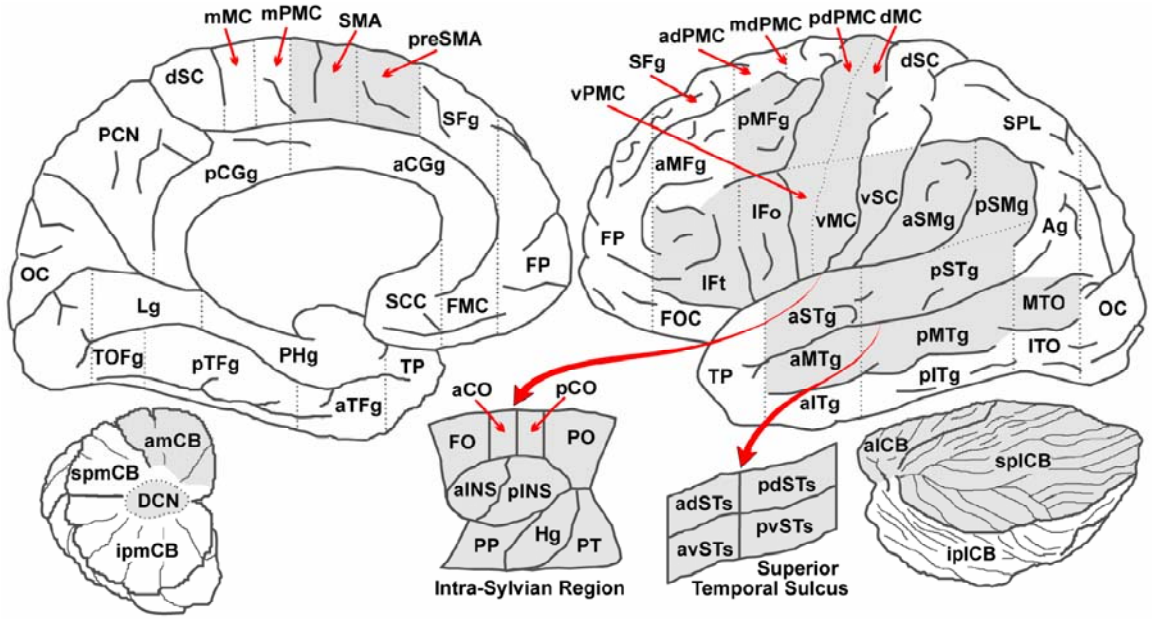


Figure 3-1. Cortical and cerebellar regions of interest for speech studies.
The cortical and cerebellar regions of interest (ROIs) are illustrated on schematics of the medial (left) and lateral (right) surfaces of the left hemisphere. Regions associated with speech production are shown in gray. Dashed lines indicate boundaries between adjacent regions. The ROIs that lie within the Sylvian fissure (Intra-Sylvian region) and the superior temporal sulcus are schematized as exposed flattened surfaces as indicated by the sweeping red arrows. Detached medial and lateral cerebellum surfaces are also shown in the lower left and lower right corners of the figure, respectively. The regional boundaries are based largely on the Caviness et al. (1996) and Schmahmann et al (2000) parcellation systems. *Abbreviations*: aCGg = anterior cingulate gyrus; adPMC = anterior dorsal premotor cortex; adSTs = anterior dorsal superior temporal sulcus; Ag = angular gyrus; aINS = anterior insula; alCB = anterior lateral cerebellum; amCB = anterior medial cerebellum; aMFg = anterior middle frontal gyrus; aMTg = anterior middle temporal gyrus; aSMg =anterior supramarginal gyrus; aSTg =anterior superior temporal gyrus; aTFg = anterior temporal fusiform gyrus; avSTs =anterior ventral superior temporal sulcus; DCN = deep cerebellar nuclei; dMC = dorsal primary motor cortex; FMC = fronto-medial cortex; FO = frontal operculum; FOC = fronto-orbital cortex; FP = frontal pole; Hg = Heschl's gyrus; IFo = inferior frontal gyrus, *pars opercularis*; IFt = inferior frontal gyrus, *pars triangularis*; ipmCB = inferior posterior medial cerebellum; iplCB = inferior posterior lateral cerebellum; Lg = lingual gyrus; mMC = medial motor cortex; mPMC = medial premotor cortex; mdPMC = middle dorsal premotor cortex; MTO = middle temporal occipital gyrus; OC = occipital cortex; pCGg = posterior cingulate gyrus; pdPMC = posterior dorsal premotor cortex; pdSTs = posterior dorsal superior temporal sulcus; PHg = parahippocampal gyrus; pINS = posterior insula; pMFg = posterior middle frontal gyrus; pMTg = posterior middle temporal gyrus; PO = parietal operculum; PP = planum polare; preSMA = pre-supplementary motor area; pSMg = posterior supramarginal gyrus; pSTg = posterior superior temporal gyrus; PT = planum temporale; pTFg = posterior temporal fusiform gyrus; pvSTs = posterior ventral superior temporal sulcus; SCC = subcallosal cortex; SFg = superior frontral gyrus; SMA = supplementary motor area; SPL = superior parietal lobule; splCB = superior posterior lateral cerebellum; spmCB = superior posterior medial cerebellum; TOF = temporal occiptial fusiform gyrus; TP = temporal pole; vMC = ventral motor cortex; vPMC = ventral premotor cortex; vSC = ventral somatosensory cortex.

within the sulcus from the gyral cortex lying on the exposed surface. A third cutting plane is made through the fundus of the superior temporal sulcus to divide the dorsal and ventral surfaces of the sulcus. The boundary separating anterior and posterior temporal lobe ROIs (superior, middle, and inferior temporal gyrus, and dorsal and ventral superior temporal sulcus) remains the anterior margin of Heschl's gyrus.

Heschl's Gyrus: To obtain a consistent, reliable definition of Heschl's gyrus, the guidelines for defining this area described by Kim et al. (2000) have been adopted. Their method addresses the difficulty encountered when multiple transverse gyri are present along the superior temporal plane. Heschl's gyrus is typically defined as lying between the first transverse fissure and Heschl's sulcus. A "double Heschl's" arises when a transverse fissure lies lateral to Heschl's sulcus, creating two "bumps" on the superior temporal plane. In the event of a "double Heschl's", if Heschl's sulcus extends caudomedially behind the insula, then it serves as the lateral border along the entire extent of Heschl's gyrus. If Heschl's sulcus terminates anterior to the posterior end of the insula, then it serves as the lateral border of Heschl's gyrus caudomedially to the point of its termination. Posterior to this point, Heschl's gyrus extends laterally to the more lateral transverse fissure. This method provides a reliable method for defining primary auditory cortex that reflects architectonic studies of this area (e.g., Rivier and Clarke, 1997, Wallace et al., 2002).

Posterior Extension of the Superior Temporal Gyrus: The posterior portion of the superior temporal gyrus extends posteriorly to the intermediate fissure of Jensen. As a

Region (ROIs)	BA	Possible Function
Heschl's gyrus (Hg)	41	Center frequency/frequency sweep encoding (Schreiner, 1995, Wang and Shamma, 1995); Sound level encoding (Brechmann et al., 2002)
Insula (aINS, pINS)	--	Articulatory planning (anterior; Dronkers, 1996, Kuriki et al., 1999, Wise et al., 1999)
Middle Temporal gyrus (aMTg, pMTg)	21	Lexical/semantic processing (Indefrey and Levelt, 2004)
Motor Cortex and anterior Central Operculum (dMC, vMC,aCO)	4,43	Primary motor cortex for speech articulators (Penfield and Roberts, 1959)
Planum Polare (PP)	52	Syntactic processing (Friederici et al., 2000)
Planum Temporale (PT)	42	Complex tone processing (Mummery et al., 1999); CV syllable perception (Jancke et al., 2002)
Inferior Frontal gyrus and Frontal Operculum (IFt, IFo, FO)	44,45	Semantic processing (Giraud and Price, 2001); Grapheme-to-phoneme conversion (Newman and Twieg, 2001)
Dorsal Premotor Cortex (adPMC, mdPMC, pdPMC)	6	Initiation and sequential planning of speech movements (Jonas, 1987)
Ventral Premotor Cortex (vPMC)	6	Planning of speech utterances at acoustic and articulatory levels
Somatosensory Cortex and posterior Central Operculum (vSC,pCO)	1,2,3,43	Primary somatosensory cortex for speech articulators (Penfield and Roberts, 1959)
Superior Temporal gyrus (aSTg, pSTg)	22	Anterior: processing of speech-like sounds (Binder et al., 2000, Scott et al., 2000); Posterior: phonological processing for speech perception and production (Hickok and Poeppel, 2000, Buchsbaum et al., 2001)
Superior Temporal sulcus (adSTs, avSTs, pdSTs, pvSTs)	22	Anterior: phoneme processing (Belin et al., 2000, Binder et al., 2000, Scott et al., 2000); Posterior: perception/retrieval of single words (Wise et al., 2001)
Supplementary Motor Area (SMA, preSMA)	6	Motor sequencing (Wildgruber et al., 1999); Initiation of articulation (Ziegler et al., 1997); Articulatory planning (Indefrey and Levelt, 2004)
Supramarginal gyrus and Parietal Operculum (aSMg, pSMg, PO)	40	Phonological processing for speech perception (Caplan et al., 1995, Celsis et al., 1999) and production(Geschwind, 1965, Damasio and Damasio, 1980); Sound localization of speech source (Weeks et al., 1999, Rauschecker and Tian, 2000)

Table 3-1. Cortical regions of interest and contributions to speech processes.
Regions are grouped according to hypothesized functional characteristics. Corresponding Brodmann areas (BA) are provided where appropriate.

result, the posterior portion of supramarginal gyrus borders superior temporal gyrus ventrally, rather than extending further ventrally to the superior temporal sulcus, as it does in the Caviness et al. system. This modification better reflects the boundary between BA 40 and BA 22.

Insular Region: The insula is divided into anterior and posterior regions along the central insular sulcus. This change is motivated by studies that suggest a role in articulatory planning within the anterior insula (e.g. Dronkers, 1996).

Motor Cortices: The precentral gyrus contains both primary motor and premotor cortices. Therefore, we divide the gyrus into anterior (premotor) and posterior (motor) regions. Since the ventral portion of the precentral gyrus is devoted to the speech articulators, we also divide the premotor and motor regions into ventral and dorsal subregions. On the medial surface, anterior to the precentral sulcus, the supplementary motor area is divided into anterior (preSMA) and posterior regions (SMA) based on recent results that suggest separate functional roles for these two regions (e.g. Boecker et al., 1998). The rostro-caudal level of the anterior commissure serves to divide the two SMA regions. The anterior region extends rostrally to the level of the interior portion of the genu of the corpus callosum, based on the parcellation system of Crespo-Faccoro et al.(2000). Two additional premotor regions are defined immediately lateral to the two SMA regions, on the dorsal surface. They extend laterally to the superior frontal sulcus and share the same boundary markers as the adjacent SMA regions.

Somatosensory Cortex: The portion of the postcentral gyrus lateral to the intraparietal sulcus is labeled ventral somatosensory cortex. This region receives sensory information

from the speech articulators (Penfield and Roberts, 1959).

Remaining cortical ROIs: Recent functional imaging work has demonstrated the involvement of a wide expanse of the cerebral cortex in speech processing. It is therefore useful to anatomically characterize the entire cerebral cortex, not simply the core speech-related areas described above. To assess activity in the remainder of the cerebral cortex,

Cortical ROIs	Caviness et al. Label	Brodmann Areas
Angular Gyrus (Ag)	AG	39
Cingulate Gyrus (aCGg, pCGg)	CGa, CGp, PAC	23, 24, 29, 30, 33
Dorsal Somatosensory Cortex (dSC)	POG	1, 2, 3, 5
Frontal Medial Cortex (FMC)	FMC	11, 12, 32
Frontal Orbital Cortex (FOC)	FOC	11, 13, 14, 47
Frontal Pole (FP)	FP, PAC	9, 10, 12
Inferior Temporal Gyrus (aITg, pITg)	T3a, T3p	20, 37
Inferior Temporal Occipital Gyrus (ITO)	TO3	37, 19
Lingual Gyrus (Lg)	LG	18, 19, 37
Middle Frontal Gyrus (aMFg, pMFg)	F2	8, 9, 46
Middle Temporal Occipital Gyrus (MTO)	TO2	19, 37
Occipital Cortex (OC)	OP, OLs, OLi, OF, LG, CALC, SCAL, CN	17, 18, 19
Parahippocampal Gyrus (PHg)	PHa, PHp	27, 28, 34, 35, 51
Precuneus Cortex (PCN)	PCN	7a, 7b, 23, 31
Subcallosal Cortex (SCC)	SC	12, 15, 24, 25, 32, 33
Superior Frontal Gyrus (SFg)	F1, PAC	8, 9
Superior Parietal Lobule (SPL)	SPL	7a, 7b
Temporal Fusiform Gyrus (aTFg, pTFg)	TFa, TFp	20, 36, 37
Temporal Occipital Fusiform Gyrus (TOF)	TOF	19, 37
Temporal Pole (TP)	TP	38

Table 3-2. Cortical ROIs outside core speech network.
ROIs covering the remainder of the cerebral cortex are listed along with approximate Caviness et al. (1996) and Brodmann area correspondence. Several of the regions listed consist of anterior and posterior segments. Note that a single ROI may consist of cortex represented by several Caviness and/or Brodmann areas. In these cases, all the areas contributing to the ROI are listed. Conversely, a single Caviness and/or Brodmann area may represent cortex in multiple ROIs.

we have largely adopted the CMA system. A few minor modifications were made to accommodate the changes to the speech-related regions described above. When necessary, the nomenclature was also made consistent with that used to describe the speech-related ROIs. For example, the post-central gyrus, labeled POG by Caviness et al.(1996), is now split into two ROIs, ventral and dorsal somatosensory cortex (vSC, dSC). Table 2 provides a list of the remaining cortical ROIs along with their approximate correspondence with the parcellation units of Caviness et al. and Brodmann areas. The schematic in Figure 3-1 shows the location of these ROIs on the cortical surface.

The principal areas of modification lie at the rostral and caudal ends of the brain. Rostrally, the paracingulate gyrus has been eliminated. The superior frontal gyrus (SFg), frontal pole (FP), and frontal medial cortex (FMC) extend ventrally, caudally, and dorsally, respectively, to the cingulate sulcus. This change allows for a more reliable parcellation of frontomedial cortex as it eliminates reliance upon the paracingulate sulcus, which is typically highly segmented and often difficult to locate. In the event of a "double cingulate" (see Ono et al., 1990), the outer cingulate sulcus serves as the rostrodorsal border of the cingulate gyrus. Also on the frontal lobe, SFg (termed F1 in the CMA system) has been truncated posteriorly to allow for the presence of the dorsal premotor ROIs (adPMC and mdPMC) laterally and the pre-supplementary motor area (preSMA) medially.

Caudally, regions of the occipital lobe have been lumped in to one ROI, the occipital cortex (OC). As in the CMA system, the occipital lobe is bordered anteriorly by the parietooccipital fissure medially and the point of opercularization of the intraparietal

sulcus laterally (Plane F). However, cortex behind these boundaries is now collapsed into a single ROI. In addition to eliminating a number of occipital ROIs, the posterior border of lingual gyrus (Lg) is moved anteriorly to Plane F. These changes allow for the elimination of a number of boundary planes in the CMA system that are difficult to define, in particular the rostral and caudal ends of the cuneal sulcus, without sacrificing anatomical specificity that is relevant to speech research.

3.2 Subcortical regions of interest

Cerebellum. The cerebellum has been shown to play a role in both speech production and speech perception (Ackermann et al., 1999, De Nil et al., 2001, Wildgruber et al., 2001, Mathiak et al., 2002). To better localize cerebellum involvement in speech related tasks, we have adopted a simplified version of the cerebellum parcellation system described by Schmahmann et al (2000; see Table 1-3). The cortex of each cerebellar hemisphere is split into six ROIs, three medial and lateral pairs (see Figure 3-2). Dividing medial from lateral regions is the sagittal plane that falls one third of the way between the midline and the lateral extent of each hemisphere, termed Plane Cb. The primary and horizontal fissures, along with the hemispheric margins provide the remainder boundaries for the cortical ROIs. Thus, the anatomical markers that define region boundaries are easily identified. The anterior medial and anterior lateral ROIs (amCB, alCB) lie anterior to the primary fissure. Behind this fissure, superior and inferior regions are divided by the horizontal fissure. The superior posterior medial and lateral ROIs (spmCB, splCB) lie dorsal to the horizontal fissure while the inferior posterior medial and lateral (ipmCB,

	Boundaries			
Cerebellum ROIs	**Anterior**	**Posterior**	**Medial**	**Lateral**
Anterior Lateral (alCB)	Anterior H.M.	Primary Fissure	Plane Cb	Lateral H.M.
Anterior Medial (amCB)	Anterior H.M.	Primary Fissure	Midline	Plane Cb
Inferior Posterior Lateral (iplCB)	Posterior H.M.	Posterior H.M.	Plane Cb	Horizontal Fissure
Inferior Posterior Medial (ipmCB)	Posterior H.M.	Posterior H.M.	Midline	Plane Cb
Superior Posterior Lateral (splCB)	Primary Fissure	Posterior H.M.	Plane Cb	Horizontal Fissure
Superior Posterior Medial (spmCB)	Primary Fissure	Posterior H.M.	Midline	Plane Cb
Deep Cerebellar Nuclei (DCN)*	Brainstem, Posterior End	alCB, Posterior End	Midline	Dentate Nuc., Lateral Border

Table 3-3. Cerebellum regions of interest and their anatomical boundaries.
Plane Cb is a sagittal plane one third of the way between the midline of the cerebellum and its lateral extent. H.M. = hemispheric margin. *Because the deep cerebellar nuclei are difficult to view on standard structural data sets acquired on 1.5T or 3T magnets, these boundaries serve as easily identified gross approximations of the extents of the deep cerebellar nuclei, provided brains are in Talairach space (oriented along the anterior commissure – posterior commisure line). The DCN ROI lies entirely within the region of the cerebellum that is enclosed by the cerebellar cortical ribbon, therefore this entire region could serve as an alternative ROI definition.

iplCB) lie ventral to it.

Finally, we define an ROI representing the deep cerebellar nuclei (DCN). The difficulty associated with localizing the deep cerebellar nuclei on MRI slices necessitates a gross definition relative to the other ROIs described here. The nuclei lie medially within cerebellar white matter (the area completely enclosed by the cortical ribbon after the cortex has been segmented). The goal of the region definition is to eliminate as much of this area as possible that does not correspond the the nuclei without discarding any portion of the nuclei. Thus, the DCN ROI is an overestimate. The lateral boundary is approximated by the lateral extent of the dentate nucleus, the only deep cerebellar nucleus that is consistently viewable in standard structural MRI data sets. The anterior

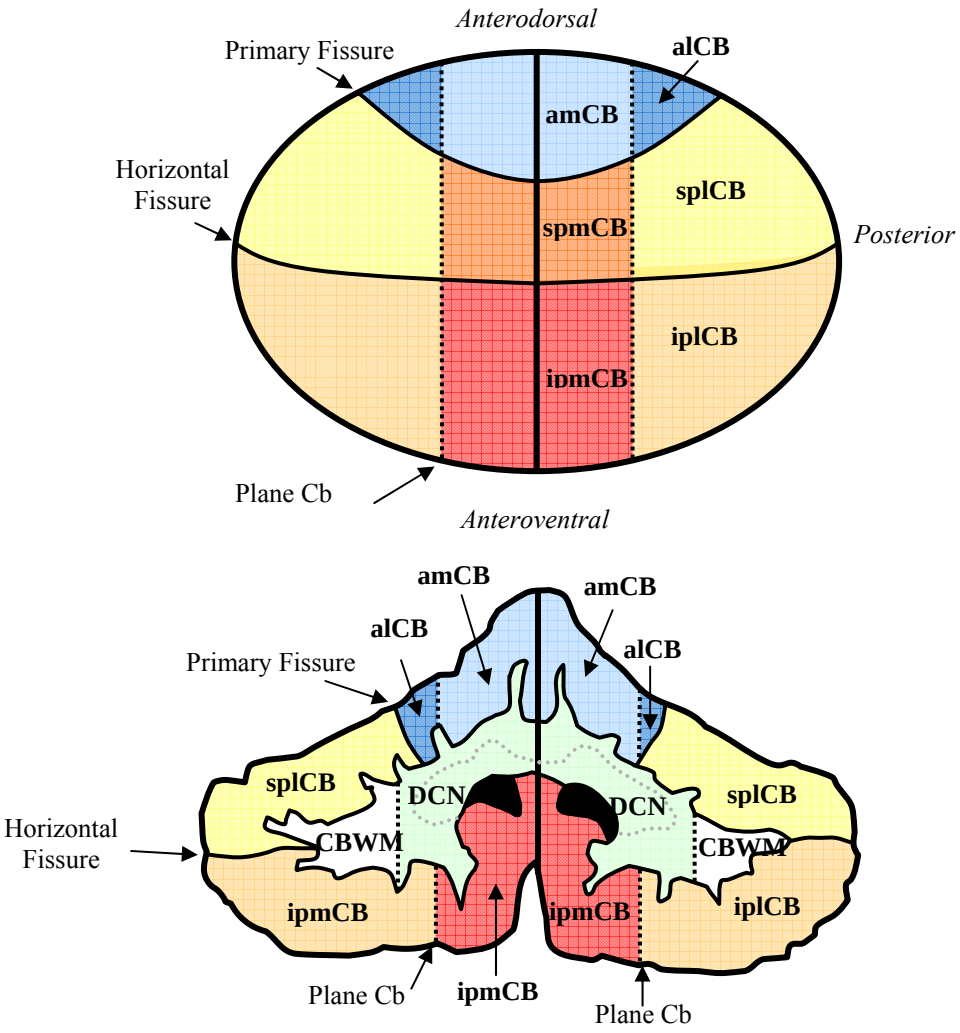


Figure 3-2. Cerebellum regions of interest.
Top: Labeled ROIs on a schematized flattened cerebellar cortical surface. The top of the figure corresponds to the anterior extent of the dorsal surface of the cerebellum and the bottom corresponds to the anterior extent of the ventral surface. Plane Cb marks the plane one third of the way between the cerebellum midline and the lateral margin. The primary and horizontal fissures serve as the other boundary markers. *Bottom*: Labeled coronal slice. The cerebellum ROIs are shown on a representative coronal slice. The DCN ROI can be seen within the cerebellum white matter (CBWM). The light gray dotted lines represent an approximate outline for the deep cerebellar nuclei within the DCN ROI. See the text for a description of the boundaries of DCN. Refer to Table 3-3 for an explanation of ROI abbreviations.

and posterior DCN borders are grossly defined using extrinsic anatomical markers. The DCN ROI begins anteriorly on the posterior-most coronal slice containing brainstem and ends posteriorly on the posterior-most slice coronal slice containing amCB. This rostro-caudal extent provides an overestimate of the range of slices containing deep cerebellar nuclei. The gray-white interface forms the dorsal and ventral boundaries. The bottom of Figure 3-2 shows a labeled coronal slice through the cerebellum. Both the cortical and DCN ROIs can be seen.

Other subcortical regions: Subcortical regions, aside from the cerebellum, are defined according to the FreeSurfer segmentation algorithm (Fischl et al., 2002). The subcortical classifier labels subcortical regions of interest based on gray – white intensity values and local inter-regional spatial probabilities derived from a set of manually segmented anatomical MRI volumes. The manual segmentation was based on the procedure developed at Massachusetts General Hospital's Center for Morphometric Analysis (Kennedy et al., 1989, Filipek et al., 1994). Classified regions that are associated with speech processing include the thalamus, caudate, putamen, globus pallidus, and brainstem. Other regions identified by the classifier include the lateral, third and fourth ventricles, hippocampus and the amygdala.

3.3 Conclusions

The analysis of functional data sets is greatly hindered by the high degree of individual anatomical variability across brains (Nieto-Castanon et al., 2003). To ensure comparison of like brain areas, regions of interest must be defined according to individual anatomical

markers prior to averaging. Here we have described a parcellation system that encompasses the entire cerebral cortex and the cerebellum based on individual anatomical markers that are discernable from standard MRI data sets. Based largely on the parcellation scheme described by Caviness et al.(1996), the system was designed to be particularly well suited for studies of speech processing. Cortical areas shown to be involved in speech production and/or perception were redefined to reflect known functional boundaries. To this end, several of the Caviness et al. parcellation units were subdivided into more discrete ROIs, particularly the superior temporal sulcus and premotor areas. These changes provide greater power for the localization and functional characterization of speech-relevant cortical regions. Conversely, for regions that have not been shown, as yet, to play a specific role in speech processing, ROIs have been combined to allow for more reliable parcellation. For instance, the posterior occipital lobe has been lumped into a single ROI, and the paracingulate gyrus has been eliminated. These changes permit the removal of boundary markers that are difficult to locate and thus make it easier to consistently define regions.

The parcellation system described here is meant to serve as a starting point. Several of the regions, even those known to play a role in speech processing, such as the cerebellar ROIs, are defined crudely. This was done either in the interest of definition reliability or because there is insufficient information to support more strictly defined regions. Advances in imaging technology will lead to greater ease in localizing boundary markers. The potential for greater advances, however, lies in well-designed functional studies of speech processing that target specific brain regions. For instance, studies that

have recently begun utilize stimulus parameterizations that will allow us to localize topographic maps along the superior temporal plane. The goal of this research is to further subdivide this core auditory area into more functionally relevant regions. Other studies are searching for specific sites within the cerebellum and premotor regions that contribute to speech production. Thus, the parcellation scheme will be continually updated according to the results of speech-related research.

CHAPTER 4

AN INVESTIGATION OF THE NEURAL SUBSTRATES OF AUDITORY FEEDBACK CONTROL OF SPEECH

4.1 Auditory feedback in the DIVA model

The DIVA model of speech production (Guenther et al., 1998, Guenther et al., 2006) is a quantitatively defined neuroanatomical model that provides a parsimonious account of how auditory feedback is used for both feedback control and for tuning feedforward commands. According to the model, feedforward and feedback commands are combined in primary motor cortex to produce the overall muscle commands for the speech articulators. Both control processes are initiated by activating cells in a *Speech Sound Map* located in the left ventral premotor areas, including Broca's area in the opercular portion of the inferior frontal gyrus. Activation of these cells leads to the readout of excitatory feedforward commands through projections to the primary motor cortex. Additional projections from the *Speech Sound Map* to higher-order auditory cortical areas located in the posterior superior temporal gyrus and planum temporale encode auditory targets for the syllable to be spoken. The *Speech Sound Map* -to-auditory error cell projections are hypothesized to have a net inhibitory effect on auditory cortex. The auditory targets encoded in these projections are compared to the incoming auditory signal by cells in the *Auditory Error Map* that respond when a mismatch is detected between the auditory target and the current auditory feedback signal. When a mismatch is detected, projections from the auditory error cells to motor cortex transform the auditory

error into a corrective motor command. The model proposes that these corrective motor commands are added (see Equation 2 above) to the feedforward command for the speech sound. The updated command is stored in projections from the *Speech Sound Map* to the *Articulator Velocity Map* so that future productions of the sound will contain the corrective command. In other words, the feedforward control system becomes tuned by incorporating the commands sent by the auditory feedback control system on earlier attempts to produce the syllable.

Because the DIVA model is defined both quantitatively and neuroanatomically, the activity of model components in computer simulations of perturbed and unperturbed speech can be compared directly to task-related blood-oxygen-level-responses (BOLD) in speakers performing the same tasks. According to the model, unexpected auditory feedback should induce activation of auditory error cells in the posterior superior temporal gyrus and planum temporale (Guenther et al., 2006). Auditory error cell activation then drives a compensatory motor response marked by increased activation of ventral motor, premotor, and superior cerebellar cortex.

4.2 Identifying the neural substrates of auditory feedback control with fMRI

The current study utilizes auditory perturbation of speech, in the form of unpredictable upward and downward shifts of the first formant frequency, to identify the neural circuit underlying auditory feedback control of speech movements and to test DIVA model predictions regarding feedback control of speech. Functional magnetic resonance imaging

(fMRI) was performed while subjects read aloud monosyllabic words projected orthographically onto a screen. A sparse sampling protocol permitted vocalization in the absence of scanner noise (Yang et al., 2000, Le et al., 2001, Engelien et al., 2002). An electrostatic microphone and headset provided subjects with auditory feedback of their vocalizations while in the scanner. On a subset of trials, an unpredictable real-time F1 shift was introduced to the subject's auditory feedback. Standard voxel-based analysis of neuroimaging data was supplemented with region of interest (ROI) analyses (Nieto-Castanon et al., 2003) to improve anatomical specificity and increase statistical power. Compensatory responses were also characterized behaviorally by comparing the formant frequency content of vocalizations made during perturbed and unperturbed feedback conditions. Structural equation modeling was used to assess changes in effective connectivity that accompanied increased use of auditory feedback control.

Eleven right handed native speakers of American English (6 female, 5 male; 23-36 years of age, mean age = 28) with no history of neurological disorder participated in the study. All study procedures, including recruitment and acquisition of informed consent, were approved by the institutional review boards of Boston University and Massachusetts General Hospital. A scanner problem that resulted in the introduction of non-biological noise in acquired scans required the elimination of imaging data from one subject.

Experimental Protocol. Scanning was performed with a Siemens Trio 3T whole-body scanner equipped with a volume transmit-receive birdcage head coil (USA Instruments, Aurora, OH) at the Athinoula A. Martinos Center for Biomedical Imaging, Charlestown,

MA. An electrostatic microphone (Shure SM93) was attached to the head coil approximately 3 inches from the subject's mouth. Electrostatic headphones (Koss EXP-900) placed on the subject's head provided acoustic feedback to the subject at the beginning of each trial. Each trial began with the presentation of a speech or control stimulus projected orthographically on a screen viewable from within the scanner. Speech stimuli consisted of 8 /CɛC/ words (beck, bet, deck, debt, peck, pep, ted, tech) and a control stimulus (the letter string 'yyy'). Subjects were instructed to read each speech stimulus as soon as it appeared on the screen and to remain silent when the control stimulus appeared. Stimuli remained onscreen for 2 seconds.

An experimental run consisted of 64 speech trials (8 presentations of each word) and 16 control trials. On a subset of speech trials, F1 of the subject's speech was altered before being fed back to the subject. Of the 8 presentations of each stimulus in an experimental run, F1 was increased by 30% on 1 presentation (*shift up* condition trial), decreased by 30% on 1 presentation (*shift down* condition trial), and unaltered in the remaining 6 presentations (*no shift* condition trials). Trial order was randomly permuted within each run; presentation of the same stimulus type on more than 2 consecutive trials was prohibited as were consecutive F1 shifts in the same direction regardless of the stimulus. To allow for robust formant tracking and to encourage the use of auditory feedback mechanisms, subjects were instructed to speak each word slowly and clearly; production of each stimulus was practiced prior to scanning until the subject was able to consistently match a sample production. Still, subject mean vowel duration across all trial types ranged from 357 to 593 ms with standard deviations (SD) ranging from 44 to 176

ms). Paired *t*-tests indicated no utterance duration differences between the mean *no shift* and mean lumped *shift* responses (combined *shift up* and *shift down* responses; $df = 10$, $p = 0.79$) or between the *shift up* and *shift down* responses ($df = 10$, $p = 0.37$). Each subject performed 3 or 4 runs in a single scanning session, depending on subject fatigue and tolerance for lying motionless in the scanner. Stimulus delivery and scanner triggering were performed by Presentation Version 0.80 (www.neurobs.com) software.

MRI Data Acquisition. A high resolution T1-weighted anatomical volume (128 slices in the sagittal plane, slice thickness = 1.33 mm, in-plane resolution = 1 mm^2, TR = 2000 ms, TE=3300 ms, flip angle = 7°, FOV = 256 mm^2) was obtained prior to functional imaging. Functional volumes consisted of 32 T2*-weighted gradient echo, echo planar images covering the whole brain in the axial plane, oriented along the bicommissural line (slice thickness = 5 mm, in-plane resolution = 3.125 mm^2, skip = 0 mm, TR = 2000 ms, TE 30 ms, flip angle = 90˚, FOV = 200 mm^2).

Functional data were obtained using an event-triggered sparse sampling technique (Yang et al., 2000, Le et al., 2001, Engelien et al., 2002). The timeline for a single trial is shown in Figure 4-1. Two consecutive volumes (each volume acquisition taking 2 seconds) were acquired beginning 5 seconds after trial onset. The 5 second delay period was inserted to allow collection of BOLD data at or near the peak of the hemodynamic response to speaking (estimated to occur approximately 4-7 seconds after vocalization). Auditory feedback to the subject was turned off during image acquisition to prevent transmittance of scanner noise over the headphones. The next trial started after another 3 second delay period, for a total trial length of 12 seconds and a total run length of 16

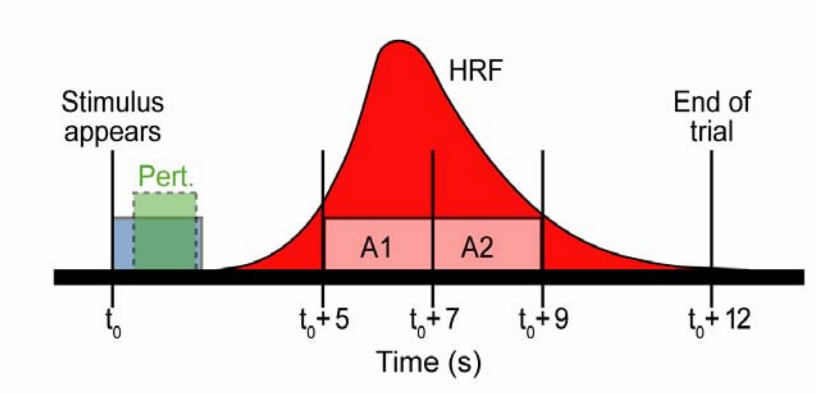


Figure 4-1. Timeline of a single trial in the event-triggered sparse sampling protocol.
At the onset of each trial, the visual stimulus appeared and remained onscreen for 2 seconds (blue rectangle). On perturbed trials, auditory feedback was shifted during the subject's response (green). 3 seconds after stimulus offset, two whole-brain volumes were acquired (A1, A2). Data acquisition was timed to cover the peak of the hemodynamic response to speech; the putative hemodynamic response function (HRF), is schematized in red. The next trial started 3 seconds after data acquisition was complete, resulting in a total trial length of 12 seconds.

minutes. The sparse sampling design afforded several important advantages. First, it allowed subjects to speak during relative silence, a more natural speaking condition than speech during loud scanner noise. Second, it allowed for online digital signal processing of the speech signal to apply the perturbation, which is not possible in the presence of scanner noise. Finally, since scanning is carried out only after speech has ceased, it eliminates artifacts due to movement of the head and changing volume of the oral cavity during speech.

4.3 Real-time auditory feedback perturbation

Subject vocalizations were transmitted to a Texas Instruments DSK6713 digital signal processor (DSP). Prior to reaching the DSP board, the original signal was amplified (Behringer Eurorack UB802 mixer) and split into two channels using a MOTU 828mkII

audio mixer. One channel was sent to the DSP board and the other to a laptop where it was recorded using Audacity 1.2.3 audio recording software (44,100 kHz sampling rate). Following processing, the DSP output was again split into two channels by the MOTU board, one channel was sent to the subject's headphones, the other to the recording laptop.

F1 tracking and perturbation, and signal resynthesis were carried out in the manner described by Villacorta and colleagues (2007). The incoming speech signal was digitized at 8 kHz then doubled buffered; data were sampled over 16 ms blocks that were incremented every 8 ms. Each 16 ms bin was sampled at 8 kHz and then pre-emphasized to compensate for the -6dB/octave high-frequency spectral slope (glottal roll-off) typically present in the speech signal. To remove onset and offset transients, a hamming window was convolved with the pre-emphasized signal. An 8^{th} order linear predictive coding (LPC) analysis was then used to identify formant frequencies. F1 was then altered according to the trial type before the signal was re-synthesized and sent to the subject. A delay of 17 ms was introduced by the DSP board. Unperturbed trials were processed through the DSP in exactly the same manner as the perturbed trials except that the original F1 value was preserved, rather than shifted, during resynthesis; this was done to limit the difference in auditory feedback between perturbed and unperturbed trials to the first formant shift. The upward F1 shift had the effect of moving the vowel sound toward /æ/ (e.g., *bet* → *bat*); a downward shift moved the vowel toward /ɪ/ (e.g., *bet* → *bit*). When questioned following scanning, subjects reported no awareness of the feedback delay or alteration.

4.3.1 Imaging data analysis

Voxel-based Analysis. Voxel-based analysis was performed to assess task-related effects in a standardized coordinate frame using conventional image data analysis techniques, thereby permitting easier comparison with results from prior investigations. Image data were preprocessed using tools from the SPM2 software package provided by the Wellcome Department of Imaging Neuroscience, University College London (Friston et al., 1995b; http://www.fil.ion.ucl.ac.uk/spm/). Functional images were realigned to the mean EPI image (Friston et al., 1995a), coregistered with the T1-weighted anatomical dataset (Collignon et al., 1995), and spatially normalized into standard stereotaxic space using the EPI template provided by the Montreal Neurological Institute (MNI ICBM-152; Evans et al., 1993, Mazziotta et al., 2001). Functional images were then spatially smoothed (12 mm full-width-half-maximum Gaussian kernel) and globally scaled. Realignment parameters were included as covariates of non-interest in the study design prior to parameter estimation. Remaining global differences between the two volume acquisitions within each trial were removed during parameter estimation by a covariate that modeled these differences. The BOLD response for each event was modeled using a single-bin finite impulse response (FIR) basis function spanning the time of acquisition of the two consecutive volumes. Responses from the *shift up* and *shift down* conditions were averaged to form a single "lumped" shifted speech condition (hereafter referred to as *shift*) for fMRI analysis. Voxel responses were fit to a set of condition (*shift*, *no shift*, *baseline*) regressors according the general linear model.

Group statistics were assessed using fixed and mixed effects procedures. In mixed effects analysis, contrast-of-interest images were first generated for each subject by comparing the relevant condition parameter estimates on a voxel-by-voxel basis. Estimates for these analyses were obtained using a general linear model where conditions are treated as fixed effects. Group effects were then assessed by treating subjects as random effects and performing one-sample *t*-tests across the individual contrast images. The resulting group parametric maps were thresholded by a corrected significance level to ensure a false discovery rate (FDR) < 5%. A map of normalized effect sizes for those voxels surpassing the significant *t* threshold was then created; suprathreshold voxel effects were divided by the mean significant ($p < 0.05$ uncorrected) effect of the *shift – baseline* contrast; this procedure permits assessment of relative activations. The contrast maps are shown in terms of effect size to provide a comparison of the how BOLD responses differed in the contrasted conditions.

ROI Analysis. Region-of-interest (ROI) analysis was performed to test hypotheses regarding the response of specific regions to the task manipulation. According to the DIVA model, shifted feedback is expected to result in increased bilateral activation of planum temporale (PT), posterior temporal gyrus (pSTg), ventral motor and premotor cortex (vMC, vPMC), and the anterior medial cerebellum (amCB). Regional effects also served as the input for post-hoc tests of laterality and structural equation modeling. Regions included in these analyses were selected based on the finding of uncorrected significance in both the voxel-based and ROI-based analyses. Delineation of ROI

boundaries was based on a set of *a priori* anatomical definitions and was independent of the results of the voxel-based analyses.

Cortical, subcortical, and cerebellar ROIs were created using Freesurfer (http://surfer.nmr.mgh.harvard.edu) image processing software. Segmentation of gray and white matter structures (Fischl et al., 2002) and cortical surface reconstruction were performed on each anatomical volume (Fischl et al., 2004). Subcortical ROIs were segmented according to the Freesurfer-supplied subcortical training set. The Freesurfer cortical classifier (Fischl et al., 2002) was trained on a set of 14 manually parcellated brains. The manual parcellations used as the classifier training set were based on the parcellation system described in Chapter 3. Cerebellar ROIs were parcellated by applying a cerebellar classifier to the segmented cerebellar gray and white matter in the same manner as that used for the subcortical ROIs (Fischl et al., 2002). The cerebellar classifier was based on manually parcellated cerebellums from the same 14 brains used to train the cortical classifier.

Characterization of BOLD responses within each ROI was performed according to the procedure described by Nieto-Castanon et al. (2003). Following spatial realignment, functional data were subjected to a rigid body transform and co-registered with the structural data set. The BOLD response averaged across all voxels within each ROI mask was then extracted. Regional noise temporal correlations were removed by whitening a fit of the estimated noise spectrum within each ROI. Average regional responses for each event were modeled using a single-bin FIR and fit to the same set of condition regressors (*shift, no shift, baseline*) used in the voxel based analyses.

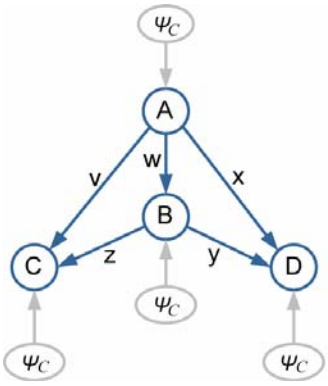
Group effects were assessed by first computing regional contrasts for each subject. The regional contrasts were then pooled and tested for significance using one-sample *t*-tests. Regional effect sizes were normalized by the mean significant ($p < 0.05$, uncorrected) effect. In a first set of tests, the *a priori* hypotheses that responses in PT, pSTg, vMC, vPMC, and amCB are greater in the *shift* than the *no shift* conditions were tested. Probabilities were corrected to ensure FDR < 5%. Subsequent tests for significance were performed on the remaining ROIs ($n = 132$; only the posterior occipital cortex was excluded from ROI analysis). Regional effects associated with brain areas that were significant (uncorrected) in the *shift – no shift* contrasts in both the ROI and voxel-based results were used as in *post-hoc* tests of laterality and structural equation modeling. Laterality effects were determined by pooling the effect for each ROI within each hemisphere across subjects ($n = 10$) and performing a paired *t*-test on the pooled data from each hemisphere.

In addition to providing average regional responses, our ROI analysis permits visualization of effects within ROIs (Nieto-Castanon et al., 2003). Briefly, voxel responses within each ROI were projected onto a spherical representation of the cortical surface. A reduced set of temporal eigenvariates was then created for each ROI by projecting the 2-D surface responses from each ROI onto a set of 15 orthogonal spatial Fourier bases and keeping only those components with low spatial frequency. The resulting set of eigenvariates for each ROI was fitted to the condition predictors and "eigenvariate contrasts" were calculated by comparing the appropriate condition effects. A spatial response profile was created for each subject by projecting the eigenvariate

contrasts back onto the spherical surface using the transpose of the original set of orthogonal spatial bases. Spatial profiles were then averaged across subjects, normalized as described above, and the resulting group profile was flattened for display.

4.3.2 Assessing effective connectivity by structural equation modeling

Structural equation modeling (SEM) provides a means to characterize the interactions between these regions. A simple example of a SEM, adapted from McIntosh & Gonzalez-Lima (1994), is shown in Figure 4-2. The path diagram on the left describes a network of 4 regions (*A,B,C,D*) with projections between regions indicated by blue arrows. The relative strength of the projections is given by the weights or path coefficients *v, w, x, y,* and *z*. On the right is a system of structural equations that models the variance of each region in the network as a function of the variance of the other regions in the network. Each equation also includes an error term (Ψ_N;) that accounts for residual variance not captured by the regions and connections in the model. The goal of SEM is to estimate the



$$A = \Psi_A$$
$$B = wA + \Psi_B$$
$$C = vA + zB + \Psi_C$$
$$D = xA + yB + \Psi_D$$

Figure 4-2. Structural equation model example.
Left: A simple path diagram for a network of 4 regions (*A, B, C, D*). The weights or path coefficients *v, w, x, y*, and *z* describe the relative strength of connections between regions in the network (blue arrows). *Right*: A set of structural equations that describes the variance of each region in the network as a function of the variance of the other regions in the network plus residual error (Ψ_N; gray circles).

path coefficients by finding a solution to the set of structural equations that minimizes the difference between the covariance matrix resulting from the experimental data and that implied by the model.

Interactions between regions involved in auditory feedback control were assessed using covariance SEM (McIntosh and Gonzalez-Lima, 1994). Network connectivity, e.g., the structural model, was constrained by interactions conceptualized within the DIVA model. This a departure from the standard application using established one-to-one anatomical connections (McIntosh and Gonzalez-Lima, 1994), typically based on studies of non-human primates, as the primary constraint on network connectivity. The DIVA model describes reciprocal connections between ventral frontal and posterior temporal cortex that enable a comparison of expected and realized acoustic consequences and drive compensatory movements in the event an error is detected. Effective connectivity was therefore assessed between those ventral frontal and posterior temporal regions found significantly more active in the *shift – no-shift* contrast.

Functional imaging data were prepared for SEM analysis by first dividing the mean regional responses from each functional run into two series, one consisting of only the first volume acquired for each trial, the other consisting of only the second volume acquired for each trial. The two series were de-trended and averaged to give a single response for each trial. Trials were then assigned to the appropriate group and concatenated within each subject. Outliers (> 3 standard deviations) were removed following mean correction, variance was unitized, and responses were concatenated across all subjects.

Covariance SEM (McIntosh and Gonzalez-Lima, 1994) was performed with AMOS 7 (http://www.spss.com/amos/index.htm) software. Path coefficients between observed variables were determined by maximum likelihood estimate. Differences in effective connectivity due to condition were assessed using the stacked model approach (Della-Maggiore et al., 2000). For a given network, the χ^2 goodness-of-fit measure was determined for a null model in which path coefficients are constrained to be equal between the two conditions and an unconstrained model in which they are allowed to vary. A comparison of the model fits, $\chi^2_{diff} = \chi^2_{null} - \chi^2_{uncon}$, was calculated using degrees of freedom equal to the difference in degrees of freedom between the two models. A significant χ^2_{diff} value was interpreted as evidence that the alternative model was a better fit than the null model and that the global effective connectivity of the network differed between the two conditions. The commonly used goodness-of-fit (GFI) and adjusted goodness-of-fit (AGFI) indices and the root mean square residual (RMR) and root mean square error of approximation (RMSEA) of differences between sampled and estimated variances and covariances were used to assess model fit (see Schumacker & Lomax (2004) and Hu and Bentler (1999) for a detailed description of these criteria). The alternative model also had to meet the AMOS 7 stability index criteria for both conditions. Connectivity between regions was constrained to meet acceptable fit and stability criteria and to produce path coefficients that were significant in at least one of the two conditions. These criteria were chosen to bias the network toward a parsimonious account of effective connectivity in the two conditions while still providing a good fit to the data. Significant connectivity was determined by converting estimated path

coefficients to *z* statistics (*z* = coefficient estimate/standard error estimate) then performing a two-tailed test that the absolute value of *z* is greater than 0 with $p > 0.05$). Comparisons of path coefficients in the two conditions were performed for the accepted model; *z* statistics were calculated by dividing the difference between the estimated coefficient in each condition by the estimated standard error of the difference.

Acoustic Data Analysis. Subject responses were identified, isolated from the remainder of the acoustic recording, and resampled at 16 kHz. MATLAB software was created to identify the first two formants in the original and shifted feedback signals using LPC analysis. Analysis was performed on 20 ms samples of the speech signal, incremented every 4 ms. Vowel onset and offset, F1 and F2 contours, and signal intensity were estimated. Formant estimates for each utterance were inspected visually. If the initial formant estimation indicated a poor fit, the number of LPC coefficients was manually changed. The LPC order typically ranged between 16-18 coefficients for male speakers and 14-16 for female speakers. Vowel onset and offset estimates were also manually adjusted as needed.

To determine whether subjects compensated for perturbations during the shift conditions, subject-specific baseline F1 traces were first created by averaging the *no shift* traces within each stimulus type. Traces were aligned to the onset of voicing. Averaging was done on a point-by-point basis and was restricted to time points that fell within the 80^{th} percentile of all utterance lengths for a given subject to ensure a sufficient number of samples at each time point. Each shifted-feedback trace was then divided by the appropriate subject- and stimulus-matched baseline *no shift* trace. The resulting F1 *shift*

up and *shift down* compensation traces were averaged across the 8 stimulus types within each subject. The shifted feedback responses were characterized with respect to normal feedback responses to account for individual formant variation.

One-sample, two-tailed *t*-tests were performed at each time step to test for differences during the shifted conditions relative to the normal feedback condition. Compensation was detected when the null hypothesis ($H_0 = 1$) was rejected ($p < 0.05$) at a given time point and at all subsequent time points. These restrictions, which provide a conservative means for detecting compensation, necessarily overestimate the latency of compensation onset. Therefore, compensation response latencies were determined by fitting the mean subject compensation traces to a piecewise non-linear model of compensation (M_{Comp}). The model consisted of a constant segment (no compensation) followed by a logistic curve (compensation segment) of the form

$$M_{Comp} = \mathrm{Max}\left[C, \frac{1 - e^{-k(t-t_0)}}{1 + e^{-k(t+t_0)}} \right],$$

where t is a vector of compensation trace time points, t_0 marks the start of the non-linear segment, i.e., the estimated onset of compensation, C is the value of the constant component, and k modulates the rate of change of the non-linear component. While C was allowed to vary during the estimation procedure, k was held constant at $k = 0.1$. C and t_0 were estimated for each subject by determining least squares fits of the model to the two F1 compensation traces, then averaging across subjects. Estimates of t_0 were constrained to a window between 20 and 250 ms after voicing onset; the lower limit corresponds to the earliest time point at which shifted feedback could be heard; the upper

limit eliminated the influence of vowel-consonant transition effects introduced by short utterances. Confidence intervals for the t_0 fits were determined by a bootstrapping procedure. Compensation onsets were estimated from 1000 random resamples of the subject data with replacement for each condition. 95% confidence intervals were determined from the resulting t_0 distributions.

The magnitude of compensation was assessed by determining the peak response of the mean subject compensation traces in the two shift conditions. The peak response was restricted to time points following each subject's estimated onset latency.

4.4 Evidence of compensation

Subjects responded to unexpected F1-shifted auditory feedback by altering the F1 of their speech in the direction opposite the induced shift. Mean F1 traces from the *shift-up* and *shift-down* conditions expressed relative to their token-and subject-matched *no shift* responses are plotted in Figure 4-3*A*. The compensation traces, averaged across subjects, demonstrate significant downward divergence in the *shift up* condition and upward divergence in the *shift down* condition compared to the *no shift* condition. One-sample *t*-tests were performed at each time point of the compensation traces to test for deviation from baseline F1 values. The *t*-tests revealed significant, sustained compensation (df = 10, $p < 0.05$) 176 and 172 ms after the onset of voicing in the *shift up* and *shift down* conditions, respectively.

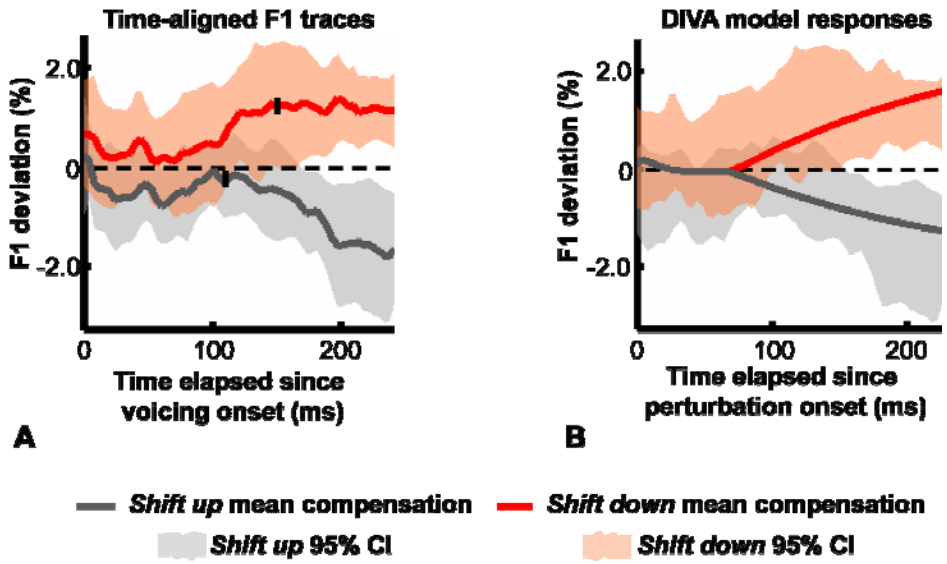


Figure 4-3. First formant response to an induced shift in F1.
(A) Mean F1 compensation plotted as percent deviation from the mean F1 value in unperturbed speech. Solid lines indicate compensation in the shift up (gray) and shift down (red) conditions averaged across subjects. Dashed lines represent the matched no shift formant value. Shaded regions indicate 95% confidence intervals at each time point. Shifting F1 by 30% resulted in a compensatory F1 response in the direction opposite the shift within an utterance. The black hash marks that intersect each compensation trace indicate the mean estimated onset of compensation (165 ms in the shift up condition, 108 ms in the shift down condition). (B) Comparison of experimental and simulated responses to unexpected F1 feedback perturbation during DIVA model production of the word /bɛd/. F1 traces produced by the model (lines) and 95% confidence intervals from the group experimental data (shaded regions) are aligned to the onset of perturbation.

Compensation response latencies were estimated by fitting subject compensation traces to a piece-wise non-linear model of compensation. The estimates are given with respect to the onset of perturbation. Mean subject latencies ranged from 87 to 235 ms (mean = 164.8 ms, SD = 43.5 ms) in the *shift up* condition and from 55 to 227 ms (mean = 107.7 ms, SD = 61.2 ms) in the *shift down* condition. Estimates of 95% confidence intervals (CI) for the mean response latencies were 145 – 186 ms and 81-139 ms in the *shift up* and *shift down* conditions, respectively. A paired *t*-test of response latencies demonstrated a significant difference between the two conditions ($df = 10$, $p = 0.01$).

Compensation magnitudes ranged from 11.1 to 59.4 Hz (mean = 30.0 Hz, SD = 14.8 Hz) in the *shift up* condition and 13.8 to 67.2 Hz (mean = 28.3 Hz, SD = 13.8 Hz) in the *shift down* condition. Expressed relative to the magnitude of the induced F1 perturbation, subjects compensated for 4.3% to 22.5% (mean = 13.6%, SD = 6.2%) of the upward shift and 6.3% to 25.5% (mean = 13.0%, SD = 6.7%) of the downward shift. A paired *t*-test comparing compensation magnitudes for the two shift directions indicated no difference (df =10, p = 0.75).

The DIVA model was trained to produce the word /bɛd/ with normal feedback (see Guenther et al., 2006 for model details). Training the model consists of repeated attempts to match a dynamic auditory target formed by extracting the first 3 formants from recorded productions of the desired speech sound by an adult male speaker. Initially, the match between the model's output and the auditory target are relatively poor, resulting in corrective feedback commands. The corrective commands are used to update the model's feedforward controller (encoded by synapatic weights projecting from a *Speech Sound Map* cell in premotor cortex to auditory error cells in superior temporal gyrus) so that with each subsequent attempt, the model's performance is improved. As performance improves, control of the vocal tract shifts from relying primarily on the feedback control system to greater reliance on feedforward control; the model is able to reliably match the dynamic formant targets of the training example.

Following successful training of the word /bɛd/, the auditory perturbation task was simulated by shifting F1 feedback to the model by 30% in the upward and downward

directions. The model's F1 response closely matched the experimental data (Figure 4-3*B*) following modification of those parameters that modulate the relative contributions of the feedforward (α_{ff}) and feedback (α_{fb}) commands to control of motor output. During training, α_{ff} and α_{fb} were set to 0.9 and 0.1, respectively. Based on simulations of the shifted-feedback conditions, these values were modified; and the F1 traces plotted in Figure 4-3*B* were obtained from simulations of the upward and downward shifted F1 conditions in which α_{ff} = 0.85 and α_{fb} = 0.15. A static parameter governing inertial damping of the feedback command, FB_{INERT}, was also modified. FB_{INERT} determines the relative influences of the previous feedback command, $\dot{M}_{Feedback}(t-1)$ on the current command, $\dot{M}_{Feedback}(t)$. The simulation results shown in Figure 4-3*B* reflect a 10% increase in this damping parameter over the training value.

4.5 The neural substrates of auditory feedback control

4.5.1 Voxel-based analysis

BOLD responses during the normal and shifted speech conditions compared to the baseline condition are shown in Figure 4-4. Voxels with peak *t*-statistic (peak-to-peak minimum distance = 6 mm) were assigned to anatomical regions based on cortical, subcortical, and cerebellar parcellation of the SPM2 canonical brain into ROIs. For each region containing a peak response, the voxel location, *t*-statistic, and normalized effect of the maximum response are provided in Table 4-1. Talairach coordinates (Talairach and

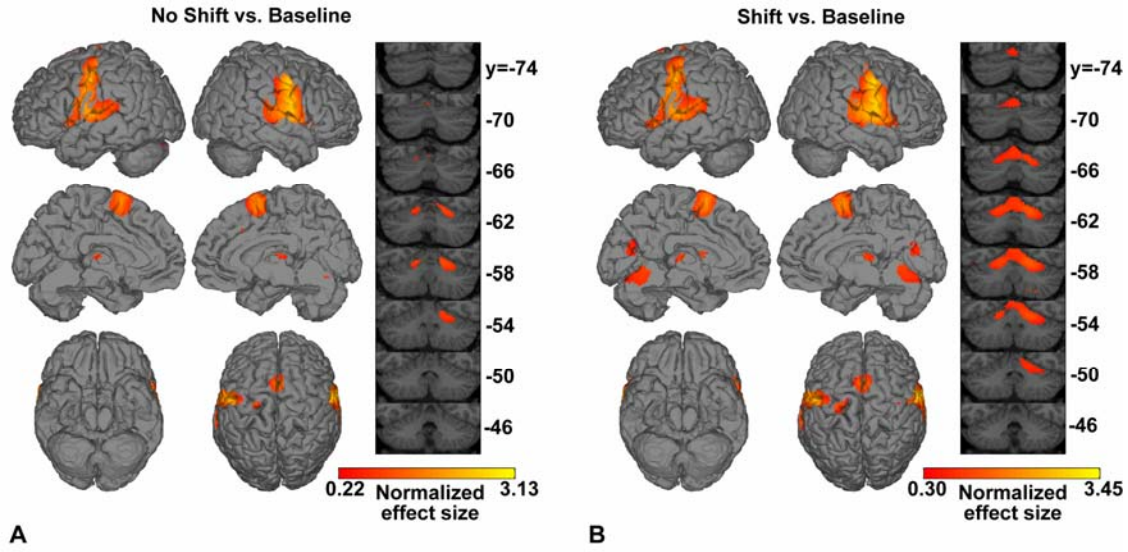


Figure 4-4. Responses for normal and shifted speech compared to baseline.
(A) Map of statistically significant normalized effect sizes ($t > 3.63$; $df = 9$; FDR < 5%) from the comparison of BOLD responses in the normal speech condition to the silent baseline condition (*no-shift – baseline*). Coronal slices through the cerebellum are shown to the right of renderings of the lateral (top), medial (middle) and ventral (bottom) cortical surfaces of the hemispheres. Activation is found in the expected speech production network including bilateral peri-Sylvian, lateral Rolandic, and medial prefrontal cortex, superior cerebellum, ventral thalamus and anterior striatum. (B) BOLD responses during shifted speech compared to baseline (*shift – baseline*; $t > 3.41$; $df = 9$; FDR < 5%). The network of active regions during shifted feedback speech is qualitatively similar to that of normal speech, with additional activation in the superior cerebellar cortex bilaterally, in the right inferior cerebellar cortex, and in the medial parietal-occipital cortex bilaterally.

Tournoux, 1988) reported in the table were determined using the MNI to Talairach mapping function described by Lancaster et al. (2007).

The *no shift – baseline* ($t > 3.63$, $df = 9$) contrast revealed a large network of active regions which has been previously implicated in speech production (Bohland and Guenther, 2006, Guenther et al., 2006) , including bilateral peri-Sylvian auditory cortex, ventral Rolandic cortex, medial prefrontal cortex, anterior striatum, ventral thalamus, and superior cerebellum (peaking in lobule 6). Activations in the *shift – baseline* contrast ($t > 3.41$, $df = 9$) overlap those of the *no shift – baseline* contrast with two notable exceptions:

		No Shift - Baseline				*Shift – Baseline*			
	Region	**Peak Voxel Location (x,y,z)**			**Norm.**	**Peak Voxel Stereotaxic Location**			**Norm.**
	Label	**MNI**	**Talairach**	***T***	**Effect**	**MNI**	**Talairach**	***T***	**Effect**
Rolandic Cortex									
Left	dMC	(-48,-8,60)	(-46,-14,56)	6.92	1.60	(-44,-8,64)	(-43,-15,60)	8.08	1.08
	vPMC	(-48,0,30)	(-46,-4,30)	7.06	1.05	(-60,4,8)	(-57,2,11)	8.89	1.42
	vSC	(-56,-10,42)	(-53,-14,40)	15.44	2.45	(-52,-14,38)	(-50,-18,36)	12.79	2.81
	pCO	(-62,-6,12)	(-59,-8,13)	12.09	2.42	(-60,-8,12)	(-57,-10,13)	11.23	2.69
Right	dMC					(44,-18,70)	(39,-25,66)	4.94	0.43
	aCO	(48,8,4)	(43,5,9)	10.63	0.96				
	vSC	(54,-8,30)	(49,-12,31)	22.67	2.15	(56,-10,30)	(50,-14,31)	16.26	2.40
Frontal Cortex									
Left	IFo					(-60,8,2)	(-57,6,6)	8.36	1.38
	FO	(-48,10,-2)	(-45,8,2)	5.98	1.45	(-48,10,-2)	(-45,8,2)	8.04	1.85
	preSMA	(-2,8,62)	(-4,0,60)	10.91	1.37	(0,4,72)	(-2,-4,69)	7.61	1.70
Right	FO					(48,12,2)	(43,9,8)	7.45	1.38
	aCg	(4,20,36)	(2,14,38)	3.89	0.68				
Parietal Cortex									
Left	PO	(-44,-34,24)	(-42,-35,22)	8.85	1.01	(-40,-30,18)	(-38,-31,17)	6.98	1.99
Right	PO	(46,-26,18)	(41,-28,19)	5.42	1.26	(44,-24,18)	(39,-26,19)	5.29	1.57
	aSMg					(72,-24,32)	(65,-27,32)	4.49	0.58
Temporal Cortex									
Left	Hg	(-58,-10,10)	(-55,-12,11)	11.79	2.27				
	pSTg	(-64,-30,14)	(-60,-31,13)	9.70	1.31	(-62,-30,14)	(-59,-31,13)	7.93	2.00
	pdSTs					(-62,-22,4)	(-58,-22,5)	4.95	1.85
	PT	(-52,-34,16)	(-49,-35,15)	5.79	1.51				
Right	pSTg	(72,-24,6)	(65,-25,8)	7.40	1.03	(68,-16,8)	(62,-18,11)	8.66	2.31
	pdSTs					(58,-28,6)	(53,-29,8)	5.17	1.50
	PT	(58,-28,12)	(52,-29,13)	5.06	1.30	(64,-10,10)	(58,-12,13)	9.33	2.57
Insular Cortex									
Left	aINS	(-44,6,2)	(-42,4,6)	5.73	1.17				
	aINS					(-34,-4,10)	(-33,-6,12)	4.49	0.85
	pINS					(-34,-20,8)	(-33,-21,9)	5.16	0.99
Cerebellum									
Left	spmCB,L6	(-16,-62,-20)	(-16,-58,-19)	5.10	0.76	(-18,-62,-20)	(-18,-58,-20)	6.72	1.08
	spmCB,V6					(-2,-78,-14)	(-3,-73,-15)	5.45	0.62
	splCB,Cr1	(-30,-86,-24)	(-29,-80,-25)	4.17	0.40				
Right	amCB,V5					(8,-64,-10)	(6,-61,-10)	5.36	0.94
	splCB,L6	(26,-58,-24)	(23,-54,-22)	6.30	0.95	(24,-56,-22)	(21,-52,-20)	9.46	1.35
	ipmCB,L8A					(28,-58,-54)	(25,-51,-49)	3.42	0.86
Subcortical Nuclei									
Left	Put	(-26,-4,4)	(-25,-6,7)	7.97	0.62	(-30,-6,4)	(-29,-8,7)	4.34	0.71
	Pal	(-24,-6,-4)	(-23,-7,-0)	6.74	0.58	(-22,-2,-2)	(-21,-3,2)	5.15	0.95
	Caud					(-10,2,10)	(-10,-1,13)	3.74	0.92
	Tha,VL	(-8,-14,6)	(-9,-15,8)	4.64	0.64	(-8,-14,6)	(-9,-15,8)	3.82	0.88
Right	Put	(30,8,4)	(27,5,9)	6.13	0.35				
	Pal	(26,-2,-4)	(23,-3,1)	5.32	0.62	(22,2,-2)	(19,0,3)	4.27	0.61
	Cau					(14,4,10)	(12,1,14)	4.11	0.84
	Tha, VL	(14,-12,10)	(12,-14,12)	3.98	0.62				
	Tha,MD	(6,-18,8)	(4,-19,10)	4.19	0.76	(4,-22,6)	(3,-23,8)	4.50	0.98
Occipital Cortex									
Left	OC					(-2,-70,16)	(-3,-68,12)	4.94	0.81

Table 4-1. Peak voxel responses for baseline contrasts.
Peak responses for the *No Shift – Baseline* and *Shift – Baseline* contrasts, defined as local *t*-statistic maxima ($p_{FDR} < 0.05$) separated by a minimum 6 mm, were mapped to anatomical ROIs. Each ROI containing a peak voxel response is listed with the location, *t*-statistic and normalized effect associated with the peak voxel. Voxel locations are provided in both MNI and Talairach stereotaxic reference frames. When possible, a more specific label is listed in addition to the ROI label for voxels that lie within well-characterized subsets of an anatomical region (e.g., cerebellar lobule 6). *n.s.* = not significant. Cerebellar vermis (V) and lobule (L) labels are given by Arabic numerals to conserve space.

the extension of the superior cerebellar activation anterior-medially to include the cerebellar vermis bilaterally, and activation at the junction of the calcarine and parietal-occipital sulci.

Figure 4-5*A* shows the results of a fixed effects analysis of the *shift – no shift* contrast ($t > 3.19$, $df = 5488$) corrected to ensure FDR < 5% (Figure 4-5*A*, *Voxel-Based*

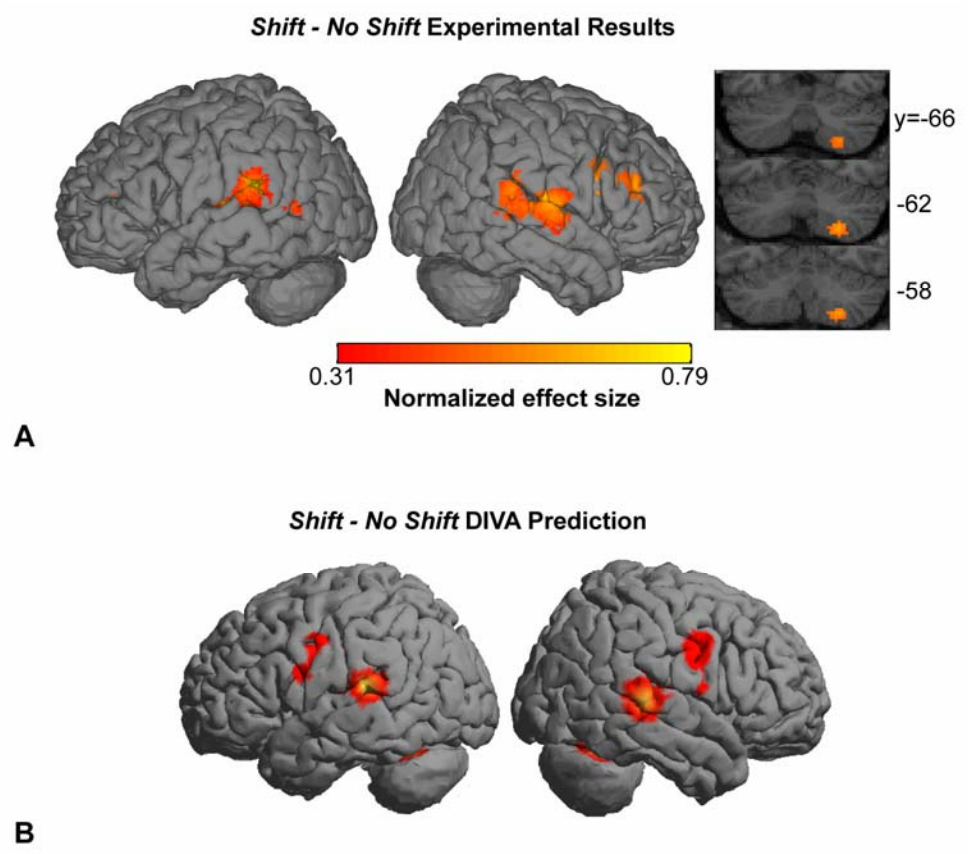


Figure 4-5. BOLD responses in the *shift – no shift* contrast.
(A) Map of statistically significant normalized effect sizes (voxel threshold: $t > 3.19$; $df = 5488$; FDR < 5%) derived from group voxel-based fixed effects analysis. Activation of the posterior peri-Sylvian region is seen bilaterally. In the right hemisphere, greater responses were found in ventral somatosensory, motor and premotor cortex, and the inferior cerebellum. (B) Simulated BOLD responses for the *shift – no shift* contrast in the DIVA model. Model cell responses from *no shift* and *shift* simulations were contrasted and normalized in the same manner applied to the experimental results and smoothed with a spherical Guassian point spread function (FWHM = 12 mm). The resulting activations were then plotted on a cortical surface rendering based on hypothesized anatomical locations (Guenther et al., 2006). The model predicts increased activation in bilateral posterior peri-Sylvian cortex (auditory error cells), ventral Rolandic cortex (motor cells and somatosensory error cells), and superior cerebellum.

		Voxel-based Results				*ROI-based Results*	
	Region Label	**Peak Voxel Stereotaxic Location MNI**	**Talairach**	***T***	**Norm. Effect**	***T***	**Norm. Effect**
Rolandic Cortex							
Left	**vMC**					***n.s***	***n.s***
	vPMC					***n.s***	***n.s***
Right	**vMC**	(48,-10,44)	(43,-15,43)	3.25	0.54	**2.10**	**0.28**
	vPMC	(60,14,34)	(54,8,37)	3.52	0.53	**2.52**	**0.36**
	vSC	(70,-2,20)	(64,-6,23)	3.20	0.40	1.87	0.24
Frontal Cortex							
Right	IFo	(58,14,28)	(52,9,31)	3.64	0.62	2.13	0.32
	IFt	(56,32,24)	(51,26,29)	4.07	0.57	2.07	0.30
Parietal Cortex							
Left	PO	(-54,-24,14)	(-51,-25,14)	3.98	0.67	4.47	0.63
Temporal Cortex							
Left	**pSTg**	(-66,-38,22)	(-62,-39,19)	5.25	0.59	**3.78**	**0.48**
	pdSTs	(-60,-30,10)	(-57,-30,9)	3.96	0.60	2.08	0.28
	PT	(-62,-24,10)	(-59,-25,10)	3.88	0.57	**3.83**	**0.65**
	MTO	(-60,-62,10)	(-57,-60,7)	3.64	0.44	2.88	0.29
Right	pSTg	(68,-36,18)	(62,-37,18)	4.13	0.56	**4.64**	**0.63**
	pdSTs	(72,-40,12)	(65,-40,12)	3.94	0.44	3.15	0.31
	PT	(68,-16,8)	(62,-18,11)	5.01	0.68	**4.32**	**0.49**
	PP	(48,-8,-8)	(43,-9,-3)	3.59	0.63	4.07	0.45
	adSTs	(56,-10,-4)	(51,-11,0)	3.31	0.55	2.18	0.39
Cerebellum							
Left	**amCB,V5**	***n.s***	***n.s.***	***n.s***	***n.s.***	***n.s***	***n.s.***
Right	**amCB, V5**	***n.s***	***n.s.***	***n.s***	***n.s.***	**2.25**	**0.34**
	ipmCB, L8A	(26,-62,-54)	(24,-55,-49)	4.06	0.62	2.14	0.24

Table 4-2. Peak voxel and regional responses for the *shift - no shift* contrast.
Peak voxel responses were based on standard voxel-based analyses and are reported as in Table 4-1 Regional *t*-statistics and normalized effects were derived from ROI-based analyses. Regions highlighted in boldface type were included in our initial ROI analysis of a priori hypotheses that each of these regions would be more active during perturbed feedback trials. *n.s.* = not significant.

Results). Bilateral activation of higher order auditory cortical areas in posterior superior temporal cortex, including the posterior superior temporal gyrus (pSTG) and planum temporale (PT), is consistent with the DIVA model prediction of auditory error cells in

these areas (see simulation results in Figure 4-3*B*). Additional temporal lobe activity was noted in middle temporal-occipital cortex in the left hemisphere.

Responses in ventral Rolandic and lateral prefrontal cortex were noted only in the right hemisphere. In addition to ventral motor and somatosensory cortical activation, a region of activity along the ventral precentral sulcus extended to inferior frontal gyrus pars opercularis (IFo) and ventral premotor cortex (vPMC). More anteriorly, activity near the inferior frontal sulcus peaked in the inferior frontal gyrus pars triangularis (IFt). Activation in the inferior cerebellar cortex (lobule VIII) was also found only in the right hemisphere.

4.5.2 Region of interest analysis

The *shift – no shift* results presented in Figure 4-5*A* reflect group fixed-effects analysis; no voxels survived a mixed-effects analysis of the contrast following threshold correction (FDR < 5%). The responses shown in Figure 4-5*A* therefore do not necessarily represent those of the general population. Failure to achieve significance at the population level can be the result of a lack of statistical power due to a limited sample size. It may also be due to anatomical variability in the subject population. To control for this possibility, we utilized a more sensitive ROI-based mixed effects analysis (Nieto-Castanon et al., 2003) on those regions to test our *a priori* hypotheses that perturbed feedback would cause increased activation in bilateral posterior auditory, ventral motor and premotor and superior cerebellar cortex.

Results from the ROI analysis of the *shift – no shift* contrast are presented in Table 4-2, *ROI-based Results.* The ROI results supported our hypotheses regarding

posterior auditory cortex (see Figure 4-5 and Table 4-2, regions in boldface type). Significant bilateral responses ($t > 2.08$, $df = 9$, FDR < 5% for tests of 10 ROIs) were noted in pSTg and PT. However, increases in vMC, vPMC and amCB were significant only in the right hemisphere.

Tests on the remaining ROIs (n = 132) found no regions surviving a corrected significance threshold for the *shift – no shift* contrast. Several regions did survive individual (i.e., uncorrected, ROI-level) significance thresholds ($p < 0.05$), however, and

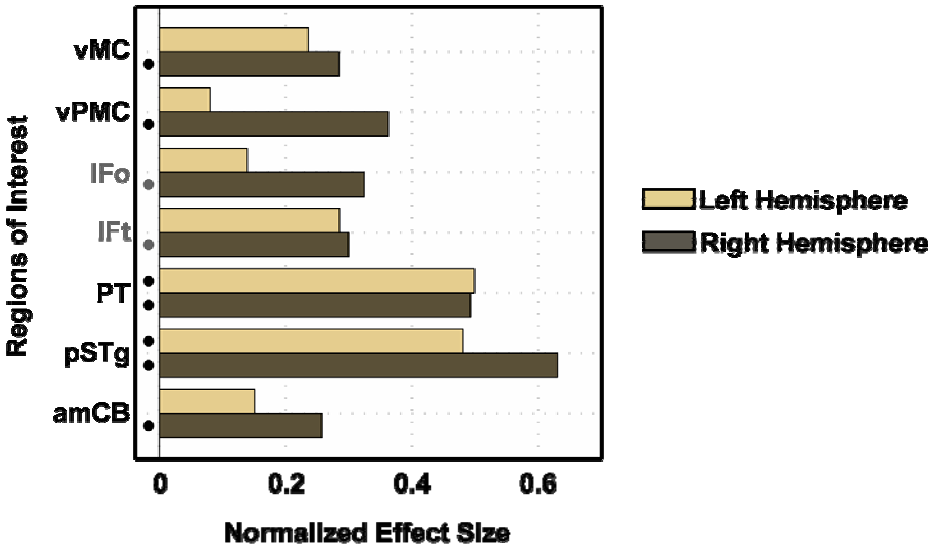


Figure 4-6. Selected regional responses from the *shift – no shift* contrast.
Normalized effects are shown for those ROIs hypothesized to be more active in the perturbed feedback condition: ventral motor cortex (vMC), ventral premotor cortex (vPMC), planum temporale (PT), posterior superior temporal gyrus (pSTg), and anterior medial cerebellum (amCB). Left and right hemisphere responses are indicated by light and dark bars, respectively. Black bullets plotted ('•') adjacent to response bars denote significant effects ($t > 2.08$, $df = 9$, corrected to ensure a false discovery rate < 5% for tests of responses in the analysis of the 10 ROIs included in *a priori* hypotheses). Responses from the inferior frontal gyrus ROIs, the *par opercularis* (IFo) and *pars triangularis* (IFt) are also shown (labeled in gray font) and listed under vMC and vPMC to highlight the consistency of ventrolateral frontal responses. The gray bullets adjacent to the IFo and IFt responses denote significant effects ($p < .05$, $df = 131$) based on an uncorrected test of all 132 ROIs.

these were consistent with the voxel-based results (Table 4-2). Activation of a wider

range of peri-Sylvian ROIs was noted bilaterally, including right inferior frontal cortex including Heschl's gyrus (Hg), planum polare (PP) and the parietal operculum (PO).

Figure 4-7 provides a more detailed illustration of activation within peri-Sylvian regions not directly visible in Figure 4-4. The figure depicts flattened representations of the spatial profiles of ROIs along the superior temporal gyrus and the parietal operculum in the left and right cerebral hemispheres. The flattened representations provide a means

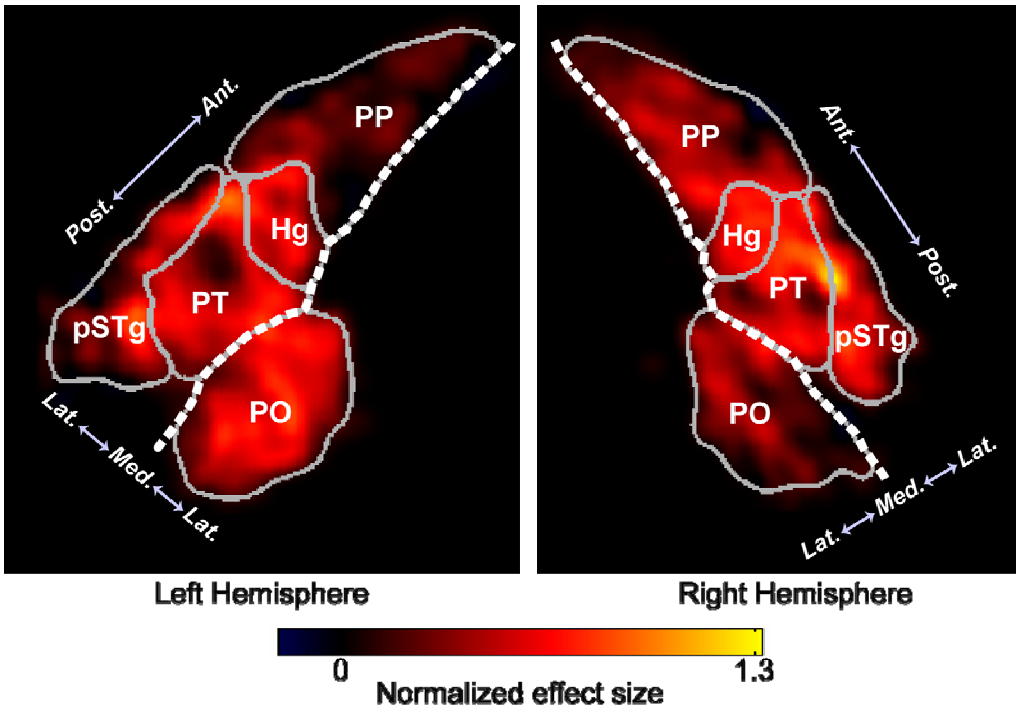


Figure 4-7. Spatial profiles of peri-Sylvian ROI responses for the *shift – no shift* contrast.
A flattened representation of normalized activity from a subset of ROIs within the Sylvian fissure is shown. For each hemisphere, solid, gray lines indicate ROI boundaries; heavy, dotted white lines represent the location of the Sylvian fissure with respect to regional boundaries. The labeled arrows indicate the relative orientation of the activation plots: the Sylvian fissure approximates the anterior (*Ant.*) – posterior (*Post.*) axis in each hemisphere and the medial extreme of the medial (*Med.*) - lateral (*Lat.*) plane, i.e., in each hemisphere, planum polare (PP) lies anterior to Heshl's gyrus (Hg), and planum temporal (PT) lies medial to posterior superior temporal gyrus (pSTg). Activation peaks are seen along the PT/pSTg boundary in both hemispheres. Greater activation was noted in the parietal operculum (PO) in the left hemisphere and in Hg in the right hemisphere.

for viewing the topography of activation along the adjacent banks of the Sylvian fissure.

The strongest responses were noted in the right hemisphere along the posterior superior temporal gyrus/planum temporale (pSTg/PT) border. An additional peak in posterior right pSTg is also visible. Peak activations in the left hemisphere were also found along the pSTg/PT border: a posterior peak in pSTg and an anterior peak in PT. Widespread activation in left PO was greater than that seen in the right hemisphere. Activation of primary auditory cortex, located in the posteromedial portion of Hg, was stronger in the right hemisphere. Other regions found active in both the voxel-based and ROI-based analyses included right inferior frontal gyrus, par opercularis and pars triangularis (IFo and IFt, respectively), right inferior posterior medial cerebellum (ipmCB), and left middle temporal-occipital cortex (MTO; see Table 4-2).

The apparent right-lateralized ventral frontal responses motivated *post-hoc* tests for hemispheric differences in the responses of ROIs shown in Figure 4-6 (those regions hypothesized to be more active in the perturbed feedback condition) and the inferior frontal gyrus. In addition to testing for regional interhemispheric differences in the *shift – no shift* condition, a measure of relative laterality in the two speaking conditions, laterality effects were also assessed for responses in two baseline contrasts (*no shift – baseline*, *shift – baseline*) to investigate the absolute laterality in the two speaking conditions.

Figure 4-8 illustrates the interhemispheric response differences for those regions tested that were found to differ significantly in at least one contrast. Significant right lateralization in the *shift – no shift* contrast was apparent only in vPMC ($p = 0.03$).

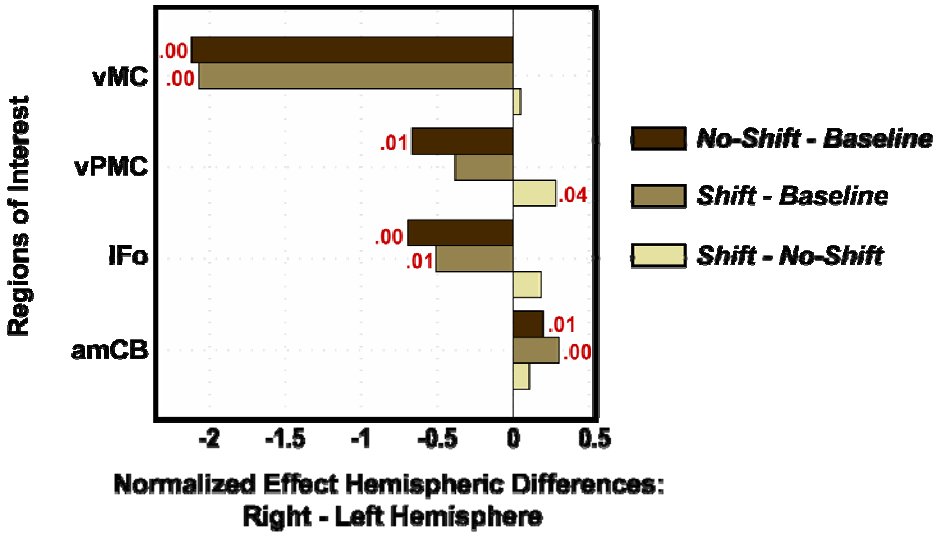


Figure 4-8. Interhemispheric response differences for selected ROIs across conditions.
Bars represent the right hemisphere response minus the left hemisphere response for a given contrast (*no shift – baseline*, *shift – baseline*, *shift – no shift*), i.e, positive values indicate a greater response in the right hemisphere. *P*-values (red font) are displayed adjacent to significant laterality effects for a given ROI and contrast. Laterality effects were assessed for all ROIs included in Figure 4-6; those ROIs that demonstrated a significant interhemispheric response difference in any of the three contrasts of interest are included above. Only vPMC was significantly more active in the right hemisphere in the *shift – no shift* contrast; all ventral frontal ROIs were more active in the left hemisphere in both speaking conditions relative to baseline.

Significant left hemisphere lateralization of vMC ($p < 0.00$), vPMC ($p = 0.01$) and IFo ($p < 0.00$) was noted in the *no shift – baseline* contrast. vMC ($p < 0.00$) and IFo ($p = 0.01$) remained left lateralized in the *shift – baseline* contrast.

4.6 Inter-regional interactions in the auditory feedback control network

Structural equation modeling was used to compare interactions between the main regions found active in the *shift – no shift* contrast in the two speech conditions. The network shown in Figure 4-9 easily met the prescribed fit and stability criteria. The unconstrained model, in which connections strengths were allowed to vary across the two speech

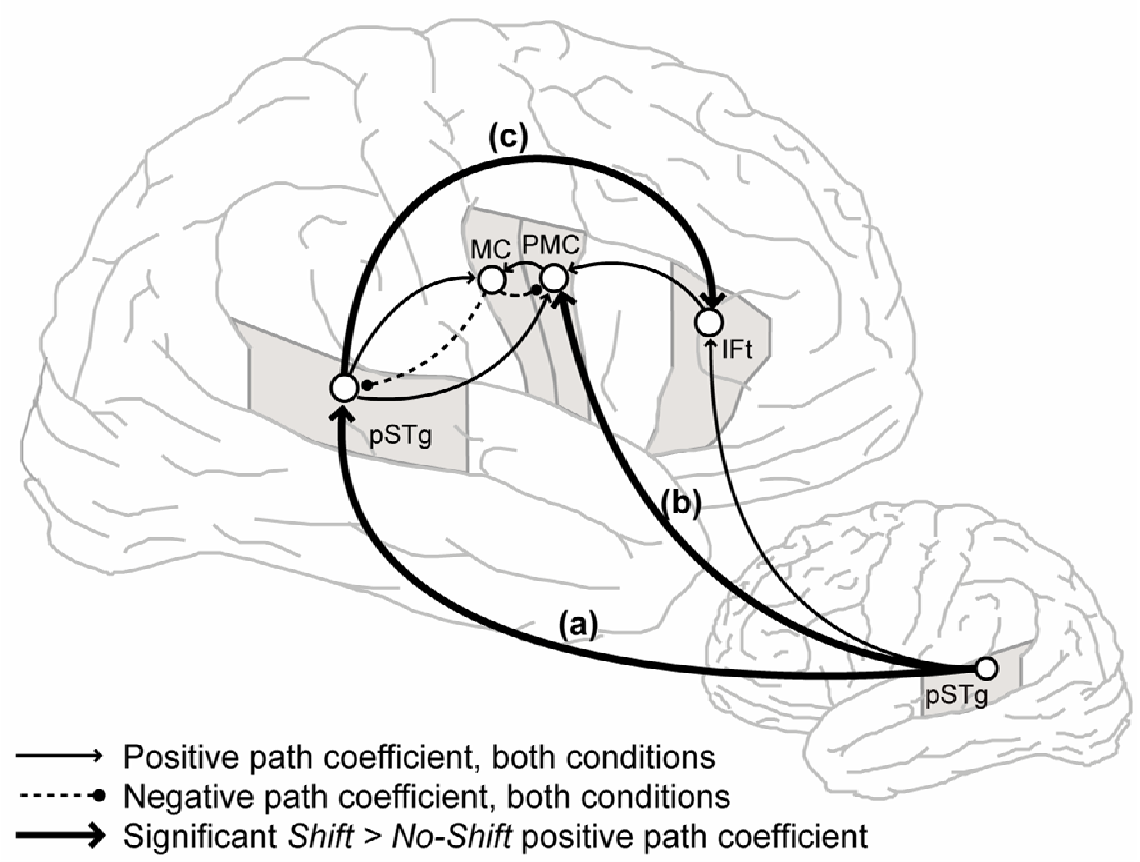


Figure 4-9. Schematic of the path diagram evaluated by structural equation modeling.
Effective connectivity within the network of regions shown was significantly modulated by the auditory feedback perturbation. Path coefficients for all connections shown were significant in both conditions except right ventral MC to right ventral PMC (*no shift* probability = 0.07; see Table 2 for a list of all estimated path coefficients). Pair-wise comparisons of path coefficients in the two conditions revealed significant increases in the positive weights from left pSTg to right pSTg (the path labeled *a* in the diagram above), from left pSTg to right ventral PMC (path *b*), and from right pSTg to right IFt (path *c*) when auditory feedback was perturbed during speech production. Abbreviations: IFt = inferior frontal gyrus, pars triangularis; pSTg = posterior superior temporal gyrus; MC = motor cortex; PMC = premotor cortex.

conditions, provided a significantly better fit to the data than did the null model, in which the connections are constrained to be the same in both conditions (χ^2_{diff} = 19.89, df_{diff} = 9, p_{diff} = 0.02). This indicates that the network was significantly modulated by the feedback perturbation. The unconstrained model provided an excellent fit to both the *no shift* and *shift* experimental data covariances (χ^2_{uncon} = 3.60, $df_{uncon.}$ = 5, $p_{uncon.}$ = 0.61, GFI_{uncon} = 1.00,

Network Path			*No Shift*				*Shift*				*Shift - No Shift*	
			Path Coeff.	*S. E.*	*C.R.*	*p*	Path Coeff.	*S. E.*	*C.R.*	*p*	*C.R.*	*p*
Left pSTg	**→**	**Right vPMC**	**0.06**	**0.03**	**2.22**	**0.03**	**0.22**	**0.05**	**4.01**	**< .001**	**2.64**	**0.01**
Left pSTg	→	Right IFt	0.25	0.02	10.61	< .001	0.26	0.04	5.84	< .001	0.12	0.90
Left pSTg	**→**	**Right pSTg**	**0.49**	**0.03**	**15.68**	**< .001**	**0.61**	**0.06**	**10.94**	**< .001**	**2.11**	**0.03**
Right IFt	→	Right vPMC	0.35	0.03	12.29	< .001	0.34	0.05	7.10	< .001	-0.03	0.98
Right pSTg	**→**	**Right IFt**	**0.27**	**0.03**	**10.58**	**< .001**	**0.37**	**0.05**	**8.13**	**< .001**	**2.08**	**0.04**
Right vMC	→	Right pSTg	-0.25	0.07	-3.53	< .001	-0.34	0.09	-3.95	< .001	-1.33	0.18
Right pSTg	→	Right vMC	0.36	0.06	5.96	< .001	0.40	0.07	5.33	< .001	0.58	0.56
Right vMC	→	Right vPMC	-0.15	0.08	-1.83	0.07	-0.26	0.09	-2.77	0.01	-1.56	0.12
Right vPMC	→	Right vMC	0.48	0.07	7.30	< .001	0.54	0.08	7.09	< .001	0.92	0.36
Right pSTg	→	Right vPMC	0.32	0.04	7.41	< .001	0.33	0.06	5.23	< .001	0.02	0.98

Table 4-3. Effective connectivity determined by structural equation modeling.
A comparison of effective connectivity within the network shown in Figure 4-9 in the *no shift* and *shift* conditions is given. Path strengths (Path Coeff.) that were significantly modulated by the F1 shift are shown in boldface type. Abbreviations: Coeff. = coefficient; C.R. = critical ratio; IFt = inferior frontal gyrus, pars triangularis; pSTg = posterior superior temporal gyrus; S.E. = standard error; vMC = ventral motor cortex; vPMC = ventral premotor cortex.

$AGFI_{uncon}$ = 1.00, RMR_{uncon} = 0.02, $RMSEA_{uncon}$ = 0.00). Pair-wise comparisons of path coefficients (Table 4-3) revealed that connection strengths from left pSTg to right pSTg (path *a* in Figure 4-9), left pSTg to right vPMC (*b*), and from right pSTg to right IFt (*c*) were significantly greater in the shift condition, indicative of greater use of these pathways when the auditory feedback control network was invoked due to the feedback shift.

4.7 Discussion

4.7.1 Formant shift compensation

As illustrated in Figure 4-3*A*, subjects responded to unexpected F1 shifts by altering the F1 of their speech in the direction opposite the induced shift. Computer simulations of the

DIVA model verified the model's ability to account for these compensatory responses (Figure 4-3*B*) following an increase in the relative contribution of auditory feedback on motor control. Adaptation to consistently applied upward or downward F1 shifts during production of /CɛC/ utterances similar to those used in the current study has been demonstrated previously (Purcell and Munhall, 2006a, Villacorta et al., 2007). In those studies, the F1 shift was presented on every trial throughout a training phase, allowing speakers to modify stored feedforward motor plans. Simulations of the F1 shift condition with the DIVA model verified that the model's interactions between auditory feedback control and feedforward control could account for the adaptation results. In the current study, the compensatory response is apparent *within an utterance* despite the unpredictable nature of the shift, and DIVA model simulations quantitatively account for this compensation.

The estimated compensation latencies (108 ms and 165 ms to the downward and upward shifts, respectively) were short enough to permit online correction within the duration of a typical /CɛC/ utterance (Hillenbrand et al., 2001, Ferguson and Kewley-Port, 2002). The estimates, particularly in the *shift down* condition, fall at the low end of ranges reported following unexpected perturbation of F0 (Hain et al., 2000, Burnett and Larson, 2002, Xu et al., 2004). Faster response times may be due to differences in formant and pitch control (discussed further below) but may also reflect differences in the method used to determine response latencies. Earlier pitch shift studies also often feature a steady-state vocal task (e.g., prolonged vowel) which may reduce the relevance of pitch

modulation. The considerable inter-subject variability noted in the latency of compensatory responses to auditory feedback manipulations has been noted previously (Xu et al., 2004). This variability may reflect the difficulty associated with real-time tracking of formants (cf., magnitude estimates noted by Purcell and Munhall, 2006b). The hearing ability and perhaps the attention spans of subjects, neither of which were controlled here, may also contribute to the latency variability.

Though relatively short, the latencies are sufficiently long to allow for a cortically mediated compensatory response (see Guenther et al., 2006 for discussion) and are much longer than brainstem-mediated auditory perioral reflex responses (McClean and Sapir, 1981). A recent study used transcranial magnetic stimulation to demonstrate motor cortical involvement in phonetically specific compensation to jaw perturbation with a latency of approximately 85 ms but not during short-latency perioral reflex responses with an 18 ms latency (Ito et al., 2005).

The faster response to downward relative to upward F1 shifts during /ɛ/ production noted here has been reported previously (Purcell and Munhall, 2006b). While not conclusive, F1 may provide a more robust cue for distinguishing /ɛ/ and /ɪ/ than for /ɛ/ and /æ/ (Clopper et al., 2005), suggesting the downward F1 shift toward /ɪ/ is more likely to produce a phonemic or lexical categorical error than the upward shift toward /æ/, even if the acoustic difference is the same. The faster response to the downward shift may therefore reflect greater lexical saliency. The impact of lexical saliency on compensatory responses is supported by a recent report describing the effects of F0 shift direction when

unexpected perturbations were delivered between Mandarin bi-tonal disyllables. Shifts in the direction opposite the intended inter-syllabic tonal transition resulted in shorter latencies and larger compensations than did shifts in the same direction (Xu et al., 2004).

4.7.2 The auditory feedback control network

According to the DIVA model, perturbation of F1 feedback causes a mismatch between the auditory expectation for the current syllable and the auditory signal fed back to the subject (Guenther et al., 2006). This mismatch leads to activation of auditory error cells in the posterior superior temporal gyrus, including the planum temporale, a prediction strongly supported by the bilateral peri-Sylvian activation noted in the *shift – no shift* contrast (Figure 4-5*A*). Increased bilateral activation of posterior temporal regions during perturbed speech is consistent with previous results from imaging studies of pitch perturbation (McGuire et al., 1996, Zarate and Zatorre, 2005, Fu et al., 2006). Imaging studies of the effects of auditory feedback disruption, including delayed auditory feedback (Hirano et al., 1997, Hashimoto and Sakai, 2003) and noise masking (Christoffels et al., 2007), have also demonstrated bilateral activation of posterior superior temporal gyrus, providing further support that auditory error is represented in these regions.

Numerous lines of evidence support the hypothesis that the expected sensory consequences of articulatory movements and the resulting auditory feedback are compared in posterior temporal cortex. Portions of posterior left planum temporale and lateral posterior superior temporal gyrus bilaterally have been shown to respond during both speech perception and speech production in several studies (Hickok et al., 2003,

Buchsbaum et al., 2005). Bi-directional functional connections between inferior frontal and posterior temporal cortex have been demonstrated *in vivo* using cortico-cortical evoked potentials in humans (Matsumoto et al., 2004). Attenuation of the posterior auditory cortex responses during self-produced speech has been shown in a number of studies (Paus et al., 1996, Numminen and Curio, 1999, Numminen et al., 1999, Curio et al., 2000, Houde et al., 2002). Similar modulation of somatosensory responses to self generated movements (Blakemore et al., 1998) has been interpreted as evidence that an efference copy from motor to sensory cortex encodes the expected sensory consequences of upcoming movements (Blakemore et al., 2000, Wolpert and Flanagan, 2001). According to these theories, an efference copy of the outgoing motor plan attenuates the regions of sensory cortex responsive to the expected sensory feedback. In doing so, the efference copy effectively "cancels" the sensory feedback response resulting from the movement. Support for this mechanism in the auditory domain has recently been demonstrated using magnetoencephalography (MEG) to measure modulations of the magnetic fields generated by auditory cortical neurons. Heinks-Maldonado et al. (2006) demonstrated attenuation of auditory cortex during self-produced speech that was modulated by how closely the auditory feedback matches expected feedback. Greater attenuation of auditory cortical responses was noted when speakers heard normal rather than pitch-shifted auditory feedback. Work in monkeys has shown how precise the efference copy-induced attenuation of auditory cortex may be. Single unit recordings from monkey auditory cortex demonstrated pre-vocalization suppression of primary auditory and lateral belt neurons that was tightly linked to the subsequent vocal output

(Eliades and Wang, 2005). Our current findings are consistent with these studies which collectively support the view that higher-level auditory cortical regions include auditory "error" cells that encode the difference between actual and expected auditory feedback during vocalization.

The DIVA model also predicts bilateral ventral precentral gyrus activation in the *shift – no shift* contrast, reflecting corrective commands sent from auditory error cells to the bilateral ventral motor cells that drive compensatory articulator movement (see Figure 2-4B; Guenther et al., 2006). The *shift – no shift* experimental results, however, revealed ventral precentral activation only in the right hemisphere in the *shift – no shift* contrast. Subsequent laterality tests on the ventral frontal responses revealed: (i) under normal auditory feedback conditions, control of the articulators is predominantly left lateralized, a finding noted previously for overt speech (Riecker et al., 2000a, Sidtis et al., 2006b, Ghosh et al., 2008), (ii) feedback-based articulator control relies on significantly greater involvement of right hemisphere ventral frontal regions, especially premotor and inferior frontal cortex (discussed further below), and (iii) in addition to the motor projections previously hypothesized (Guenther et al., 2006), auditory error cells appear to project to premotor and inferior prefrontal cortex in the right hemisphere. These conclusions were supported by structural equation modeling which revealed increased effective connectivity from left posterior temporal cortex to right posterior temporal and ventral premotor cortex. Reciprocal connectivity between right posterior superior temporal gyrus and ventral motor cortex, though significant in both conditions, increased only modestly. Right posterior temporal cortex may exert additional influence over motor output via a

connection through right inferior frontal gyrus, *par triangularis* (BA 45), which increased significantly in response to shifted feedback.

4.7.3 Implications for the study of normal and disordered speech production

The current findings shed light on a perplexing issue in speech and language neuroscience. Functional imaging of speech production typically results in bilateral pre-frontal and sensorimotor activation (e.g. Wise et al., 1999, Bohland and Guenther, 2006, Ozdemir et al., 2006, Soros et al., 2006). These findings appear to conflict with the large body of lesion data that supports the traditional view of left hemisphere dominance of speech production (Duffy, 1995, Dronkers, 1996, Kent and Tjaden, 1997, Hillis et al., 2004). The present results reconcile these findings: while both hemispheres contribute to speech production, feedforward control is predominantly subserved by the left hemisphere whereas auditory feedback control is subserved by right hemisphere frontal regions. Inferior frontal lesions in the left hemisphere are therefore more likely to disrupt stored feedforward speech motor commands, resulting in disordered speech since the feedforward control system is more crucial for fluent speech than the auditory feedback control system (cf. Neilson and Neilson, 1987).

The conclusion that auditory feedback control relies more heavily on contributions from the right hemisphere is consistent with findings from neuroimaging studies of by pitch perturbation studies that have shown greater activation in inferior frontal cortex Fu et al. (2006) and right posterior temporal cortex (McGuire et al., 1996). In another recent fMRI study, Toyomura et al. (2007) used a sparse image acquisition paradigm similar to the one described in this report to investigate the neural correlates of

auditory feedback control of F0. Twelve right-handed subjects could hear their own auditory feedback through headphones while vocalizing /a/ for five seconds. On a subset of trials a two semi-tone F0 shift was induced. The laterality of activations resulting from a comparison of the shifted and normal feedback conditions is striking: the *shift – no shift* contrast feedback condition resulted in significant responses in dorsal inferior frontal gyrus, *pars opercularis*, frontal operculum, posterior superior temporal gyrus (probably planum temporale), anterior supramarginal gyrus, and intraparietal sulcus *only in right hemisphere*. A portion of the precentral gyrus dorsal to its junction with the inferior frontal sulcus was the only activation noted in the left hemisphere. Since specific tests of laterality were not performed, it cannot be determined whether the hemispheric differences were significant. The authors concluded that the feedback control of pitch is primarily influenced by the right hemisphere. However, the current findings suggest that auditory feedback control of speech, in general, involves a greater contribution from the right hemisphere than does feedforward control.

In the absence of baseline contrasts for the two speaking conditions and explicit tests for hemispheric differences, it is difficult to draw conclusions regarding the discrepancy between the laterality of precentral activation reported by Toyomura et al. (2007) study and results from the auditory perturbation experiment described above. It should be noted that the right dorsal inferior frontal activation noted by Toyomura et al. (2007) overlaps with the right precentral/posterior inferior frontal activation in the *shift – no shift* contrast noted here (see Figure 4-5), making comparisons of ventral premotor responses particularly difficult. However, further evidence for a specific, right-lateralized

premotor/inferior frontal cortex contribution to auditory feedback control is given by recent data from another study of brain responses to real-time F0 shifts during speech production using MEG (Houde and Nagarajan, 2007). Correlations between compensation magnitude and cortical magnetic fields revealed the strongest coupling in the dorsal premotor cortex of the right hemisphere, overlapping with the region of activation shown above in Figure 4-5. Finally, asymmetric premotor contributions to feedforward and feedback motor control were also recently noted in a visuomotor tracking experiment involving arm movements (Grafton et al., 2008). BOLD responses were measured while subjects learned to control movements of a cursor on a screen via an arm rotation. The function relating arm rotation to cursor movements alternated with each successive trial. Brain responses were correlated with measures of feedforward and feedback learning and motor error. The data revealed a strong correlation between feedforward learning and activation in the left dorsal premotor cortex. Right ventral premotor cortex, on the other hand, was strongly correlated with measures of feedback learning. Interestingly, BOLD response correlations with motor error demonstrated activation of the *right*, but not left, cerebellar cortex. This was accompanied by activation in right, but not left, dorsolateral prefrontal and premotor cortex, frontal/Rolandic operculum and the anterior insula. Thus there is evidence that the laterality of feedforward and feedback learning in motor control may generalize beyond the speech domain.

The implications of feedforward and feedback motor control of speech may be relevant to the study and treatment of stuttering. Neuroimaging studies of speech

production in persons who stutter consistently demonstrate increased right hemisphere activation in the precentral and inferior frontal gyrus regions identified in the *shift – no shift* contrast of the current study relative to normal speakers (see Brown et al., 2005 for review). It has been hypothesized that stuttering involves excessive reliance upon auditory feedback control due to poor feedforward commands (Max et al., 2004). The current findings provide support for this view: auditory feedback control during the perturbed feedback condition, clearly demonstrated by the behavioral results, was associated with increased activation of right precentral and inferior frontal cortex. According to this view, the right hemisphere inferior frontal activation is a secondary consequence of the root problem, which is aberrant performance in the feedforward system. Poor feedforward performance leads to auditory errors that in turn activate the right-lateralized auditory feedback control system in an attempt to correct for the errors This hypothesis is consistent with the affects of fluency-inducing therapy on BOLD responses; successful treatment has been associated with a shift toward more normal, left-lateralized frontal activation (De Nil et al., 2003, Neumann et al., 2005).

Left-lateralized feedforward control may similarly impact the study and treatment of apraxia of speech (AOS). Lesions associated with AOS are predominantly located in the left hemisphere (Duffy, 2005), and particularly affect ventral BA 6 and 44 (ventral precentral gyrus, posterior inferior frontal gyrus, frontal operculum) and the underlying white matter. The current findings corroborate characterizations of AOS as a disruption of the use and development of speech motor programs (Ballard et al., 2000, McNeil et al., 2007)

4.7.4 Implications for the DIVA model

The results of the auditory perturbation study provide strong evidence that ventral motor, premotor and adjacent inferior frontal activation is left-lateralized during speech production under normal conditions (see Figure 4-6 and Figure 4-8). In another fMRI study by our group, left-lateralized ventral premotor activation was also demonstrated during production of simple /V/ and /CV/ syllable productions (Ghosh et al., 2008). When auditory feedback was perturbed, a right-lateralized increase of ventral premotor cortex activity was noted. These findings are interpreted as evidence that the left hemisphere plays a predominant role in feedforward control of speech production whereas auditory feedback control is predominantly a function of the right hemisphere premotor and adjacent inferior frontal cortex. There is mounting evidence from studies of compensatory motor control of speech (Houde and Nagarajan, 2007) and reaching movements (Grafton et al., 2008) that supports this interpretation, particularly with respect to asymmetric ventral premotor/posterior inferior frontal involvement in feedforward and feedback control.

Based on these findings, lateralized premotor control maps have been added to the DIVA model (Figure 4-10, boxes highlight in red). A *Feedback Control Map* is hypothesized to lie in the dorsal portion of right posterior inferior frontal gyrus and ventral premotor cortex and (Figure 4-11). The *Feedback Control Map* receives projections from bilateral sensory association cortex that encode sensory error (cf. Figure 4-9). Projections from the *Feedback Control Map* to bilateral *Articulator Velocity Maps* in ventral motor cortex transform these error signals into feedback-based articulator

velocity signals (Equation 11 in Section 2.3.3) that contribute to the outgoing motor command (Equation 2 in Section 2.3.2). The location of the *Feedback Control Map* in MNI space (Figure 4-11) was based on the peak ventral premotor activation noted in the *shift – no shift* contrast (see Figure 4-5). The addition of the *Feedforward Control Map*, in its current form, does not impact the functional behavior of the DIVA model.

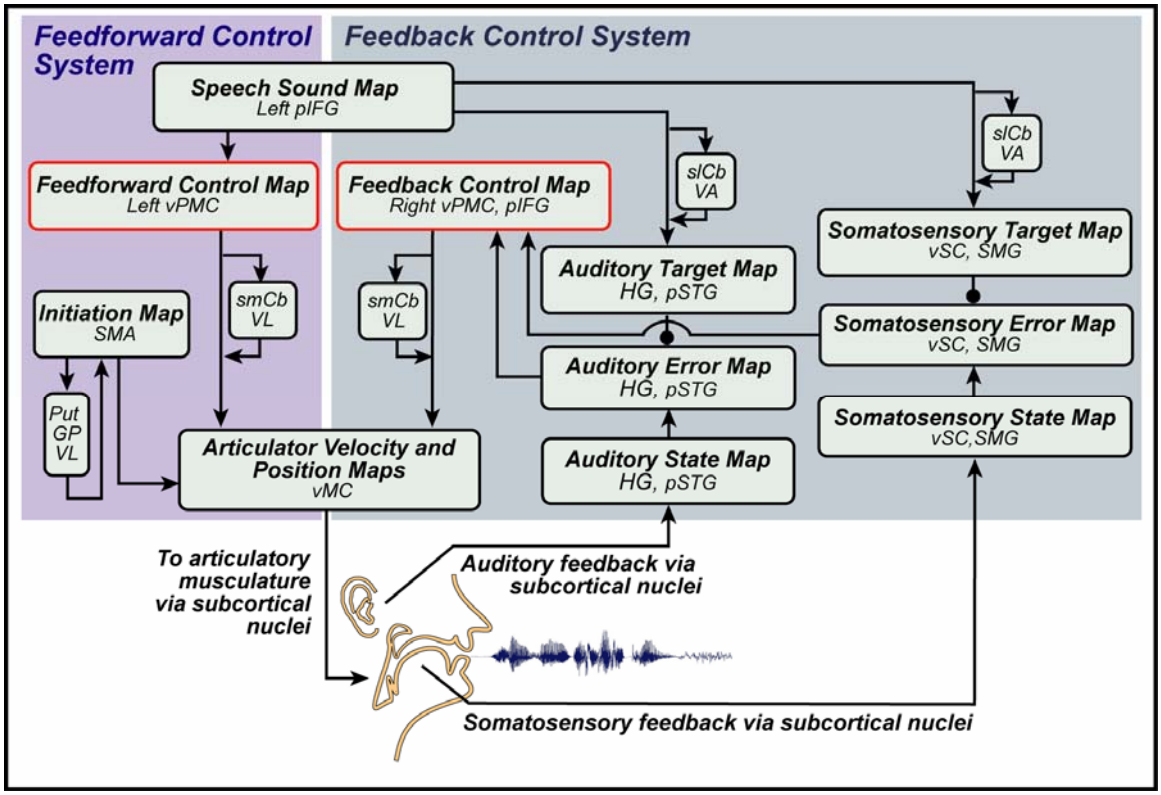


Figure 4-10. Updated DIVA model.
Based on the results of imaging studies of unexpected feedback perturbations, *Feedforward* and *Feedback Control Maps* (indicated by red outlines) located in left ventral premotor and right ventral premotor/inferior frontal cortex , respectively, were added to the model.

A *Feedforward Control Map* is hypothesized to be distributed within left ventral premotor cortex (Figure 4-11). Cells of the *Feedforward Control Map* are hypothesized to represent articulatory gestures(cf. Browman and Goldstein, 1989). Projections from the *Speech Sound Map,* to the *Feedforward Control Map* represent a gestural score (see Section 2.3.1) for a learned speech sound. The *Speech Sound Map* inputs result in activation of cells in the *Feedforward Control Map* that collectively encode the motor programs for a speech sound. Projections from *Feedforward Control Map* cells to *Articulator Velocity Maps* in bilateral ventral motor cortex transform the motor gesture representations into articulator velocity signals (Equation 12 in Section 2.3.4) that contribute to the feedforward component of the outgoing motor command (Equation 2 in Section 2.3.2).

4.7.5 Additional regions to be incorporated

The activation of right inferior frontal gyrus *pars triangularis* (IFt), left posterior middle temporal gyrus, and right inferior cerebellum noted in the *shift – no shift* contrast were not predicted and did not reach a corrected significance threshold. However, their activation is noteworthy when considered with respect to other findings. The greater right IFt response during shifted feedback is consistent with the finding of increased BOLD response in this region when auditory feedback was delayed during speech production (Hashimoto and Sakai, 2003). The IFt increase was accompanied by increased right pSTg activation. A strikingly similar pattern of right hemisphere activation was also noted when listening to unfamiliar vs. familiar voices (Kriegstein and Giraud, 2004). Greater activation was noted near the right inferior frontal sulcus and posterior superior temporal

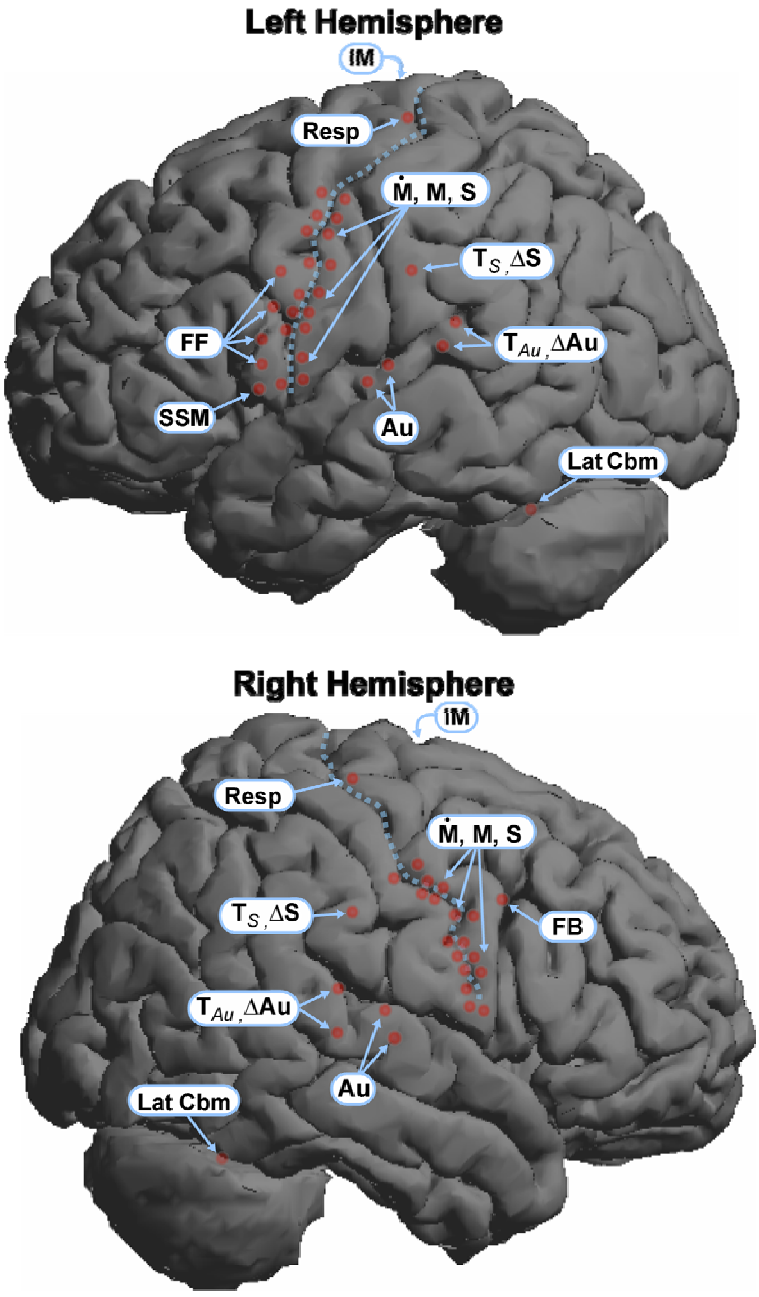


Figure 4-11. Neuranatomical locations of updated DIVA model.
Based on the results of the auditory feedback perturbation experiment, *Feedback* (FB) and *Feedforward* (FF) *Motor Control Maps* were add to right ventral premotor/posterior inferior frontal and left ventral premotor cortex, respectively. To accommodate additional sites, the motor and somatosensory articulator representations are labeled as a group by the symbols representing the *Articulator Velocity, Motor Position*, and *Somatosensory State Maps*, $\dot{M}$, M, and S, respectively. See the caption of Figure 2-6 for further description of the plotting method and additional abbreviations.

gyrus/sulcus during the unfamiliar voice condition. A functional interaction between these regions was demonstrated only when unfamiliar voices were presented. Other studies have shown right inferior frontal activation when sensory input dictates alteration of a pre-set motor response, e.g., successful response inhibition or rapid switching of an ongoing task (see Aron et al., 2004 for review) and detection of rare ("oddball") sensory stimuli (Stevens et al., 2000). In general, right inferior frontal cortex appears to respond when sensory inputs dictate the need for increased sensorimotor processing, either due to the task definition as in the Stevens et al. (2000) oddball study or to the detection of performance errors (as in perturbed auditory feedback during speech). In the current study, this activity may contribute to auditory feedback control by increasing the influence of sensory input on the motor output system.

Activation of right cerebellar lobule VIII has been associated with increased sequence complexity (Bohland and Guenther, 2006) and limb motor task complexity (Habas et al., 2004, Habas and Cabanis, 2006). This area has also been associated specifically with motor error correction, becoming active when unexpected execution errors were induced during a reaching task (Diedrichsen et al., 2005). This result is consistent with the present finding of increased lobule VIII activation when sensory error was introduced.

Activation in the left hemisphere posterior middle temporal cortex was found near the temporal-occipital junction in the *shift – no shift* comparison. This region has been hypothesized to serve as a lexical store (Indefrey and Levelt, 2004, Prabhakaran et al., 2006). Increased activation during the perturbation condition may therefore be the result

of the F1 shift causing the speaker to hear an unanticipated word (e.g., /bɪd/ instead of /bɛd/). Further study is required to incorporate this region and the inferior cerebellum into a comprehensive model of auditory feedback control.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

This dissertation described efforts to advance our understanding of speech motor control by combining computational modeling and functional neuroimaging. The DIVA model of speech acquisition and production has proven to be a valuable tool for the study of the mechanisms underlying normal (Callan et al., 2000, Perkell et al., 2000, Perkell et al., 2004a, Perkell et al., 2004b, Lane et al., 2007, Perkell et al., 2007, Villacorta et al., 2007) and disordered speech (Max et al., 2004, Robin et al., 2008, Terband et al., 2008). The utility of the model is rooted in its expression as a neural network. As such, the model provides a substrate for generating predictions that are well-suited for empirical testing. Importantly, the model's development has been constrained to biologically plausible mechanisms. Thus, as DIVA has come to account for a wide range of speech production phenomena (e.g., Guenther, 1994, Guenther, 1995, Guenther et al., 1998, Callan et al., 2000, Nieto-Castanon et al., 2005), it does so from a unified *and neurobiologically realistic* framework.

The study of speech motor control has also benefited greatly from the advent of non-invasive functional neuroimaging techniques (Belliveau et al., 1992, Kwong et al., 1992, Ogawa et al., 1993). Still in its relative infancy, functional imaging has already provided some consensus regarding the network of brain regions involved in speech production (e.g. Turkeltaub et al., 2002, Indefrey and Levelt, 2004). An understanding of the role played by each region, how they interact, and the breakdowns that result in

speech disorders, like imaging itself, however, is in its infancy. An obstacle to further advancement has been the lack of a cohesive theoretical framework from which to interpret the tremendous amount of imaging data currently available and to guide the design of future investigation. The DIVA model, because it is a neurobiologically plausible neural network, is particularly well-suited for this purpose. The principal aim of work described in this report was to advance our understanding of the auditory feedback control of speech. In pursuing this goal, the foundations necessary for the efficient application of the DIVA model to the design and interpretation of neuroimaging studies of speech motor control were developed.

The first step in this effort, described in Chapter 2, was the neuroanatomical specification of DIVA model components. This was guided by relevant input from behavioral, neuroanatomical, neurophysiological, and neuroimaging data. The model components were mapped to appropriate locations in the MNI stereotactic reference frame and a method for plotting model cell activities in MNI space was developed. As a result, simulations of various speaking conditions with the DIVA model can now be used to generate both behavioral and neural response predictions that can be compared to experimental data. An additional consequence of this work was the emergence of a functional neuroanatomical atlas of speech motor control, represented in Figure 4-10 and Figure 4-11. Future work will address missing elements of this functional atlas (e.g., regions involved in chunking and sequencing speech motor programs), and it will continue to evolve with additional experimental and theoretical work.

The second step in applying the predictive power of the DIVA model to functional imaging experiments was the development of a neuroanatomical parcellation system geared toward studies of speech processes. A set of anatomically-defined regions of interest (ROIs) that encompass the entire brain is described in Chapter 3. This development permits the comparison of functional responses from like anatomical regions in each subject. This ROI-based method of functional imaging data, by accounting for inter-subject anatomical variability, greatly improves statistical sensitivity over standard voxel-based techniques. To the degree possible, regional boundaries were chosen to approximate putative functional distinctions.

An important goal of future work is to further refine and improve upon the functional relevance of these distinctions. For instance, there is growing evidence that dorsal inferior frontal gyrus *par opercularis*, specifically the area within the inferior frontal sulcus, plays a role in speech motor control that is distinct from ventral portions of the inferior frontal gyrus (Bohland and Guenther, 2006). In its present form, the ROI system fails to capture this functional distinction. However, the designation of an ROI representing the inferior frontal sulcus is a straightforward modification to the current system. (Grafton et al., 2008)Similarly, studies of motor learning (Imamizu et al., 2000, Nezafat et al., 2001, Miall and Jenkinson, 2005, O'Reilly et al., 2008) suggest that the designation of additional cerebellar ROIs is necessary to capture functional distinctions.

The methods described above were applied to the study of auditory feedback control of speech. An fMRI experiment was conducted to measure the brain’s response to unexpected perturbations of auditory feedback during speech production. These

experimental responses were compared to predicted responses based on simulations of the DIVA model. According to the model, unexpected auditory feedback during speech production should result in activation of *Auditory Error Map* cells in bilateral posterior superior temporal gyrus. The resulting error signals in turn should lead to increased activation of *Motor Velocity* and *Position Map* cells in bilateral ventral precentral gyrus that drive compensatory articulator movements. These predictions, illustrated by the activation shown in Figure 4-5*B,* served as a guide for the interpretation of imaging results.

Collectively, the behavioral and imaging results presented in Chapter 4 provide important advancements to our understanding of the role of sensory feedback in on-line control of vocalization and the network of brain regions that support this control. Several key aspects of the DIVA model were supported by this investigation: (i) the brain contains auditory error cells that signal differences between a speaker's auditory target and the incoming auditory signal during speech; (ii) these error cells are located in the posterior superior temporal gyrus bilaterally; and (iii) unexpected perturbation of a speaker's auditory feedback results in a compensatory articulatory response within approximately 108-165 ms of the perturbation onset. The behavioral finding is notable as it is the first demonstration of a segmental compensatory response that is fast enough to correct ongoing speech production. The correction of abrupt, transient perturbation of suprasegmental feedback content on a similar time scale has been demonstrated previously (e.g., Burnett et al., 1998, Natke and Kalveram, 2001, Natke et al., 2003). While segmental compensation has been reported, (Purcell and Munhall, 2006b), the

response latency was too long to enable on-line corrections. The slow response reported by Purcell and Munhall was expected consequence of an experimental procedure that maximized response magnitude. The findings reported here, then, are the first to evidence closed-loop auditory feedback-based control of segmental parameters within the normal time scale of a single speech segment.

The fMRI results indicated that, in the absence of feedback error, articulator control was left-lateralized in the frontal cortex. When an auditory feedback error was introduced via the F1 shift, right hemisphere frontal regions, particularly ventral precentral and inferior frontal cortex, were recruited to participate in corrective articulator movements. These findings suggest that while both hemispheres contribute to speech production, feedforward control is predominantly subserved by the left hemisphere whereas auditory feedback-based control is subserved by right hemisphere frontal regions. The demonstration of lateralized feedforward and feedback motor control is supported by the results of recent studies of speech (e.g., Houde and Nagarajan, 2007, Tourville and Guenther, 2007, Toyomura et al., 2007, Grafton et al., 2008) and limb movement (Grafton et al., 2008) control. Notably, each of these studies also found evidence for a right-lateralized lateral premotor/posterior inferior frontal contribution to feedback-based motor control. Accordingly, lateralized *Feedforward* and *Feedback Control Maps* within left and right premotor/inferior frontal cortex, respectively, have been added to the DIVA model (see Figure 4-11).

An exciting consequence of these results is the potential impact on the study and treatment of speech disorders. For instance, the current findings support the theory that

stuttered speech, which is associated with increased right lateral frontal brain activity, is due to an over-reliance on auditory feedback-based control. Likewise, the findings also support the assertion that adult-onset apraxia of speech, a disorder typically associated with damage to the left posterior inferior frontal region, is the result of disrupted feedforward speech motor programs (Ballard et al., 2000, McNeil et al., 2007). Further study of the contributions of the left and right lateral frontal regions during both normal and disordered speech could accelerate the development of efficient treatment paradigms.

Questions also remain regarding the interaction between lateralized frontal regions and bilateral sensory regions. As discussed above, the data are consistent with DIVA-predicted projections from left-lateralized premotor cells to bilateral auditory cortex that encode sensory expectations. The data also suggest that projections from bilateral auditory cells to right premotor cortex encode auditory error. The anatomical pathways that support these mechanisms are not fully understood. Further study of the information conveyed by those projections is also necessary. Studies have begun to explore the putative sensory expectation projections from lateral frontal cortex to auditory cortex, establishing an inhibitory effect linked to ongoing articulator movements (e.g., Heinks-Maldonado et al., 2006). A clear understanding of the units of this inhibitory input, (e.g., is it an acoustic, articulatory, phonological, etc., representation) nor those of the error maps themselves is yet to be established and begs further research.

BIBLIOGRAPHY

Ackermann, H., Graber, S., Hertrich, I. and Daum, I. (1999). Phonemic vowel length contrasts in cerebellar disorders. *Brain and Language,* 67(2): 95-109.

Ackermann, H. and Hertrich, I. (1994). Speech rate and rhythm in cerebellar dysarthria: an acoustic analysis of syllabic timing. *Folia Phoniatrica et Logopaedica,* 46(2): 70-78.

Ackermann, H., Vogel, M., Petersen, D. and Poremba, M. (1992). Speech deficits in ischaemic cerebellar lesions. *Journal of Neurology,* 239(4): 223-227.

Adams, R.D. (1989). *Principles of Neurology*. New York: McGraw-Hill.

Alario, F.X., Chainay, H., Lehericy, S. and Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production. *Brain Research,* 1076(1): 129-143.

Albin, R.L., Young, A.B. and Penney, J.B. (1995). The functional anatomy of disorders of the basal ganglia. *Trends in Neurosciences,* 18(2): 63-64.

Alexander, G.E. and Crutcher, M.D. (1990). Neural representations of the target (goal) of visually guided arm movements in three motor areas of the monkey. *Journal of Neurophysiology,* 64(1): 164-178.

Alexander, G.E., DeLong, M.R. and Strick, P.L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Reviews of Neuroscience,* 9: 357-381.

Andersen, R.A. (1997). Multimodal integration for the representation of space in the posterior parietal cortex. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences,* 352(1360): 1421-1428.

Aron, A.R., Robbins, T.W. and Poldrack, R.A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Science,* 8(4): 170-177.

Ballard, K.J., Granier, J.P. and Robin, D.A. (2000). Understanding the nature of apraxia of speech: Theory, analysis, and treatment. *Aphasiology,* 14(10): 969-995.

Bauer, J.J., Mittal, J., Larson, C.R. and Hain, T.C. (2006). Vocal responses to unanticipated perturbations in voice loudness feedback: an automatic mechanism for stabilizing voice amplitude. *Journal of the Acoustical Society of America,* 119(4): 2363-2371.

Bays, P.M., Flanagan, J.R. and Wolpert, D.M. (2006). Attenuation of self-generated tactile sensations is predictive, not postdictive. *Public Library of Science Biology,* 4(2): e28.

Becker, J.T., MacAndrew, D.K. and Fiez, J.A. (1999). A comment on the functional localization of the phonological storage subsystem of working memory. *Brain and Cognition,* 41(1): 27-38.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P. and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature,* 403(6767): 309-312.

Belliveau, J.W., Kwong, K.K., Kennedy, D.N., Baker, J.R., Stern, C.E., Benson, R., Chesler, D.A., Weisskoff, R.M., Cohen, M.S., Tootell, R.B. and et al. (1992). Magnetic resonance imaging mapping of brain function. Human visual cortex. *Investigative Radiology,* 27 Suppl 2: S59-65.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S., Springer, J.A., Kaufman, J.N. and Possing, E.T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex,* 10(5): 512-528.

Binkofski, F. and Buccino, G. (2004). Motor functions of the Broca's region. *Brain and Language,* 89(2): 362-369.

Blakemore, S.J., Frith, C.D. and Wolpert, D.M. (2001). The cerebellum is involved in predicting the sensory consequences of action. *Neuroreport,* 12(9): 1879-1884.

Blakemore, S.J., Wolpert, D. and Frith, C. (2000). Why can't you tickle yourself? *Neuroreport,* 11(11): R11-16.

Blakemore, S.J., Wolpert, D.M. and Frith, C.D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience,* 1(7): 635-640.

Blakemore, S.J., Wolpert, D.M. and Frith, C.D. (1999). The cerebellum contributes to somatosensory cortical activity during self-produced tactile stimulation. *Neuroimage,* 10(4): 448-459.

Boecker, H., Dagher, A., Ceballos-Baumann, A.O., Passingham, R.E., Samuel, M., Friston, K.J., Poline, J., Dettmers, C., Conrad, B. and Brooks, D.J. (1998). Role of the human rostral supplementary motor area and the basal ganglia in motor sequence control: investigations with H2 15O PET. *Journal of Neurophysiology,* 79(2): 1070-1080.

Bohland, J.W. and Guenther, F.H. (2006). An fMRI investigation of syllable sequence production. *Neuroimage,* 32: 821-841.

Boling, W., Reutens, D.C. and Olivier, A. (2002). Functional topography of the low postcentral area. *Journal of Neurosurgery,* 97(2): 388-395.

Brechmann, A., Baumgart, F. and Scheich, H. (2002). Sound-level-dependent representation of frequency modulations in human auditory cortex: A low-noise fMRI study. *Journal of Neurophysiology,* 87(1): 423-433.

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

Brown, J.W., Bullock, D. and Grossberg, S. (2004). How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Networks,* 17(4): 471-510.

Brown, S., Ingham, R.J., Ingham, J.C., Laird, A.R. and Fox, P.T. (2005). Stuttered and fluent speech production: an ALE meta-analysis of functional neuroimaging studies. *Human Brain Mapping,* 25(1): 105-117.

Buchsbaum, B.R., Hickok, G. and Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science,* 25(5): 663-678.

Buchsbaum, B.R., Olsen, R.K., Koch, P.F., Kohn, P., Kippenhan, J.S. and Berman, K.F. (2005). Reading, hearing, and the planum temporale. *Neuroimage,* 24(2): 444-454.

Bullock, D., Cisek, P. and Grossberg, S. (1998). Cortical networks for control of voluntary arm movements under variable force conditions. *Cerebral Cortex,* 8(1): 48-62.

Burnett, T.A., Freedland, M.B., Larson, C.R. and Hain, T.C. (1998). Voice F0 responses to manipulations in pitch feedback. *Journal of the Acoustical Society of America,* 103(6): 3153-3161.

Burnett, T.A. and Larson, C.R. (2002). Early pitch-shift response is active in both steady and dynamic voice pitch control. *Journal of the Acoustical Society of America,* 112(3 Pt 1): 1058-1063.

Callan, D.E., Kent, R.D., Guenther, F.H. and 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(3): 721-736.

Caplan, D., Gow, D. and Makris, N. (1995). Analysis of Lesions by Mri in Stroke Patients with Acoustic-Phonetic Processing Deficits. *Neurology,* 45(2): 293--298.

Caviness, V.S., Meyer, J., Makris, N. and Kennedy, D. (1996). MRI-based topographic parcellation of human neocortex: an anatomically specified method with estimate of reliability. *Journal of Cognitive Neuroscience,* 8(6): 566-587.

Celsis, P., Boulanouar, K., Doyon, B., 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(1): 135-144.

Chen, H., Smith, M. and Shadmehr, R. (2005). Effects of deep brain stimulation on adaptive control of reaching. *Conference Proceedingsof the Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* 5: 5445-5448.

Christoffels, I.K., Formisano, E. and Schiller, N.O. (2007). Neural correlates of verbal feedback processing: An fMRI study employing overt speech. *Human Brain Mapping,* 28(9): 868-879.

Clopper, C.G., Pisoni, D.B. and de Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America,* 118(3 Pt 1): 1661-1676.

Clower, D.M., Dum, R.P. and Strick, P.L. (2005). Basal ganglia and cerebellar inputs to 'AIP'. *Cerebral Cortex,* 15(7): 913-920.

Clower, D.M., West, R.A., Lynch, J.C. and Strick, P.L. (2001). The inferior parietal lobule is the target of output from the superior colliculus, hippocampus, and cerebellum. *Journal of Neuroscience,* 21(16): 6283-6291.

Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P. and Marchal, G. (1995). Automated multi-modality image registration based on information theory. In: *Information Processing in Medical Imaging* (Y. Bizais, Y., Barillot, C. and Di Paola, R., eds), pp 263–274 Dordrecht, The Netherlands: Kluwer Academic Publishers.

Corfield, D.R., Murphy, K., Josephs, O., Fink, G.R., Frackowiak, R.S., 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,* 86(5): 1468-1477.

Cowie, R.I. and Douglas-Cowie, E. (1983). Speech production in profound post-lingual deafness. In: *Hearing Science and Hearing Disorders* (Lutman, M. E. and Haggard, M. P., eds), pp 183-231 New York: Academic Press.

Crespo-Facorro, B., Kim, J., Andreasen, N.C., Spinks, R., O'Leary, D.S., Bockholt, H.J., Harris, G. and Magnotta, V.A. (2000). Cerebral cortex: a topographic segmentation method using magnetic resonance imaging. *Psychiatry Research,* 100(2): 97-126.

Crosson, B. (1992). *Subcortical functions in language and memory.* New York: Guilford.

Cullen, K.E. (2004). Sensory signals during active versus passive movement. *Current Opinions in Neurobiology,* 14(6): 698-706.

Curio, G., Neuloh, G., Numminen, J., Jousmaki, V. and Hari, R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Human Brain Mapping,* 9(4): 183-191.

Damasio, H. and Damasio, A.R. (1980). The anatomical basis of conduction aphasia. *Brain,* 103(2): 337-350.

Darley, F.L., Aronson, A.E. and Brown, J.R. (1975). *Motor speech disorders*. Philadelphia: Saunders.

Davidson, P.R. and Wolpert, D.M. (2005). Widespread access to predictive models in the motor system: a short review. *Journal of Neural Engineering,* 2(3): S313-319.

De Nil, L.F., Kroll, R.M. and Houle, S. (2001). Functional neuroimaging of cerebellar activation during single word reading and verb generation in stuttering and nonstuttering adults. *Neuroscience Letters,* 302(2-3): 77-80.

De Nil, L.F., Kroll, R.M., Lafaille, S.J. and Houle, S. (2003). A positron emission tomography study of short- and long-term treatment effects on functional brain activation in adults who stutter. *Journal of Fluency Disorders,* 28(4): 357-379; quiz 379-380.

Della-Maggiore, V., Sekuler, A.B., Grady, C.L., Bennett, P.J., Sekuler, R. and McIntosh, A.R. (2000). Corticolimbic interactions associated with performance on a short-term memory task are modified by age. *Journal of Neuroscience,* 20(22): 8410-8416.

Desmond, J.E., Gabrieli, J.D., Wagner, A.D., Ginier, B.L. and Glover, G.H. (1997). Lobular patterns of cerebellar activation in verbal working-memory and finger-tapping tasks as revealed by functional MRI. *Journal of Neuroscience,* 17(24): 9675-9685.

Desmurget, M. and Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Science,* 4(11): 423-431.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V. and Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research,* 91(1): 176-180.

Diedrichsen, J., Hashambhoy, Y., Rane, T. and Shadmehr, R. (2005). Neural correlates of reach errors. *Journal of Neuroscience,* 25(43): 9919-9931.

Donath, T.M., Natke, U. and Kalveram, K.T. (2002). Effects of frequency-shifted auditory feedback on voice F0 contours in syllables. *Journal of the Acoustical Society of America,* 111(1 Pt 1): 357-366.

Dronkers, N.F. (1996). A new brain region for coordinating speech articulation. *Nature,* 384(6605): 159-161.

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

Duffy, J.R. (2005). *Motor speech disorders: Substrates, differential diagnosis, and management*, 2nd ed. St. Louis, MO: Elsevier Mosby.

Duvernoy, H.M. (1999). *The human brain: Surface, blood supply, and three dimensional anatomy*, 2nd ed. New York: Springer-Verlag Wien.

Eliades, S.J. and Wang, X. (2003). Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *Journal of Neurophysiology,* 89(4): 2194-2207.

Eliades, S.J. and Wang, X. (2005). Dynamics of auditory-vocal interaction in monkey auditory cortex. *Cerebral Cortex,* 15(10): 1510-1523.

Engelien, A., Yang, Y., Engelien, W., Zonana, J., Stern, E. and Silbersweig, D.A. (2002). Physiological mapping of human auditory cortices with a silent event-related fMRI technique. *Neuroimage,* 16(4): 944-953.

Evans, A.C., Collins, D.L., Mills, S.R., Brown, E.D., Kelly, R.L. and Peters, T.M. (1993). 3D statistical neuroanatomical models from 305 MRI Volumes. *Proceedings of the IEEE Nuclear Science Symposium on Medical Imaging,* 3: 1813-1817.

Evans, K.C., Shea, S.A. and Saykin, A.J. (1999). Functional MRI localisation of central nervous system regions associated with volitional inspiration in humans. *Journal of Physiology,* 520 Pt 2: 383-392.

Fang, P.C., Stepniewska, I. and Kaas, J.H. (2005). Ipsilateral cortical connections of motor, premotor, frontal eye, and posterior parietal fields in a prosimian primate, Otolemur garnetti. *Journal of Comparative Neurology,* 490(3): 305-333.

Ferguson, S.H. and Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America,* 112(1): 259-271.

Ferrari, P.F., Gallese, V., Rizzolatti, G. and Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience,* 17(8): 1703-1714.

Fesl, G., Moriggl, B., Schmid, U.D., Naidich, T.P., Herholz, K. and Yousry, T.A. (2003). Inferior central sulcus: variations of anatomy and function on the example of the motor tongue area. *Neuroimage,* 20(1): 601-610.

Fiez, J.A. and Petersen, S.E. (1998). Neuroimaging studies of word reading. *Proceedings of the National Academy of Sciences of the United States of America,* 95(3): 914-921.

Filipek, P.A., Richelme, C., Kennedy, D.N. and Caviness, V.S., Jr. (1994). The young adult human brain: an MRI-based morphometric analysis. *Cerebral Cortex,* 4(4): 344-360.

Fink, G.R., Corfield, D.R., Murphy, K., Kobayashi, I., Dettmers, C., Adams, L., Frackowiak, R.S. and Guz, A. (1996). Human cerebral activity with increasing inspiratory force: a study using positron emission tomography. *Journal of Applied Physiology,* 81(3): 1295-1305.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B. and Dale, A.M. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron,* 33(3): 341-355.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B. and Dale, A.M. (2004). Automatically parcellating the human cerebral cortex. *Cerebral Cortex,* 14(1): 11-22.

Flament, D., Ellermann, J.M., Kim, S.G., Ugurbil, K. and Ebner, T.J. (1996). Functional magnetic resonance imaging of cerebellar activation during the learning of visuomotor dissociation task. *Human Brain Mapping,* 4(3): 210-226.

Foerstner, O. (1936). The motor cortex of man in the light of Hughlings Jackson' doctrines. *Brain,* 59: 135-159.

Fox, P.T. (2003). Brain imaging in stuttering: where next? *Journal of Fluency Disorders,* 28(4): 265-272.

Fox, P.T., Huang, A., Parsons, L.M., Xiong, J.H., 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(1): 196-209.

Friederici, A.D., Meyer, M. and von Cramon, D.Y. (2000). Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information. *Brain and Language,* 74(2): 289-300.

Friston, K.J., Frith, C.D., Frackowiak, R.S. and Turner, R. (1995a). Characterizing dynamic brain responses with fMRI: a multivariate approach. *Neuroimage,* 2(2): 166-172.

Friston, K.J., Holmes, A.P., Poline, J.B., Grasby, P.J., Williams, S.C., Frackowiak, R.S. and Turner, R. (1995b). Analysis of fMRI time-series revisited. *Neuroimage,* 2(1): 45-53.

Fu, C.H., Vythelingum, G.N., Brammer, M.J., Williams, S.C., Amaro, E., Jr., Andrew, C.M., Yaguez, L., van Haren, N.E., Matsumoto, K. and McGuire, P.K. (2006). An fMRI study of verbal self-monitoring: neural correlates of auditory verbal feedback. *Cerebral Cortex,* 16(7): 969-977.

Fulton, J.F. (1938). *Physiology of the nervous system*. London: Oxford University Press.

Geschwind, N. (1965). Disconnexion syndromes in animals and man. II. *Brain,* 88(3): 585--644.

Ghosh, S.S., Bohland, J.W. and Guenther, F.H. (2003). Comparisons of brain regions involved in overt production of elementary phonetic units. *Neuroimage (9th Meeting of the Organization for Human Brain Mapping, New York),* 19(2): S57.

Ghosh, S.S., Tourville, J.A. and Guenther, F.H. (2008). An fMRI study of the overt production of simple speech sounds. *Journal of Speech, Language, and Hearing Research,* in press.

Giraud, A.L. and Price, C.J. (2001). The constraints functional neuroimaging places on classical models of auditory word processing. *Journal of Cognitive Neuroscience,* 13(6): 754-765.

Goodglass, H. (1993). *Understanding aphasia*. New York: Academic Press.

Grafton, S.T., Schmitt, P., Van Horn, J. and Diedrichsen, J. (2008). Neural substrates of visuomotor learning based on improved feedback and prediction. *Neuroimage,* 39: 1383-1395.

Graziano, M.S., Taylor, C.S., Moore, T. and Cooke, D.F. (2002). The cortical control of movement revisited. *Neuron,* 36(3): 349-362.

Greenlee, J.D., Oya, H., Kawasaki, H., Volkov, I.O., Kaufman, O.P., Kovach, C., Howard, M.A. and Brugge, J.F. (2004). A functional connection between inferior frontal gyrus and orofacial motor cortex in human. *Journal of Neurophysiology,* 92(2): 1153-1164.

Greenlee, J.D., Oya, H., Kawasaki, H., Volkov, I.O., Severson, M.A., 3rd, Howard, M.A., 3rd and Brugge, J.F. (2007). Functional connections within the human inferior frontal gyrus. *Journal of Comparative Neurology,* 503(4): 550-559.

Guenther, F.H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics,* 72(1): 43-53.

Guenther, F.H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review,* 102(3): 594-621.

Guenther, F.H. and Ghosh, S.S. (2003) A neural model of speech production. In: *Proceedings of the 6th International Seminar on Speech Production*, pp 85-90 Sydney, Australia.

Guenther, F.H., Ghosh, S.S. and Tourville, J.A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language,* 96(3): 280-301.

Guenther, F.H., Hampson, M. and Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review,* 105(4): 611-633.

Guenther, F.H., Nieto-Castanon, A., Ghosh, S.S. and Tourville, J.A. (2004). Representation of sound categories in auditory cortical maps. *Journal of Speech, Language, and Hearing Research,* 47(1): 46-57.

Habas, C., Axelrad, H., Nguyen, T.H. and Cabanis, E.A. (2004). Specific neocerebellar activation during out-of-phase bimanual movements. *Neuroreport,* 15(4): 595-599.

Habas, C. and Cabanis, E.A. (2006). Cortical areas functionally linked with the cerebellar second homunculus during out-of-phase bimanual movements. *Neuroradiology,* 48(4): 273-279.

Hain, T.C., Burnett, T.A., Kiran, S., Larson, C.R., Singh, S. and Kenney, M.K. (2000). Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Experimental Brain Research,* 130(2): 133-141.

Hashimoto, Y. and Sakai, K.L. (2003). Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: an fMRI study. *Human Brain Mapping,* 20(1): 22-28.

Heinks-Maldonado, T.H. and Houde, J.F. (2005). Compensatory responses to brief perturbations of speech amplitude. *Acoustics Research Letters Online,* 6(3): 131-137.

Heinks-Maldonado, T.H., Mathalon, D.H., Gray, M. and Ford, J.M. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology,* 42(2): 180-190.

Heinks-Maldonado, T.H., Nagarajan, S.S. and Houde, J.F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport,* 17(13): 1375-1379.

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(5): 673-682.

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(2): 156-160.

Hickok, G. and Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Science,* 4(4): 131-138.

Hickok, G. and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition,* 92(1-2): 67-99.

Hickok, G. and Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience,* 8(5): 393-402.

Hillenbrand, J.M., Clark, M.J. and Nearey, T.M. (2001). Effects of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America,* 109(2): 748-763.

Hillis, A.E., Work, M., Barker, P.B., Jacobs, M.A., Breese, E.L. and Maurer, K. (2004). Re-examining the brain regions crucial for orchestrating speech articulation. *Brain,* 127(Pt 7): 1479-1487.

Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., Ishizu, K., Yonekura, Y., Nagahama, Y., Fukuyama, H. and Konishi, J. (1997). Cortical processing mechanism for vocalization with auditory verbal feedback. *Neuroreport,* 8(9-10): 2379-2382.

Houde, J.F. and Jordan, M.I. (1998). Sensorimotor adaptation in speech production. *Science,* 279(5354): 1213-1216.

Houde, J.F. and Jordan, M.I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research,* 45(2): 295-310.

Houde, J.F. and Nagarajan, S.S. (2007) How is auditory feedback processed during speaking? In: *Program of the 154th Meeting of the Acoustical Society of America, Journal of the Acoustical Society of America*, vol. 122(5, Pt. 2), p 3087 New Orleans, Louisiana.

Houde, J.F., Nagarajan, S.S., Sekihara, K. and Merzenich, M.M. (2002). Modulation of the auditory cortex during speech: an MEG study. *Journal of Cognitive Neuroscience,* 14(8): 1125-1138.

Hu, L. and Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling,* 6(1): 1-55.

Iacoboni, M., Woods, R.P., Brass, M., Bekkering, H., Mazziotta, J.C. and Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science,* 286(5449): 2526-2528.

Imamizu, H., Higuchi, S., Toda, A. and Kawato, M. (2007). Reorganization of brain activity for multiple internal models after short but intensive training. *Cortex,* 43(3): 338-349.

Imamizu, H., Miyauchi, S., Tamada, T., Sasaki, Y., Takino, R., Putz, B., Yoshioka, T. and Kawato, M. (2000). Human cerebellar activity reflecting an acquired internal model of a new tool. *Nature,* 403(6766): 192-195.

Indefrey, P. and Levelt, W.J. (2004). The spatial and temporal signatures of word production components. *Cognition,* 92(1-2): 101-144.

Ito, M. (2000). Mechanisms of motor learning in the cerebellum. *Brain Research,* 886(1-2): 237-245.

Ito, T., Kimura, T. and Gomi, H. (2005). The motor cortex is involved in reflexive compensatory adjustment of speech articulation. *Neuroreport,* 16(16): 1791-1794.

Jancke, L., Wustenberg, T., Scheich, H. and Heinze, H.J. (2002). Phonetic perception and the temporal cortex. *Neuroimage,* 15(4): 733-746.

Johnson, M.D. and Ojemann, G.A. (2000). The role of the human thalamus in language and memory: evidence from electrophysiological studies. *Brain and Cognition,* 42(2): 218-230.

Jonas, S. (1981). The supplementary motor region and speech emission. *Journal of Communication Disorders,* 14: 349-373.

Jonas, S. (1987). The supplementary motor region and speech. In: *The frontal lobes revisited* (Perecman, E., ed), pp 241–250 New York: IRBN Press.

Jones, J.A. and Munhall, K.G. (2002). The role of auditory feedback during phonation: studies of Mandarin tone production. *Journal of Phonetics,* 30: 303-320.

Jones, J.A. and Munhall, K.G. (2005). Remapping auditory-motor representations in voice production. *Current Biology,* 15(19): 1768-1772.

Jonides, J., Schumacher, E.H., Smith, E.E., Koeppe, R.A., Awh, E., Reuter-Lorenz, P.A., Marshuetz, C. and Willis, C.R. (1998). The role of parietal cortex in verbal working memory. *Journal of Neuroscience,* 18(13): 5026-5034.

Jurgens, U. (1984). The efferent and efferent connections of the supplementary motor area. *Brain Research,* 300: 63-81.

Kaas, J.H. and Hackett, T.A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences of the United States of America,* 97(22): 11793-11799.

Kalaska, J.F., Cohen, D.A., Hyde, M.L. and Prud'homme, M. (1989). A comparison of movement direction-related versus load direction-related activity in primate motor cortex, using a two-dimensional reaching task. *Journal of Neuroscience,* 9(6): 2080-2102.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinions in Neurobiology,* 9(6): 718-727.

Kawato, M. and Gomi, H. (1992). A computational model of four regions of the cerebellum based on feedback-error learning. *Biological Cybernetics,* 68(2): 95-103.

Kawato, M., Kuroda, T., Imamizu, H., Nakano, E., Miyauchi, S. and Yoshioka, T. (2003). Internal forward models in the cerebellum: fMRI study on grip force and load force coupling. *Progress in Brain Research,* 142: 171-188.

Kelly, R.M. and Strick, P.L. (2003). Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. *Journal of Neuroscience,* 23(23): 8432-8444.

Kennedy, D.N., Filipek, P.A. and Caviness, V.R. (1989). Anatomic segmentation and volumetric calculations in nuclear magnetic resonance imaging. *IEEE Transactions on Medical Imaging,* 8(1): 1-7.

Kent, R.D. and Tjaden, K. (1997). Brain functions underlying speech. In: *Handbook of phonetic sciences* (Hardcastle, W. J. and Laver, J., eds), pp 220-255 Oxford: Blackwell.

Keuhn, D.P., Lemme, M.L. and Baumgaertner, J.M. (1989). *Neural bases of speech, hearing, and language*. Austin, TX: Pro-ed.

Kim, J.J., Crespo-Facorro, B., Andreasen, N.C., O'Leary, D.S., Zhang, B., Harris, G. and Magnotta, V.A. (2000). An MRI-based parcellation method for the temporal lobe. *Neuroimage,* 11(4): 271-288.

Kohler, E., Keysers, C., Umilta, M.A., Fogassi, L., Gallese, V. and Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science,* 297(5582): 846-848.

Krakauer, J.W. and Ghez, C. (1999). Voluntary movement. In: *Principles of Neural Science* (Kandel, E. R., Schwartz, J. J. and Jessell, T. M., eds), pp 756-781 New York: McGraw Hill.

Kriegstein, K.V. and Giraud, A.L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage,* 22(2): 948-955.

Kropotov, J.D. and Etlinger, S.C. (1999). Selection of actions in the basal ganglia-thalamocortical circuits: review and model. *Int J Psychophysiol,* 31(3): 197-217.

Kuriki, S., Mori, T. and Hirata, Y. (1999). Motor planning center for speech articulation in the normal human brain. *Neuroreport,* 10(4): 765-769.

Kwong, K.K., Belliveau, J.W., Chesler, D.A., Goldberg, I.E., Weisskoff, R.M., Poncelet, B.P., Kennedy, D.N., Hoppel, B.E., Cohen, M.S., Turner, R. and et al. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences of the United States of America,* 89(12): 5675-5679.

Lancaster, J.L., Tordesillas-Gutierrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., Mazziotta, J.C. and Fox, P.T. (2007). Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Human Brain Mapping,* 28(11): 1194-1205.

Lane, H., Matthies, M.L., Guenther, F.H., Denny, M., Perkell, J.S., Stockmann, E., Tiede, M., Vick, J. and Zandipour, M. (2007). Effects of short- and long-term changes in auditory feedback on vowel and sibilant contrasts. *Journal of Speech, Language, and Hearing Research,* 50(4): 913-927.

Lane, H. and Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *J Speech Lang Hear Res,* 14: 677-709.

Lane, H. and Webster, J.W. (1991). Speech deterioration in postlingually deafened adults. *Journal of the Acoustical Society of America,* 89(2): 859-866.

Larson, C.R., Burnett, T.A., Kiran, S. and Hain, T.C. (2000). Effects of pitch-shift velocity on voice F0 responses. *Journal of the Acoustical Society of America,* 107(1): 559-564.

Le, T.H., Patel, S. and Roberts, T.P. (2001). Functional MRI of human auditory cortex using block and event-related designs. *Magnetic Resonance in Medicine,* 45(2): 254-260.

Lehericy, S., Ducros, M., Krainik, A., Francois, C., Van de Moortele, P.F., Ugurbil, K. and Kim, D.S. (2004). 3-D diffusion tensor axonal tracking shows distinct SMA and pre-SMA projections to the human striatum. *Cerebral Cortex,* 14(12): 1302-1309.

Levelt, W.J., Roelofs, A. and Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences,* 22(1): 1-38; discussion 38-75.

Levelt, W.J. and Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition,* 50(1-3): 239-269.

Lombard, E. (1911). Le signe de l'elevation de la voix. *Annales des Maladies de l'Oreille du Larynx,* 37: 101-119.

Lotze, M., Erb, M., Flor, H., Huelsmann, E., Godde, B. and Grodd, W. (2000a). fMRI evaluation of somatotopic representation in human primary motor cortex. *Neuroimage,* 11(5 Pt 1): 473-581.

Lotze, M., Seggewies, G., Erb, M., Grodd, W. and Birbaumer, N. (2000b). The representation of articulation in the primary sensorimotor cortex. *Neuroreport,* 11(13): 2985-2989.

Ludlow, C.L. (2005). Central nervous system control of the laryngeal muscles in humans. *Respiratory Physiology and Neurobiology,* 147(2-3): 205-222.

Luppino, G., Matelli, M., Camarda, R. and Rizzolatti, G. (1993). Corticocortical connections of area F3 (SMA-proper) and area F6 (pre-SMA) in the macaque monkey. *Journal of Comparative Neurology,* 338(1): 114--140.

Luppino, G., Murata, A., Govoni, P. and Matelli, M. (1999). Largely segregated parietofrontal connections linking rostral intraparietal cortex (areas AIP and VIP) and the ventral premotor cortex (areas F5 and F4). *Experimental Brain Research,* 128(1-2): 181-187.

Maeda, S. (1990). Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. In: *Speech Production and Speech Modeling* (Hardcastle, W. J. and Marchal, A., eds) Boston: Kluwer Academic Publishers.

Martin, J.H. (1996). *Neuroanatomy: Text and atlas*, 2nd ed. Stamford, CT: Appleton & Lange.

Mathiak, K., Hertrich, I., Grodd, W. and Ackermann, H. (2002). Cerebellum and speech perception: a functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience,* 14(6): 902-912.

Matsumoto, R., Nair, D.R., LaPresto, E., Bingaman, W., Shibasaki, H. and Luders, H.O. (2007). Functional connectivity in human cortical motor system: a cortico-cortical evoked potential study. *Brain,* 130(Pt 1): 181-197.

Matsumoto, R., Nair, D.R., LaPresto, E., Najm, I., Bingaman, W., Shibasaki, H. and Luders, H.O. (2004). Functional connectivity in the human language system: a cortico-cortical evoked potential study. *Brain,* 127(Pt 10): 2316-2330.

Max, L., Guenther, F.H., Gracco, V.L., Ghosh, S.S. and Wallace, M.E. (2004). Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: A theoretical model of stuttering. *Contemporary Issues in Communication Science and Disorders,* 31: 105-122.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N., Le Goualher, G., Feidler, J., Smith, K., Boomsma, D., Hulshoff Pol, H., Cannon, T., Kawashima, R. and Mazoyer, B. (2001). A four-dimensional probabilistic atlas of the human brain. *Journal of the American Medical Informatics Association,* 8(5): 401-430.

Mazziotta, J.C., Toga, A.W., Evans, A.C., Fox, P.T. and Lancaster, J.L. (1995). Digital brain atlases. *Trends in Neurosciences,* 18(5): 210-211.

McCarthy, G. and Allison, T. (1995). Trigeminal evoked potentials in somatosensory cortex of the Macaca mulatta. *Journal of Neurosurgery,* 82(6): 1015-1020.

McCarthy, G., Allison, T. and Spencer, D.D. (1993). Localization of the face area of human sensorimotor cortex by intracranial recording of somatosensory evoked potentials. *Journal of Neurosurgery,* 79(6): 874-884.

McClean, M.D. and Sapir, S. (1981). Some effects of auditory stimulation on perioral motor unit discharge and their implications for speech production. *Journal of the Acoustical Society of America,* 69(5): 1452-1457.

McGuire, P.K., Silbersweig, D.A. and Frith, C.D. (1996). Functional neuroanatomy of verbal self-monitoring. *Brain,* 119 (Pt 3): 907-917.

McIntosh, A.R. and Gonzalez-Lima, F. (1994). Structural equation modeling and its application to network analysis in functional imaging. *Human Brain Mapping,* 2: 2-22.

McLeod, S. and Bleile, K. (2003) Neurological and developmental foundations of speech acquisition. In: *American Speech-Language-Hearing Association* Chicago.

McNeil, M.R., Robin, D.A. and Schmidt, R.A. (2007). *Apraxia of speech: Definition and differential diagnosis*, 2nd. ed. New York: Thieme.

Miall, R.C. and Jenkinson, E.W. (2005). Functional imaging of changes in cerebellar activity related to learning during a novel eye-hand tracking task. *Experimental Brain Research,* 166(2): 170-183.

Miall, R.C. and Wolpert, D.M. (1996). Forward models for physiological motor control. *Neural Networks,* 9(8): 1265-1279.

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

Middleton, F.A. and Strick, P.L. (2000). Basal ganglia and cerebellar loops: motor and cognitive circuits. *Brain Res Brain Res Rev,* 31(2-3): 236-250.

Middleton, F.A. and Strick, P.L. (2001). Cerebellar projections to the prefrontal cortex of the primate. *Journal of Neuroscience,* 21(2): 700-712.

Mink, J.W. (1996). The basal ganglia: focused selection and inhibition of competing motor programs. *Progress in Neurobiology,* 50(4): 381-425.

Mink, J.W. and Thach, W.T. (1993). Basal ganglia intrinsic circuits and their role in behavior. *Current Opinions in Neurobiology,* 3(6): 950-957.

Mochizuki, H. and Saito, H. (1990). Mesial frontal lobe syndromes: correlations between neurological deficits and radiological localizations. *Tohoku Journal of Experimental Medicine,* 161 Suppl: 231-239.

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T. and Zilles, K. (2001). Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage,* 13(4): 684-701.

Morosan, P., Schleicher, A., Amunts, K. and Zilles, K. (2005). Multimodal architectonic mapping of human superior temporal gyrus. *Anatomy and Embryology,* 210(5-6): 401-406.

Mummery, C.J., Ashburner, J., Scott, S.K. and Wise, R.J.S. (1999). Functional neuroimaging of speech perception in six normal and two aphasic subjects. *Journal of the Acoustical Society of America,* 106(1): 449-457.

Nakamura, A., Yamada, T., Goto, A., Kato, T., Ito, K., Abe, Y., Kachi, T. and Kakigi, R. (1998). Somatosensory homunculus as drawn by MEG. *Neuroimage,* 7(4 Pt 1): 377-386.

Natke, U., Donath, T.M. and Kalveram, K.T. (2003). Control of voice fundamental frequency in speaking versus singing. *Journal of the Acoustical Society of America,* 113(3): 1587-1593.

Natke, U. and Kalveram, K.T. (2001). Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables. *Journal of Speech Language and Hearing Research,* 44(3): 577-584.

Neilson, M. and Neilson, P. (1987). Speech motor control and stuttering: A computational model of adaptive sensory-motor processing. *Speech Communication,* 6: 325-333.

Nemeth, G., Hegedus, K. and Molnar, L. (1988). Akinetic mutism associated with bicingular lesions: clinicopathological and functional anatomical correlates. *European Archives of Psychiatry and Neurological Sciences,* 237(4): 218-222.

Neumann, K., Preibisch, C., Euler, H.A., von Gudenberg, A.W., Lanfermann, H., Gall, V. and Giraud, A.L. (2005). Cortical plasticity associated with stuttering therapy. *Journal of Fluency Disorders,* 30(1): 23-39.

Newman, S.D. and Twieg, D. (2001). Differences in auditory processing of words and pseudowords: An fMRI study. *Human Brain Mapping,* 14(1): 39-47.

Nezafat, R., Shadmehr, R. and Holcomb, H.H. (2001). Long-term adaptation to dynamics of reaching movements: a PET study. *Experimental Brain Research,* 140(1): 66-76.

Nieto-Castanon, A., Ghosh, S.S., Tourville, J.A. and Guenther, F.H. (2003). Region of interest based analysis of functional imaging data. *Neuroimage,* 19(4): 1303-1316.

Nieto-Castanon, A., Guenther, F.H., Perkell, J.S. and Curtin, H.D. (2005). A modeling investigation of articulatory variability and acoustic stability during American English /r/ production. *Journal of the Acoustical Society of America,* 117(5): 3196-3212.

Numminen, J. and Curio, G. (1999). Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neuroscience Letters,* 272(1): 29-32.

Numminen, J., Salmelin, R. and Hari, R. (1999). Subject's own speech reduces reactivity of the human auditory cortex. *Neuroscience Letters,* 265(2): 119-122.

O'Reilly, J.X., Mesulam, M.M. and Nobre, A.C. (2008). The cerebellum predicts the timing of perceptual events. *Journal of Neuroscience,* 28(9): 2252-2260.

Ogawa, S., Menon, R.S., Tank, D.W., Kim, S.G., Merkle, H., Ellermann, J.M. and Ugurbil, K. (1993). Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. A comparison of signal characteristics with a biophysical model. *Biophysical Journal,* 64(3): 803-812.

Ohyama, T., Nores, W.L., Murphy, M. and Mauk, M.D. (2003). What the cerebellum computes. *Trends in Neurosciences,* 26(4): 222-227.

Ojemann, G.A. (1991). Cortical organization of language. *Journal of Neuroscience,* 11(8): 2281-2287.

Okada, K. and Hickok, G. (2006). Left posterior auditory-related cortices participate both in speech perception and speech production: Neural overlap revealed by fMRI. *Brain and Language,* 98(1): 112-117.

Okada, K., Smith, K.R., Humphries, C. and Hickok, G. (2003). Word length modulates neural activity in auditory cortex during covert object naming. *Neuroreport,* 14(18): 2323-2326.

Oller, D.K. (1980). The emergence of the sounds of speech in infancy. In: *Child Phonology, Volume 1: Production*, vol. 1 (Yeni-Komishian, G. H., Kavanagh, J. F. and Ferguson, C. A., eds), pp 93-112 New York: Academic Press.

Oller, D.K. and Eilers, R.E. (1988). The role of audition in infant babbling. *Child Development,* 59(2): 441-449.

Ono, M., Kubik, S. and Abernathy, C.D. (1990). *Atlas of the cerebral sulci*. Stuttgart: Georg Thieme Verlag.

Ozdemir, E., Norton, A. and Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *Neuroimage,* 33(2): 628-635.

Parent, A. and Hazrati, L.N. (1995). Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. *Brain Research Brain Research Reviews,* 20(1): 91-127.

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

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(11): 2236-2246.

Penfield, W. and Rasmussen, T. (1950). *The cerbral cortex of man. A clinical study of localization of function*. New York: Macmillan.

Penfield, W. and Roberts, L. (1959). *Speech and brain mechanisms*. Princeton, N.J: Princeton University Press.

Penfield, W. and Welch, K. (1951). The supplementary motor area of the cerebral cortex; a clinical and experimental study. *American Medical Association Archives of Neurology and Psychiatry,* 66(3): 289-317.

Penhune, V.B. and Doyon, J. (2005). Cerebellum and M1 interaction during early learning of timed motor sequences. *Neuroimage,* 26(3): 801-812.

Perkell, J.S., Denny, M., Lane, H., Guenther, F., Matthies, M.L., Tiede, M., Vick, J., Zandipour, M. and Burton, E. (2007). Effects of masking noise on vowel and sibilant contrasts in normal-hearing speakers and postlingually deafened cochlear implant users. *Journal of the Acoustical Society of America,* 121(1): 505-518.

Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Perrier, P., Vick, J., Wilhems-Tricarico, R. and Zandipour, M. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and profound hearing loss. *Journal of Phonetics,* 28: 232-272.

Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Stockmann, E., Tiede, M. and Zandipour, M. (2004a). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *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. and Guenther, F.H. (2004b). The distinctness of speakers' /s/-/S/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *Journal of Speech, Language, and Hearing Research,* 47(6): 1259-1269.

Picard, C. and Olivier, A. (1983). Sensory cortical tongue representation in man. *Journal of Neurosurgery,* 59(5): 781-789.

Pickett, E., Kuniholm, E., Protopapas, A., Friedman, J. and Lieberman, P. (1998). Selective speech motor, syntax and cognitive deficits associated with bilateral damage to the putamen and the head of the caudate nuclues: a case study. *Neuropsychologia,* 36(173-188).

Pittman, A.L. and Wiley, T.L. (2001). Recognition of speech produced in noise. *Journal of Speech, Language, and Hearing Research,* 44(3): 487-496.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication,* 41(1): 245-255.

Prabhakaran, R., Blumstein, S.E., Myers, E.B., Hutchison, E. and Britton, B. (2006). An event-related fMRI investigation of phonological-lexical competition. *Neuropsychologia,* 44(12): 2209-2221.

Prescott, T.J., Montes Gonzalez, F.M., Gurney, K., Humphries, M.D. and Redgrave, P. (2006). A robot model of the basal ganglia: behavior and intrinsic processing. *Neural Networks,* 19(1): 31-61.

Pugh, K.R., Mencl, W.E., Jenner, A.R., Katz, L., Frost, S.J., Lee, J.R., Shaywitz, S.E. and Shaywitz, B.A. (2001). Neurobiological studies of reading and reading disability. *Journal of Communication Disorders,* 34(6): 479-492.

Purcell, D.W. and Munhall, K.G. (2006a). Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. *Journal of the Acoustical Society of America,* 120(2): 966-977.

Purcell, D.W. and Munhall, K.G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. *Journal of the Acoustical Society of America,* 119(4): 2288-2297.

Rauschecker, J.P. and Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America,* 97(22): 11800-11806.

Redgrave, P., Prescott, T.J. and Gurney, K. (1999). The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience,* 89(4): 1009--1023.

Reppas, J.B., Usrey, W.M. and Reid, R.C. (2002). Saccadic eye movements modulate visual reponses in the lateral geniculate nucleus. *Neuron,* 35: 961-074.

Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G. and Grodd, W. (2000a). Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *Neuroreport,* 11(9): 1997-2000.

Riecker, A., Ackermann, H., Wildgruber, D., Meyer, J., Dogil, G., Haider, H. and Grodd, W. (2000b). Articulatory/phonetic sequencing at the level of the anterior perisylvian cortex: a functional magnetic resonance imaging (fmri) study. *Brain and Language,* 75(2): 259-276.

Riecker, A., Kassubek, J., Groschel, K., Grodd, W. and Ackermann, H. (2006). The cerebral control of speech tempo: opposite relationship between speaking rate and BOLD signal changes at striatal and cerebellar structures. *Neuroimage,* 29(1): 46-53.

Riecker, A., Wildgruber, D., Dogil, G., Grodd, W. and Ackermann, H. (2002). Hemispheric lateralization effects of rhythm implementation during syllable repetitions: An fMRI Study. *Neuroimage,* 16(1): 169--176.

Rivier, F. and Clarke, S. (1997). Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: evidence for multiple auditory areas. *Neuroimage,* 6(4): 288-304.

Rizzolatti, G. and Arbib, M.A. (1998). Language within our grasp. *Trends in Neurosciences,* 21(5): 188-194.

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G. and Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Experimental Brain Research,* 71(3): 491-507.

Rizzolatti, G., Fadiga, L., Gallese, V. and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Research Cognitive Brain Research,* 3(2): 131-141.

Rizzolatti, G., Fogassi, L. and Gallese, V. (1997). Parietal cortex: from sight to action. *Current Opinions in Neurobiology,* 7(4): 562-567.

Rizzolatti, G. and Luppino, G. (2001). The cortical motor system. *Neuron,* 31(6): 889-901.

Robin, D.A., Guenther, F.H., Narayana, S., Jacks, A., Tourville, J.A., Ramage, A.E., Lancaster, J.L., Franklin, C., Ghosh, S. and Fox, P.T. (2008) A transcranial magnetic stimulation virtual lesion study of speech. In: *Proceedings of the Conference on Motor Speech* Monterey, California.

Roy, J.E. and Cullen, K.E. (2004). Dissociating self-generated from passively applied head motion: neural mechanisms in the vestibular nuclei. *Journal of Neuroscience,* 24(9): 2102-2111.

Schaltenbrand, G. (1965). The effects of stereotactic electrical stimulation in the depth of the brain. *Brain,* 88(4): 835-840.

Schaltenbrand, G. (1975). The effects on speech and language of stereotactical stimulation in thalamus and corpus callosum. *Brain and Language,* 2(1): 70-77.

Schmahmann, D., Doyon, J., Toga, A.W., Petrides, M. and Evans, A.C. (2000). *MRI atlas of the human cerebellum*. San Diego.

Schmahmann, J. and Pandya, D. (2006). *Fiber pathways of the brain*. New York: Oxford University Press.

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

Schreiner, C.E. (1995). Order and disorder in auditory cortical maps. *Current Opinions in Neurobiology,* 5(4): 489-496.

Schreurs, B.G., McIntosh, A.R., Bahro, M., Herscovitch, P., Sunderland, T. and Molchan, S.E. (1997). Lateralization and behavioral correlation of changes in regional cerebral blood flow with classical conditioning of the human eyeblink response. *Journal of Neurophysiology,* 77(4): 2153-2163.

Schumacker, R. and Lomax, R. (2004). *A beginner's guide to structural equation modeling*, 2nd ed. Mahwah, N.J., USA: Lawrence Elrbaum.

Scott, S.K., Blank, C.C., Rosen, S. and Wise, R.J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain,* 123 Pt 12: 2400-2406.

Scott, S.K. and Johnsrude, I.S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences,* 26(2): 100-107.

Sidtis, J.J., Gomez, C., Groshong, A., Strother, S.C. and Rottenberg, D.A. (2006a). Mapping cerebral blood flow during speech production in hereditary ataxia. *Neuroimage,* 31(1): 246-254.

Sidtis, J.J., Gomez, C., Groshong, A., Strother, S.C. and Rottenberg, D.A. (2006b). Mapping cerebral blood flow during speech production in hereditary ataxia. *Neuroimage,* 31(1): 246-254.

Silveri, M.C., Di Betta, A.M., Filippini, V., Leggio, M.G. and Molinari, M. (1998). Verbal short-term store-rehearsal system and the cerebellum. Evidence from a patient with a right cerebellar lesion. *Brain,* 121 (Pt 11): 2175--2187.

Simonyan, K. and Jurgens, U. (2003). Efferent subcortical projections of the laryngeal motorcortex in the rhesus monkey. *Brain Research,* 974(1-2): 43-59.

Soros, P., Sokoloff, L.G., Bose, A., McIntosh, A.R., Graham, S.J. and Stuss, D.T. (2006). Clustered functional MRI of overt speech production. *Neuroimage,* 32(1): 376-387.

Sperry, R.W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology,* 43(6): 482-489.

Stark, R.E. (1980). Stages of speech development in the first year of life. In: *Child Phonology, Volume 1: Production* (Yeni-Komishian, G. H., Kavanagh, J. F. and Ferguson, C. A., eds), pp 73-92 New York: Academic Press.

Stevens, A.A., Skudlarski, P., Gatenby, J.C. and Gore, J.C. (2000). Event-related fMRI of auditory and visual oddball tasks. *Magnetic Resonance Imaging,* 18(5): 495-502.

Stuart, A., Kalinowski, J., Rastatter, M.P. and Lynch, K. (2002). Effect of delayed auditory feedback on normal speakers at two speech rates. *Journal of the Acoustical Society of America,* 111(5 Pt 1): 2237-2241.

Tai, Y.F., Scherfler, C., Brooks, D.J., Sawamoto, N. and Castiello, U. (2004). The human premotor cortex is 'mirror' only for biological actions. *Current Biology,* 14(2): 117-120.

Talairach, J. and Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain.*. New York, NY.: Thieme Medical Publishers.

Tallal, P., Miller, S. and Fitch, R.H. (1993). Neurobiological basis of speech: A case for the preeminence of temporal processing. In: *Temporal information processing in the nervous system with special reference to dyslexia and dysphasia* (Tallal, P., Galaburda, A. M., Llinas, R. and von Euler, C., eds), pp 27-47 New York: New York Academy of Sciences.

Tardif, E. and Clarke, S. (2001). Intrinsic connectivity of human auditory areas: a tracing study with DiI. *European Journal of Neuroscience,* 13(5): 1045-1050.

Terband, H., Maassen, B., Brumberg, J. and Guenther, F.H. (2008) A transcranial magnetic stimulation virtual lesion study of speech. In: *Proceedings of the Conference on Motor Speech* Monterey, California.

Tesche, C.D. and Karhu, J.J. (2000). Anticipatory cerebellar responses during somatosensory omission in man. *Human Brain Mapping,* 9(3): 119-142.

Tomaiuolo, F., MacDonald, J.D., Caramanos, Z., Posner, G., Chiavaras, M., Evans, A.C. and Petrides, M. (1999). Morphology, morphometry and probability mapping of the pars opercularis of the inferior frontal gyrus: an in vivo MRI analysis. *European Journal of Neuroscience,* 11(9): 3033-3046.

Tourville, J.A. and Guenther, F.H. (2007). Neural mechanisms underlying sensory feedback control of speech. *Program of the 154th Meeting of the Acoustical Society of America, Journal of the Acoustical Society of America,* 122(5, Pt.2): 3087.

Toyomura, A., Koyama, S., Miyamaoto, T., Terao, A., Omori, T., Murohashi, H. and Kuriki, S. (2007). Neural correlates of auditory feedback control in human. *Neuroscience,* 146(2): 499-503.

Tsao, Y.C. and Weismer, G. (1997). Interspeaker variation in habitual speaking rate: evidence for a neuromuscular component. *Journal of Speech, Language, and Hearing Research,* 40(4): 858-866.

Tseng, Y.W., Diedrichsen, J., Krakauer, J.W., Shadmehr, R. and Bastian, A.J. (2007). Sensory prediction errors drive cerebellum-dependent adaptation of reaching. *Journal of Neurophysiology*.

Turkeltaub, P.E., Eden, G.F., Jones, K.M. and Zeffiro, T.A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *Neuroimage,* 16(3 Pt 1): 765-780.

Urasaki, E., Uematsu, S., Gordon, B. and Lesser, R.P. (1994). Cortical tongue area studied by chronically implanted subdural electrodes--with special reference to parietal motor and frontal sensory responses. *Brain,* 117 (Pt 1): 117-132.

Urban, P.P., Marx, J., Hunsche, S., Gawehn, J., Vucurevic, G., Wicht, S., Massinger, C., Stoeter, P. and Hopf, H.C. (2003). Cerebellar speech representation: lesion topography in dysarthria as derived from cerebellar ischemia and functional magnetic resonance imaging. *Archives of Neurology,* 60(7): 965-972.

Van Buren, J.M. (1963). Confusion and disturbance of speech from stimulation in the vicinity of the head of the caudate nucleus. *Journal of Neurosurgery,* 20: 148-157.

Villacorta, V.M., Perkell, J.S. and Guenther, F.H. (2007). Sensorimotor adaptation to feedback perturbations on vowel acoustics and its relation to perception. *Journal of the Acoustical Society of America,* 122(4): 2306-2319.

von Holst, E. and Mittelstaedt, H. (1950). Das reafferenzprinzip. *Naturewissenschaften,* 37: 464-476.

Voogd, J. (2003). The human cerebellum. *Journal of Chemical Neuroanatomy,* 26(4): 243-252.

Voss, M., Ingram, J.N., Haggard, P. and Wolpert, D.M. (2006). Sensorimotor attenuation by central motor command signals in the absence of movement. *Nature Neuroscience,* 9(1): 26-27.

Wallace, M.N., Johnston, P.W. and Palmer, A.R. (2002). Histochemical identification of cortical areas in the auditory region of the human brain. *Experimental Brain Research,* 143(4): 499-508.

Wang, K.S. and Shamma, S.A. (1995). Spectral Shape-Analysis in the Central Auditory-System. *Ieee Transactions on Speech and Audio Processing,* 3(5): 382-395.

Wang, Y., Isoda, M., Matsuzaka, Y., Shima, K. and Tanji, J. (2005). Prefrontal cortical cells projecting to the supplementary eye field and presupplementary motor area in the monkey. *Neuroscience Research,* 53(1): 1-7.

Wang, Y., Shima, K., Isoda, M., Sawamura, H. and Tanji, J. (2002). Spatial distribution and density of prefrontal cortical cells projecting to three sectors of the premotor cortex. *Neuroreport,* 13(10): 1341-1344.

Weeks, R.A., Aziz-Sultan, A., Bushara, K.O., Tian, B., Wessinger, C.M., Dang, N., Rauschecker, J.P. and Hallett, M. (1999). A PET study of human auditory spatial processing. *Neuroscience Letters,* 262(3): 155-158.

Wildgruber, D., Ackermann, H. and Grodd, W. (2001). Differential contributions of motor cortex, basal ganglia, and cerebellum to speech motor control: effects of syllable repetition rate evaluated by fMRI. *Neuroimage,* 13(1): 101-109.

Wildgruber, D., Kischka, U., Ackermann, H., Klose, U. and Grodd, W. (1999). Dynamic pattern of brain activation during sequencing of word strings evaluated by fMRI. *Cognitive Brain Research,* 7(3): 285-294.

Wise, R.J., Greene, J., Buchel, C. and Scott, S.K. (1999). Brain regions involved in articulation. *Lancet,* 353(9158): 1057-1061.

Wise, R.J., Scott, S.K., Blank, S.C., Mummery, C.J., Murphy, K. and Warburton, E.A. (2001). Separate neural subsystems within 'Wernicke's area'. *Brain,* 124(Pt 1): 83-95.

Wolpert, D., Miall, C. and Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Science,* 2(9): 338-347.

Wolpert, D.M. and Flanagan, J.R. (2001). Motor prediction. *Current Biology,* 11(18): R729-732.

Wolpert, D.M. and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks,* 11(7-8): 1317-1329.

Xu, Y., Larson, C.R., Bauer, J.J. and Hain, T.C. (2004). Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences. *Journal of the Acoustical Society of America,* 116(2): 1168-1178.

Yang, Y., Engelien, A., Engelien, W., Xu, S., Stern, E. and Silbersweig, D.A. (2000). A silent event-related functional MRI technique for brain activation studies without interference of scanner acoustic noise. *Magnetic Resonance in Medicine,* 43(2): 185-190.

Yates, A.J. (1963). Delayed auditory feedback. *Psychological Bulletin,* 60: 213-232.

Zarate, J.M. and Zatorre, R.J. (2005). Neural substrates governing audiovocal integration for vocal pitch regulation in singing. *Annals of the New York Academy of Science,* 1060: 404-408.

Zatorre, R.J., Belin, P. and Penhune, V.B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Science,* 6(1): 37-46.

Zatorre, R.J., Evans, A.C., Meyer, E. and Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science,* 256(5058): 846-849.

Zemlin, W.R. (1998). *Speech and hearing science anatomy and physiology*. Needham Heights, Massachusetts: Allyn and Bacon.

Ziegler, W., Kilian, B. and Deger, K. (1997). The role of the left mesial frontal cortex in fluent speech: evidence from a case of left supplementary motor area hemorrhage. *Neuropsychologia,* 35(9): 1197-1208.

CURRICULUM VITAE

Jason A. Tourville

677 Beacon St.,

Boston, MA 02472

(617) 913 9816

e-mail: jaytour@gmail.com

EDUCATION

2000 – Present	PhD. Cognitive and Neural Systems, Boston University, Boston, MA.
1992 - 1996	B.A., Neuroscience, Oberlin College, Oberlin, OH

EXPERIENCE

1999 – Present	Research Assistant, Department of Cognitive and Neural Systems Boston University, Boston, MA
2001 – 2002	Teaching Fellow, Department of Cognitive and Neural Systems Boston University, Boston, MA
1999 – 2000	Research Assistant, Center for Morphometric Analysis, Massachusetts General Hospital, Boston, MA
1996 – 1999	Research Assistant, Department of Psychiatric Epidemiology, Harvard Medical School, Boston, MA

AWARDS

2000 – Present	Boston University Graduate Research Fellowship
2001 – 2002	Boston University Outstanding Teaching Fellow Award
2004	Best Student Paper in Speech Communication, 148th Meeting of the Acoustical Society of America, San Diego, CA

PUBLICATIONS

Ghosh, S.S., Tourville, J.A., and Guenther, F.H. (2008). An fMRI study of the overt production of simple speech sounds. *Journal of Speech, Language, and Hearing Research*. In press.

Tourville, J.A., Reilly, K.J., and Guenther, F.H., (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*. 39(3):1429-1443

Tourville, J.A., Reilly, K.J., and Guenther, F.H., (2007). Neural mechanisms underlying sensory feedback control of speech. Program of the 154th Meeting of the Acoustical Society of America, *Journal of the Acoustical Society of America*, 122(5, Pt. 2):3087.

Reilly, K.J., Guenther, F.H., Tourville, J.A. (2007). A neuroimaging investigation of auditory-motor learning. Program of the 154th Meeting of the Acoustical Society of America, *Journal of the Acoustical Society of America*, 122(5, Pt. 2):3087.

Reilly, K.J., Guenther, F.H., Tourville, J.A., and Bohland, J.W. (2006). Brain activations during learning of a novel speech sensorimotor mapping. Conference on Motor Speech: Motor Speech Disorders. Austin, Texas.

Guenther, F.H., Ghosh, S.S., and Tourville, J.A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*. 96(3):280-301.

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

Klein A, Mensh B, Ghosh S, Tourville J, Hirsch J. (2005). Mindboggle: automated brain labeling with multiple atlases. *BMC Med Imagin*g. 5:7.

Tourville, J.A., Guenther, F.H., Ghosh, S.S., and Bohland, J.W. (2004). Effects of jaw perturbation on cortical activity during speech production. *Journal of the Acoustical Society of America (148th Meeting of the Acoustical Society of America, San Diego)*, 116(4):2631.

Guenther, F.H., Nieto-Castanon, A., Ghosh, S.S., and Tourville, J.A. (2004). Representation of sound categories in auditory cortical maps. *Journal of Speech, Language, and Hearing Research*, 47(1):46-57.

Nieto-Castanon, A., Ghosh, S.S., Tourville, J.A., and Guenther, F.H. (2003). Region-of-interest based analysis of functional imaging data. *Neuroimage*, 19:1303-1316.

Guenther, F.H., Tourville, J.A., and Bohland, J. (2003). Modeling the representation of speech sounds in auditory cortical areas. Program of the 145th Meeting of the Acoustical Society of America, *Journal of the Acoustical Society of America*, 113(4, Pt. 2):2210.

Tourville, J.A. and Guenther, F.H. (2003). A cortical and cerebellar parcellation system for speech studies. Boston University Technical Report CAS/CNS-03-022. Boston, MA: Boston University.

Guenther, F.H., Nieto-Castanon, A., Tourville, J.A., and Ghosh, S.S. (2001). The effects of categorization training on auditory perception and cortical representations. *Proceedings of the Speech Recognition as Pattern Classification (SPRAAC) Workshop*, Nijmegen, The Netherlands, July 11-13.

Ghosh, S., Nieto-Castanon, A., Tourville, J., and Guenther, F. (2001). ROI-based analysis of fMRI data incorporating individual differences in brain anatomy. *Neuroimage*, 13(6, part2):125.

Guenther, F.H., Nieto-Castanon, A., Tourville, J.A., and Ghosh, S.S. (2000). The representation of prototypical and non-prototypical vowels in peri-sylvian cortical areas. *Society for Neuroscience Abstracts*, 26(2):1971.

Seidman, L.J., Faraone, S.V., Goldstein, J.M., Goodman, J.M., Kremen, W.S., Toomey, R., Tourville, J., Kennedy, D., Makris, N., Caviness, V.S., and Tsuang, M.T. (1999). Thalamic and Amygdala-Hippocampal Volume Reductions in First Degree Relatives of Schizophrenic Patients: An MRI-Based Morphometric Analysis. *Biological Psychiatry*. 46:941-954.

Goldstein, J.M., Goodman, J.M., Seidman, L.J., Kennedy, D., Makris, N., Hang, L., Tourville, J., Caviness, V.S., Faraone, S.V., and Tsuang, M.T. (1999). Cortical Abnormalities in Schizophrenia Identified by Structural Magnetic Resonance Imaging. *Archives of General Psychiatry*. 56:537-547.

INVITED LECTURES

"Neural Modeling and Imaging of Speech Production." University of Connecticut Communication Disorders Research Colloquium Series. Storrs, Connecticut, April 6, 2005.

PROFESSIONAL MEMBERSHIPS AND SERVICE

Student Member, The Society for Neuroscience
Student Member, The Organization for Human Brain Mapping
Ad-Hoc Reviewer, *Neural Networks*